Introduction to Electronic Structure Methods

Prof. Ursula Röthlisberger

Fall Semester 2018

EPFL, Lausanne

Contents

Bibliography

- On Quantum Mechanics and Quantum Chemistry:
  - C. Cohen-Tannoudji, B. Diu, F. Laloe *Quantum Mechanics Volumes I & II*, Wiley-Interscience (1977).
  - L. D. Landau and E. M. Lifshitz, *Quantum mechanics: non-relativistic theory* , Pergamon.
  - J.J. Sakurai, *Modern Quantum Mechanics*, Addison Wesley.
  - A. Szabo, N.S. Ostlund, Modern Quantum Chemistry, McGraw-Hill (1982).
  - P. W. Atkins, R. S. Friedman *Molecular Quantum Mechanics*, Oxford (1997).
- On Computational Chemistry:
  - F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons (1999).
  - C. J. Cramer, *Essentials of Computational Chemistry*, Wiley (2002).
- On Density Functional Theory (DFT):
  - R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford (1989).
  - W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH (2000).
- On numerical methods:
  - W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes. The Art of Scientific Computing.*
    The basic handbook on algorithms. Available everywhere and also fully available online at http://www.nr.com.
- More advanced books:
  - R. L. Martin, *Electronic Structure. Basic theory and practical methods*, Cambridge University Press (2004).
  - R.M. Dreizler and E. K. U. Gross, *Density Functional Theory*, Springer (1990).

Chapter 1

Introduction to computational quantum chemistry

Computational Quantum Chemistry is a branch of Theoretical Chemistry whose major goal is to create efficient mathematical approximations and computer programs to calculate the properties of molecules such as their **total energy**, **dipole** and **quadrupole moments**, **potential energy (free energy) sur-faces**, **vibrational frequencies**, **excitation energies** and other diverse **spectroscopic quantities**, **reactivity behavior**, the involved **reaction mechanism** and reaction **dynamics**. The term is also sometimes used to cover the areas of overlap between computer science and chemistry.

Theoretical vs. Computational Chemistry. The term theoretical chemistry may be defined as a mathematical description of chemistry, whereas computational chemistry is usually used when a mathematical method is sufficiently well developed that it can be automated for implementation on a computer. Note that the word exact does not appear here, as very few aspects of chemistry can be computed exactly. Almost every aspect of chemistry, however, can be and has been described in a qualitative or approximate quantitative computational scheme.

Accuracy vs. efficiency. It is, in principle, possible to use a very accurate method and apply it to all molecules. Although such methods are well-known and available in many computer programs, the computational cost of their use grows exponentially with the number of electrons. Therefore, a great number of approximate methods strive to achieve the best trade-off between accuracy and computational cost. Present computational chemistry can routinely and very accurately calculate the properties of molecules that contain no more than 40-50 electrons. The treatment of molecules that contain up to 1'000 electrons or more is computationally tractable by some relatively accurate methods like Density Functional Theory (DFT).

The main target of this course is to introduce the basic notions

of Computational Quantum Chemistry, which allow to explore the properties of molecular systems to understand chemical and biochemical reaction mechanisms and predict and rationalize observations of laboratory experiments. To this end, we will discuss the basic theoretical concepts and approximations of different quantum chemical methods and we will learn how to apply them to run calculations on computers.

Computational quantum chemistry methods can provide use with a lot of useful information. They can for example be applied to:

- compute physical and chemical properties of molecules (structure and dynamics)
- identify correlations between chemical structures and properties
- understand reaction mechanisms and thermodynamic properties
- determine potential energy surfaces and quantum forces on the nuclei to perform *ab initio* molecular dynamics simulations.
- help in the efficient synthesis of compounds
- design molecules that interact in specific ways with other molecules (e.g. drug design)

Quantum chemical methods can also be applied to solid state physics problems. The electronic structure of a crystal is in general described by a band structure, which defines the energies of electron orbitals for each point in the Brillouin zone.

1.1 Ab Initio methods

The programs used in Quantum Chemistry are based on different quantum-mechanical methods that solve the molecular Schrödinger equation associated with the molecular Hamiltonian. Methods that do not include empirical or semi-empirical parameters in their equations - are derived directly from theoretical principles, with no inclusion of experimental data - are generally called ***ab initio*** methods. However, the same term is also used to design theories that are derived from exact quantum mechanical equations but which also assume a certain level of approximation. The approximations made in these cases are usually mathematical in nature, such as using a simpler functional form or getting an approximate solution for a complicated differential equation.

1.1.1 Electronic structure

One of the primary goals of quantum chemical methods is to determine the electronic structure, e.g. the probability distribution of electrons in chemical systems. The electronic structure is determined by solving the Schrödinger equation associated with the electronic molecular Hamiltonian. In this process, the molecular geometry is considered as a fixed parameter. Once the optimal electronic wavefunction is determined, one can access the gradient on each nuclei (as the derivative of the total energy with respect to the positions of the nuclei) and update their positions accordingly until the process reaches convergence. This is how we obtain optimized molecular geometries. Usually the basis set (which is usually built from the LCAO ansatz) used to solve the Schrödinger equation is not complete and does not span the full Hilbert space associated to the system. However, this approximation allows one to treat the Schrödinger equation as a "simple" eigenvalue equation of the electronic molecular Hamiltonian with a discrete set of solutions. The problem of dealing with function (functionals) and operators can be therefore translated into a linear algebra calculation based on energy matrices and state vectors. The obtained eigenvalues are functions of the molecular geometry and describe the **potential energy surfaces**.
Many optimized linear algebra packages have been developed for this purpose (*e.g.*, LAPACK – Linear Algebra PACKage: http://www.netlib.org/lapack).

One of the most basic ab initio electronic structure approaches is called the **Hartree-Fock (HF)** method, in which the Coulombic electron-electron repulsion is taken into account in an averaged way (mean field approximation). This is a variational calculation, therefore the obtained approximate energies, expressed in terms of the system's wavefunction, are always equal to or greater than the exact energy, and tend to a limiting value called the Hartree-Fock limit. Many types of calculations begin with a HF calculation and subsequently correct for the missing electronic correlation. **Møller-Plesset perturbation theory** (MP) and **Coupled Cluster (CC)** are examples of such methods.

An alternative stochastic approach is **Quantum Monte Carlo (QMC)**, in its variational, diffusion, and Green's functions flavors. These methods work with an explicitly correlated wavefunction and evaluate integrals numerically using a Monte Carlo integration. Such calculations can be very time consuming, but they are probably the most accurate methods known today.

Density Functional Theory (DFT) methods are often considered to be *ab initio* methods for determining the molecular electronic structure, even though one utilizes functionals usually derived from empirical data, properties of the electron gas, calculations on rare gases or more complex higher level approaches. In DFT, the total energy is expressed in terms of the total electron density rather than the wavefunction.

The Kohn-Sham formulation of DFT introduces a set of non-interacting *fictitious* molecular orbitals, which makes the chemical interpretation of DFT simpler, even though their nature is different from the *real*-HF ones. In addition, KS based DFT has the advantage to be easily translated into an efficient computer code.

The most popular classes of ab initio electronic structure methods are:

- Hartree-Fock
- Generalized Valence Bond
- Møller-Plesset Perturbation Theory
- Multi-Configurations Self Consistent Field (MCSCF)
- Configuration Interaction
- Multi-Reference Configuration Interaction
- Coupled Cluster
- Quantum Monte Carlo
- Reduced Density Matrix Approaches
- Density Functional Theory

Ab initio electronic structure methods have the advantage that they can be made to converge systematically to the exact solution (by improving the level of approximation). The convergence, however, is usually not monotonic, and sometimes - for a given property - a less accurate calculation can give a better result (for instance the total energy in the series HF, MP2, MP4, . . .).

The drawback of ab initio methods is their cost. They often take enormous amounts of computer time, memory, and disk space. The HF method scales theoretically as N^4 (N being the number of basis functions) and correlated calculations often scale much less favorably (correlated DFT calculations being the most efficient of this lot).

1.1.2 Quantum dynamics

Once the electronic and nuclear motions are separated (within the Born-Oppenheimer approximation), the wave packet corresponding to the nuclear degrees of freedom can be propagated via the time evolution operator associated with the time-dependent Schrödinger equation (for the full molecular Hamiltonian). The most popular methods for propagating the wave packet associated to the molecular geometry are

- the split operator technique for the time evolution of a phase-space distribution of particles (nuclei) subject to the potential generated by the electrons (Liouville dynamics),

- the Multi-Configuration Time-Dependent Hartree method (MCTDH), which deals with the propagation of nuclear wavepackets of molecular system evolving on one or several coupled electronic potential energy surfaces.

- Feynman path-integral methods. In this approach the quantum partition function describing the dynamics of N nuclei is mapped into a classical configurational partition function for a $N \times P$-particle system, where P are discrete points along a cyclic path.

The semiclassical approach. In this case one adopts a classical molecular mechanics propagation of the nuclear dynamics (where the equation of motion is given by Newton's law) combined with any treatment of the time-dependent Schrödinger equation for the electrons. The combination of classical molecular dynamics with an efficient scheme for the computation of the potential energy surface, like for instance DFT, allows for *ab initio* MD simulations of systems of the size of few hundreds of atoms. Nonadiabatic effects[1] can also be introduced by means of the *fewest switches surface hopping* scheme of Tully et. al. or a *mean-field* approximation (Ehrenfest dynamics).

Different numerical schemes for the update of the electronic state at each molecular step are possible

- wavefunction optimization at the new nuclear configuration. This method is often called Born-Oppenheimer Molecular Dynamics (BO MD).

- within DFT, the electronic density can be propagated using the "fictitious" electronic dynamics obtained from the Car-Parrinello extended Lagrangian scheme. In this way one saves the cost of a full wavefunctions optimization at each MD step. The method is referred to as Car-Parrinello Molecular Dynamics (CP MD).

- propagation of the electronic wavefunction using an approximate solution of the time-dependent Schrödinger equation. Due to the small mass of the electrons (compared to the ones of the nuclei) the time step for the propagation of the electrons is necessary much smaller than the one required for the solution of the Newton's equations for the ions.

1.2 Semiempirical methods

1.2.1 Electronic structure

The most expensive term in a Hartree-Fock calculation, are the socalled two electron integrals. Semiempirical methods are based on the HF method

---

[1] Nonadiabatic effects describe the quantum phenomena that arise when the dynamics reaches regions in phase space where two or more adiabatic potential energy surfaces are approaching each other.

but the costly two electrons integrals are approximated by empirical data or omitted completely. In order to correct for this approximation, semiempirical methods introduce additional empirical terms in the Hamiltonian, which are weighted by a set of *a priori* undefined parameters. In a second step, these are fitted in order to reproduce results in best agreement with experimental data.

Semiempirical calculations are much faster than their ab initio counterparts. Their results, however, can be badly wrong if the molecule being computed is not similar enough to the molecules in the database used to parameterize the method.

Semiempirical calculations have been very successful in the description of organic chemistry, where only a few elements are used extensively and molecules are of moderate size.

Semiempirical methods also exist for the calculation of electronically excited states. These methods, such as the Pariser-Parr-Pople (PPP) method, can provide good estimates of the electronic excited states when parameterized well. Indeed, for many years, the PPP method outperformed ab initio excited state calculations. Large scale ab initio calculations have confirmed many of the approximations of the PPP model and explain why the PPP-like models work so well in spite of their simple formulation.

1.3 Mixed Quantum Mechanical/Molecular Mechanical (QM/MM) methods

The combination of quantum chemical methods (QM) with a classical molecular mechanical (MM) treatment of the environment enables the investigation of larger systems, in which the chemical active region (QM) is confined in a localized volume surrounded by MM atoms. This approach is particularly suited for the analysis of enzymatic reactions in proteins.

1.4 Software packages

A number of software packages that are self-sufficient and include many quantum chemical methods are available. The following is a (non complete) table illustrating the capabilities of various software packages:

Package	MM	Semi-Empirical	HF	Post-HF	DFT	Ab-inito MD	Periodic	QM/MM
ACES	N	N	Y	Y	N	N	N	N
ADF	N	N	N	N	Y	N	Y	Y
CPMD	Y	N	N	N	Y	Y	Y	Y
DALTON	N	N	Y	Y	Y	N	N	N
GAUSSIAN	Y	Y	Y	Y	Y	Y(?)	Y	Y
GAMESS	N	Y	Y	Y	Y	N	N	Y
MOLCAS	N	N	Y	Y	N	N	N	N
MOLPRO	N	N	Y	Y	Y	N	N	N
MOPAC	N	Y	N	N	N	N	Y	N
NWChem	Y	N	Y	Y	Y	Y(?)	Y	N
PLATO	Y	N	N	N	Y	N	Y	N
PSI	N	N	Y	Y	N	N	N	N
Q-Chem	?	N	Y	Y	Y	N	N	N
TURBOMOL	N	N	Y	Y	Y	Y	Y	N

Chapter 2

Basic principles of Quantum Mechanics

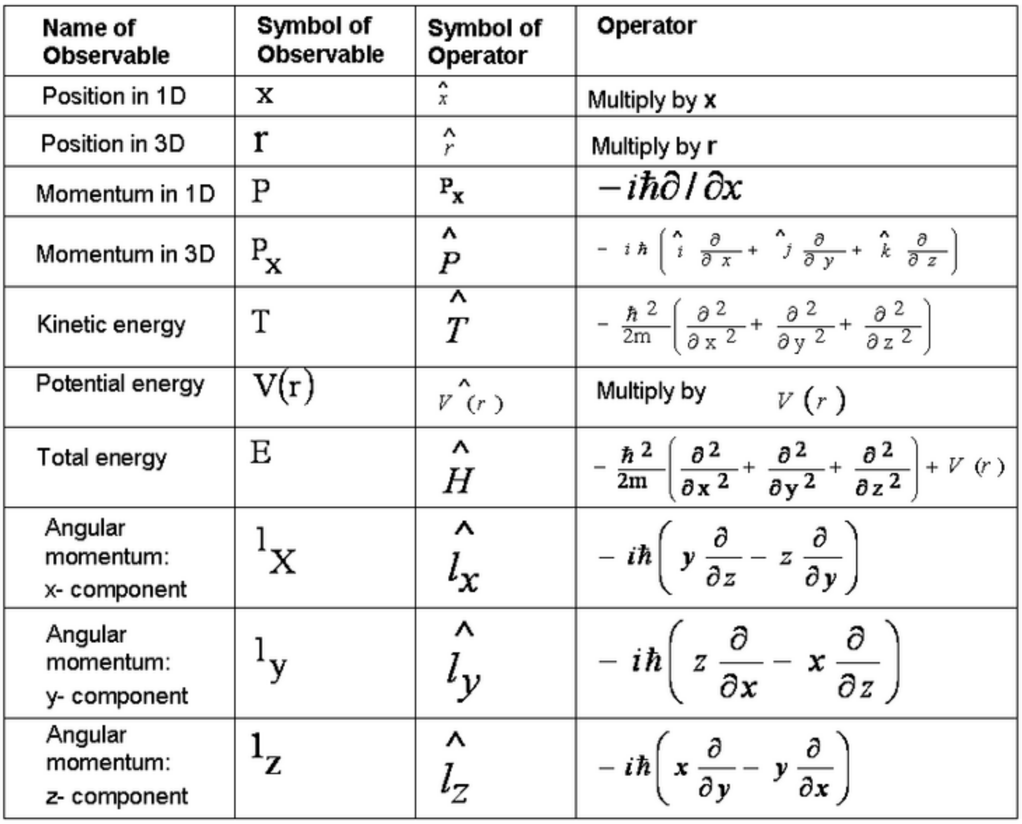
2.1 Postulates of Quantum Mechanics

- *Postulate 1.* The state of a quantum mechanical system is completely specified by a function $\Psi(\mathbf{r}, t)$ that depends on the coordinates of all particles $\mathbf{r}$ and on time. This function, called the wave function or state function, has the important property that $\Psi^*(\mathbf{r}, t)\Psi(\mathbf{r}, t)d\tau$ is the probability that the particle lies in the volume element $d\tau$ located at $\mathbf{r}$ at time t. The wavefunction must satisfy certain mathematical conditions because of this probabilistic interpretation. For the case of a single particle, the probability of finding it somewhere in space is 1, so that we have the normalization condition

$$\int_{-\infty}^{\infty} \Psi^*(\mathbf{r}, t)\Psi(\mathbf{r}, t)d\tau = 1 \tag{2.1}$$

It is customary to also normalize many-particle wavefunctions to 1. The wavefunction must also be single-valued, continuous, and finite.

- Postulate 2. To every observable in classical mechanics there corresponds a linear, Hermitian operator in quantum mechanics. If we require that the expectation value of an operator $\hat{A}$ is real, then $\hat{A}$ must be a Hermitian operator. Some common operators occurring in quantum mechanics are collected in the following table.



Name of Observable	Symbol of Observable	Symbol of Operator	Operator
Position in 1D	x	$\hat{x}$	Multiply by **x**
Position in 3D	$\mathbf{r}$	$\hat{r}$	Multiply by **r**
Momentum in 1D	P	P_x	$-i\hbar\partial/\partial x$
Momentum in 3D	P_x	$\hat{P}$	$-i\hbar\left(\hat{i}\frac{\partial}{\partial x}+\hat{j}\frac{\partial}{\partial y}+\hat{k}\frac{\partial}{\partial z}\right)$
Kinetic energy	T	$\hat{T}$	$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)$
Potential energy	$V(r)$	$\hat{V}(r)$	Multiply by $V(r)$
Total energy	E	$\hat{H}$	$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V(r)$
Angular momentum: x- component	l_x	$\hat{l}_x$	$-i\hbar\left(y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}\right)$
Angular momentum: y- component	l_y	$\hat{l}_y$	$-i\hbar\left(z\frac{\partial}{\partial x}-x\frac{\partial}{\partial z}\right)$
Angular momentum: z- component	l_z	$\hat{l}_z$	$-i\hbar\left(x\frac{\partial}{\partial y}-y\frac{\partial}{\partial x}\right)$

Figure 2.1: Physical observables and their corresponding quantum operators (from wikipedia).

- Postulate 3. In any measurement of the observable associated with operator $\hat{A}$, the only values that will ever be observed are the eigenvalues a, which satisfy the eigenvalue equation

$$\hat{A}\Psi(\mathbf{r},t)=a\Psi(\mathbf{r},t) \tag{2.2}$$

This postulate captures the central point of quantum mechanics–the values of dynamical variables can be quantized (although it is still possible to have a continuum of eigenvalues in the case of unbound states). If the system is in an eigenstate of $\hat{A}$ with a single eigenvalue a, then any measurement of the quantity A will yield a.

Although measurements must always yield an eigenvalue, the state does not have to be an eigenstate of $\hat{A}$ initially. An arbitrary state can be expanded in the complete set of eigenvectors of $\hat{A}$ ($\hat{A}\Psi_i(\mathbf{r},t) = a_i\Psi_i(\mathbf{r},t)$) as

$$\Psi(r,t) = \sum_{i=1}^{N} c_i\Psi_i(\mathbf{r},t) \tag{2.3}$$

where N may go to infinity. In this case we only know that the measurement of A will yield one of the values a_i, but we don't know which one. However, we do know the probability that eigenvalue a_i will occur: it is the absolute value squared of the coefficient, $|c_i|^2$, leading to the fourth postulate below.
An important second half of the third postulate is that, after measurement of Ψ yields some eigenvalue a_i, the wavefunction immediately "collapses" into the corresponding eigenstate Ψ_i. Thus, measurement affects the state of the system. This fact is used in many elaborate experimental tests of quantum mechanics.

- Postulate 4. If a system is in a state described by a normalized wavefunction Ψ, then the expectation value of the observable corresponding to $\hat{A}$ is given by

  $$\langle A\rangle = \int_{-\infty}^{\infty} \Psi^*(\mathbf{r},t)\hat{A}\Psi(\mathbf{r},t)\,d\tau \tag{2.4}$$

- Postulate 5. The wavefunction or state function of a system evolves in time according to the time-dependent Schrödinger equation

  $$\hat{H}\Psi(\mathbf{r},t) = i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} \tag{2.5}$$

  The central equation of quantum mechanics must be accepted as a postulate.

- Postulate 6. The total wavefunction must be antisymmetric with respect to the interchange of all coordinates of one fermion[1] with those of another. Electronic spin must be included in this set of coordinates. The Pauli exclusion principle is a direct result of this antisymmetry principle. We will later see that Slater determinants provide a convenient means of enforcing this property on electronic wavefunctions.

2.2 The molecular Hamiltonian and the Born-Oppenheimer approximation

This chapter was adapted from the lecture notes "A Short Summary of Quantum Chemistry", MITOPENCOURSEWARE, December 2004.

---

[1] Fermions: particles with half-integer spins. Electrons are fermions.
Bosons: particles with integer spin, e.g. the nucleus of a C-12 atom.

Quantum Chemistry is (typically) based on the non-relativistic Schrödinger equation within the Born-Oppenheimer approximation. The Schrödinger equation is (we use atomic units: $\hbar = 1, m_{elec} = 1, e_{elec} = 1$)

$$\hat{H}_{tot}(\mathbf{R}, \mathbf{P}, \mathbf{r}, \mathbf{p})\Psi_{tot}(\mathbf{R}, \mathbf{P}, \mathbf{r}, \mathbf{p}) = E(\mathbf{R}, \mathbf{P})\,\Psi_{tot}(\mathbf{R}, \mathbf{P}, \mathbf{r}, \mathbf{p}) \tag{2.6}$$

where $\mathbf{r}, \mathbf{p} = \partial/\partial\mathbf{r}$ are the electronic collective coordinates and $\mathbf{R}, \mathbf{P} = \partial/\partial\mathbf{R}$ are the nuclear collective coordinates, and

- E is an allowed energy of the system (the system is usually a molecule).
- Ψ_{tot} is a function of the positions of all the electrons and nuclei (we drop all spin dependencies).
- H_{tot} is a differential operator constructed from the classical Hamiltonian $H(\mathbf{P}, \mathbf{R}, \mathbf{p}, \mathbf{r}) = E$ by replacing all the momenta $\mathbf{p}_i$ (resp.$\mathbf{P}_I$) with $(i)\partial/\partial\mathbf{r}_i$ $((i)\partial/\partial\mathbf{R}_i)$ as long as all the $\mathbf{p}$ ($\mathbf{P}$) and $\mathbf{r}$ ($\mathbf{R}$) are Cartesian.

For a system of nuclei and electrons in vacuum with no external fields, neglecting magnetic interactions, using atomic units:

$$\hat{H}_{tot} = -\frac{1}{2}\sum_I \frac{1}{M_I}\nabla_I^2 - \frac{1}{2}\sum_n \nabla_n^2 + \sum_{I<J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{In}\frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_n|} + \sum_{n<m}\frac{1}{|\mathbf{r}_m - \mathbf{r}_n|} \tag{2.7}$$

The Born-Oppenheimer approximation is to neglect some of the terms coupling the electrons and nuclei, so one can write:

$$\Psi_{tot}(\mathbf{R}, \mathbf{r}) = \Psi_{nucl}(\mathbf{R})\,\Psi_{elec}(\mathbf{r}; \mathbf{R}) \tag{2.8}$$

and

$$\hat{H}_{tot} = \hat{T}_{nucl}(\mathbf{P}, \mathbf{R}) + \hat{H}_{elec}(\mathbf{p}, \mathbf{r}; \mathbf{R}) \tag{2.9}$$

which ignores the dependence of $\hat{H}_{elec}$ on the momenta of the nuclei $\mathbf{P}$. One can then solve the Schrödinger equation for the electrons (with the nuclei fixed, indicated by $(;\mathbf{R})$). The energy we compute will depend on the positions $\mathbf{R}$ of those fixed nuclei, call it $E(\mathbf{R})$:

$$\hat{H}_{elec}(\mathbf{p}, \mathbf{r}; \mathbf{R})\Psi_{elec}(\mathbf{r}; \mathbf{R}) = E(\mathbf{R})\,\Psi_{elec}(\mathbf{r}; \mathbf{R}) \tag{2.10}$$

The collection of all possible nuclear configurations, $\mathbf{R}$ together with the associated energies, $E(\mathbf{R})$, defines a potential energy surface, $V(\mathbf{R})$ for the nuclei.
Now we can go back to the total Hamiltonian, and integrate over all the electron positions $\mathbf{r}$, ignoring any inconvenient term, to obtain an approximate Schrödinger equation for the nuclei:

$$\langle\Psi_{elec}(\mathbf{r}, \mathbf{R})|\hat{H}_{tot}|\Psi_{elec}(\mathbf{r}, \mathbf{R})\rangle \cong \hat{H}_{nucl} = \hat{T}_{nucl}(\mathbf{P}, \mathbf{R}) + V(\mathbf{R}) \tag{2.11}$$

with

$$\big(\hat{T}_{nucl}(\mathbf{P}, \mathbf{R}) + V(\mathbf{R})\big)\Psi_{nucl}(\mathbf{R}) = E_{nucl}\Psi_{nucl}(\mathbf{R}) \tag{2.12}$$

2.2.1 The nuclear Schrödinger equation

Both approximate Schrödinger equations (for electrons eq. 2.10 and for nuclei eq. 2.12) are still much too hard to solve exactly (they are partial differential equations in $3N$ particle coordinates), so we have to make more approximations.
$V(\mathbf{R})$ is usually expanded to second order $\mathbf{R}$ about a stationary point $\mathbf{R}_0$:

$$V(\mathbf{R}) \cong V(\mathbf{R}_0) + \frac{1}{2}\sum_{i,j}\Big(\frac{\partial^2 V(\mathbf{R})}{\partial \mathbf{R}_i \partial \mathbf{R}_j}\Big)(\mathbf{R}_i - \mathbf{R}_{0i})(\mathbf{R}_j - \mathbf{R}_{0j}) \tag{2.13}$$

and then the translations, rotations, and vibrations are each treated separately, neglecting any inconvenient terms that couple the different coordinates. In this famous "rigid-rotor-harmonic-oscillator (RRHO)" approximation, analytical formulas are known for the energy eigenvalues, and for the corresponding partition functions Q (look in any Phys.Chem. textbook).
This approximate approach has the important advantage that we do not need to solve the Schrödinger equation for the electrons at very many $\mathbf{R}$'s: we just need to find a stationary point $\mathbf{R}_0$, and compute the energy and the second derivatives at that $\mathbf{R}_0$. Many computer programs have been written that allow one to compute the first and second derivatives of $V(\mathbf{R})$ almost as quickly as you can compute V. For example, for a system with 10 atoms and $3 * 10 = 30$ coordinates $\mathbf{R}_I$, it takes about half a minute on a PC to compute $V(\mathbf{R}_0)$ and only about 13 more minutes to compute the $30 * 30 = 900$ second derivatives $\left(\frac{\partial^2 V(\mathbf{R})}{\partial \mathbf{R}_i \partial \mathbf{R}_j}\right)$. If you tried to do this naively by finite differences, it would take about 15 hours to arrive at the same result (and it would probably be less accurate because of finite differencing numerical errors.) The analytical first derivatives are used to speed the search for the stationary point (e.g. the equilibrium geometry) $\mathbf{R}_0$. Often the geometry and the second derivatives are calculated using certain approximations, but the final energy $V(\mathbf{R}_0)$ is computed more accurately (since thermodynamics data and reaction rates are most sensitive to errors in $V(\mathbf{R}_0)$, and even poor approximations often get geometry and frequencies close to correct).
Therefore, as long as a second-order Taylor expansion approximation for V is adequate we are in pretty good shape. Molecules and transition states with "large amplitude motions" (i.e. the Taylor expansion is not adequate) are much more problematic, dealing with them is an active research area. Fortunately, there are many systems where the conventional second-order V, RRHO approximation is accurate.

2.2.2 The electronic Schrödinger equation

The question is now how to compute the required potential $V(\mathbf{R})$ which acts on the nuclei at a given geometry $\mathbf{R}$. What we need to solve is 2.10,

$$\hat{H}_{elec}(\mathbf{p}, \mathbf{r}; \mathbf{R})\Psi_{elec}(\mathbf{r}; \mathbf{R}) = V(\mathbf{R})\Psi_{elec}(\mathbf{r}; \mathbf{R})$$

where in a vacuum, in the absence of fields, and neglecting magnetic effects

$$\hat{H}_{elec}(\mathbf{R}) = -\frac{1}{2}\sum_n \nabla_n^2 + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{In} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_n|} + \sum_{n<m} \frac{1}{|\mathbf{r}_m - \mathbf{r}_n|} \tag{2.14}$$

and because the electrons are indistinguishable fermions any permutation of two electrons must change the sign of the wavefunction $\Psi_{elect}(\mathbf{r};\mathbf{R})$ (this is a really important constraint called the Pauli exclusion principle, it is the reason for the specific structure of the periodic table).
In addition, because the spin is a good quantum number we can chose the electronic wavefunction to be simultaneously an eigenfunction of the spin operator:

$$S^2|\Psi_{elec}\rangle = S(S+1)|\Psi_{elec}\rangle \tag{2.15}$$

$$S_z|\Psi_{elec}\rangle = S_z|\Psi_{elec}\rangle \tag{2.16}$$

We can write Ψ_{elec} in a form that will guarantee it satisfies the Pauli principle, namely using the Slater determinant many-electron wavefunctions:

$$\Psi_{el}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \sum_{m_1, m_2, \ldots, m_N} C_{m_1, m_2, \ldots, m_N} |\phi_{m_1}(\mathbf{r}_1)\phi_{m_2}(\mathbf{r}_2)\ldots\phi_{m_N}(\mathbf{r}_N)|$$

where

$$|\phi_{m_1}(\mathbf{r}_1)\phi_{m_2}(\mathbf{r}_2)\ldots\phi_{m_N}(\mathbf{r}_N)| = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{m_1}(\mathbf{r}_1) & \phi_{m_2}(\mathbf{r}_1) & \cdots & \phi_{m_N}(\mathbf{r}_1) \\ \phi_{m_1}(\mathbf{r}_2) & \phi_{m_2}(\mathbf{r}_2) & \cdots & \phi_{m_N}(\mathbf{r}_2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \phi_{m_1}(\mathbf{r}_N) & \phi_{m_2}(\mathbf{r}_N) & \cdots & \phi_{m_N}(\mathbf{r}_N) \end{vmatrix}$$

The components of the Slater determinant, $\phi_{m_i}(\mathbf{r}_i)$, are one-electron molecular orbitals which are usually given as an expansion in "atomic orbitals", χ_n:

$$\phi_m(\mathbf{r}, s) = \sum_n D_{mn}\chi_n(\mathbf{r}) \otimes s \tag{2.17}$$

($\mathbf{r}$ stays for the Cartesian coordinates (x, y, z) and s is the spin variable $(s \in \{\alpha, \beta\})$)
The collection of coefficients $D_{\ldots}$ and $C_{\ldots}$ fully characterizes the solution of the electronic Schrödinger equation for atoms and molecules.

The main subject of this course is a discussion of the approximation methods for the solution of the Schrödinger equation for the electrons (given by the coefficients $C_{\ldots}$ and $D_{\ldots}$), which provides the potential for the nuclei dynamics $V(\mathbf{R})$. IF time allows, towards the endof the course, we will discuss some of the semiclassical adiabatic approaches for the nuclear dynamics such as the Car-Parrinelllo method and QM/MM molecular dynamics.

Before starting, we have however to translate this problem into a formulation suited for computation. Using appropriate basis function it is possible to translate 2.10 into a simple linear algebra problem, which can be solved using efficient computer software (see for instance the parallel package for the solution of linear algebra problem LAPACK: Linear Algebra PACKage (www.netlib.org/lapack) [2]).

2.3 Basis sets, linear algebra and the secular equation

2.3.1 Basis kets and matrix representation

Given an Hermitian operator A, its eigenkets (eigenfunctions),$|\varphi_a\rangle$ form a complete orthonormal set. An arbitrary ket, $|\alpha\rangle$ can be expanded in terms of the eigenkets of A.

$$|\alpha\rangle = \sum_a c_a |\varphi_a\rangle \tag{2.18}$$

Multiplying with $\langle\varphi_{a'}|$ on the left and using the orthogonality property $\langle\varphi_{a'}|\varphi_a\rangle$, we can immediately find the coefficient,

$$c_a = \langle\varphi_a|\alpha\rangle \tag{2.19}$$

In other words, we have

$$|\alpha\rangle = \sum_{a'} |\varphi_{a'}\rangle\langle\varphi_{a'}|\alpha\rangle\,, \tag{2.20}$$

---

[2] LAPACK is written in Fortran90 and provides routines for solving systems of simultaneous linear equations, least-squares solutions of linear systems of equations, eigenvalue problems, and singular value problems. The associated matrix factorizations (LU, Cholesky, QR, SVD, Schur, generalized Schur) are also provided, as are related computations such as reordering of the Schur factorizations and estimating condition numbers. Dense and banded matrices are handled, but not general sparse matrices. In all areas, similar functionality is provided for real and complex matrices, in both single and double precision.
LAPACK routines are written so that as much as possible of the computation is performed by calls to the Basic Linear Algebra Subprograms (BLAS). LAPACK was designed at the outset to exploit Level 3 BLAS – a set of specifications for Fortran subprograms that do various types of matrix multiplication and the solution of triangular systems with multiple right-hand sides. Because of the coarse granularity of the Level 3 BLAS operations, their use promotes high efficiency on many high-performance computers, particularly if specially coded implementations are provided by the manufacturer.
Highly efficient machine-specific implementations of the BLAS are available for many modern high-performance computers. For details of known vendor- or ISV-provided BLAS, consult the BLAS FAQ. Alternatively, the user can download ATLAS to automatically generate an optimized BLAS library for the architecture. A Fortran77 reference implementation of the BLAS is available from netlib; however, its use is discouraged as it will not perform as well as a specially tuned implementations.

which is analogous to an expansion of a vector $\vec{v}$ in the (real) Euclidean space:

$$\vec{v} = \sum_i \hat{e}_i \left(\hat{e}_i \cdot \vec{v}\right), \tag{2.21}$$

where $\{\hat{e}_i\}$ form an orthogonal set of unit vectors.

An important operator is the projection operator Λ_a, which acting on a ket state $|\alpha\rangle$ gives the component of the ket parallel to $|\varphi_a\rangle$,

$$\left(|\varphi_a\rangle\langle\varphi_a|\right)|\alpha\rangle = |\varphi_a\rangle\langle\varphi_a|\alpha\rangle = c_a|\varphi_a\rangle\,. \tag{2.22}$$

Since the sum of all projections of a ket $|\alpha\rangle$ gives back the same ket,

$$\sum_a |\varphi_a\rangle\langle\varphi_a| = 1 \tag{2.23}$$

where 1 is the identity operator. This representation of the unity operator is called the **completness relation**.

Having specified the base ket, we now show how to represent an operator X, by a square matrix. Using the completness relation twice, we can write the operator X as

$$X = \sum_{a'}\sum_a |\varphi_{a'}\rangle\langle\varphi_{a'}|X|\varphi_a\rangle\langle\varphi_a| \tag{2.24}$$

There are alltogether N^2 numbers of the form $\langle\varphi_a'|X|\varphi_a\rangle$, where N is the dimensionality of the ket space. We may arrange them into a $N \times N$ square matrix where the column and row indices appear as

$$X \doteq \begin{pmatrix} \langle\varphi_1|X|\varphi_1\rangle & \langle\varphi_1|X|\varphi_2\rangle & \dots \\ \langle\varphi_2|X|\varphi_1\rangle & \langle\varphi_2|X|\varphi_2\rangle & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{2.25}$$

where the symbol $\doteq$ stands for "is represented by".
Knowing all (infinite many) matrix elements $\langle\varphi_a'|X|\varphi_a\rangle$ of the operator X is equivalent to the knowledge of the operator itself (in the same way as knowing the 3 componets of a vector in the Euclidean space is sufficient to determine its orientation and length).

In the same way we describe operators by matrices, kets can be described by colum vectors,

$$|\alpha\rangle \doteq \begin{pmatrix} \langle\varphi_1|\alpha\rangle \\ \langle\varphi_2|\alpha\rangle \\ \langle\varphi_3|\alpha\rangle \\ \vdots \end{pmatrix} \tag{2.26}$$

and bras as row vectors,

$$\begin{aligned}\langle\beta| &\doteq (\langle\beta|\varphi_1\rangle \quad \langle\beta|\varphi_2\rangle \quad \langle\beta|\varphi_3\rangle \quad \cdots) = \\ &= (\langle\varphi_1|\beta\rangle^* \quad \langle\varphi_2|\beta\rangle^* \quad \langle\varphi_3|\beta\rangle^* \quad \cdots)\ .\end{aligned} \tag{2.27}$$

Therefore, the action of an operator on a ket can be represented as a matrix multiplication with a vector (link to your linear algebra course).

2.3.2 Basis functions in quantum chemistry

In one of the most frequent approximations used in quantum chemistry, the complex one-electron or even many-electron molecular wavefunctions are described in basis of atom centered functions. These simplified atomic orbitals are often taken to have the form of sums of Gaussians centered on the atoms times a polynomial,P_l, in the electron coordinates relative to that atom:

$$\chi_n(\mathbf{r}) = \sum_l N_l^n \exp\big(-\alpha_l^n(|\mathbf{r}-\mathbf{R}_I^n|^2)\big)P_l(\mathbf{r}-\mathbf{R}_I^n) \quad . \tag{2.28}$$

There are conventional sets of these atomic orbitals that are used, that cover the polynomials up to a certain order with certain choices of "α"; these are called "basis sets" and are given names like "6-31G*" , "TZ2P", and "cc-pVQZ". The general procedure is to pick one of these basis sets, and then to vary the $C's$ and the $D's$ in

$$\begin{aligned}&\Psi_{el}(\mathbf{r}_1, s_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, s_N) = \\ &\quad = \sum_{m_1,m_2,\ldots,m_N} C_{m_1,m_2,\ldots,m_N}|\phi_{m_1}(\mathbf{r}_1,s_1)\phi_{m_2}(\mathbf{r}_2,s_2)\ldots\phi_{m_N}(\mathbf{r}_N,s_N)|\end{aligned} \tag{2.29}$$

with

$$\phi_m(\mathbf{r},s) = \sum_n D_{mn}\chi_n(\mathbf{r}) \otimes s \tag{2.30}$$

to try to find an approximate Ψ_{elec} that solves the Schrödinger equation as closely as possible. If your basis set has a very good overlap with the true wavefunction, you will be able to achieve good accuracy only varying a few $C's$ and $D's$.[3]

2.3.3 The variational principle and the secular equation

The variational problem consists in varying the $C's$ and $D's$ to minimize

$$E[\Psi_{elec}] = E(C_{\ldots}, D_{\ldots}) = \frac{\langle\Psi_{elec}|\hat{H}_{elec}|\Psi_{elec}\rangle}{\langle\Psi_{elec}|\Psi_{elec}\rangle} \tag{2.31}$$

---

[3] More about the specific basis functions used in computational quantum chemistry will follow in Chapter 3.

For any trial wavefunction,Ψ_{elec}^{trial}, the following inequality holds

$$E[\Psi_{elec}] \leq \frac{\langle \Psi_{elec}^{trial} | \hat{H}_{elect} | \Psi_{elec}^{trial} \rangle}{\langle \Psi_{elec}^{trial} | \Psi_{elec}^{trial} \rangle} \quad . \tag{2.32}$$

This is called the variational principle. The evaluation of the integral requires $\mathcal{O}(N_{\text{basis}}^3)$ operations. (Gaussian functions are used because they allow the integrals to be computed analytically.) Typically a basis set might include 15 atomic orbitals for each atom (except H atoms which do not need so many) and you would vary the $(15 * N_{\text{atoms}})^2$ coefficients D_{mn}. The number of possible coefficients C is much larger, something like N_{basis} raised to the $N_{\text{electrons}}$ power, so it is almost always impossible to do anything with a complete expansion. Often people don't bother to vary the $C's$, or only allow a small fraction of the $C's$ to vary independently, to reduce the number of parameters. By allowing the $C's$ to vary, you are allowing to account for the fact that the different electrons are correlated with each other: when one is close to the nucleus the others are likely to be far away.

2.3.4 Linear variational calculus

In variational calculus, stationary states of the energy functional are found within a subspace of the Hilbert space. An important example is **linear variational calculus**, in which the subspace is spanned by a set of basis vectors $|\Xi_m\rangle, m = 1, ..., M$, that we take to be orthonormal. Here we consider the case of fixed atomic orbital expansion coefficients ($D_{...}$) and to-be-optimized Slater expansion coefficients ($C_{...}$) (for example a set of M Slater determinants, $|\Xi_m\rangle = |\phi_{m_1}(\mathbf{r}_1)\phi_{m_2}(\mathbf{r}_2)\ldots\phi_{m_N}(\mathbf{r}_N)|\rangle$). For a state

$$\begin{aligned} \Psi_{el}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) &= \sum_{m_1, m_2, \ldots, m_N} C_{m_1, m_2, \ldots, m_N} |\phi_{m_1}(\mathbf{r}_1)\phi_{m_2}(\mathbf{r}_2)\ldots\phi_{m_N}(\mathbf{r}_N)| \\ &\doteq \sum_{m=1}^{M} c_m |\Xi_m\rangle \end{aligned} \tag{2.33}$$

the energy functional is given by

$$E = \frac{\sum_{p,q=1}^{M} c_p^* c_q \, H_{p,q}}{\sum_{p,q=1}^{M} c_p^* c_q \, \delta_{p,q}} \tag{2.34}$$

with

$$H_{p,q} = \langle \Xi_p | \hat{H}_{elec} | \Xi_q \rangle \tag{2.35}$$

The **stationary states** follow from the condition that the derivative of this functional with respect to the c_p vanishes, which leads to

$$\sum_{q=1}^{M} (H_{p,q} - E\,\delta_{p,q})\, c_q = 0, \quad \text{for} \quad p = 1, \ldots, M\,. \tag{2.36}$$

Equation (2.36) is an eigenvalue problem which can be written in the matrix notation

$$\mathbf{H}\,\mathbf{C} = E\,\mathbf{C} \tag{2.37}$$

This is the Schrödinger equation formulated for a finite, orthonormal basis. Although in principle it is possible to use nonlinear parameterizations of the wave function, linear parameterizations are used in the large majority of cases because of the simplicity of the resulting method, allowing for numerical matrix diagonalization techniques. **The lowest eigenvalue of (2.37) is always higher than or equal to the exact ground state energy, as the ground state is the minimal value assumed by the energy functional in the full Hilbert space**. If we restrict ourselves to a part of this space, then the minimum value of the energy functional must always be higher than or equal to the ground state of the full Hilbert space. Including more basis functions into our set, the subspace becomes larger, and consequently the minimum of the energy functional will decrease (or stay the same). For the specific case of linear variational calculus, this result can be generalized to higher stationary states: they are always higher than the equivalent solution to the full problem, but approximate the latter better with increasing basis set size.

Because the computer time needed for matrix diagonalization scales with the third power of the linear matrix size (it is called a $\mathcal{O}(M^3)$ process), the basis should be kept as small as possible. Therefore, it must be chosen carefully: it should be possible to approximate the solutions to the full problem with a small number of basis functions [4].

In the case in which the basis consists of **nonorthonormal basis functions**, as is often the case in practical calculations, we must reformulate (2.37), taking care of the fact that the *overlap matrix* $\mathbf{S}$, whose elements S_{pq} are given by

$$S_{p,q} = \langle \Xi_p | \Xi_q \rangle \tag{2.38}$$

is not the unit matrix. This means that in Eq. (2.34) the matrix elements δ_{pq} of the unit matrix, occurring in the denominator, have to be replaced by S_{pq}, and we obtain (for the derivation see the next section)

$$\mathbf{H}\,\mathbf{C} = E\,\mathbf{S}\,\mathbf{C}\,. \tag{2.39}$$

This looks like an ordinary eigenvalue equation, the only difference being the matrix $\mathbf{S}$ in the right hand side. It is called a generalized eigenvalue equation and there exist computer programs for solving such a problem.

---

[4] The fact that the basis in (continuous) variational calculus can be chosen so much smaller than the number of grid points in a finite difference approach implies that even though the latter can be solved using special $\mathcal{O}(N)$ methods for sparse systems, they are still far less efficient than variational methods with continuous basis functions in most cases. This is the reason why, in most electronic structure calculations, variational calculus with continuous basis functions is used to solve the Schrödinger equation.

2.4 Overview of possible approximate solutions of the electronic Schrödinger equation

The most commonly used approximate methods for the solution of the electronic molecular Schrödinger equation are:

- Semi-empirical (MNDO, AM1, PM3, etc.): use a single Slater determinant (only one C is equal 1 while all the others are set to 0). Vary the coefficients D, but just use empirical estimates rather than the true integrals. *Very cheap, but only accurate for molecule similar to those used to develop the empirical estimates.*

- DFT (B3LYP, BLYP, PW91, etc.): slightly empirical, but much more reliable than semi-empirical methods. CPU: cheap, same as HF $\mathcal{O}(N^3)$. Errors $\sim$ 4 kcal/mole (comparable accuracy to MP2 but much cheaper). Preferred method for geometries, second derivatives, transition-metal containing systems.

- HF (Hartree-Fock, SCF): only one many-electrons Slater determinant is used. Vary the D's. All terms calculated 'ab-initio' within the mean field approximation, no empirical parameters.
CPU: cheap $\mathcal{O}(N^3)$ *errors* ~ 15 *kcal/mol.*

- MP2, MP4 (Moller-Plesset, MBPT): Vary the D's first, then set the C's to the values given by perturbation theory (you don't freely vary these C's, saving CPU).
MP2: medium CPU: $\mathcal{O}(N^5)$*, errors* ~ 5 *kcal/mol.*

- CI, CISD, QCISD (Configuration Interaction): Vary the coefficients D first, freeze them, then vary a lot of the coefficients C.
Expensive. Not used much anymore, CCSD is preferred.

- MCSCF, CASSCF: vary a finite set of C's and all the D's simultaneously. *Expensive. Good for understanding cases where several electronic states have comparable energies. User expertise required to select which C's to vary.*

- CAS-PT2: Determine the D's and some C's by CASSCF, then determine more C's by perturbation theory.
Not much more expensive than CASSCF. Sometimes very good, but not reliable.

- MRCI (multi reference CI): Determine the D's and some C's by CASSCF or MCSCF, freeze these, then allow many of the C's to vary.
Super expensive. Very high accuracy for small systems.

- CCSD, CCSD(T) (Coupled Cluster): Vary the D's, fix them, then vary a lot of the C's, but constraining certain relationships between the C's.

This allows you to effectively use a longer expansion without increasing the number of adjustable parameters so much. The constraints force the solution to be "size-consistent", i.e. two molecules calculated simultaneously have exactly the same energy as two molecules calculated separately.
Expensive. Often very accurate.
- Extrapolations ("Composite Methods"): G2, G3, CBS-q, CBS-Q, CBS-QB3, CBS-RAD Run a series of the above calculations with different size basis sets, following some recipe. The results from all these calculations are extrapolated to an estimate of the true potential $V(R)$. These methods give excellent accuracy in less CPU time than CCSD or MRCI. However, the multiple steps involved provide many opportunities for something to go wrong.*Accuracy: usually 1-2 kcal/mol.*

Some Practical Warnings

1) The optimization (SCF/HF/DFT/CASSCF/MRSCF) problem required to solve for the $D's$ is nonlinear and has multiple solutions, only one of which is the one you want (usually you want the lowest energy solution). So you may end up converging to a wavefunction which is qualitatively incorrect, perhaps it corresponds to an electronically excited state.

2) Most of the quantum chemistry methods have problems (convergence, accuracy) with systems where there are low-lying electronic states (close to the ground state). In these cases, sometimes the numbers computed are completely nuts, other times they are subtly wrong. This is particularly a problem for transition states and where there are several lone pair electrons in the system. If you must study these systems, get expert assistance.

3) Many molecules have multiple geometrical conformations (local minima in $V(\mathbf{R})$), and sometimes there are multiple saddle points that might be confused with the transition state (TS). Look at your structures, if they are not what you expected, investigate. Also, it is worth some effort to make sure your initial guess at the molecular geometry is quite good, otherwise the geometry-optimization algorithm may get lost and waste a lot of CPU time to no avail. If you are having troubles, you can constrain some of the coordinates to make things easier for the optimizer.

4) For radicals and other open-shell systems, compare your computed solutions $\langle S^2 \rangle$ with the theoretical value $S(S+1)$. If your number is way off, chances are you have other problems as well. Sometimes you can use "restricted" methods like ROHF and RMP2, or spin-projection methods to fix this "spin-contamination" problem.

5) Every method runs into problems sometimes, and sometimes they are quite subtle. It is a good idea to double check your calculation with another calculation done using a very different method. If they both agree you can be pretty confident that your result is real.

Chapter 3

Basis functions in quantum chemistry

This chapter is adapted from Chapter 5 of Jensen's book: F. Jensen, 'Introduction to Computational Chemistry', Wiley.

In the derivation in the previous chapter, we have introduced the concept of basis function for the expansion of the one-electron molecular orbitals used for the generation of the many-electrons wave functions (Slater determinants or linear combination of Slater determinants).
There we derived the following expansion (eq. 2.17):

$$\phi_m(\mathbf{r}, s) = \sum_n D_{mn}\chi_n(\mathbf{r}) \tag{3.1}$$

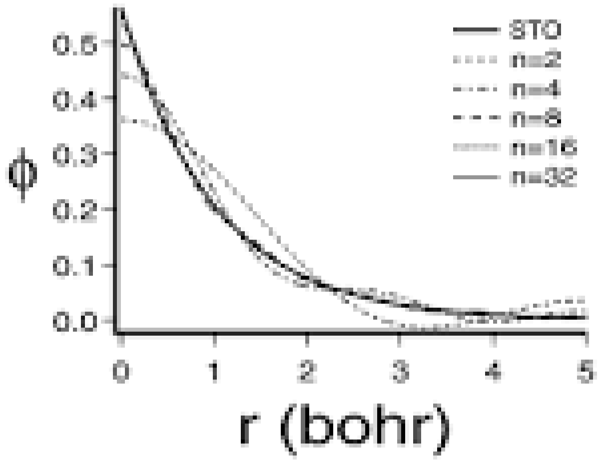
(where χ_n is an atom centered basis function and the spin dependent part of the wavefunctions is left out).
In this chapter, we introduce the different basis functions,χ_n commonly used in computational quantum chemistry.

Finiteness of Basis Sets: Approximations

One of the approximations inherent in essentially all *ab initio* methods is the introduction of a finite basis set. Expanding an unknown function, such as a molecular orbital, in a set of known functions is not an approximation, if the basis is complete. However, a complete basis means that an infinite number of functions must be used, which is impossible in actual calculations. An unknown MO can be thought of as a function in the infinite coordinate system spanned by the complete basis set. When a finite basis is used, only the components of the MO along those coordinate axes corresponding to the selected basis can be represented. The smaller the basis, the poorer the representation. The type of basis functions used also influences the accuracy. The better a single basis function is able to reproduce the unknown function, the fewer basis functions are necessary for achieving a given level of accuracy. Knowing that the computational effort of ab initio methods scales formally at least as M^4, it is of course of prime importance to make the basis set as small as possible without compromising the accuracy.

3.1 Slater and Gaussian Type Orbitals

There are two types of basis functions (also called Atomic Orbitals, AO, although in general they are not solutions to an atomic Schrödinger equation) commonly used in electronic structure calculations: Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO).
A procedure that has come into wide use is to fit a Slater-type orbital (STO) to a linear combination of $n = 1, 2, 3, \ldots$ primitive Gaussian functions. This is the STO-nG procedure. In particular, STO-3G basis sets are often used in polyatomic calculations, in preference to evaluating integrals with Slater functions.



1

Figure 3.1: Comparison of Slater function with Gaussian function: least squares fits of a $1s$ Slater function ($\zeta = 1.0$) by a n GTOs

1. Slater type orbitals have the functional form

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = N\, Y_{l,m}(\theta,\varphi)\, r^{n-1}\, e^{-\zeta r} \tag{3.2}$$

N is a normalization constant and $Y_{l,m}$ are the usual spherical harmonic functions. The exponential dependence on the distance between the nucleus and the electron mirrors the exact decay behavior of the orbitals for the hydrogen atom. However, since STOs do not have any radial nodes, nodes in the radial part are introduced by making linear combinations of STOs. The exponential dependence ensures a fairly rapid convergence with increasing number of functions, however, the calculation of three- and four-centre two electron integrals cannot be performed analytically. STOs are primarily used for atomic and diatomic systems where high accuracy is required, and in semi-empirical methods where all three- and four- center integrals are neglected.

2. Gaussian type orbitals can be written in terms of polar or Cartesian coordinates

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = N\,Y_{l,m}(\theta,\varphi)\,r^{2n-2-l}\,e^{-\zeta r^2} \tag{3.3}$$

$$\chi_{\zeta,l_x,l_y,l_z}(x,y,z) = N\,x^{l_x}\,y^{l_y}\,y^{l_z}\,e^{-\zeta r^2} \tag{3.4}$$

where the sum of l_x, l_y and l_z determines the type of orbital (for example $l_x + l_y + l_z = 1$ is a p-orbital)[1].

3. Comparison between STO and GTO

i. The r^2 dependence in the exponent makes the GTOs inferior to the STOs in two aspects. At the nucleus the GTO has zero slope, in contrast to the STO which has a "**cusp**" (discontinuous derivative), and GTOs have problems representing the proper behavior near the nucleus.

ii. The other problem is that the GTO falls off too rapidly far from the nucleus compared with an STO, and the "tail" of the wave function is consequently represented poorly.

iii. Both STOs and GTOs can be chosen to form a complete basis, but the above considerations indicate that more GTOs are necessary for achieving a certain accuracy compared with STOs. A rough guideline says that three times as many GTOs as STOs are required for reaching a given level of accuracy. The increase in number of basis functions, however, is more than compensated for by the ease by which the required

---

[1] Although a GTO appears similar in the two sets of coordinates, there is a subtle difference. A d-type GTO written in terms of the spherical functions has five components $(Y_{2,2}, Y_{2,1}, Y_{2,0}, Y_{2,-1}, Y_{2,-2})$, but there appear to be six components in the Cartesian coordinates $(x_2, y_2, z_2, xy, xz, yz)$. The latter six functions, however, may be transformed to the five spherical d-functions and one additional s-function $(x^2 + y^2 + z^2)$. Similarly, there are 10 Cartesian "f-functions" which may be transformed into seven spherical f-functions and one set of spherical p-functions. Modern programs for evaluating two-electron integrals are geared to Cartesian coordinates, and they generate pure spherical d-functions by transforming the six Cartesian components to the five spherical functions. When only one d-function is present per atom the saving by removing the extra s-function is small, but if many d-functions and/or higher angular moment functions (f-, g-, h- etc. functions) are present, the saving can be substantial. Furthermore, the use of only the spherical components reduces the problems of linear dependence for large basis sets, as discussed below.

integrals can be calculated. In terms of computational efficiency, GTOs are therefore preferred, and used almost universally as basis functions in electronic structure calculations.

3.2 Classification of Basis Sets

Having decided on the type of basis function (STO/GTO) and their location (nuclei), the most important factor is the number of functions to be used. The smallest number of functions possible is a **minimum basis set**. Only enough atomic orbital functions are employed to contain all the electrons of the neutral atom(s).

3.2.1 Minimum basis sets. Examples

For hydrogen (and helium) this means a single s-function. For the first row in the periodic table it means two s-functions ($1s$ and $2s$) and one set of p-functions ($2p_x$, $2p_y$ and $2p_z$). Lithium and beryllium formally only require two s-functions, but a set of p-functions is usually also added. For the second row elements, three s-functions ($1s$, $2s$ and $3s$) and two sets of p-functions ($2p$ and $3p$) are used.

3.2.2 Improvements

1. The first improvement in the basis sets is a doubling of all basis functions, producing a **Double Zeta** (DZ) type basis. The term zeta stems from the fact that the exponent of STO basis functions is often denoted by the greek letter ζ.
   A DZ basis thus employs two s-functions for hydrogen (1s and 1s'), four s-functions (1s, 1s', 2s and 2s') and two p-functions (2p and 2p') for first row elements, and six s-functions and four p-functions for second row elements. Doubling the number of basis functions allows for a much better description of the fact that the electron distribution in molecules can differ significantly from the one in the atoms and the chemical bond may introduce directionalities which can not be described by a minimal basis.
   The chemical bonding occurs between valence orbitals. Doubling the $1s$-functions in for example carbon allows for a better description of the $1s$-electrons. However, the 1s orbital is essentially independent of the chemical environment, being very close to the atomic case. A variation of the DZ type basis only doubles the number of valence orbitals, producing a **split valence basis**. [2].

---

[2] In actual calculations a doubling of the core orbitals would rarely be considered, and the term DZ basis is also used for split valence basis sets (or sometimes VDZ, for valence double zeta)

2. The next step up in basis set size is a **Triple Zeta** (TZ) basis. Such a basis contains three times as many functions as the minimum basis, i.e. six s-functions and three p-functions for the first row elements. Some of the core orbitals may again be saved by only splitting the valence, producing a triple zeta split valence basis set. The names **Quadruple Zeta** (QZ) and **Quintuple Zeta** (5Z, not QZ) for the next levels of basis sets are also used, but large sets are often given explicitly in terms of the number of basis functions of each type.

3. In most cases higher angular momentum functions are also important, these are denoted **polarization functions**. Consider for example a C-H bond which is primarily described by the hydrogen s-orbital(s) and the carbon s- and p_z-orbitals. It is clear that the electron distribution along the bond will be different than that perpendicular to the bond. If only s-functions are present on the hydrogen, this cannot be described. However, if a set of p-orbitals is added to the hydrogen, the p component can be used for improving the description of the H-C bond. The p-orbital introduces a polarization of the s-orbital(s). Similarly, d-orbitals can be used for polarizing p-orbitals, f-orbitals for polarizing d-orbitals etc. Once a p-orbital has been added to a hydrogen s-orbital, it may be argued that the p-orbital now should be polarized by adding a d-orbital, which should be polarized by an f-orbital, etc. For single determinant wave functions, where electron correlation is not considered, the first set of polarization functions (i.e. p-functions for hydrogen and d-functions for heavy atoms) is by far the most important, and will in general describe all the important charge polarization effects.
   Adding a single set of polarization functions (p-functions on hydrogens and d-functions on heavy atoms) to the DZ basis forms a Double Zeta plus Polarization (DZP) type basis [3]. Similarly to the sp-basis sets, multiple sets of polarization functions with different exponents may be added. If two sets of polarization functions are added to a TZ sp-basis, a Triple Zeta plus Double Polarization (TZ2P) type basis is obtained. For larger basis sets with many polarization functions the explicit composition in terms of number and types of functions is usually given. At the HF level there is usually little gained by expanding the basis set beyond TZ2P, and even a DZP type basis set usually gives "good" results (compared to the HF limit).

---

[3] There is a variation where polarization functions are only added to non-hydrogen atoms. This does not mean that polarization functions are not important on hydrogens. However, hydrogens often have a "passive" role, sitting at the end of bonds which does not take an active part in the property of interest. The errors introduced by not including hydrogen polarization functions are often rather constant and, as the interest is usually in energy differences, they tend to cancel out. As hydrogens often account for a large number of atoms in the system, a saving of three basis functions for each hydrogen is significant. If hydrogens play an important role in the property of interest, it is of course not a good idea to neglect polarization functions on hydrogens.

3.3 Basis set balance

In principle many sets of polarization functions may be added to a small sp-basis. This is not a good idea. If an insufficient number of sp-functions bas been chosen for describing the fundamental electron distribution, the optimization procedure used in obtaining the wave function (and possibly also the geometry) may try to compensate for inadequacies in the sp-basis by using higher angular momentum functions, producing artefacts. A rule of thumb says that the number of functions of a given type should at most be one less than the type with one lower angular momentum. A 3s2p1d basis is balanced, but a 3s2p2d2f1g basis is too heavily polarized.

Another aspect of basis set balance is the occasional use of **mixed basis sets**, for example a DZP quality on the atoms in the "interesting" part of the molecule and a minimum basis for the "spectator" atoms. Another example would be addition of polarization functions for only a few hydrogens which are located "near" the reactive part of the system. For a large molecule this may lead to a substantial saving in the number of basis functions. It should be noted that this may bias the results and can create artefacts. For example, a calculation on the H_2 molecule with a minimum basis at one end and a DZ basis at the other end will predict that H_2 has a dipole moment, since the variational principle will preferentially place the electrons near the center with the most basis functions. The majority of calculations are therefore performed with basis sets of the same quality (minimum, DZP, TZ2P, . . .) on all atoms, possibly cutting polarization and/or diffuse (small exponent) functions on hydrogens.

Except for very small systems it is impractical to saturate the basis set so that the absolute error in the energy is reduced below chemical accuracy, for example 1 kcal/ mol. The important point in choosing a balanced basis set is to keep the error as constant as possible. The use of mixed basis sets should therefore only be done after careful consideration. Furthermore, the use of small basis sets for systems containing elements with substantially different numbers of valence electrons (like LiF) may produce artefacts.

3.4 How do we choose the exponents in the basis functions?

The values for s- and p-functions are typically determined by performing variational HF calculations for atoms, using the exponents as variational parameters. The exponent values which give the lowest energy are the "best", at least for the atom. In some cases the optimum exponents are chosen on the basis of minimizing the energy of a wave function which includes electron correlation. The HF procedure cannot be used for determining exponents of polarization functions for atoms. By definition these functions are unoccupied in atoms, and therefore make no contribution to the energy. Suitable polarization exponents may be chosen by performing variational calculations on molecular systems (where the HF energy does depend on polarization functions) or on atoms with correlated wave functions. Since the main function of higher angular momentum functions is to recover electron correlation, the latter approach is usually preferred. Often only the optimum exponent is determined for a single polarization function, and multiple polarization functions are generated by splitting the exponents symmetrically around the optimum value for a single function. The splitting factor is typically taken in the range 2-4. For example if a single d-function for carbon has an exponent value of 0.8, two polarization functions may be assigned with exponents of 0.4 and 1.6 (splitting factor of 4).

3.5 Contracted Basis functions

One disadvantage of all **energy optimized basis sets** is the fact that they primarily depend on the wave function in the region of the inner shell electrons. The 1s-electrons account for a large part of the total energy, and minimizing the energy will tend to make the basis set optimal for the core electrons, and less than optimal for the valence electrons. However, chemistry is mainly dependent on the valence electrons. Furthermore, many properties (for example polarizability) depend mainly on the wave function "tail" (far from the nucleus), which energetically is unimportant. An energy optimized basis set which gives a good description of the outer part of the wave function needs therefore to be very large, with the majority of the functions being used to describe the 1s-electrons with an accuracy comparable to that for the outer electrons in an energetic sense. This is not the most efficient way of designing basis sets for describing the outer part of the wave function. Instead energy optimized basis sets are usually augmented explicitly with **diffuse functions** (basis functions with small exponents). Diffuse functions are needed whenever loosely bound electrons are present (for example in anions or excited states) or when the property of interest is dependent on the wave function tail (for example polarizability).
The fact that many basis functions go into describing the energetically important, but chemically unimportant, core electrons is the foundation for **contracted basis sets**.

1

An example. The carbon atom Consider for example a basis set consisting of 10 s-functions (and some p-functions) for carbon. Having optimized these 10 exponents by variational calculations on the carbon atom, maybe six of the 10 functions are found primarily to be used for describing the 1s orbital, and two of the four remaining describe the "inner" part of the 2s-orbital. The important chemical region is the outer valence. Out of the 10 functions, only two are actually used for describing the chemically interesting phenomena. Considering that the computational cost increases as the fourth power (or higher) of the number of basis functions, this is very inefficient. As the core orbitals change very little depending on the chemical bonding situation, the MO expansion coefficients in front of these inner basis functions also change very little. The majority of the computational effort is therefore spent describing the chemically uninteresting part of the wave function, which furthermore is almost constant. Consider now making the variational coefficients in front of the inner basis functions constant, i.e. they are no longer parameters to be determined by the variational principle. The 1s-orbital is thus described by a fixed linear combination of say six basis functions. Similarly the remaining four basis functions may be contracted into only two functions, for example by fixing the coefficient in front of the inner three functions. In doing this the number of basis functions to be handled by the variational procedure has been reduced from 10 to three.

Combining the full set of basis functions, known as the **primitive GTOs (PGTOs)**, into a smaller set of functions by forming fixed linear combinations is known as **basis set contraction**, and the resulting functions are called **contracted GTOs (CGTOs)**

$$\chi(\text{CGTO}) = \sum_{i}^{k} a_i\, \chi_i(\text{PGTO}) \tag{3.5}$$

The previously introduced acronyms DZP, TZ2P etc., refer to the number of contracted basis functions. Contraction is especially useful for orbitals describing the inner (core) electrons, since they require a relatively large number of functions for representing the wave function cusp near the nucleus, and furthermore are largely independent of the environment. Contracting a basis set will always increase the energy, since it is a restriction of the number of variational parameters, and makes the basis set less flexible, but will also reduce the computational cost significantly. The decision is thus how much loss in accuracy is acceptable compared to the gain in computational efficiency.

3.5.1 The degree of contraction

The degree of contraction is the number of PGTOs entering the CGTO, typically varying between 1 and 10. The specification of a basis set in terms of primitive and contracted functions is given by the notation

$$(\text{10s4p1d/4s1p}) \longrightarrow [\text{3s2p1d/2s1p}] \quad . \tag{3.6}$$

The basis in parentheses is the number of primitives with heavy atoms (first row elements) before the slash and hydrogen after. The basis in the square brackets is the number of contracted functions. Note that this does not tell how the contraction is done, it only indicates the size of the final basis (and thereby the size of the variational problem in HF calculations).

3.6 Example of Contracted Basis Sets; Pople Style Basis Sets

There are many different contracted basis sets available in the literature or built into programs, and the average user usually only needs to select a suitable quality basis for the calculation. For short description of some basis sets which often are used in routine calculations (see for instance the book of Frank Jensen, *Introduction to Computational Chemistry*, Wiley, 2002. Chapter 5).

STO-*n*G basis sets n PGTOs fitted to a 1 STO. This is a minimum type basis where the exponents of the PGTO are determined by fitting to the STO, rather than optimizing them by a variational procedure. Although basis sets with $n = 2 - 6$ have been derived, it has been found that using more than three PGTOs to represent the STO gives little improvement, and the STO-3G basis is a widely used minimum basis. This type of basis set has been determined for many elements of the periodic table. The designation of the carbon/hydrogen STO-3G basis is (6s3p/3s) $\longrightarrow$ [2s1p/1s].

***k*-*nlm*G basis sets** These basis sets have been designed by Pople and coworkers, and are of the split valence type, with the k in front of the dash indicating how many PGTOs are used for representing the core orbitals. The nlm after the dash indicate both how many functions the valence orbitals are split into, and how many PGTOs are used for their representation. Two values (e.g. nl) indicate a split valence, while three values (e.g. nlm) indicate a triple split valence. The values before the G (for Gaussian) indicate the s- and p-functions in the basis; the polarization functions are placed after the G. This type of basis sets has the further restriction that the same exponent is used for both the s- and p-functions in the valence. This increases the computational efficiency, but of course decreases the flexibility of the basis set. The exponents in the PGTO have been optimized by variational procedures.

3-21G This is a split valence basis, where the core orbitals are a contraction of three PGTOs, the inner part of the valence orbitals is a contraction of two PGTOs and the outer part of the valence is represented by one PGTO. The designation of the carbon/hydrogen 3-21G basis is (6s3p/3s) $\longrightarrow$ [3s2p/2s]. Note that the 3-21G basis contains the same number of primitive GTOs as the STO-3G, however, it is much more flexible as there are twice as many valence functions which can combine freely to make MOs.

6-31G This is also a split valence basis, where the core orbitals are a contraction of six PGTOs, the inner part of the valence orbitals is a contraction of three PGTOs and the outer part of the valence represented by one PGTO. The designation of the carbon/hydrogen 6-31G basis is (10s4p/4s) $\longrightarrow$ [3s2p/2s]. In terms of contracted basis functions it contains the same number as 3-21G, but the representation of each functions is better since more PGTOs are used.

6-311G This is a triple zeta split valence basis, where the core orbitals are a contraction of six PGTOs and the valence split into three functions, represented by three, one, and one PGTOs, respectively.

To each of these basis sets one can add **diffuse** and/or **polarization functions**.

- **Diffuse functions** are normally s- and p-functions and consequently go before the G. They are denoted by + or ++, with the first + indicating one set of diffuse s- and p-functions on heavy atoms, and the second + indicating that a diffuse s-function is also added to hydrogens. The arguments for adding only diffuse functions on non-hydrogen atoms is the same as that for adding only polarization functions on non-hydrogens.

- **Polarization functions** are indicated after the G, with a separate designation for heavy atoms and hydrogens. The *6-31+G(d)* is a split valence basis with one set of diffuse sp-functions on heavy atoms only and a single d-type polarization function on heavy atoms.
  A *6-311++G(2df,2pd)* is similarly a triple zeta split valence with additional diffuse sp-functions, and two d- and one f-functions on heavy atoms and diffuse s- and two p- and one d-functions on hydrogens. The largest standard Pople style basis set is *6-311 ++G(3df, 3pd).* These types of basis sets have been derived for hydrogen and the first row elements, and same of the basis sets have also been derived for second and higher row elements.
  If only one set of polarization functions is used, an alternative notation in terms of * is also widely used. The *6-31G** basis is identical to *6-31G(d)*, and **6-31G**** is identical to *6-31G(d,p).* A special note should be made for the *3-21G** basis. The 3-21G basis is basicly too small to support polarization functions (it becomes unbalanced). However, the *3-21G* basis by itself performs poorly for hypervalent molecules, such as sulfoxides and sulfones. This can be substantially improved by adding a set of d-functions. The *3-21G** basis has only d-functions on second row elements (it is sometimes denoted *3-21G(*)* to indicate this), and should not be considered a polarized basis. Rather, the addition of a set of d-functions should be considered an ad hoc repair of a known flaw.

Chapter 4

An Introduction to Hartree Fock Theory

Adapted from C. D. Sherrill's notes: "An Introduction to Hartree-Fock Molecular Orbital Theory".

4.1 Introduction

Hartree-Fock theory is fundamental to much of electronic structure theory. It is the basis of molecular orbital (MO) theory, which posits that each electron's motion can be described by a single-particle function (orbital) which does not depend explicitly on the instantaneous motions of the other electrons. Some of you have probably learned about (and maybe even solved problems with) Hückel MO theory, which takes Hartree-Fock MO theory as an implicit foundation and throws away most of the terms to make it tractable for simple calculations. The ubiquity of orbital concepts in chemistry is a testimony to the predictive power and intuitive appeal of Hartree-Fock MO theory. However, it is important to remember that these orbitals are mathematical constructs which only approximate reality. Only for the hydrogen atom (or other one electron systems, like He^+) are orbitals exact eigenfunctions of the full electronic Hamiltonian. As long as we are content to consider molecules near their equilibrium geometry, Hartree-Fock theory often provides a good starting point for more elaborate theoretical methods which are better approximations to the electronic Schrödinger equation (e.g., many-body perturbation theory, single-reference configuration interaction). So, how do we calculate molecular orbitals using Hartree-Fock theory? This is the subject of these notes; we will explain Hartree-Fock theory at an introductory level.

4.2 What Problem Are We Solving?

It is always important to remember the context of a theory. Hartree-Fock theory was developed to solve the electronic Schrödinger equation that results from the time-independent Schrödinger equation after invoking the Born-Oppenheimer approximation. In atomic units, and with r denoting electronic and R denoting nuclear degrees of freedom, the electronic Schrödinger equation is

$$\Big[-\frac{1}{2}\sum_i \nabla_i^2 - \sum_{I,i}\frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{I>J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} + \sum_{i>j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}\Big]\Psi(\mathbf{r};\mathbf{R}) = E_{el}\Psi(\mathbf{r};\mathbf{R}) \tag{4.1}$$

or, in a more compact notation,

$$\big[\hat{T}_e(\mathbf{r}) + \hat{V}_{eN}(\mathbf{r},\mathbf{R}) + \hat{V}_{NN}(\mathbf{R}) + \hat{V}_{ee}(\mathbf{r})\big]\Psi(\mathbf{r};\mathbf{R}) = E_{el}\Psi(\mathbf{r};\mathbf{R}) \tag{4.2}$$

Recall from the Born-Oppenheimer approximation that E_{el} (plus or minus $\hat{V}_{NN}(\mathbf{R})$, which we include here) will give us the potential energy experienced by the nuclei. In other words, $E_{el}(\mathbf{R})$ is the potential energy surface (from which we can get, for example, the equilibrium geometry and the vibrational frequencies). That's one good reason why we want to solve the electronic Schrödinger equation. The other is that the electronic wavefunction $\Psi(\mathbf{r};\mathbf{R})$ contains lots of useful information about molecular properties such as dipole (and multipole) moments, polarizability, etc.

4.3 The many-electron wavefunction: the Slater determinant

The basic idea of Hartree-Fock theory is as follows. We know how to solve the electronic problem for the simplest atom, hydrogen, which has only one electron. We imagine that perhaps if we added another electron to hydrogen, to obtain H^-, then maybe it might be reasonable to start off pretending that the electrons don't interact with each other (i.e., that $\hat{V}_{ee} = 0$). If that was true, then the Hamiltonian would be separable, and the total electronic wavefunction $\Psi(\mathbf{r}_1, \mathbf{r}_2)$ describing the motions of the two electrons would just be the product of two hydrogen atom wavefunctions (orbitals), $\Psi_H(\mathbf{r}_1)\,\Psi_H(\mathbf{r}_2)$ (you should be able to prove this easily).

However, we have already mentioned in Chapter 2 that a correct many-electron wavefunction must fulfill the *antisymmetry principle* together with the principle of *indistinguishability* of the electrons. The solution of this problem is the Slater determinant made of one-electron molecular orbitals (MOs). In the following we will derive a theory, the Hartree Fock theory, for the calculation of such single-electron MOs.

4.4 Simplified Notation for the Hamiltonian

Hartree-Fock theory is based on the assumption that the many-electron wavefunction of the system can be described by a single Slater determinant made of one-electron molecular orbitals. Let's re-examine the Hamiltonian to make it look as simple as possible. In the process, we will bury some complexity that would have to be taken care of later (in the evaluation of integrals).

We will define a one-electron operator $\hat{h}$ as follows

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \tag{4.3}$$

and a two-electron operator $\hat{v}(i,j)$ as

$$\hat{v}(i,j) = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{4.4}$$

Now we can write the electronic Hamiltonian much more simply, as

$$\hat{H}_{el} = \sum_i \hat{h}(i) + \sum_{i<j} \hat{v}(i,j) + V_{NN} \tag{4.5}$$

Since V_{NN} is just a constant for the fixed set of nuclear coordinates $\{\mathbf{R}\}$, it doesn't change the eigenfunctions, and only shifts the eigenvalues.

4.5 Energy Expression

Now that we have a form for the wavefunction and a simplified notation for the Hamiltonian, we have a good starting point to tackle the problem. Still, how do we obtain the molecular orbitals?

We state that the Hartree-Fock wavefunction will have the form of a Slater determinant, and that the energy will be given by the usual quantum mechanical expression (assuming the wavefunction is normalized):

$$E_{el} = \langle \Psi | \hat{H}_{el} | \Psi \rangle. \tag{4.6}$$

where $\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ is a Slater determinant, $|\phi_1, \phi_2, \ldots, \phi_N|$.

We can employ the variational theorem, which states that the energy is always an upper bound to the true energy. Hence, we can obtain better approximate wavefunctions Ψ by varying their parameters until we minimize the energy within the given functional space. Hence, the correct molecular orbitals in the Slater determinant, $\{\phi_i\}_{i=1}^N$, are those which minimize the elec-tronic energy E_{el}! As we have seen in Chapter 3, the molecular orbitals can be obtained numerically using integration over a grid, or (much more commonly) as a linear combination of a set of given basis functions (so-called "atomic orbital" basis functions, usually atom-centered Gaussian type functions).

Now, using some tricks (you can find a full derivation in Appendix C), we can re-write the Hartree-Fock energy E_{el} in terms of integrals of the one- and two-electron operators:

$$E_{HF} = \sum_i \langle i|h|i\rangle + \frac{1}{2}\sum_{ij}[ii|jj] - [ij|ji], \tag{4.7}$$

where the one electron integral is ($\mathbf{x}_i = (\mathbf{r}_i, s_i)$) [1]

$$\langle i|h|j\rangle = \int d\mathbf{x}_1 \phi_i^*(\mathbf{x}_1) h(\mathbf{r}_1) \phi_j(\mathbf{x}_1) \tag{4.8}$$

and a two-electron integral (Chemists' notation) is

$$[ij|kl] = \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i^*(\mathbf{x}_1)\phi_j(\mathbf{x}_1)\frac{1}{r_{12}}\phi_k^*(\mathbf{x}_2)\phi_l(\mathbf{x}_2). \tag{4.9}$$

There exist efficient computer algorithms for computing such one- and two-electron integrals.

4.6 The Hartree-Fock Equations

Again, the Hartree-Fock method seeks to approximately solve the electronic Schrödinger equation, and it assumes that the wavefunction can be approximated by a single Slater determinant made up of one spin orbital per electron. According to the variational theorem we know that the Slater determinant with the lowest energy is as close as we can get to the true wavefunction for the assumed functional form of a single Slater determinant. The Hartree-Fock method determines the set of spin orbitals which minimize the energy and give us this "best single determinant."

So, we need to minimize the Hartree-Fock energy expression with respect to changes in the orbitals $\phi_i \longrightarrow \phi_i + \delta\phi_i$. We have also been assuming that the orbitals ϕ are orthonormal, and we want to ensure that our variational procedure leaves them orthonormal. We can accomplish this by Lagrange's method of undetermined multipliers, where we employ a functional $\mathcal{L}$ defined as

$$\mathcal{L}[\{\phi_i\}] = E_{HF}[\{\phi_i\}] - \sum_{ij}\epsilon_{ij}(\langle i|j\rangle - \delta_{ij}) \tag{4.10}$$

where ϵ_{ij} are the undetermined Lagrange multipliers and $\langle i|j\rangle$ is the overlap between spin orbitals i and j, i.e.,

$$\langle i|j\rangle = \int \phi_i^*(\mathbf{x})\phi_j(\mathbf{x}) d\mathbf{x}. \tag{4.11}$$

---

[1] $\int d\mathbf{x}_i = \sum_{s\in\{\alpha,\beta\}} \int d\mathbf{r}_i$

Setting the first variation $\delta\mathcal{L} = 0$, and working through some algebra (shown explicitly in Appendix D), we eventually arrive at the Hartree-Fock equations defining the orbitals:

$$h(\mathbf{x}_1)\phi_i(\mathbf{x}_1) + \sum_{j\neq i}\left[\int d\mathbf{x}_2|\phi_j(\mathbf{x}_2)|^2 r_{12}^{-1}\right]\phi_i(\mathbf{x}_1) - \tag{4.12}$$

$$-\sum_{j\neq i}\left[\int d\mathbf{x}_2\phi_j^*(\mathbf{x}_2)\phi_i(\mathbf{x}_2)r_{12}^{-1}\right]\phi_j(\mathbf{x}_1) = \epsilon_i\phi_i(\mathbf{x}_1), \tag{4.13}$$

where ϵ_i is the energy eigenvalue associated with orbital ϕ_i.

The Hartree-Fock equations can be solved numerically (exact Hartree-Fock), or they can be solved in the space spanned by a set of basis functions (Hartree-Fock-Roothan equations). In either case, note that the solutions depend on the orbitals. Hence, we need to guess some initial orbitals and then refine our guesses iteratively. For this reason, Hartree-Fock is called a self-consistent-field (SCF) approach.

The first term above in square brackets,

$$\sum_{j\neq i}\left[\int d\mathbf{x}_2|\phi_j(\mathbf{x}_2)|^2 r_{12}^{-1}\right]\phi_i(\mathbf{x}_1), \tag{4.14}$$

($r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$) gives the Coulomb interaction of an electron in spin orbital ϕ_i with the average charge distribution of the other electrons. Here we see in what sense Hartree-Fock is a "mean field" theory. This is called the *Coulomb term*, and it is convenient to define a Coulomb operator as

$$\mathcal{J}_j(\mathbf{x}_1) = \int d\mathbf{x}_2|\phi_j(\mathbf{x}_2)|^2 r_{12}^{-1}, \tag{4.15}$$

which gives the average local potential at point $\mathbf{x}_1$ due to the charge distribution from the electron in orbital ϕ_j.

The other term in brackets in eq.(4.13) is harder to explain and does not have a simple classical analog. It arises from the antisymmetry requirement of the wavefunction. It looks much like the Coulomb term, except that it switches or exchanges spin orbitals ϕ_i and ϕ_j. Hence, it is called the *exchange term*:

$$\sum_{j\neq i}\left[\int d\mathbf{x}_2\phi_j^*(\mathbf{x}_2)\phi_i(\mathbf{x}_2)r_{12}^{-1}\right]\phi_j(\mathbf{x}_1). \tag{4.16}$$

We can define an exchange operator in terms of its action on an arbitrary spin orbital ϕ_i

$$\mathcal{K}_j(\mathbf{x}_1)\phi_i(\mathbf{x}_1) = \left[\int d\mathbf{x}_2\phi_j^*(\mathbf{x}_2)r_{12}^{-1}\phi_i(\mathbf{x}_2)\right]\phi_j(\mathbf{x}_1). \tag{4.17}$$

In terms of these Coulomb and exchange operators, the Hartree-Fock equations become considerably more compact.

$$\left[h(\mathbf{x}_1) + \sum_{j\neq i}\mathcal{J}_j(\mathbf{x}_1) - \sum_{j\neq i}\mathcal{K}_j(\mathbf{x}_1)\right]\phi_i(\mathbf{x}_1) = \epsilon_i\phi_i(\mathbf{x}_1). \tag{4.18}$$

Perhaps now it is more clear that the Hartree-Fock equations are eigenvalue equations. If we realize that

$$[\mathcal{J}_i(\mathbf{x}_1) - \mathcal{K}_i(\mathbf{x}_1)]\,\phi_i(\mathbf{x}_1) = 0, \tag{4.19}$$

then it becomes clear that we can remove the restrictions $j \neq i$ in the summations, and we can introduce a new operator, the ***Fock operator***, as

$$f(\mathbf{x}_1) = h(\mathbf{x}_1) + \sum_j \mathcal{J}_j(\mathbf{x}_1) - \mathcal{K}_j(\mathbf{x}_1). \tag{4.20}$$

And now the Hartree-Fock equations are just

$$f(\mathbf{x}_1)\phi_i(\mathbf{x}_1) = \epsilon_i\phi_i(\mathbf{x}_1). \tag{4.21}$$

4.6.1 Matrix representation of the Hartree-Fock equation: the Roothaan equation

Introducing a basis set transforms the Hartree-Fock equations into the Roothaan equations. Denoting the atomic orbital basis functions as χ, we have the expansion

$$\phi_i = \sum_{\mu=1}^{K} D_{\mu i}\chi_\mu \tag{4.22}$$

for each spin orbital i. This leads to

$$f(\mathbf{x}_1)\sum_{\nu\nu} D_{\nu i}\chi_\nu(\mathbf{x}_1) = \epsilon_i \sum D_{\nu i}\chi_\nu(\mathbf{x}_1). \tag{4.23}$$

Left multiplying by $\chi^{\star}{}_{\mu}(\mathbf{x}_1)$ and integrating yields a matrix equation

$$\sum_\nu D_{\nu i}\int d\mathbf{x}_1\chi_\mu^*(\mathbf{x}_1)f(\mathbf{x}_1)\chi_\nu(\mathbf{x}_1) = \epsilon_i\sum_\nu D_{\nu i}\int d\mathbf{x}_1\chi_\mu^*(\mathbf{x}_1)\chi_\nu(\mathbf{x}_1). \tag{4.24}$$

This can be simplified by introducing the matrix element notation

$$S_{\mu\nu} = \int d\mathbf{x}_1\chi_\mu^*(\mathbf{x}_1)\chi_\nu(\mathbf{x}_1), \tag{4.25}$$

$$F_{\mu\nu} = \int d\mathbf{x}_1\chi_\mu^*(\mathbf{x}_1)f(\mathbf{x}_1)\chi_\nu(\mathbf{x}_1). \tag{4.26}$$

Now the Hartree-Fock-Roothaan equations can be written in matrix form as

$$\sum_\nu F_{\mu\nu}D_{\nu i} = \epsilon_i\sum_\nu S_{\mu\nu}D_{\nu i} \tag{4.27}$$

or even more simply as matrices

$$\mathbf{FD} = \mathbf{SD}\epsilon \tag{4.28}$$

where ϵ is a diagonal matrix of the orbital energies ϵ_i. This is like an eigenvalue equation except for the overlap matrix $\mathbf{S}$. One performs a transformation of basis to go to an orthogonal basis to make $\mathbf{S}$ vanish. Then it's just a matter of solving an eigenvalue equation (or, equivalently, diagonalizing $\mathbf{F}$!). Well, not quite. Since $\mathbf{F}$ depends on it's own solution (through the orbitals), the process must be done iteratively. This is why the solution of the Hartree-Fock-Roothaan equations are often called the *self-consistent-field* procedure.

Computational Aspects: Variational Optimization of Orbitals

The variational theorem states that for a time-independent Hamiltonian operator, any trial wavefunction will have an energy expectation value that is greater than or equal to the true ground state wavefunction corresponding to the given Hamiltonian.

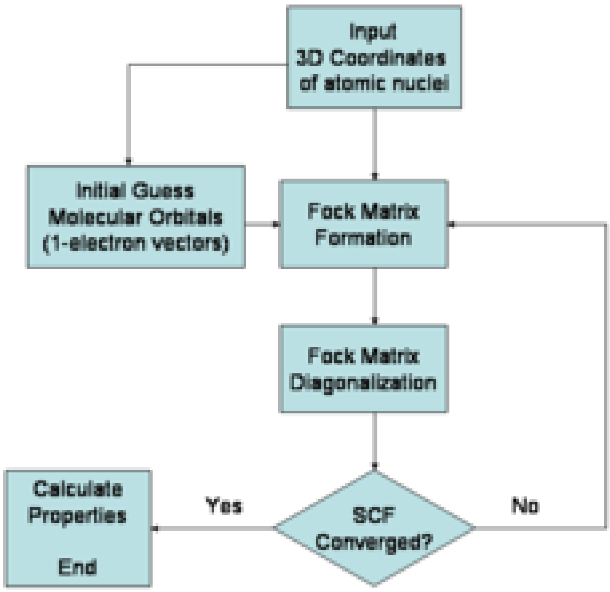
$$E = \frac{\langle\Psi|\hat{\mathcal{H}}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \geq E_0 \tag{4.29}$$

Because of this, the Hartree-Fock energy is an upper bound to the true ground state energy of a given molecule. In the context of the Hartree-Fock method, the best possible solution is at the Hartree-Fock limit, i.e. the limit of the Hartree-Fock energy as the basis set approaches completeness. (The other is the full-CI limit, where the last two approximations of the Hartree-Fock theory as described above are completely undone. It is only when both limits are attained that the exact solution is obtained.)

The starting point for the Hartree-Fock method is a set of approximate one-electron wavefunctions known as orbitals. For an atomic calculation, these are typically the orbitals for a hydrogenic atom (an atom with only one electron, but the appropriate nuclear charge). For a molecular or crystalline calculation, the initial approximate one-electron wavefunctions are typically a linear combination of atomic orbitals (LCAO).

The orbitals above only account for the presence of other electrons in an average manner. In the Hartree-Fock method, the effect of other electrons are accounted for in a mean-field theory context. The orbitals are optimized by requiring them to minimize the energy of the respective Slater determinant. This operation leads to the Hartree-Fock equation described above.

Since the Fock operator depends on the orbitals used to construct the corresponding Fock matrix, the eigenfunctions of the Fock operator are in turn new orbitals which can be used to construct a new Fock operator. In this way, the Hartree-Fock orbitals are optimized iteratively until the change in total electronic energy falls below a predefined threshold. In this way, a set of self-consistent one-electron orbitals are calculated. The Hartree-Fock electronic wavefunction is then the Slater determinant constructed out of these orbitals. Following the basic postulates of quantum mechanics, the Hartree-Fock wavefunction can then be used to compute any desired chemical or physical property within the framework of the Hartree-Fock method and the approximations employed.



Flow diagram of a Hartree-Fock calculation (from wikipedia.

Types of Hartree-Fock Calculations

We can perform different types of HF calculations: *restricted (RHF)*, *unrestricted (UHF)* or *restricted open-shell (ROHF)* depending on how much variational freedom is given to the spatial wavefunctions of electrons with different spins.

Restricted Hartree-Fock

For closed-shell system with an even number of electrons, it is often possible to describe the spatial part of the wavefunction of pairs of electrons with opposite spin with the same function. This leads to a substantial reduction of the overall computational cost since we only have to determine N/2 HF orbitals.

Unrestricted Hartree-Fock (UHF)

For sytems with an odd number of electrons or open-shell systems with a spin multiplicity that is higher than that of a singlet state, the spatial distribution of corresponding α and β spin electrons can start to differ and it might be necessary to include one HF orbital for every single one of the N electrons. Unfortunately, this procedure (though in principle more accurate) can also lead to some artifacts since in this case the resulting HF orbitals may no longer be eigenfunctions of the total spin operator S^2. This can lead to the so-called '*spin-contamination*'.

Restricted Open-Shell Hartree Fock (ROHF)

ROHF is a HF variant midway between a fully restricted and a fully unrestricted description. In ROHF, paired electrons are described with a single HF orbital whereas the unpaired electrons are described in an unrestricted way, i.e. by allowing for separate spatial wavefunctions for each electron. This procedure has the advantage that the total spin remains properly defined.

Chapter 5

An Introduction to Configuration Interaction Theory

Adapted from C. D. Sherrill's notes: "An Introduction to Configuration Interaction Theory".

5.1 Introduction

These notes attempt to present the essential ideas of Configuration Interaction (CI) theory in a fairly detailed mathematical framework. Of all the *ab initio* methods, CI is probably the easiest to understand, and perhaps one of the hardest to implement efficiently on a computer! The next two sections explain what the CI method is: the matrix formulation of the Schrödinger equation $\hat{H}\Psi = E\,\Psi$. The remaining sections describe various simplifications, approximations, and computational techniques. Much of the notation used in these notes is consistent with that of Szabo and Ostlund, *Modern Quantum Chemistry* (see references) and the rest of this script. Below several of the commonly-used symbols and their meanings are listed once more for convenience:

N	The number of electrons in the system.
n_α	The number of α electrons.
n_β	The number of β electrons.
n	The number of orbitals in the one-particle basis set.
δ_{ij}	kronecker delta function, equal to one if $i = j$ and zero otherwise.
$\hat{H}$	The exact nonrelativistic electronic Hamiltonian operator.
$\hat{h}$	The one-particle part of $\hat{H}$, $\hat{H} = \sum_i \hat{h}_i + \frac{1}{2}\sum_{ij} 1/r_{ij}$

$\mathbf{H}$	The Hamiltonian matrix, i.e. the matrix of $\hat{H}$, in whatever N-electron basis is currently being used.
H_{ij}	The i,j-th element of $\mathbf{H}$, equal to $\langle\Phi_i\vert\hat{H}\vert\Psi_j\rangle$, where Φ_i and Φ_j are N-electron CI basis functions.
$\mathbf{x}_i$	The space and spin coordinates of electron i.
$\mathbf{r}_i$	The spatial coordinates of electron i.
ϕ_i	The i-th one-particle basis function (orbital). Usually denotes a spin-orbital obtained from a Hartree-Fock procedure. May also be written simply as i.
χ_i	The i-th one-particle basis function (orbital) used to expand the HF orbitals, ϕ_i. Usually denotes an atom spin-orbital.
$\vert\Phi_i\rangle$	The i-th N-electron basis function. Usually denotes a single Slater determinant, but may also be a configuration state function (CSF).
$\vert\Psi_i\rangle$	Usually denotes an eigenfunction of $\mathbf{H}$. The exact nonrelativistic wave-function if a complete basis is used in the expansion of $\hat{H}$
$\vert\Phi_a^r\rangle$	An N-electron basis function which differs from some reference function $\vert\Phi_0\rangle$ by the replacement of spin-orbital a by spin orbital r. Usually implies a single Slater determinant.
$\vert ab\ldots c\rangle$	A Slater determinant with spin-orbitals $a, b, \ldots c$ occupied, $$\vert\phi_a\phi_b\ldots\phi_c\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_a(\mathbf{x}_1) & \phi_b(\mathbf{x}_1) & \cdots & \phi_c(\mathbf{x}_1) \\ \phi_a(\mathbf{x}_2) & \phi_b(\mathbf{x}_2) & \cdots & \phi_c(\mathbf{x}_2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \phi_a(\mathbf{x}_N) & \phi_b(\mathbf{x}_N) & \cdots & \phi_c(\mathbf{x}_N) \end{vmatrix}$$
$\langle i\vert\hat{h}\vert j\rangle$	One-electron integral in the physicists' (bra-ket) notation (i and j) are spin-orbitals). More explicitly, $$\langle i\vert\hat{h}\vert j\rangle = \int \phi_i^*(\mathbf{x}_1)\hat{h}(\mathbf{x}_1)\phi_j(\mathbf{x}_1)\, d\mathbf{x}_1$$
$[i\vert\hat{h}\vert j]$	The same one-electron integral in the chemists' notation (i and j) are spin-orbitals).
$(i\vert\hat{h}\vert j)$	One-electron integral in the chemists' notation (i and j) are spatial orbitals).

$\langle ij\|kl\rangle$	A simple two-electron integral, in physicists' (bra-ket) notation, where i, j, k and l are spin-orbitals. $\langle ij\|kl\rangle = \int \phi_i^*(\mathbf{x}_1)\phi_j^*(\mathbf{x}_2)\frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|}\phi_k(\mathbf{x}_1)\phi_l(\mathbf{x}_2)\, d\mathbf{x}_1 d\mathbf{x}_2$
$\big[ij\|kl\big]$	A simple two-electron integral, in chemists' notation, where i, j, k and l are spin-orbitals. $\big[ij\|kl\big] = \int \phi_i^*(\mathbf{x}_1)\phi_j(\mathbf{x}_1)\frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|}\phi_k^*(\mathbf{x}_2)\phi_l(\mathbf{x}_2)\, d\mathbf{x}_1 d\mathbf{x}_2$
$(ij\|kl)$	A simple two-electron integral, in chemists' notation, where i, j, k and l are spatial orbitals. $(ij\|kl) = \int \phi_i^*(\mathbf{r}_1)\phi_j(\mathbf{r}_1)\frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|}\phi_k^*(\mathbf{r}_2)\phi_l(\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2$
$\langle ij\|\|kl\rangle$	Asymmetrized two-electron integral, equal to $\langle ij\|kl\rangle - \langle ij\|lk\rangle$.

5.2 Fundamental Concepts

5.2.1 Scope of the Method

The scope of CI is to improve the HF solution by increasing the space of all possible many-electron wavefunctions from a single Slater determinant (in Hartree-Fock theory) to a set of, in principle infinite, Slater determinants. In the following we will learn a method for a systematic generation of Slater determinants starting from the one-electron orbitals obtained from the HF calculation[1].

5.2.2 What is a Configuration Interaction

We learned in chapter 2.1 that any solution of the eigenvalue equation

$$\hat{H}\Psi_j(\mathbf{r};\mathbf{R}) = E_j\Psi_j(\mathbf{r};\mathbf{R}) \tag{5.1}$$

can be expanded as a linear combination of wavefunctions that belong to a complete basis set $\{\Phi_i\}$,

$$|\Psi_j\rangle = \sum_{i=1}^{M} c_{ij}|\Phi_i\rangle \tag{5.2}$$

In the case of HF theory, this expansion had a single element, the Slater determinant made of the occupied HF one-electron orbitals.
More generally, an arbitrary N-electron wavefunction can be expressed exactly as a linear combination of all possible N-electron Slater determinants formed from a complete set of spin orbitals $\{\chi_i(\mathbf{x})\}$. If we solve the matrix equation $\mathbf{H}|\Psi\rangle = E|\Psi\rangle$ in a complete basis of N-electron functions as just described, we will obtain all electronic eigenstates of the system exactly. If our N-electron basis functions are denoted $|\Phi_i\rangle$, the eigenvectors of $\mathbf{H}$ are given as

$$|\Psi_j\rangle = \sum_{i}^{M} c_{ij}|\Phi_i\rangle$$

if there are M possible N-electron basis functions (M will be infinite if we actually have a complete set of one electron spin orbitals χ_i. The matrix $\mathbf{H}$ is constructed so that $H_{ij} = \langle\Phi_i|H|\Phi_j\rangle$ for $i, j = 1, 2, \ldots, M$. The matrix elements H_{ij} may be written in terms of one- and two-electron integrals according to "Slater's rules", as discussed in section 5.4.
The N-electron basis functions $|\Phi_i\rangle$ can be written as substitutions or "excitations" from the Hartree-Fock "reference" determinant, i.e.

$$|\Psi\rangle = c_0|\Phi_0\rangle + \sum_{ra} c_a^r|\Phi_a^r\rangle + \sum_{a<b,r<s} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \sum_{r<s<t,a<b<c} c_{abc}^{rst}|\Phi_{abc}^{rst}\rangle + \ldots \tag{5.3}$$

---

[1] Remember that the HF equation can be solved, in principle, for an infinite number of HF orbitals, N occupied plus an infinite number of unoccupied (virtual) orbitals.

where $|\Phi_a^r\rangle$ means the Slater determinant formed by replacing spin-orbital a in $|\Phi_0\rangle$ with spin orbital r, etc. Every N-electron Slater determinant can be described by the set of N spin orbitals from which it is formed, and this set of orbital occupancies is often referred to as a **configuration**. Thus the **configuration interaction** method is, in its most straightforward implementation, nothing more than the matrix solution of the time-independent non-relativistic electronic Schrödinger equation $\hat{H}\Psi = E\Psi$. One of the great strengths of the CI method is its generality; the formalism applies to excited states, to open-shell systems, and to systems far from their equilibrium geometries [2].

In practice, one does not have a complete set of one-particle basis functions $\{\chi_i(\mathbf{x})\}$; typically one assumes that the incomplete one-electron basis set is large enough to give useful results, and the CI procedure is not modified. The quality of the one-particle basis set can be checked by comparing the results of calculations using progressively larger basis sets.

It is also possible to reduce the size of the N-electron basis set. If we desire only wavefunctions of a given spin and/or spatial symmetry, as is usually the case, we need include only those N-electron basis functions of that symmetry, since the Hamiltonian matrix is block-diagonal according to space and spin symmetries. If one performs the matrix mechanics calculation using a given set of one-particle functions $\{\chi_i(\mathbf{x})\}$ and all possible N-electron basis functions $\{|\Phi_i\rangle\}$ (possibly symmetry-restricted), the procedure is called **full CI**. The full CI corresponds to solving Schrödinger's equation exactly *within* the space spanned by the specified one-electron basis. If the one-electron basis is complete (it never is in practice, but it may be in theory), then the procedure is called a **complete CI**.

Unfortunately, even with an incomplete one-electron basis, a full CI is computationally intractable for any but the smallest systems, due to the vast number of N-electron basis functions required. The CI space must be reduced somehow-hopefully in such a way that the approximate CI wavefunction and energy are as close as possible to the exact values. The effective reduction of the CI space is a major concern in CI theory, and we will discuss some of the more popular approaches in these notes.

By far the most common CI approximation is the truncation of the CI space expansion according to excitation level relative to the reference state. The widely-employed CI singles and doubles (**CISD**) wavefunction includes only those N-electron basis functions which represent single or double excitations relative to the reference state. Since the Hamiltonian operator includes only one- and two-electron terms, only singly and doubly excited configurations can interact directly with the reference, and they typically account for about 95% of the correlation energy in small molecules at their equilibrium geometries.

---

[2] By contrast, traditional single-reference perturbation theory and coupled-cluster approaches generally *assume* that the reference configuration is dominant, and they may fail when it is not.

5.3 The Correlation Energy

Approximate CI methods can be judged according to what fraction of the correlation energy they recover. The correlation energy is defined as the difference between the energy in the Hartree-Fock limit (E_{HF}) and the exact nonrelativistic energy of a system ($\mathcal{E}_0$)

$$E_{\text{corr}} = \mathcal{E}_0 - E_{HF} \tag{5.4}$$

This energy will always be negative because the Hartree-Fock energy is an upper bound to the exact energy (this is guaranteed by the variational theorem). The exact nonrelativistic energy $\mathcal{E}_0$ could, in principle, be calculated by performing a full CI in a complete one-electron basis set. If we have an incomplete one-electron basis set, then we can only compute the *basis set correlation energy*, which is the correlation energy for a given one-electron basis. For convenience, the basis set correlation energy is often simply referred to as the correlation energy.

The correlation energy is the energy recovered by fully allowing the electrons to avoid each other; Hartree-Fock improperly treats interelectron repulsions in an averaged way. However, there is some inconsistency in this line of thinking. When a molecule is pulled apart, the electrons shouldn't need to avoid each other as much, so the magnitude of the correlation energy should decrease. In fact, the opposite is true, as shown by the basis set correlation energies given in the following Table for H_2O at three different geometries.

Geometry	E_{corr} (hartree)
R_e	-0.148028
$1.5R_e$	-0.210992
$2.0R_e$	-0.310067

The correlation energy increases at stretched geometries, because our definition of the correlation energy, $E_{\text{corr}} = \mathcal{E}_0 - E_{HF}$, includes not only the concept of electrons avoiding each other, which is called the *dynamical correlation energy*, but also a more subtle effect called the *nondynamical, static correlation energy.* Nondynamical correlation energy reflects the inadequacy of a single reference in describing a given molecular state, and is due to nearly degenerate states or rearrangement of electrons within partially filled shells.

Siegbahn [3] offers the following explanation of the difference between dynamical and nondynamical correlation energies:

"In many situations it is further convenient to subdivide the correlation energy into two parts with different physical origins. For chemical reactions where bonds are broken and formed, and for most excited states, the major part of the correlation energy can be obtained by adding only a few extra configurations besides the Hartree-Fock configuration. This part of the

---

[3] P.E. Siegbahn, The direct CI method, in Methods in *Computational Molecular Physics*, edited by G.H.F. Diercksen and S. Wilson, pages 189-207, D. Reidel, Dordrecht, 1983.

correlation energy is due to near degeneracy between different configurations and has its origin quite often in artifacts of the Hartree-Fock approximation. The physical origin of the second part of the correlation energy is the dynamical correlation of the motion of the electrons and is therefore sometimes called the dynamical correlation energy. Since the Hamiltonian operator contains only one- and two-particle operators this part of the correlation energy can be very well described by single and double replacements from the leading, near degenerate, reference configurations."

5.4 Slater's Rules

Whether we perform a full CI or only a limited CI, we must be able to express $\hat{H}$ in matrix form so that we can diagonalize it and obtain the eigenvectors and eigenvalues of interest. In this section we discuss Slater's rules (or the Slater-Condon rules), which allow us to express matrix elements

$$H_{ij} = \langle \Phi_i | \hat{H} | \Phi_j \rangle$$

in terms of one- and two-electron integrals.

At the moment, we will express these results in terms of spin-orbitals using the physicist's notation. The one-electron integrals are written as

$$\langle i|\hat{h}|j\rangle = \int \phi_i^*(\mathbf{r}_1)\hat{h}(\mathbf{r}_1)\phi_j(\mathbf{r}_1)d\mathbf{r}_1 \tag{5.5}$$

and the two-electron integrals are written as

$$\langle ij||kl\rangle = \langle ij|kl\rangle - \langle ij|lk\rangle \tag{5.6}$$

$$\langle ij|kl\rangle = \int \phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\frac{1}{\mathbf{r}_{12}}\phi_k(\mathbf{r}_1)\phi_l(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{5.7}$$

Before Slater's rules can be used, the two Slater determinants must be arranged in *maximum coincidence.* Remember that switching columns in a determinant introduces a minus sign. For instance, to calculate $\langle \Phi_1|\hat{H}|\Phi_2\rangle$ where we have

$$|\Phi_1\rangle = |abcd\rangle$$
$$|\Phi_2\rangle = |crds\rangle$$

we must *first* interchange columns of $|\Phi_1\rangle$ or $|\Phi_2\rangle$ to make the two determinants look as much alike as possible. For example, we may rearrange $|\Phi_2\rangle$ as

$$|\Phi_2\rangle = |crds\rangle = -|crsd\rangle = |srcd\rangle \tag{5.8}$$

After the determinants are in maximum coincidence, we see how many spin orbitals they differ by, and we then use the following rules:

1. **Identical Determinants:** If the determinants are identical, then

$$\langle\Phi_1|\hat{H}|\Phi_1\rangle = \sum_m^N \langle m|\hat{h}|m\rangle + \sum_{m>n}^N \langle mn||mn\rangle \tag{5.9}$$

2. **Determinants that Differ by One Spin Orbital:**

$$\begin{aligned} |\Phi_1\rangle &= |\cdots mn \cdots\rangle \\ |\Phi_2\rangle &= |\cdots pn \cdots\rangle \\ \langle\Phi_1|\hat{H}|\Phi_2\rangle &= \langle m|\hat{h}|p\rangle + \sum_n^N \langle mn||pn\rangle \end{aligned} \tag{5.10}$$

3. **Determinants that Differ by Two Spin Orbitals:**

$$\begin{aligned} |\Phi_1\rangle &= |\cdots mn \cdots\rangle \\ |\Phi_2\rangle &= |\cdots pq \cdots\rangle \\ \langle\Phi_1|\hat{H}|\Phi_2\rangle &= \langle mn||pq\rangle \end{aligned} \tag{5.11}$$

4. **Determinants that Differ by More than Two Spin Orbitals:**

$$\begin{aligned} |\Phi_1\rangle &= |\cdots mno \cdots\rangle \\ |\Phi_2\rangle &= |\cdots pqr \cdots\rangle \\ \langle\Phi_1|\hat{H}|\Phi_2\rangle &= 0 \end{aligned} \tag{5.12}$$

The derivation of these rules can be found in the book of Szabo and Ostlund, section 2.3.4 (pp. 77-81).

5.5 The Solution of the CI equation: The Variational Equation

One particularly nice feature of the CI method is that the calculated lowest energy eigenvalue is always an upper bound to the exact ground state energy.
This is a direct consequence of the variational theorem (see sections 2.3.3 and 2.3.4) applied to the total energy expectation value,

$$E = \frac{\langle\Psi_0|\hat{H}|\Psi_0\rangle}{\langle\Psi_0|\Psi_0\rangle} \tag{5.13}$$

where $|\Psi_0\rangle$ is an approximate solution given as a linear combination of N-electron wavefunctions,

$$|\Psi_0\rangle = c_0|\Phi_0\rangle + \sum_{ra} c_a^r|\Phi_a^r\rangle + \sum_{a<b,r<s} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \sum_{r<s<t,a<b<c} c_{abc}^{rst}|\Phi_{abc}^{rst}\rangle + \dots$$

or, using a more compact notation,

$$|\Psi_0\rangle = \sum_{i=0} c_{0i}|\Phi_i\rangle \tag{5.14}$$

Once the expansion basis is chosen, $\{|\Phi_i\rangle\}$, the resulting secular equation (see sections 2.3.3 and 2.3.4)

$$\mathbf{HC} = E\mathbf{C} \tag{5.15}$$

can be solved numerically. The outcome of this calculation is the **ground state energy** and the coefficients, $\{c_{0i}\}$ for the expansion of the ground state N-electron wavefunction. Note that the quality of the results depends on the quality and the number of the chosen basis functions $\{|\Phi_i\rangle\}$.
At this point it is reasonable to ask *why* we wish to minimize the energy by varying the coefficients in equation 5.14. How do we know that this will give us the best estimate of the wavefunction? There are two answers to this. First, as we have just shown, minimizing the energy by variation of the linear expansion coefficients gives the Schrödinger equation in matrix form; thus the procedure is justified a posteriori by the validity of its result. The other reason is that, for the ground state, the linear expansion in equation (5.14) gives an expectation value for the energy E, which is always an upper bound to the exact ground state energy (this is called the Variational Theorem). The best estimate of E, then, is the minimum value which can be obtained by varying the coefficients in equation 5.14 (while also maintaining normalization). These arguments also hold for excited states, so long as each excited state is made orthogonal to all lower states.

5.6 Classification of Basis Functions by Excitation Level

Now we will discuss the importance of various excitation classes to the CI wavefunction. As noted in the equation

$$|\Psi\rangle = c_0|\Phi_0\rangle + \sum_{ra} c_a^r|\Phi_a^r\rangle + \sum_{a<b,r<s} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \sum_{r<s<t,a<b<c} c_{abc}^{rst}|\Phi_{abc}^{rst}\rangle + \ldots$$

the CI expansion is typically truncated according to excitation level; in the vast majority of CI studies, the expansion is truncated (for computational tractability) at doubly-excited configurations. Since the Hamiltonian contains only two-body terms, only singles and doubles can interact directly with the reference. This is a direct result of Slater's Rules. The structure of the CI matrix with respect to excitation level is given below (adapted from Szabo and Ostlund, p. 235), where $|S\rangle, |D\rangle, |T\rangle$, represent blocks of singly, doubly, triply, and quadruply excited determinants, respectively. The Hamiltonian matrix **H** is Hermitian; if only real orbitals are used, as is usually the case, then the Hamiltonian is also symmetric. Thus only the lower triangle of **H** is shown below.

$$\mathbf{H} = \begin{array}{r} \langle\Phi_0| \\ \langle S| \\ \langle D| \\ \langle T| \\ \langle Q| \\ \vdots \end{array} \begin{bmatrix} \langle\Phi_0|H|\Phi_0\rangle & & & & & & \cdots \\ 0 & \langle S|H|S\rangle & & & & & \cdots \\ \langle D|H|\Phi_0\rangle & \langle D|H|S\rangle & \langle D|H|D\rangle & & & & \cdots \\ 0 & \langle T|H|S\rangle & \langle T|H|D\rangle & \langle T|H|T\rangle & & & \cdots \\ 0 & 0 & \langle Q|H|D\rangle & \langle Q|H|T\rangle & \langle Q|H|Q\rangle & & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Note that the matrix elements $\langle S|H|\Phi_0\rangle$ are given as 0. This is due to Brillouin's theorem, which is valid when our reference function $|\Phi_0\rangle$ is obtained by the Hartree-Fock method (Hartree-Fock guarantees that off-diagonal elements of the Fock matrix are zero, and it turns out that the matrix element between two Slater determinants which differ by one spin orbital is equal to an off-diagonal element of the Fock matrix).
Furthermore, the blocks $\langle X|H|Y\rangle$ which are not necessarily zero may still be sparse; for example, the matrix element $\langle\Phi_{ab}^{rs}|H|\Phi_{cdef}^{tuvw}\rangle$ which belongs to the block $\langle D|H|Q\rangle$, will be nonzero only if a and b are contained in the set $\{c, d, e, f\}$ and if r and s are contained in the set $\{t, u, v, w\}$.
Since only the doubles interact directly with the Hartree-Fock reference, we expect double excitations to make the largest contributions to the CI wavefunction, after the reference state. Indeed, this is what is observed. Even though singles, triples, etc. do not interact directly with the reference, they can still become part of the CI wavefunction (i.e. have non-zero coefficients) because they mix with the doubles, directly or indirectly. Although singles are much less important to the energy than doubles, they are generally included in CI treatments because of their relatively small number and because of their greater importance in describing one-electron properties (dipole moment, etc.).

5.7 Energy Contributions of the Various Excitation Levels

The following table demonstrates the importance of various excitation classes in obtaining CI energies. We see that singles and doubles account for 95% of the correlation energy at the equilibrium geometries of the molecules listed. We see that quadruple excitations are more important than triples, at least as far as the energy is concerned. At stretched geometries, the CISD and CISDT methods become markedly poorer, yet the CISDTQ method still recovers a very high (and nearly constant) fraction of the correlation energy, suggesting that CISDTQ should give reliable results for energy differences across potential energy surfaces for molecules of this size.

	Percent	Correlation Energy	
Molecule	CISD	CISDT	CISDTQ
BH	94.91	n/a	99.97
$H_2O(R_e)$	94.70	95.47	99.82
$H_2O(1.5R_e)$	89.39	91.15	99.48
$H_2O(2.0R_e)$	80.51	83.96	98.60
NH_3	94.44	95.43	99.84
HF	95.41	96.49	99.86
H_7^+	96.36	96.87	99.96

5.8 Size of the CI Space as a Function of Excitation Level

From the next table, we can see that the number of N-electron basis functions increases dramatically with increasing excitation level. It should be pointed out that while the calculations on BH, HF, and H_7^+ used DZP basis sets, those on H_2O and NH_3 used only DZ basis sets. A DZP basis should be considered the minimum adequate basis for a truly meaningful benchmark study, and even then it is desirable to use a high-quality basis such as an Atomic Natural Orbital (ANO) set.

		CFS's required		
Molecule	CISD	CISDT	CISDTQ	FCI
BH	568	n/a	28'698	132'686
H_2O	361	3'203	17'679	256'473
NH_3	461	4'029	19'925	137'321
HF	552	6'712	48'963	944'348
H_7^+	1'271	24'468	248'149	2'993'933

While it is generally possible to perform CISD calculations on small molecules with a good one-electron basis, the CISDTQ method is limited to molecules containing very few heavy atoms, due to the extreme number of N-electron basis functions required. Full CI calculations are of course even more difficult to perform, so that despite their importance as benchmarks, few full CI energies using flexible one-electron basis sets have been obtained.

Chapter 6

Many-Body Perturbation Theory

The idea in perturbation methods is that the problem under investigation only differs slightly from a problem which has already been solved (exactly or approximately). The solution to the given problem should therefore in some sense be close to the solution of the already known system. This is described mathematically by defining a Hamiltonian operator which consists of two parts, a reference ($\mathcal{H}^{(0)}$ and a perturbation ($\mathcal{H}'$).

The basic assumption of perturbation methods is that the $\mathcal{H}'$ operator in some sense is "small" compared to $\mathcal{H}^{(0)}$. In quantum mechanics, perturbation methods can be used for adding corrections to solutions which employ an independent particle approximation, and the theoretical framework is then called Many-Body Perturbation Theory (MBPT).

6.1 Perturbation Theory in Quantum Mechanics

Let us assume that the Schrödinger equation for the reference Hamiltonian operator is

$$\begin{aligned} &\mathcal{H} = \mathcal{H}^{(0)} + \lambda\,\mathcal{H}' \\ &\mathcal{H}^{(0)}\Phi_i{}^{(0)} = E_i^{(0)}\,\Phi_i{}^{(0)}\,,\, i = 0, 1, 2, 3, \ldots, \infty \end{aligned} \tag{6.1}$$

The solutions for the unperturbed Hamiltonian operator form a complete set (since $\mathcal{H}^{(0)}$ is Hermitian) which can be chosen to be orthonormal, and λ is a (variable) parameter determining the strength of the perturbation. The parameter λ can be varied systematically to switch the system from the unperturbed case ($\lambda = 0$) to the fully perturbed ($\lambda = 1$) case. At present we will only consider cases where the perturbation is time-independent, and the reference wave function is nondegenerate. To keep the notation simple, we will furthermore only consider the lowest energy state ($i = 0$) and therefore will leave out the index i.

The perturbed Schrödinger equation is

$$\mathcal{H}\,\Psi(\lambda) = E(\lambda)\,\Psi(\lambda) \tag{6.2}$$

If $\lambda = 0$, then $\mathcal{H} = \mathcal{H}^{(0)}$, $\Psi^{(0)} = \Phi_0$ and $E^{(0)} = E_0$. As the perturbation is increased from zero to a finite value, the new energy and wavefunction must also change continuously, and they can be written as a Taylor expansion in powers of the perturbation parameter λ (from now on, we will consider the perturbation of the ground state orbital only, $\Phi_0 \equiv \Phi$).

$$\begin{aligned} E(\lambda) &= \lambda^0\, E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \dots \\ \Psi(\lambda) &= \lambda^0\, \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \dots \end{aligned} \tag{6.3}$$

For $\lambda = 0$, $\lambda^0 = 1$, $\Psi^{(0)} = \Phi$ and $E^{(0)} = E_0$, these are the *unperturbed, or zeroth-order* wavefunction and energy. The $\Psi^{(1)}$, $\Psi^{(2)}$, ... and $E^{(1)}$, $E^{(2)}$, ... are the *first-, second-*, etc. *order* **corrections**. The λ parameter will eventually be set equal to 1, and the nth order energy or wavefunction will become a sum of all terms up to order n.

6.1.1 Normalization condition

It is convenient to choose the perturbed wavefunction so that the overlap with the unperturbed wavefunction is equal 1. This has the consequence that all correction terms are orthogonal to the reference wavefunction. We therefore have

$$\begin{aligned} \langle \Psi | \Phi \rangle &= 1 \\ \langle \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} \dots | \Phi \rangle &= 1 \\ \langle \Psi^{(0)} | \Phi \rangle + \lambda^1 \langle \Psi^{(1)} | \Phi \rangle + \lambda^2 \langle \Psi^{(2)} | \Phi \rangle + \dots &= 1 \end{aligned} \tag{6.4}$$

which implies

$$\langle \Psi^{(i \neq 0)} | \Phi \rangle = 0 \tag{6.5}$$

6.1.2 The nth-order perturbation equation

Once all the correction terms have been calculated, it is trivial to normalize the total wavefunction.

With the expansions (Eq.6.3) the Schrödinger equation (6.2) becomes

$$\begin{aligned} &(\mathcal{H}^{(0)} + \lambda \mathcal{H}')(\lambda^0 \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \dots) = \\ &(\lambda^0 E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \dots)\, (\lambda^0 \Psi^{(0)} + \lambda^1 \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \dots) \end{aligned} \tag{6.6}$$

Since this holds for any value of λ, we can collect terms with the same power

power of λ to give

$$\begin{aligned}
&\lambda^{(0)}:\ \mathcal{H}^{(0)}\,\Psi^{(0)} = E^{(0)}\,\Psi^{(0)} \\
&\lambda^{(1)}: \mathcal{H}^{(0)}\,\Psi^{(1)} + \mathcal{H}'\Psi^{(0)} = E^{(0)}\,\Psi^{(1)} + E^{(1)}\,\Psi^{(0)} \\
&\lambda^{(2)}: \mathcal{H}^{(0)}\,\Psi^{(2)} + \mathcal{H}'\Psi^{(1)} = E^{(0)}\,\Psi^{(2)} + E^{(1)}\,\Psi^{(1)} + E^{(2)}\,\Psi^{(0)} \qquad (6.7)\\
&\quad \dots \\
&\lambda^{(n)}: \mathcal{H}^{(0)}\Psi^{(n)} + \mathcal{H}'\Psi^{(n-1)} = \sum_{j=1}^{n} E^{(j)}\,\Psi^{(n-j)}
\end{aligned}$$

These are zeroth-, first-, second-, nth-order perturbation equations. The zeroth-order equation is just the Schrödinger equation for the unperturbed problem. The first-order equation contains two unknowns, the first-order correction to the energy, $E^{(1)}$, and the first-order correction to the wavefunction, $\Psi^{(1)}$.

6.1.3 Rayleigh-Schrödinger perturbation formula

Up to this point we are still dealing with undetermined quantities, energy and wavefunction corrections at each order. The first-order equation is one equation with two unknowns. Since the solutions to the unperturbed Schrödinger equation generates a complete set of orthogonal functions, the unknown first-order correction to the wave function, $\Psi^{(1)}$, can be expanded in these functions, $\{\Phi_i\}_{i=0}^{\infty}$. This is known as **Rayleigh-Schrödinger perturbation theory**.

First order perturbation. The λ^1 equation in (6.7) becomes

$$(\mathcal{H}^{(0)} - E^{(0)})\,\Psi^{(1)} + (\mathcal{H}' - E^{(1)})\Phi^{(0)} = 0 \qquad (6.8)$$

with

$$\Psi^{(1)} = \sum_{i=0}^{\infty} c_i \Phi_i \qquad (6.9)$$

and yields (exercise)

$$E^{(1)} = \langle \Phi_0 | \mathcal{H}' | \Phi_0 \rangle \qquad (6.10)$$

This shows that the first-order correction to the energy is an average of the perturbation operator over the unperturbed wave function.

The first-order correction to the wavefunction can be obtained by multiplying Eq. 6.7 for λ^1 from the left by the function corresponding to a given c_j (Φ_j) and integrating to give

$$c_j = \frac{\langle \Phi_j | \mathcal{H}' | \Phi_0 \rangle}{E_0 - E_j} \qquad (6.11)$$

The expansion coefficients determine the first-order correction to the perturbed wavefunction (Eq.6.8), and they can be calculated using the known unperturbed wavefunctions and energies. The coefficient in front of Φ_0 for $\Psi^{(1)}$ (Eq.6.9) cannot be determined from the above formula, but the assumption of intermediate normalization (Eq.6.4) makes $c_0 = 0$.

Second order perturbation. Starting from the second-order perturbation equation (Eq. 6.7) and using intermediate normalization ($c_0 = d_0 = 0$), the second order energy correction is expressed as

$$\Psi^{(2)} = \sum_i d_i \Phi_i \tag{6.12}$$

with

$$E^{(2)} = \sum_i c_i \langle \Phi | \mathcal{H}' | \Phi_i \rangle = \sum_{i \neq 0} \frac{\langle \Phi_0 | \mathcal{H}' | \Phi_i \rangle \langle \Phi_i | \mathcal{H}' | \Phi_0 \rangle}{E_0 - E_i} \tag{6.13}$$

and second-order wavefunction expansion coefficients

$$d_j = \sum_{i \neq 0} \frac{\langle \Phi_j | \mathcal{H}' | \Phi_i \rangle \langle \Phi_i | \mathcal{H}' | \Phi_0 \rangle}{(E_0 - E_j)(E_0 - E_i)} - \frac{\langle \Phi_j | \mathcal{H}' | \Phi_0 \rangle \langle \Phi_0 | \mathcal{H}' | \Phi_j \rangle}{(E_0 - E_j)^2} \tag{6.14}$$

The formulas for higher-order corrections become increasingly complex. The main point, however, is that all corrections can be expressed in terms of matrix elements of the perturbation operator over the unperturbed wavefunctions, and the unperturbed energies.

6.2 Møller-Plesset Perturbation Theory

So far the theory has been completely general. In order to apply perturbation theory to the calculation of correlation energy, the unperturbed Hamiltonian operator must be selected.
The most common choice is to take the sum of the single particle Fock operators as the upperturbed Hamiltonian, and the difference between the full Hamiltonian and the HF Hamiltonian as perturbation. This is called **Møller-Plesset** (MP) perturbation theory.
The sum of Fock operators counts the (average) electron-electron repulsion twice (Eqs. 4.7) and the perturbation becomes the exact $r_{ij}^{-1} \equiv \hat{v}_{ij}$ operator minus twice the $\hat{v}_{ij}^{HF}$ operator,

$$\hat{v}_{ij}^{HF} = \hat{J}_{ij} - \hat{K}_{ij} \tag{6.15}$$

$$= \sum_k \langle ik|jk \rangle - \langle ik|kj \rangle = \sum_k \langle ik||jk \rangle \, . \tag{6.16}$$

The operator associated with this difference is often referred to as the fluctuation potential.

The starting point is the Hartree-Fock operator, Eq. 4.20, (i and j are electron indices)

$$\hat{\mathcal{F}}_i = \hat{h}_i + \sum_j^N (\hat{\mathcal{J}}_j - \hat{\mathcal{K}}_j) \tag{6.17}$$

with

$$\hat{\mathcal{F}}_i \phi_i(x_i) = \varepsilon_i \phi_i(x_i) \tag{6.18}$$

The Coulomb and the exchange operators are defined as

$$\hat{\mathcal{J}}_j |\phi_i(x_i)\rangle = \langle \phi_j(x) | \hat{v}_{ij} | \phi_j(x) \rangle |\phi_i(x_i)\rangle \tag{6.19}$$

$$\hat{\mathcal{K}}_j |\phi_i(x_i)\rangle = \langle \phi_j(x) | \hat{v}_{ij} | \phi_i(x) \rangle |\phi_j(x_i)\rangle \tag{6.20}$$

Starting from $\hat{\mathcal{F}}_i$ (the operator for the electron with index i) we can construct the "N-electrons" operator $\sum_{i=1}^N \hat{\mathcal{F}}_i$, which is taken as reference Hamiltonian

$$\mathcal{H}^{(0)} = \sum_{i=1}^N \hat{\mathcal{F}}_i = \sum_{i=1}^N \hat{h}_i + \sum_{i=1}^N \sum_{j=1}^N \left(\hat{J}_{ij} - \hat{K}_{ij} \right) \equiv \sum_{i=1}^N \hat{h}_i + {\sum_{i,j=1}^N}' \hat{v}_{ij}^{HF} \,. \tag{6.21}$$

The perturbation Hamiltonian is therefore given by

$$\hat{\mathcal{H}}' = \mathcal{H}\hat{} - \mathcal{H}^{(0)} = \sum_{i<j}^N \hat{v}_{ij} - {\sum_{i,j=1}^N}' \hat{v}_{ij}^{HF} \tag{6.22}$$

where for a given Slater determinant many-electron wavefunction Φ_0

expressed in the HF orbital basis $\{\phi_i\}$ (compare with Eq. 4.7 [1])

$$\langle\Phi_0|\sum_{i<j}^{N}\hat{v}_{ij}|\Phi_0\rangle = \frac{1}{2}\langle\Phi_0|\sum_{i,j}^{N}{}' \hat{v}_{ij}^{HF}|\Phi_0\rangle \equiv \langle\mathbf{V}_{ee}\rangle \tag{6.23}$$

The **zeroth-order** wave function is the HF determinant, and the zeroth-order energy is just a sum of MO energies.

$$E^{(0)} = \sum_{i}^{N}\langle\phi_i|\hat{\mathcal{F}}_i|\phi_i\rangle = \sum_{i}^{N}\varepsilon_i^{\mathrm{HF}} \tag{6.24}$$

The **first-order energy correction** is the average of the perturbation operator over the zero-order wave function (Eq.6.10):

$$\begin{aligned} E^{(1)} = \langle\Phi_0|\hat{\mathcal{H}}'|\Phi_0\rangle &= \langle\Phi_0|\sum_{i<j}^{N}\hat{v}_{ij}|\Phi_0\rangle - \langle\Phi_0|\sum_{i,j=1}^{N}{}' \hat{v}_{ij}^{HF}|\Phi_0\rangle \\ &= \langle\mathbf{V}_{ee}\rangle - 2\,\langle\mathbf{V}_{ee}\rangle = -\langle\mathbf{V}_{ee}\rangle \end{aligned} \tag{6.25}$$

This yields a correction for the double counting of the electron-electron repulsion at zeroth order. Comparing Eq.6.22 with the expression for the total energy in Eq.4.7, it is seen that the first-order energy (sum of $E^{(0)}$ and $E^{(1)}$) is exactly the HF energy.
Using the notation E(MPn) to indicate the correction at order n, and MPn to indicate the total energy up to order n, we have

$$\begin{aligned} \mathrm{MP0}: &\quad E(\mathrm{MP0}) = \sum_{a}^{N}\varepsilon_a^{\mathrm{HF}} \\ \mathrm{MP1}: &\quad E(\mathrm{MP0}) + E(\mathrm{MP1}) = E(\mathrm{HF}) \end{aligned}$$

---

[1] For the set of HF orbitals, $\{\phi_i\}$, we have

$$\begin{aligned} \langle\hat{\mathcal{H}}_0\rangle = \sum_i\langle\hat{F}_i\rangle &= \sum_i\langle i|h|i\rangle + \sum_{ij}[ii|jj] - [ij|ji] \\ &= \sum_i\langle i|h|i\rangle + \sum_{ij}\langle ij|ij\rangle - \langle ij|ji\rangle \\ \langle\hat{\mathcal{H}}\rangle &= \sum_i\langle i|h|i\rangle + \frac{1}{2}\sum_{ij}[ii|jj] - [ij|ji] \\ &= \sum_i\langle i|h|i\rangle + \frac{1}{2}\sum_{ij}\langle ij|ij\rangle - \langle ij|ji\rangle \\ &= \sum_i\langle i|h|i\rangle + \frac{1}{2}\sum_i J_{ii} - K_{ii} \end{aligned}$$

Electron correlation energy thus starts at order 2 with this choice of $\mathcal{H}^{(0)}$.

In developing perturbation theory it was assumed that the solutions to the unperturbed problem formed a complete set. This in general means that there must be an infinite number of functions, which is impossible in actual calculations. The lowest energy solution to the unperturbed problem is the HF wave function, additional higher energy solutions are **excited Slater determinants**, analogously to the CI method. When a finite basis set is employed it is only possible to generate a finite number of excited determinants. The expansion of the many-electron wave function is therefore **truncated**.

The second-order correction to the energy, which is the first contribution to the correlation energy, involves a sum over doubly excited determinants (singly excited Slater determinants give no contribution to the energy. This is known as Brillouin's theorem). These can be generated by promoting two electrons from **occupied orbitals** a **and** b **to virtual orbitals** r **and** s. The summation must be restricted so that each excited state is only counted once

$$E^{(2)} = \sum_{a<b}^{\text{occ.}} \sum_{r<s}^{\text{virt.}} \frac{\langle \Phi_0 | \hat{\mathcal{H}}' | \Phi_{ab}^{rs} \rangle \langle \Phi_{ab}^{rs} | \hat{\mathcal{H}}' | \Phi_0 \rangle}{E_0 - E_{ab}^{rs}} \tag{6.26}$$

The matrix elements between the HF ground state and a doubly excited state are given by two electron integrals over MOs. The difference in total energy between two Slater determinants becomes a difference in MO energies (essentially Koopmans' theorem), and the explicit formula for the second-order Møller-Plesset correction is

$$E(\text{MP2}) = \sum_{a<b}^{\text{occ.}} \sum_{r<s}^{\text{virt.}} \frac{\left[\langle \phi_a \phi_b | \hat{v} | \phi_r \phi_s \rangle - \langle \phi_a \phi_b | \hat{v} | \phi_s \phi_r \rangle \right]^2}{(\varepsilon_a + \varepsilon_b - \varepsilon_r - \varepsilon_s)} \tag{6.27}$$

Once the two-electron integrals over MOs are available, the second-order energy correction can be calculated as a sum over such integrals. There are of the order of M^4 integrals, thus the calculation of the energy (only) increases as M^4 with the system size. However, the transformation of the integrals from the AO to the MO basis grows as M^5. MP2 is an M^5 method, but fairly inexpensive as not all two-electron integrals over MOs are required. Only those corresponding to the combination of two occupied and two virtual MOs are needed. In practical calculations this means that the MP2 energy for systems with 100-150 basis functions can be calculated at a cost similar to or less than what is required for calculating the HF energy. MP2 typically accounts for 80-90% of the correlation energy, and it is the most economical wavefunction-based method for including electron correlation.

Some remarks

The main limitation of perturbation methods is the assumption that the zeroth-order wavefunction is a reasonable approximation to the real wavefunction, i.e. the perturbation operator is sufficiently "small". The poorer the HF wavefunction describes the system, the larger are the correction terms, and more terms must be included to achieve a given level of accuracy. If the reference state is a poor description of the system, the convergence may be so slow or erratic that perturbation methods cannot be used. Actually it is difficult to prove that the perturbation expansion is convergent, although many systems show a behavior which suggests that it is the case. This may to some extent be deceptive, as it has been demonstrated that the convergence properties depend on the size of the basis set and the majority of studies have employed small or medium sized basis sets. A convergent series in a DZP type basis for example may become divergent or oscillating in a larger basis, especially if diffuse functions are present. In the ideal case the HF, MP2, MP3 and MP4 results show a monotonic convergence towards a limiting value, with the corrections being of the same sign and numerically smaller as the order of perturbation increases. Unfortunately, this is not the typical behavior. Even in systems where the reference is well described by a single determinant, oscillations in a given property as a function of perturbation order are often observed. This is not completely understood, but may at least partly be due to the fact that the choice of the unperturbed Hamilton operator does not make the perturbation particularly small.

In practice only low orders of perturbation theory can be carried out, and it is often observed that the HF and MP2 results differ considerably, the MP3 result moves back towards the HF and the MP4 moves away again. For "well-behaved" systems the correct answer is normally somewhere between the MP3 and MP4 results. MP2 typically overshoots the correlation effect, but often gives a better answer than MP3, at least if medium sized basis sets are used. Just as the first term involving doubles (MP2) tends to overestimate the correlation effect, it is often observed that MP4 overestimates the effect of the singles and triples contributions, since they enter the series for the first time at fourth order.

Chapter 7

Coupled Cluster

Coupled cluster (CC) is nowadays one of the most prevalent methods in quantum chemistry that includes electronic correlation.
Perturbation methods (like MPn) add all **type of corrections** (S,D,T, etc) to the reference wavefunction up to a given **order** (2,3,4, etc) [1].
The idea in Coupled Cluster (CC) methods is to include **all corrections of a given type** (S,D,T, etc) to **infinite order**.
The wavefunction and the ground state energy are denoted by $|\Psi\rangle$ and E, respectively. Other variants of the coupled-cluster theory, such as *equation-of-motion coupled cluster* and *multi-reference coupled cluster* may also produce approximate solutions for the excited states of the system.
Properties:

- CC is non-variational
- CC is size-extensive
- CC is size-consistent if the reference wavefunction is size consistent

The wavefunction of the coupled-cluster theory is written as an exponential ansatz:

$$|\Psi\rangle = e^{\hat{T}}|\Phi_0\rangle \tag{7.1}$$

where $|\Phi_0\rangle$ is a Slater determinant usually constructed from Hartree-Fock molecular orbitals. The so called **cluster operator**,$\hat{T}$, is an excitation operator which, when acting on $|\Phi_0\rangle$, produces a linear combination of excited Slater determinants.
The choice of the exponential ansatz is opportune because (unlike other ansatzes, for example, configuration interaction) it guarantees the size

---

[1] The order of a correction corresponds to the exponent in the perturbation expansion.

consistency [2] and size extensivity [3] of the solution.

7.1 The cluster operator

The cluster operator is written in the form

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \tag{7.2}$$

where $\hat{T}_1$ is the operator of all single excitations, $\hat{T}_2$ is the operator of all double excitations and so forth. In the formalism of second quantization these excitation operators are conveniently expressed as

$$\hat{T}_1 = \sum_a \sum_r t_a^r \hat{a}_a \hat{a}_r^\dagger \tag{7.3}$$

$$\hat{T}_2 = \frac{1}{4} \sum_{a,b} \sum_{r,s} t_{ab}^{rs} \hat{a}_a \hat{a}_b \hat{a}_r^\dagger \hat{a}_s^\dagger \tag{7.4}$$

In the above formulae $\hat{a}^\dagger$ and $\hat{a}$ denote the creation and annihilation operators, respectively and a, b stand for occupied and r, s for unoccupied orbitals. The creation and annihilation operators in the coupled cluster terms above are written in canonical form, where each term is in normal order (i.e. with all creation operators on the right hand side). The one-particle excitation operator and the two-particle excitation operator, $\hat{T}_1$ and $\hat{T}_2$ convert the reference function $|\Phi_0\rangle$ into a linear combination of the singly- and doubly-excited Slater determinants, respectively. Solving for the unknown coefficients t_a^r and t_{ab}^{rs} is necessary for finding the approximate solution $|\Psi\rangle$.

Taking into consideration the structure of $\hat{T}$, the exponential operator $e^{\hat{T}}$ may be expanded into a Taylor series:

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \dots = 1 + \hat{T}_1 + \hat{T}_2 + \frac{\hat{T}_1^2}{2!} + \hat{T}_1\hat{T}_2 + \frac{\hat{T}_2^2}{2!} + \dots \tag{7.5}$$

This series is finite in practice because the number of molecular orbitals is finite, as is the number of excitations. In order to simplify the task for finding the coefficients $t_{ij\dots}^{ab\dots}$, the expansion into individual excitation

---

[2] Size consistency (or strict separability) is a property that guarantees the consistency of the energy behavior when interaction between the involved molecular system is nullified (for example, by increasing the distance between two moieties: $E(AB)$ has the right limit, $E(A) + E(B)$ for large separation of the two molecules, A and B.)

[3] In physics and chemistry an intensive property (also called a bulk property) of a system is a physical property of the system that does not depend on the system size or the amount of material in the system. By contrast, an extensive property of a system does depend on the system size or the amount of material in the system.

operators is terminated at the second or slightly higher level of excitation (rarely exceeding four). This approach is warranted by the fact that even if the system admits more than four excitations, the contribution of $\hat{T}_5$, $\hat{T}_6$ etc to the operator $\hat{T}$ is small. Furthermore, if the highest excitation level in the $\hat{T}$ operator is n,

$$\hat{T} = 1 + \hat{T}_1 + \cdots + \hat{T}_n \tag{7.6}$$

then Slater determinants excited more than n times may (and usually do) still contribute to the wave function $|\Psi\rangle$ because of the non-linear nature of the exponential ansatz (i.e. for $n = 2$ the expansion contains terms like $\hat{T}_1\hat{T}_2$ and $\hat{T}_2^2$, which correspond to triple and quadruple excitations, respectively). Therefore, coupled cluster terminated at $\hat{T}_n$ usually recovers more correlation energy than configuration interaction with maximum n excitations.

7.2 The coupled cluster energy

With the coupled cluster wave function the Schrödinger equation becomes

$$\hat{H}e^{\hat{T}}|\Phi_0\rangle = Ee^{\hat{T}}|\Phi_0\rangle \tag{7.7}$$

Multiplying from the left by $\langle\Phi_0|$ and integrating gives

$$\begin{aligned} \langle\Phi_0|He^{\hat{T}}|\Phi_0\rangle &= E_{cc}\langle\Phi_0|e^{\hat{T}}\Phi_0\rangle & (7.8)\\ &= E_{cc}\langle\Phi_0|(\hat{1} + \hat{T}_1 + \hat{T}_2 + \dots)\Phi_0\rangle & (7.9) \end{aligned}$$

and therefore

$$E_{cc} = \langle\Phi_0|He^{\hat{T}}|\Phi_0\rangle \tag{7.10}$$

(all terms $\langle\Phi_0|\hat{T}_i|\Phi_0\rangle$ are zero. Why?).
In the case in which excitations are limited to single and double excited Slater determinants, the cc energy simplifies to

$$E_{cc} = E_0 + \sum_a^{\text{occ}}\sum_r^{\text{vir}} t_a^r\langle\Phi_0|\hat{H}|\Phi_a^r\rangle + \sum_{ab(a<b)}^{\text{occ}}\sum_{rs(r<s)}^{\text{vir}}(t_{ab}^{rs} + t_a^r t_b^s - t_a^s t_b^r)\langle\Phi_0|\hat{H}|\Phi_{ab}^{rs}\rangle \tag{7.11}$$

When using HF orbitals for constructing the Slater determinants, the first matrix elements are zero (Brillouin's theorem) and the second matrix elements are just two-electron integrals over molecular orbitals,

$$E_{cc} = E_0 + \sum_{ab(a<b)}^{\text{occ}}\sum_{rs(r<s)}^{\text{vir}}(t_{ab}^{rs} + t_a^r t_b^s - t_a^s t_b^r)(\langle\phi_a\phi_b|\hat{v}|\phi_r\phi_s\rangle - \langle\phi_a\phi_b|\hat{v}|\phi_s\phi_r\rangle) \tag{7.12}$$

What is still missing are the expansion amplitudes, $t_{ab\dots}^{rs\dots}$.

7.3 The coupled cluster equations for amplitudes, $t_{ab\ldots}^{rs\ldots}$

Coupled-cluster equations are equations whose solution is the set of coefficients $t_{ab\ldots}^{rs\ldots}$. There are several ways of writing such equations but the standard formalism results in a closed set of equations which may be solved iteratively. The interested students are invited to read the literature (see for example the book of Szabo and Ostlund, pp. 286-319).

7.4 Types of coupled-cluster methods

The classification of traditional coupled-cluster methods rests on the highest number of excitations allowed in the definition of $\hat{T}$. The abbreviations for coupled-cluster methods usually begin with the letters "CC" (for coupled cluster) followed by

S - for single excitations (shortened to singles in coupled-cluster terminology)

D - for double excitations (doubles)

T - for triple excitations (triples)

Q - for quadruple excitations (quadruples)

Thus, the $\hat{T}$ operator in CCSDT has the form

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$$

Terms in round brackets indicate that these terms are calculated based on perturbation theory. For example, a CCSD(T) approach simply means:

1. A coupled-cluster method
2. It includes singles and doubles fully
3. Triples are calculated with perturbation theory.

Chapter 8

Density functional theory

The content of this chapter is partially adapted from:
Klaus Capelle, *A Bird's-Eye View of Density-Functional Theory*,
arXiv:cond-mat/0211443v5

The electron density is a more attractive quantity [1], than the many particle wave function which depends on all coordinates of all particles, i.e., for N electrons, it depends on $3N$ variables (or $4N$ if you count in spin).

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = M_k \det |\psi_A(\mathbf{x}_1)\psi_B(\mathbf{x}_2) \ldots \psi_X(\mathbf{x}_N)|$$

$$\Big\downarrow$$

$$\rho(\mathbf{x}_1) = M \int \cdots \int \Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)\, \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)\, d\mathbf{x}_2 \ldots d\mathbf{x}_N$$

where M is a normalization factor.

1

Why DFT?

A practical reason

The complete oxygen wavefunction, $\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, (8 electrons) depends on 24 coordinates. To store this wavefunction we would need

10 entries for each coordinate $\longrightarrow$ 10^{24} entries

8 byte per entry $\longrightarrow$ 10^{25} bytes

5×10^9 bytes per DVD $\longrightarrow$ 2×10^{14} DVDs

10 g per DVD $\longrightarrow$ 2×10^9 t DVDs

Instead, in DFT we will need about 8×10^3 bytes to store the density of the system, which requires less than a floppy disk!

---

[1] depends only on x, y, z, and eventually, there may be two densities for spin polarized systems, one for spin up electrons $\rho_\uparrow(\mathbf{r})$ and one for spin down electrons $\rho_\downarrow(\mathbf{r})$

A physical reason

DFT has many advantages compared to other *ab initio* techniques

- DFT is computationally very efficient.
- DFT is conceptually simple.
  "A great strength of DFT language is its appropriateness for defining and elucidating important universal concepts of molecular structure and molecular reactivity. In traditional quantum chemistry this has, of course, also been a major goal, but it is tortuous to try to conceptualize how many-body wavefunctions are related to structure and behavior. In DFT not only is the electron density itself very easy to visualize, but there is the big advantage that the electron number N has a central place in the theory. After all, much of the chemistry is about transfer of electrons from one place to another." (W. Kohn,A.D.Becke, R.G. Parr, *J. Phys. Chem*, 12974-12980,**1996**.)
  Concepts like *chemical potential* (*electronegativity*), *hardness*, *polarizability*, *response functions* and *reactivity functions (Fukui functions)* emerge naturally from the DFT energy functional, $E[\rho]$, and its derivatives, and their computation becomes possible.
- DFT can be easily combined with molecular dynamics of the nuclei.
  Forces on classical ions can be computed using the Hellmann-Feynman theorem starting from the electronic density.

8.1 What is density functional theory

Density functional theory is one of the most popular and successful quantum mechanical approaches to matter. It is nowadays routinely applied for calculating, e.g., the binding energy of molecules in chemistry and the band structure of solids in physics. Also applications relevant for fields traditionally considered more distant from quantum mechanics, such as biology and mineralogy are beginning to appear. Superconductivity, atoms in the focus of strong laser pulses, relativistic effects in heavy elements and in atomic nuclei, classical liquids, and magnetic properties of alloys have all been studied with DFT.

In quantum mechanics we learn that all information we can possibly have about a given system is contained in the system's wave function, Ψ, which is the lowest energy solution of the many-electron Schrödinger equation

$$\Big[-\frac{1}{2}\sum_i \nabla_i^2 - \sum_{I,i}\frac{Z_I}{|\mathbf{R}_I-\mathbf{r}_i|} + \sum_{I>J}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|} + \sum_{i>j}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}\Big]\Psi(\mathbf{r},\mathbf{R}) = E_{el}\Psi(\mathbf{r},\mathbf{R}) \tag{8.1}$$

or, in a more compact notation,

$$\big[\hat{T}_e(\mathbf{r}) + \hat{V}_{eN}(\mathbf{r},\mathbf{R}) + \hat{V}_{NN}(\mathbf{R}) + \hat{V}_{ee}(\mathbf{r})\big]\Psi(\mathbf{r},\mathbf{R}) = E_{el}\Psi(\mathbf{r},\mathbf{R}) \tag{8.2}$$

Convention. In the following we use $v(\mathbf{r},\mathbf{R})$ for the external potential generated by the nuclei at position $\mathbf{R}_I$ ($\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{Nu})$. Since we usually consider the nuclei kept fixed in space, we will sometimes drop the dependence on $\mathbf{R}$, and write simply $v(\mathbf{r}) = v(\mathbf{r},\mathbf{R})$.

The usual quantum mechanical approach to the Schrödinger equation (SE) can be summarized by the following sequence

$$v(\mathbf{r},\mathbf{R}) \overset{SE}{\Longrightarrow} \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) \overset{\langle\Psi|\ldots|\Psi\rangle}{\Longrightarrow} \text{observables}$$

i.e., one specifies the system by choosing $v(\mathbf{r},\mathbf{R})$, plugs it into the Schrödinger equation, solves that equation for the wave function Ψ, and then calculates observables by taking expectation values of operators with this wave function. One among the observables that are calculated in this way is the electron density, $\rho(\mathbf{r})$.

In the previous chapters, we have investigated methods for the solution of the many-electron Schrödinger equation using a systematic expansion in Slater determinants. The problem with these methods is the great demand they place on one's computational resources: it is simply impossible to apply them efficiently to large and complex systems. Nobody has ever calculated the chemical properties of a 100-atom molecule with full CI [2]. DFT provides a viable and accurate [3] alternative to post-HF methods for the computation of medium (100 electrons) to big size (thousands of electrons) systems.

DFT explicitly recognizes that nonrelativistic Coulomb systems differ only by their potential $v(\mathbf{r},\mathbf{R})$, and supplies a prescription for dealing with the universal operators $\hat{T}_e$ and $\hat{V}_{ee}$ once and for all. Furthermore, DFT provides

---

[2] A simple estimate of the computational complexity of this task is to imagine a real space representation of Ψ on a mesh, in which each coordinate is discretized by using 20 mesh points (which is not very much). For N electrons, Ψ becomes a function of $3N$ coordinates (ignoring spin, and taking Ψ to be real), and 20^{3N} values are required to describe Ψ on the mesh. The density $n(\mathbf{r})$ is a function of three coordinates, and requires 20^3 values on the same mesh. CI and the Kohn-Sham formulation of DFT additionally employ sets of single-particle orbitals. N such orbitals, used to build the density, require 20^3N values on the same mesh. (A CI calculation employs also unoccupied orbitals, and requires more values.) For $N = 10$ electrons, the many-body wave function thus requires $20^{30}/20^3 \approx 10^{35}$ times more storage space than the density, and $20^{30}/(10 \times 20^3) \approx 10^{34}$ times more than a set of single-particle orbitals. Clever use of symmetries can reduce these ratios, but the full many-body wave function remains inaccessible for real systems with more than a few electrons.

[3] Accuracy is a relative term. As a theory, DFT is formally exact. Its performance in actual applications depends on the quality of the approximate density functionals employed. For small numbers of particles, or systems with special symmetries, essentially exact solutions of the Schrödinger equation can be obtained, and no approximate functional can compete with exact solutions. For more realistic systems, modern (2007) sophisticated density functionals attain rather high accuracy. Bond lengths of molecules can be predicted with an average error of less than 0.01 Å, lattice constants of solids with an average error of less than 0.05 Å, and molecular energies to within less than 5 kcal/mol. (For comparison: already a small molecule, such as water, has a total energy of $48'000$ kcal/mol).

a way to systematically map the many-body problem, with $\hat{V}_{ee}$, onto a single-body problem, without $\hat{V}_{ee}$. All this is done by promoting the particle density $\rho(\mathbf{r})$ from just one among many observables to the status of key variable, on which the calculation of all other observables can be based. This approach forms the basis of the large majority of electronic-structure calculations in physics and chemistry. Much of what we know about the electrical, magnetic, and structural properties of materials has been calculated using DFT, and the extent to which DFT has contributed to the science of molecules is reflected by the 1998 Nobel Prize in Chemistry, which was awarded to Walter Kohn, the founding father of DFT, and John Pople, who was instrumental in implementing DFT in computational chemistry. The density-functional approach can be summarized by the sequence

$$\rho(\mathbf{r}) \Longrightarrow \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) \Longrightarrow v(\mathbf{r}) \tag{8.3}$$

i.e., knowledge of $\rho(\mathbf{r})$ implies knowledge of the wave function and the potential, and hence of all other observables. This is valid for a given fixed position of the nuclei and therefore we remove the explicit dependence on $\mathbf{R}$ from all quantities). Although this sequence describes the conceptual structure of DFT, it does not really represent what is done in actual applications of it, which typically proceed along rather different lines, and do not make explicit use of many-body wave functions.

The literature on DFT is large, and rich in excellent reviews and overviews. Some representative examples of full reviews and systematic collections of research papers are:

1. R. M. Dreizler and E. K. U. Gross, *Density Functional Theory*, Springer, Berlin, 1990.
2. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, 1989.
3. W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, John Wiley & Sons, New York, 2001.
4. R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* 61, 689 (1989).
5. J. M. Seminario (Ed.), *Recent Developments and Applications of Modern DFT*, Elsevier, Amsterdam, 1996.

8.2 Functionals and their derivatives

Before we discuss density-functional theory more carefully, let us introduce a useful mathematical tool. Since according to the above sequence the wave function is determined by the density, we can write $\Psi = \Psi[\rho](\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ which indicates that Ψ is a function of its N spatial variables, but a functional of $\rho(\mathbf{r})$. A functional $F[\rho]$ can be defined (in a mathematically sloppy way) as a rule for going from a function to a number, just as a function $y = f(x)$ is a rule (f) for going from a number (x) to another number (y). A simple example of a functional is the particle number,

$$N = \int d^3r\, \rho(\mathbf{r}) = N[\rho] \tag{8.4}$$

which is a rule for obtaining the number N, given the function $\rho(\mathbf{r})$. Note that the name given to the argument of ρ is completely irrelevant, since the functional depends on the function itself, not on its variable. Hence we do not need to distinguish $F[\rho(\mathbf{r})]$ from, e.g., $F[\rho(\mathbf{r}')]$. Another important case is that in which the functional depends on a parameter, such as in

$$v_H[\rho](\mathbf{r}) = \int d^3r'\, \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{8.5}$$

which is a rule that for any value of the parameter $\mathbf{r}$ associates a value $v_H[\rho](\mathbf{r})$ with the function $\rho(\mathbf{r}')$. This term is the so-called Hartree potential, which we have already encountered in the HF theory.

8.2.1 Functional variation

Given a function of one variable, $y = f(x)$, one can think of two types of variations of y, one associated with x, the other with f. For a fixed functional dependence $f(x)$, the ordinary differential dy measures how y changes as a result of a variation $x \to x + dx$ of the variable x. This is the variation studied in ordinary calculus. Similarly, for a fixed point x, the functional variation δy measures how the value y at this point changes as a result of a variation in the functional form $f(x)$. This is the variation studied in variational calculus.

8.2.2 Functional derivative

The derivative formed in terms of the ordinary differential, df/dx, measures the first-order change of $y = f(x)$ upon changes of x, i.e., the slope of the function $f(x)$ at x:

$$f(x + dx) = f(x) + \frac{df}{dx}dx + \mathcal{O}(dx^2) \tag{8.6}$$

The functional derivative measures, similarly, the first-order change in a functional upon a functional variation of its argument:

$$F[f(x) + \delta f(x)] = F[f(x)] + \int dxs(x)\delta f(x)dx + O(\delta f^2)\,, \tag{8.7}$$

where the integral arises because the variation in the functional F is determined by variations in the function at all points in space. The first-order coefficient (or 'functional slope') $s(x)$ is defined to be the functional derivative $\delta F[f]/\delta f(x)$,

$$s(x) = \frac{\delta F[f]}{\delta f(x)} \tag{8.8}$$

The functional derivative allows us to study how a functional changes upon changes in the form of the function it depends on. Detailed rules for calculating functional derivatives are described in Appendix A of the book of Parr and Yang.

A general expression for obtaining functional derivatives with respect to $\rho(r)$ of a functional $F[\rho] = \int dr f(\rho, \rho', \rho'', \rho''', \ldots; r)$, where primes indicate ordinary derivatives of $\rho(r)$ with respect to r, is

$$\frac{\delta F[\rho]}{\delta \rho(x)} = \frac{\partial f}{\partial \rho} - \frac{d}{dx}\frac{\partial f}{\partial \rho'} + \frac{d^2}{dx^2}\frac{\partial f}{\partial \rho''} - \frac{d^3}{dx^3}\frac{\partial f}{\partial \rho'''} + \ldots \tag{8.9}$$

This expression is frequently used in DFT to obtain exchange-correlation (**xc**) potentials from xc energies.

8.3 The Hohenberg-Kohn theorem

At the heart of DFT is the Hohenberg-Kohn (HK) theorem. This theorem states that for ground states the equation

$$\rho(\mathbf{r}) = M \int \cdots \int \Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)\, \Psi^*(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)\, d\mathbf{r}_2 \ldots d\mathbf{r}_N \tag{8.10}$$

can be inverted: given a ground state density $\rho_0(\mathbf{r})$ it is possible, in principle, to calculate the corresponding ground state wave function $\Psi_0(\mathbf{r}_1, \mathbf{r}_2 ..., \mathbf{r}_N)$. This means that Ψ_0 is a functional of ρ_0. Consequently, all ground state observables are functionals of ρ_0, too. If Ψ can be calculated from ρ_0 and vice versa, both functions are equivalent and contain exactly the same information. At first sight this seems impossible: how can a function of one (vectorial) variable $\mathbf{r}$ be equivalent to a function of N (vectorial) variables $\mathbf{r}_1, \mathbf{r}_2 ..., \mathbf{r}_N$? How can one arbitrary variable contain the same information as N arbitrary variables?

The crucial fact which makes this possible is that knowledge of $\rho_0(\mathbf{r})$ implies implicit knowledge of much more than that of an arbitrary function $f(\mathbf{r})$. The ground-state wave function Ψ_0 must not only reproduce the ground-state density, but also minimize the energy. For a given ground-state density $\rho_0(\mathbf{r})$, we can write this requirement as

$$E_{v,0} = \underset{\Psi \to \rho_0}{min} \langle \Psi | \hat{T} + \hat{V}_{ee} + \hat{V}_{eN} | \Psi \rangle \tag{8.11}$$

where $E_{v,0}$ denotes the ground-state energy in potential $v(\mathbf{r})$. The preceding equation tells us that for a given density $\rho_0(\mathbf{r})$ the ground state wave function Ψ_0 is that which reproduces this $\rho_0(\mathbf{r})$ and minimizes the energy. For an arbitrary density $\rho(\mathbf{r})$, we define the functional

$$E_v[\rho] = \underset{\Psi\to\rho}{min}\langle\Psi|\hat{T} + \hat{V}_{ee} + \hat{V}_{eN}|\Psi\rangle\,. \tag{8.12}$$

If ρ is a density different from the ground state density ρ_0 in the external potential $v(\mathbf{r})$, then the Ψ that produce this density ρ is different from the ground state wave function Ψ_0, and according to the variational principle the minimum obtained from $E_v[\rho]$ is higher than (or equal to) the ground state energy $E_{v,0} = E_v[\rho_0]$. Thus, the functional $E_v[\rho]$ is minimized by the ground state density ρ_0, and its value at the minimum is $E_{v,0}$.

The total energy functional can be written as

$$\begin{aligned} E_v[\rho] &= \underset{\Psi\to\rho}{min}\langle\Psi|\hat{T} + \hat{V}_{ee}|\Psi\rangle + \int d^3r\,\rho(\mathbf{r})v(\mathbf{r}) && (8.13)\\ &=: F[\rho] + V[\rho] && (8.14) \end{aligned}$$

where the internal energy functional $F[\rho] = \underset{\Psi\to\rho}{min}\langle\Psi|\hat{T} + \hat{V}_{ee}|\Psi\rangle$ is independent of the potential $v(\mathbf{r})$, and thus determined only by the structure of the operators $\hat{T}$ and $\hat{V}_{ee}$. This universality of the internal energy functional allows us to define the ground-state wave function Ψ_0 as that antisymmetric N-particle function that delivers the minimum of $F[\rho]$ and reproduces ρ_0. If the ground state is nondegenerate, this double requirement uniquely determines Ψ_0 in terms of $\rho_0(\mathbf{r})$, without having to specify $v(\mathbf{r})$ explicitly. [4] Equations 8.11 to 8.14 constitute the constrained-search proof of the Hohenberg-Kohn theorem, given independently by M. Levy and E. Lieb. The original proof by Hohenberg and Kohn proceeded by assuming that Ψ_0 was not determined uniquely by ρ_0 and showed that this produced a contradiction to the variational principle. Both proofs, by constrained search and by contradiction, are elegant and simple. In fact, it is a bit surprising that it took 38 years from Schrödinger's first papers on quantum mechanics to Hohenberg and Kohn's 1964 paper containing their famous theorem.

8.3.1 Meaning and implications of the Hohenberg-Kohn theorem

Here we provide a commented summary of the content of the Hohenberg-Kohn (HK) theorem.

---

[4] Note that this is exactly the opposite of the conventional prescription to specify the Hamiltonian via $v(\mathbf{r})$, and obtain Ψ from solving Schrödinger equation, without having to specify $\rho(\mathbf{r})$ explicitly.

(1) The nondegenerate ground-state (GS) wave function is a unique functional of the ground state (GS) density [5]

$$\Psi_0(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \Psi[\rho_0(\mathbf{r})]\,. \tag{8.15}$$

This is the essence of the HK theorem (often called the **first Hohenberg-Kohn theorem**). As a consequence, the GS expectation value of any observable $\hat{O}$ is a functional of $\rho_0(\mathbf{r})$, too

$$O_0 = O[\rho_0] = \langle \Psi[\rho_0] | \hat{O} | \Psi[\rho_0] \rangle\,. \tag{8.16}$$

(2) For the special case in which the observable is the energy of the system,

$$E_{v,0} = O[\rho_0] = E_v[\rho_0] = \langle \Psi[\rho_0] | \hat{H} | \Psi[\rho_0] \rangle\,. \tag{8.17}$$

where, $\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V}_{eN}$, the following variational property holds

$$E_v[\rho_0] \leq E_v[\rho'] \tag{8.18}$$

where ρ_0 is GS density in potential $\hat{V}$ and ρ' is some other density. This is very similar to the usual variational principle for wave functions. From a calculation of the expectation value of a Hamiltonian with a trial wave function Ψ' that is not its GS wave function Ψ_0 one can never obtain an energy below the true GS energy,

$$E_{v,0} = E_v[\Psi_0] = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \leq \langle \Psi' | \hat{H} | \Psi' \rangle = E_v[\Psi'] \tag{8.19}$$

Similarly, in exact DFT, if $E[\rho]$ for fixed v_{ext} is evaluated for a density that is not the GS density of the system in potential v_{ext}, one never finds a result below the true GS energy. This is what 8.18 says, and it is so important for practical applications of DFT that it is sometimes called the **second Hohenberg-Kohn theorem**.

In performing the minimization of $E_v[\rho]$ the constraint that the total particle number N is an integer is taken into account by means of a Lagrange multiplier, replacing the constrained minimization of $E_v[\rho]$ by an unconstrained one of $E_v[\rho] - \mu N$. Since $N = \int d^3r\, \rho(\mathbf{r})$, this leads to

$$\frac{\delta E_v[\rho]}{\delta \rho(\mathbf{r})} = \mu = \frac{\partial E}{\partial N} \tag{8.20}$$

where μ is the chemical potential.

---

[5] If the ground state is degenerate, several of the degenerate ground-state wave functions may produce the same density, so that a unique functional $\Psi[\rho]$ does not exist, but by definition these wave functions all yield the same energy, so that the functional $E_v[\rho]$ continues to exist and to be minimized by ρ_0. A universal functional $F[\rho]$ can also still be defined.

(3) Recalling that the kinetic and interaction energies of a nonrelativistic Coulomb system are described by universal operators, we can also write E_v as

$$E_v[\rho] = T[\rho] + V_{ee}[\rho] + V_{eN}[\rho] = F[\rho] + V_{eN}[\rho]\,, \tag{8.21}$$

where $T[\rho]$ and $V_{ee}[\rho]$ are universal functionals (defined as expectation values of the type of Eq. 8.17), independent of $v_{eN}(\mathbf{r}) = v_{ext}(\mathbf{r})$ (sometimes simply called $v(\mathbf{r})$). On the other hand, the potential energy in a given potential $v(\mathbf{r})$ is the expectation value of the potential

$$\hat{V}_{ext}(\mathbf{r}) = \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \tag{8.22}$$

which reads

$$V_{ext}[\rho] = \int d^3r\, \rho(\mathbf{r}) v_{ext}(\mathbf{r}) \tag{8.23}$$

and is obviously nonuniversal (it depends on $v_{ext}(\mathbf{r})$, i.e., on the system under study), but very simple: once the system is specified, i.e., $v_{ext}(\mathbf{r})$ is known, the functional $V_{ext}[\rho]$ is known explicitly.

8.3.2 Density-Functional Theory in Practice

After these abstract considerations let us now consider one way in which one can make practical use of DFT. Assume we have specified our system (i.e., $v(\mathbf{r})$ is known). Assume further that we have reliable approximations for $V_{ee}[\rho]$ and $T[\rho]$. In principle, all one has to do then is to minimize the sum of kinetic, interaction and potential energies

$$E_v[\rho] = T[\rho] + V_{ee}[\rho] + V_{eN}[\rho] = T[\rho] + V_{ee}[\rho] + \int d^3r\, \rho(\mathbf{r}) v_{ext}(\mathbf{r}) \tag{8.24}$$

with respect to $\rho(\mathbf{r})$. The minimizing function $\rho_0(\mathbf{r})$ is the system's GS charge density and the value $E_{v,0} = E_v[\rho_0]$ is the GS energy. Assume now that $v_{ext}(\mathbf{r})$ depends on a parameter a. This can be, for example, the lattice constant in a solid or the angle between two atoms in a molecule. Calculation of $E_{v,0}$ for many values of a allows one to plot the curve $E_{v,0}(a)$ and to find the value of a that minimizes it. This value, a_0, is the GS lattice constant or angle. In this way one can calculate quantities like molecular geometries, lattice constants, unit cell volumes, charge distributions, total energies, etc.

By looking at the change of $E_{v,0}(a)$ with a one can, moreover, calculate compressibilities, phonon spectra and bulk moduli (in solids) and vibrational frequencies (in molecules). By comparing the total energy of a composite system (e.g., a molecule) with that of its constituent systems (e.g., individual atoms) one obtains dissociation energies. By calculating the total energy for systems with one electron more or less one obtains electron affinities and ionization energies. Using the Hellman-Feynman theorem [6] one can calculate forces on atoms from the derivative of the total energy with respect to the nuclear coordinates. All this follows from DFT without having to solve the many-body Schrödinger equation and without having to make a single-particle approximation.

In theory it should be possible to calculate all observables, since the HK theorem guarantees that they are all functionals of ρ_0. In practice, one does not know how to do this explicitly. Another problem is that the minimization of $E_v[\rho]$ is, in general, a tough numerical problem on its own. And, moreover, one needs reliable approximations for $T[\rho]$ and $V_{ee}[\rho]$ to begin with. In the next section, on the Kohn-Sham equations, we will see one widely used method for solving these problems. Before looking at that, however, it is worthwhile to recall an older, but still occasionally useful, alternative: the Thomas-Fermi approximation.

8.3.3 The Thomas-Fermi approximation and the local density approximation (LDA)

In this approximation one sets

$$V_{ee} \approx V_H = \frac{1}{2}\int d^3r \int d^3r' \, \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{8.25}$$

i.e., approximates the full interaction energy by the Hartree energy, the electrostatic interaction energy of the charge distribution $\rho(\mathbf{r})$. One further approximates, initially,

$$T[\rho] \approx T^{LDA}[\rho] = \int t^{hom}(\rho(\mathbf{r}))d^3r \tag{8.26}$$

where $t^{hom}(\rho)$ is the kinetic-energy density of a homogeneous interacting system with (constant) density ρ. Since it refers to interacting electrons $t^{hom}(\rho)$ is not known explicitly, and further approximations are called for. As it stands, however, this formula is already a first example of a local density approximation (LDA). In this type of approximation one imagines

---

[6] For the case of $a = \mathbf{R}_I$ (the coordinate of a nucleus I), the Hellmann-Feynman theorem says that, for the GS wavefunction or GS density,

$$\begin{aligned} F_I &= \frac{\partial \langle E_{GS}\rangle}{\partial \mathbf{R}_I} \\ &= \int d^{3N} \Psi^*(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \frac{\partial \hat{H}}{\partial \mathbf{R}_I} \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \\ &= \frac{\partial E_{v,0}[\rho](\mathbf{R}_I)}{\partial \mathbf{R}_I}\,. \end{aligned}$$

the real inhomogeneous system (with density $\rho(\mathbf{r})$ in potential $v(\mathbf{r})$) to be decomposed in small cells in each of which $\rho(\mathbf{r})$ and $v(\mathbf{r})$ are approximately constant. In each cell (i.e., locally) one can then use the per-volume energy of a homogeneous system to approximate the contribution of the cell to the real inhomogeneous one. Making the cells infinitesimally small and summing over all of them yields Eq. 8.26.

For a noninteracting system (specified by subscript s, for 'single-particle') the function $t_s^{hom}(\rho)$ is known explicitly,

$$t_s^{hom}(\rho) = 3\hbar^2(3\pi^2)^{2/3}\rho^{5/3}/(10m_e) \tag{8.27}$$

We can therefore use the following approximation

$$T[\rho] \approx T^{LDA}[\rho] \approx T_s^{LDA}[\rho] = \int t_s^{hom}(\rho(\mathbf{r}))d^3r \tag{8.28}$$

where $T_s^{LDA}[\rho]$ is the local-density approximation to $T[\rho]$. The Thomas-Fermi approximation then consists in combining Eq. 8.25 with Eq. 8.26 and writing

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{eN}[\rho] \approx E^{TF}[\rho] = T_s^{LDA}[\rho] + V_H[\rho] + V_{eN}[\rho]. \tag{8.29}$$

A major defect of the Thomas-Fermi approximation is that within it molecules are unstable: the energy of a set of isolated atoms is lower than that of the bound molecule. This fundamental deficiency, and the lack of accuracy resulting from neglect of correlations in 8.25 and from using the local approximation 8.28 for the kinetic energy, limit the practical use of the Thomas-Fermi approximation in its own right. However, it is a most useful starting point for a large body of work on improved approximations in chemistry and physics. More recent approximations for $T[\rho]$ can be found in the context of orbital-free DFT. The extension of the local-density concept to the exchange-correlation energy is at the heart of many modern density functionals.

8.4 Kohn-Sham Density Functional Theory

Density-functional theory can be implemented in many ways. The minimization of an explicit energy functional, discussed up to this point, is not normally the most efficient among them. Much more widely used is the Kohn-Sham approach. Interestingly, this approach owes its success and popularity partly to the fact that it does not exclusively work in terms of the particle (or charge) density, but brings a special kind of wave functions (single-particle orbitals) back into the game. As a consequence, DFT then looks formally like a single-particle theory, although many-body effects are still included via the so-called **exchange-correlation functional**. We will now see in some detail how this is done.

8.4.1 The exchange-correlation energy

The Thomas-Fermi approximation for $T[\rho]$ is not very good. A more accurate scheme for treating the kinetic-energy functional of interacting electrons, $T[\rho]$, is based on decomposing $T[\rho]$ into one part that represents the kinetic energy of noninteracting particles of density ρ, i.e., the quantity called above $T_s[\rho]$, and one that represents the remainder, denoted $T_c[\rho]$ (the sub-scripts s and c stand for 'single-particle' and 'correlation', respectively) [7].

$$T[\rho] = T_s[\rho] + T_c[\rho]\,. \tag{8.30}$$

$T_s[\rho]$ is not known exactly as a functional of ρ, but it is easily expressed in terms of the **single-particle orbitals** $\phi_i(\mathbf{r})$ of a noninteracting system with the density ρ, as

$$T_s[\rho] = -\frac{1}{2}\sum_i^N \int d^3r\, \phi_i^*(\mathbf{r})\nabla^2\phi_i(\mathbf{r}) \tag{8.31}$$

because for noninteracting particles the total kinetic energy is just the sum of the individual kinetic energies!

Since all $\phi_i(\mathbf{r})$ are functionals of ρ ($\rho(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2$), this expression for T_s is an explicit orbital functional but an implicit density functional, $T_s[\rho] = T_s[\phi_i[\rho]]$, where the notation indicates that T_s depends on the full set of occupied orbitals ϕ_i, each of which is a functional of ρ. We now rewrite the exact energy functional as

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{eN}[\rho] = T_s[\{\phi_i[\rho]\}] + V_H[\rho] + E_{xc}[\rho] + V_{eN}[\rho], \tag{8.32}$$

where by definition E_{xc} contains the differences $T - T_s$ (i.e. T_c) and $V_{ee} - V_H$. This definition shows that a part of the correlation energy E_c is due to the difference T_c between the noninteracting and the interacting kinetic energies. Unlike the Thomas-Fermi equation, Eq. 8.36 is formally exact, but of course the functional E_{xc} is unknown.

The functional $E_{xc}[\rho]$ is called the **exchange-correlation (xc) energy**. It is often decomposed into $E_{xc} = E_x + E_c$, where E_x is due to the Pauli principle (exchange energy) and E_c is due to electron correlation. (T_c is then a part of E_c.) The exchange energy can be written explicitly in terms of the single-particle orbitals as

$$E_x[\{\phi_i[\rho]\}] = -\frac{1}{2}\sum_{ij}\int d^3r \int d^3r' \frac{\phi_j(\mathbf{r})\phi_k(\mathbf{r})\phi_j(\mathbf{r})\phi_k(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} \tag{8.33}$$

---

[7] T_s is defined as the expectation value of the kinetic-energy operator $\hat{T}$ with the Slater determinant (Ψ) defined by the density ρ, i.e., $T_s[\rho] = \langle\Phi[\rho]|\hat{T}|\Phi[\rho]\rangle$. Similarly, the full kinetic energy is defined as $T[n] = \langle\Psi[\rho]|\hat{T}|\Psi[\rho]\rangle$. All consequences of antisymmetrization (i.e., exchange) are described by employing a determinantal wave function in defining T_s. Hence, T_c, the difference between T_s and T is a pure correlation effect.

8.4.2 The Kohn-Sham equations

Since T_s is now written as an orbital functional one cannot directly minimize Eq. 8.36 with respect to ρ. Instead, one commonly employs a scheme suggested by Kohn and Sham for performing the minimization indirectly.

This scheme starts by writing the minimization as

$$\begin{aligned} 0 &= \frac{\delta E[\rho]}{\delta\rho(\mathbf{r})} && (8.34)\\ &= \frac{\delta T_s[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta V_{ext}[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta V_H[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})} && (8.35)\\ &= \frac{\delta T_s[\rho]}{\delta\rho(\mathbf{r})} + v_{ext}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}) && (8.36) \end{aligned}$$

As a consequence of Eq. 8.22, $\delta V_{ext}/\delta\rho(\mathbf{r}) = v_{ext}(\mathbf{r})$, the 'external' potential the electrons move in (we use V_{ext} for V_{eN}). The term $\delta V_H/\rho(\mathbf{r})$ simply yields the Hartree potential. For the term $\delta E_{xc}/\delta\rho$, which can only be calculated explicitly once an approximation for E_{xc} has been chosen, one commonly writes v_{xc}.

Consider now a system of noninteracting particles moving in the potential $v_s(\mathbf{r})$. For this system the minimization condition is simply

$$\begin{aligned} 0 &= \frac{\delta E_s[\rho]}{\delta\rho(\mathbf{r})} && (8.37)\\ &= \frac{\delta T_s[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta V_s[\rho]}{\delta\rho(\mathbf{r})} && (8.38)\\ &= \frac{\delta T_s[\rho]}{\delta\rho(\mathbf{r})} && (8.39) \end{aligned}$$

since there are no Hartree and xc terms in the absence of interactions. The density solving this Euler equation is $\rho_s(\mathbf{r})$. Comparing this with Eq. 8.36 we find that both minimizations have the same solution $\rho_s(\mathbf{r}) \equiv \rho(\mathbf{r})$, if v_s is chosen to be

$$v_s(\mathbf{r}) = v_{ext}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r})\,. \qquad (8.40)$$

Consequently, one can calculate the density of the interacting (many-body) system in potential $v_{ext}(\mathbf{r})$, described by a many-body Schrödinger equation of the form 8.1, by solving the equations of a noninteracting (single-body) system in potential $v_s(\mathbf{r})$. [8]

---

[8] The question whether such a potential $v_s(\mathbf{r})$ always exists in the mathematical sense is called the noninteracting v-representability problem. It is known that every interacting ensemble v-representable density is also noninteracting ensemble v-representable, but only in discretized systems has it been proven that all densities are interacting ensemble v-representable. It is not known if interacting ensemble-representable densities may be noninteracting pure-state representable (i.e, by a single determinant), which would be convenient (but is not necessary) for Kohn-Sham calculations.

In particular, the Schrödinger equation of this auxiliary system,

$$\Big[-\frac{\nabla^2}{2}+v_s(\mathbf{r})\Big]\phi_i(\mathbf{r})=\varepsilon_i\phi_i(\mathbf{r}) \tag{8.41}$$

yields orbitals that reproduce the density $\rho(\mathbf{r})$ of the original system (these are the same orbitals employed in Eq. 8.31),

$$\rho(\mathbf{r})\equiv\rho_s(\mathbf{r})=\sum_i^N f_i|\phi_(\mathbf{r})|^2\,, \tag{8.42}$$

where f_i is the occupation of the i^{th} orbital. [9] Eqs. 8.40 to 8.42 are the celebrated **Kohn-Sham (KS) equations**. They replace the problem of minimizing $E[\rho]$ by that of solving the Schrödinger equation for a non-interacting system.

Since both v_H and v_{xc} depend on ρ, which depends on the ϕ_i, which in turn depend on v_s, the problem of solving the KS equations is a nonlinear one. The usual way of solving such problems is to start with an initial guess for $\rho(\mathbf{r})$, calculate the corresponding $v_s(\mathbf{r})$, and then solve the differential equation 8.41 for the ϕ_i. From these one calculates a new density, using 8.42, and starts again.

The process is repeated until it converges. The technical name for this procedure is **self-consistency cycle**, which is the same we have encountered in the solution of the HF SC field equations. Different convergence criteria (such as convergence in the energy, the density, or some observable calculated from these) and various convergence-accelerating algorithms (such as mixing of old and new effective potentials) are in common use. Only rarely it requires more than a few dozen iterations to achieve convergence, and even rarer are cases where convergence seems unattainable, i.e., a self-consistent solution of the KS equation cannot be found.

Once one has a converged solution ρ_0, one can calculate the total energy from Eq. 8.36 or, equivalently and more conveniently, from

$$E_0=\sum_{i=1}^N\varepsilon_i-\frac{1}{2}\int d^3\mathbf{r}\int d^3\mathbf{r}'\frac{\rho_0(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}-\int d^3r v_{xc}(\mathbf{r})\rho_0(\mathbf{r})+E_{xc}[\rho_0] \tag{8.43}$$

Equation 8.43 follows from writing $V_{ext}[\rho]$ in 8.36 by means of 8.40 as

$$\begin{aligned}V_{ext}[\rho]&=\int d^3\,v_{ext}(\mathbf{r})\rho(\mathbf{r})\\&=\int d^3r\,\big[v_s(\mathbf{r})-v_H(\mathbf{r})-v_{xc}(\mathbf{r})\big]\rho(\mathbf{r})\\&=V_r[\rho]-\int d^3r\big[v_H(\mathbf{r})+v_{xc}(\mathbf{r})\big]\end{aligned}$$

---

[9] Normally, the occupation numbers f_i follows an *Aufbau* principle (Fermi statistics) with $f_i=1$ for $i<N$, $f_i=0$ for $i>N$, and $0\leq f_i\leq 1$ for $i=N$ (i.e., at most the highest occupied orbital can have fractional occupation).

and identifying the energy of the noninteracting (Kohn-Sham) system as $E_s = \sum_{i=1}^{N} \varepsilon_i = T_s + V_s$.

The eigenvalues of the Kohn-Sham equations

Equation 8.43 shows that E_0 is not simply the sum of all ε_i. In fact, it should be clear from our derivation of Eq. 8.41 that the ε_i are introduced as completely artificial objects: they are the eigenvalues of an auxiliary one-body equation whose eigenfunctions (orbitals) yield the correct density. It is only this density that has strict physical meaning in the KS equations. The KS eigenvalues, on the other hand, in general bear only a semiquantitative resemblance with the true energy spectrum, but are not to be trusted quantitatively.

The main exception to this rule is the highest occupied KS eigenvalue. Denoting by $\varepsilon_N(M)$ the N^{th} eigenvalue of a system with M electrons, one can show rigorously that $\varepsilon_N(M) = -I$, the negative of the first ionization energy of the N-body system, and $\varepsilon_{N+1}(N+1) = -A$, the negative of the electron affinity of the same N-body system. These relations hold for the exact functional only. When calculated with an approximate functional of the LDA or GGA type (see below), the highest eigenvalues usually do not provide good approximations to the experimental I and A. Better results for these observables are obtained by calculating them as total energy differences, according to $I = E_0(N-1) - E_0(N)$ and $A = E_0(N) - E_0(N+1)$, where $E_0(N)$ is the ground state energy of the N-body system. Alternatively, self-interaction corrections (SIC) can be used to obtain improved ionization energies and electron affinities from Kohn-Sham eigenvalues.

%endspacing %endminipage

Given the auxiliary nature of the other Kohn-Sham eigenvalues, it comes as a great (and welcome) surprise that in many situations (typically characterized by the absence of strong correlations) the Kohn-Sham eigenvalues ε_i do, empirically, provide a reasonable first approximation to the actual energy levels of extended systems and molecules.

8.5 Making DFT practical: Approximations

The basic approximation in DFT consists in the construction of an expression for the unknown xc functional $E_{xc}[\rho]$, which contains all many-body aspects of the problem. This chapter is intended to give you an idea of what types of functionals exist, and to describe what their main features are, separately for local functionals (TF, LDA and X_α), semilocal, or gradient-dependent functionals (GEA and GGA), and nonlocal functionals (hybrids, orbital functionals such as meta-GGAs, EXX and self interaction corrected SIC ones). This chapter does deal only most superficially with the actual construction of these functionals. For more details about functional construction and performance, the students are referred to the reviews on this subject (see for instance the literature given in chapter 8.1).

8.5.1 Local functionals: LDA

Historically (and in many applications also practically) the most important type of approximation is the local-density approximation (LDA). To understand the concept of an LDA recall first how the noninteracting kinetic energy $T_s[\rho]$ is treated in the Thomas-Fermi approximation: In a homogeneous system one knows that, per volume

$$t_s^{hom}(\rho) = \frac{3\hbar}{10m_e}(3\pi^2)^{2/3}\,\rho^{5/3} \tag{8.44}$$

where $\rho = \text{const}$. In an inhomogeneous system, with $\rho = \rho(\mathbf{r})$, one approximates locally

$$t_s(\mathbf{r}) \approx t_s^{hom}(\rho(\mathbf{r})) = \frac{3\hbar}{10m_e}(3\pi^2)^{2/3}\,\rho(\mathbf{r})^{5/3} \tag{8.45}$$

and obtains the full kinetic energy by integration over all space

$$T_s^{LDA}[\rho] = \int d^3r\, t_s^{hom}(\rho(\mathbf{r})) = \frac{3\hbar}{10m_e}(3\pi^2)^{2/3} \int d^3r\,\rho(\mathbf{r})^{5/3} \tag{8.46}$$

For the kinetic energy the approximation $T_s[\rho] \approx T_s^{TLDA}[\rho]$ is much inferior to the exact treatment of T_s in terms of orbitals, offered by the Kohn-Sham equations, but the LDA concept turned out to be highly useful for another component of the total energy, the exchange-correlation energy $E_{xc}[\rho]$. For the exchange energy $E_x[\rho]$ the procedure is very simple, since the per volume exchange energy of the homogeneous electron liquid is known exactly

$$e_x^{hom}(\rho) = -\frac{3q_e^2}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{4/3} \tag{8.47}$$

so that

$$E_x^{LDA}[\rho] = -\frac{3q_e^2}{4}\left(\frac{3}{\pi}\right)^{1/3}\int d^3r\,\rho^{4/3} \tag{8.48}$$

This is the LDA for E_x.

For the correlation energy $E_c[\rho]$ the situation is more complicated since $e_c^{hom}(\rho)$ is not known exactly: the determination of the correlation energy of a homogeneous interacting electron system (an electron liquid) is already a difficult many-body problem on its own! Early approximate expressions for $e_c^{hom}(\rho)$ were based on applying perturbation theory (e.g. the random phase approximation) to this problem. These approximations became outdated with the advent of highly precise Quantum Monte Carlo (QMC) calculations for the electron liquid, by Ceperley and Alder (D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.*, **45**, 566 (1980)). Modern expressions for $e_c^{hom}(\rho)$ are parametrizations of these data. These expressions are implemented in most standard DFT program packages and in typical applications give almost identical results. On the other hand, the earlier parametrizations of the LDA, based on perturbation theory, can occasionally deviate substantially from the QMC ones, and are better avoided.

Independently of the parametrization, the LDA for E_{xc} consists in

$$E_{xc}[\rho] \approx E_{xc}^{LDA}[\rho] = \int e_{xc}^{hom}(\rho)|_{\rho \to \rho(\mathbf{r})} \, d^3r = \int e_{xc}^{hom}(\rho(\mathbf{r})) \, d^3r \tag{8.49}$$

where $e_{xc}^{hom} = e_x^{hom} + e_c^{hom}$. The corresponding xc potential

$$v_{xc}^{LDA}[\rho](\mathbf{r}) = \frac{\partial e_{xc}^{hom}(\rho)}{\partial \rho}\Big|_{\rho \to \rho(\mathbf{r})} \tag{8.50}$$

This approximation for $E_{xc}[\rho]$ has proved amazingly successful, even when applied to systems that are quite different from the electron liquid that forms the reference system for the LDA. A partial explanation for this success of the LDA is systematic *error cancellation*: typically, LDA underestimates E_c but overestimates E_x, resulting in unexpectedly good values of E_{xc}.

For many decades the LDA has been applied in, e.g., calculations of band structures and total energies in solid-state physics. In quantum chemistry it is much less popular, because it fails to provide results that are accurate enough to permit a quantitative discussion of the chemical bond in molecules (so-called '**chemical accuracy**' requires calculations with an error of not more than about $1kcal/mol = 0.04336eV/particle$). At this stage it may be worthwhile to recapitulate what practical DFT does, and where the LDA enters its conceptual structure: What real systems, such as atoms, molecules, clusters and solids, have in common, is that they are simultaneously inhomogeneous (the electrons are exposed to spatially varying electric fields produced by the nuclei) and interacting (the electrons interact via the Coulomb interaction). The way density-functional theory, in the local-density approximation, deals with this inhomogeneous many-body problem is by decomposing it into two simpler (but still highly nontrivial) problems: the solution of a spatially homogeneous interacting problem (the homogeneous electron liquid) yields the uniform xc energy $e_{xc}^{hom}(\rho)$, and the solution of a spatially inhomogeneous noninteracting problem (the inhomogeneous electron gas described by the KS equations) yields the particle density. Both steps are connected by the local-density potential (8.50), which shows how the xc energy of the uniform interacting system enters the equations for the inhomogeneous noninteracting system.

8.5.2 Semilocal functionals: GEA, GGA and beyond

In the LDA one exploits knowledge of the density at point $\mathbf{r}$. Any real system is spatially inhomogeneous, i.e., it has a spatially varying density $\rho(\mathbf{r})$, and it would clearly be useful to also include information on the rate of this variation in the functional. A first attempt at doing this were the so-called **gradient-expansion approximations** (GEA). In this class of approximation one tries to systematically calculate gradient-corrections of the form $|\nabla\rho(\mathbf{r})|$, $|\nabla\rho(\mathbf{r})|^2$, $\nabla^2\rho(\mathbf{r})$, etc., to the LDA. A famous example is the lowest-order gradient correction to the Thomas-Fermi approximation for $T_s[\rho]$,

$$T_s[\rho] \approx T_s^W[\rho] = T_s^{LDA}[\rho] + \frac{\hbar^2}{8m_e}\int d^3r\, \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} \tag{8.51}$$

This second term on the right-hand side is called the Weizsäcker term. Similarly, in

$$E_x[\rho] \approx E_x^{GEQ(2)}[\rho] = E_x^{LDA}[\rho] - \frac{10q_e^2}{432\,\pi(3\pi^2)^{1/3}}\int d^3r\, \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})^{4/3}} \tag{8.52}$$

the second term on the right-hand side is the lowest-order gradient correction to E_x^{LDA}. In practice, the inclusion of low-order gradient corrections almost never improves on the LDA, and often even worsens it. Higher-order corrections (e.g., $\propto |\nabla\rho(\mathbf{r})|^\alpha$ or $\nabla^{beta}\rho(\mathbf{r})$ with $\alpha, \beta > 2$), on the other hand, are exceedingly difficult to calculate, and little is known about them.

In this situation it was a major breakthrough when it was realized, in the early eighties, that instead of power-series-like systematic gradient expansions one could experiment with more general functions of $\rho(\mathbf{r})$ and $\nabla\rho(\mathbf{r})$, which need not proceed order by order. Such functionals, of the general form

$$E_{xc}^{GGA}[\rho] = \int d^3\, f\big(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\big) \tag{8.53}$$

have become known as **generalized-gradient approximations** (GGAs). Different GGAs differ in the choice of the function $f(\rho, \nabla\rho)$. Note that this makes different GGAs much more different from each other than the different parametrizations of the LDA: essentially there is only one correct expression for $e_{xc}^{hom}(\rho)$, and the various parametrizations of the LDA are merely different ways of writing it. On the other hand, depending on the method of construction employed for obtaining $f(\rho, \nabla\rho)$ one can obtain very different GGAs. In particular, GGAs used in quantum chemistry typically proceed by fitting parameters to test sets of selected molecules. On the other hand, GGAs used in physics tend to emphasize exact constraints. Nowadays the most popular (and most reliable) GGAs are PBE (denoting the functional proposed in 1996 by Perdew, Burke and

Ernzerhof) in physics, and BLYP (denoting the combination of Becke's 1988 exchange functional with the 1988 correlation functional of Lee, Yang and Parr) in chemistry. Many other GGA-type functionals are also available, and new ones continue to appear.

method	-E [a.u.]
Thomas-Fermi	625.7
Hartree-Fock	526.818
LDA (exchange only)	524.517
LDA (exch. and corr.)	525.94
LDA-SIC(PZ)	528.393
GGA BLYP	527.551
experiment	527.6

Table 8.1: Ground state energy in atomic units (1 a.u.= 1 Hartree = 2 Rydberg = 27.21 eV = 627.5 kcal/mol) of the *Ar* atom (Z=18), obtained with HF and DFT for some representative functionals.

Quite generally, current GGAs seem to give reliable results for all main types of chemical bonds (covalent, ionic, metallic and hydrogen bonds). For van der Waals interactions, however, common GGAs and LDA fail. To describe these very weak interactions several more specialized approaches have been developed within DFT. Both in physics and in chemistry the widespread use of GGAs has lead to major improvements as compared to LDA. *Chemical accuracy*, as defined above, has not yet been attained, but is not too far away either. A useful collection of explicit expressions for some GGAs can be found in the literature (see for instance P. Ziesche, S. Kurth and J.P. perdew, *Comp. Mat. Sci*, **11**, 122 (1998)).
No systematic attempt at comparing explicit functionals can be made here, but many detailed comparisons are available in the literature. For purely illustrative purposes only, Table 8.1 contains ground-state energies of the *Ar* atom, obtained with several of the methods discussed in this and in the following sections.

8.5.3 Orbital functionals and other nonlocal approximations: hybrids, Meta-GGA, SIC, etc.

In spite of these advances, the quest for more accurate functionals goes ever on, and both in chemistry and physics various beyond-GGA functionals have appeared. Perhaps the most popular functional in quantum chemistry is B3LYP. This is a combination of the LYP GGA for correlation with Becke's three-parameter hybrid functional B3 for exchange. Common **hybrid functionals**, such as B3, mix a fraction of Hartree-Fock exchange

$$E_x[\{\phi_i[\rho]\}] = -\frac{1}{2}\sum_{ij}\int d^3r\int d^3r' \frac{\phi_j(\mathbf{r})\phi_k(\mathbf{r})\phi_j(\mathbf{r}')\phi_k(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \qquad (8.54)$$

into the DFT exchange functional (other mixtures are also possible). The construction of hybrid functional involves a certain amount of empiricism in the choice of functionals that are mixed and in the optimization of the weight factors given to the HF and DFT terms. Formally, this might be considered a drawback, but in practice B3 has proven to be the most successful exchange functional for chemical applications, in particular when combined with the LYP GGA functional for E_c. A more extreme example of this semiempirical mode of construction of functionals is Becke's 1997 hybrid functional, which contains 10 adjustable parameters.

Another recent beyond-GGA development is the emergence of so-called **Meta-GGAs**, which depend, in addition to the density and its derivatives, also on the Kohn-Sham kinetic-energy density $\tau(\mathbf{r})$

$$\tau(\mathbf{r}) = \frac{\hbar^2}{2m_e} \sum_i |\nabla \phi_i(\mathbf{r})|^2 \,, \tag{8.55}$$

so that E_{xc} can be written as $E_{xc}[\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \tau(\mathbf{r})]$. The additional degree of freedom provided by τ is used to satisfy additional constraints on E_{xc}, such as a self-interaction-corrected correlation functional and a finite exchange potential at the nucleus. In several recent tests Meta-GGAs have given favorable results, even when compared to the best GGAs, but the full potential of this type of approximation is only beginning to be explored systematically.

As we have seen in the case of T_s, it can be much easier to represent a functional in terms of single-particle orbitals (like the Kohn-Sham orbitals) than directly in terms of the density. Such functionals are known as orbital functionals, like for instance T_s in Eq. 8.31. Another important orbital-dependent functional is the exchange energy (Fock term) of Eq. 8.54. The Meta-GGAs and hybrid functionals mentioned above are also orbital functionals, because they depend on the kinetic energy density (8.55), and on a combination of the orbital functional (Eq. 8.54) with ordinary GGAs, respectively. Still another type of orbital functional is the **self-interaction correction** (SIC). Most implementations of SIC make use of the following expression proposed by J. P. Perdew and A. Zunger (PZ-SIC),

$$E_{xc}^{GGA-SIC}[\rho_\alpha, \rho_\beta] = E_{xc}^{GGA}[\rho_\alpha, \rho_\beta] - \sum_{i\sigma} \Big(E_H[\rho_{i\sigma}] - E_{xc}^{GGA}[\rho_{i\sigma}, 0] \Big) \tag{8.56}$$

which subtracts, orbital by orbital, the contribution the Hartree and the xc functionals would make if there was only one electron in the system. Here $\sigma \in \{\alpha, \beta\}$. This correction can be applied on top of any approximate density functional, and ensures that the resulting corrected functional satisfies $E_{xc}^{GGA-SIC}[\rho_{i\sigma}^{(1)}, 0] = -E_H[\rho^{(1)}]$ for a one-electron system with density $\rho^{(1)}$. The LDA is exact for a completely uniform system, and thus is self-interaction free in this limit, but neither it nor common GGAs satisfy the requirement of freedom from self-interaction in general, and even Meta-GGAs have a remaining self-interaction error in their exchange part. This self-interaction is particularly critical for localized states, such as the d

states in transition-metal oxides. For such systems PZ-SIC has been shown to greatly improve the uncorrected LDA and GGA functionals, but for thermochemistry PZ-SIC does not seem to be significant. Unfortunately the PZ-SIC approach, which minimizes the corrected energy functional with respect to the orbitals, does not lead to Kohn-Sham equations of the usual form, because the resulting effective potential is different for each orbital. As a consequence, various specialized algorithms for minimizing the PZ-SIC energy functional have been developed. For finite systems, PZ-SIC has also been implemented by means of the "optimized effective potential" OEP method, which produces a common local potential for all orbitals. The study of this theory is however beyond the scope of this lecture. The interested students will find more about this subject and other more advanced topics in the master course.

Chapter 9

Ab-initio Molecular Dynamics in the Ground State

Since an electron is much lighter than a nucleus, one may usually assume the Born-Oppenheimer (BO) adiabatic approximation. That is, one ignores the coupling between the motions of electrons and of nuclei and performs the integrations of these two motions separately at each time step. When the motion of nuclei is not very fast, and when the temperature is very low, one assumes also that the electronic states are at the ground state. The term, "adiabatic potential surface" or "Born-Oppenheimer surface" (BO surface) is often used to indicate an energy surface, in the $3N$-dimensional (atomic) configuration space (N refers to the number of atoms in the system), calculated using the BO approximation.

Here we should make some remarks on the use of the BO approximation. First, for systems composed of light atoms such as hydrogen, there are cases in which, even in the adiabatic approximation, the zero-point vibration energy of the nuclei may become important in a discussion of the subtle energy differences between stable atomic configurations. In such a case, one has to include the contribution from the zero-point energy. Second, the BO approximation itself may break down if, for example, the electrons are not in the ground state or the atoms move at very high speeds.

The resulting Schrödinger equation for the ionic motion is given by Eq. **??**

$$\left[\hat{T}_N + \varepsilon_n(\mathbf{R})\right] \varphi_n^{\mathrm{n}}(\mathbf{R}) = E\, \varphi_n^{\mathrm{n}}(\mathbf{R})\,,$$

where $\varphi_n^{\mathrm{n}}(\mathbf{R})$ is the amplitude for the nuclei to have positions $\mathbf{R}$ when the electrons are in the state $\Psi_n^{\mathrm{e}}\left(\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \dots)\right)$[1].

The total electronic energy, $\varepsilon_n(\mathbf{R})$, which includes the interaction among the nuclei, $\hat{V}_{NN}$, plays the role of an effective potential energy for the nuclei. In the *quantum classical approximation* only the electrons are quantized, while the nuclei are treated at a classical level (point charges). As a

---

[1] The index n labels the different solutions of the electronic Scrödinger equation. $n = 0$ stays for the ground state.

consequence the above Schrödinger equation for the nuclear dynamics is replaced by its classical equivalent, the Newton equation

$$\mathbf{F}_I = m_I\,\ddot{\mathbf{R}}_I\,, \tag{9.1}$$

where the forces, $\mathbf{F}_I$, are computed from the total electronic potential, $\varepsilon_n(\mathbf{R})$

$$\mathbf{F}_I = -\frac{\partial \varepsilon_n(\mathbf{R})}{\partial \mathbf{R}_I} = -\nabla_I \varepsilon_n(\mathbf{R}) \tag{9.2}$$

On the basis of density functional theory, the right-hand side of this equation can be written as

$$-\nabla_I \varepsilon_n(\mathbf{R}) = -\nabla_I \sum_{J(\neq I)} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \int \rho_n(\mathbf{r})\,\nabla_I v_I(|\mathbf{r} - \mathbf{R}_I|)\,d\mathbf{r} \tag{9.3}$$

where the first and the second terms represent, respectively, the Coulomb repulsive force between the nuclei and the Coulomb attractive force exerted on the nuclei by the electron cloud. Here, $v_I(r)$ denotes either a pseudopotential, or $-Z_I/r$ in a all electron approach ($r = |\mathbf{r}|$), and $\rho_n(\mathbf{r})$ is the electronic density of the ground state (if $n = 0$) or a selected excited state ($n \geq 1$).

9.1 The Hellmann-Feynman Forces

An important ingredient in all dynamics methods is the efficient calculation of the forces acting on the nuclei. The straightforward numerical evaluation of the derivative

$$\mathbf{F}_I = -\nabla_I \langle \Psi_0 | \mathcal{H}_{\mathrm{el}} | \Psi_0 \rangle \tag{9.4}$$

in terms of finite-difference approximation of the total electronic energy is both too costly and too inaccurate for dynamical simulations. What happens if the gradients are evaluated analytically?
In addition to the derivative of the Hamiltonian itself

$$\nabla_I \langle \Psi_0 | \mathcal{H}_{\mathrm{el}} | \Psi_0 \rangle = \langle \Psi_0 | \nabla_I \mathcal{H}_{\mathrm{el}} | \Psi_0 \rangle + \langle \nabla_I \Psi_0 | \mathcal{H}_{\mathrm{el}} | \Psi_0 \rangle + \langle \Psi_0 | \mathcal{H}_{\mathrm{el}} | \nabla_I \Psi_0 \rangle \tag{9.5}$$

there are in general also contributions from the variations of the wavefunctions ($\nabla_I \Psi_0$). In general means here that these contributions vanish exactly if the wavefunction is an exact eigenfunction (or stationary state wavefunction) of the Hamiltonian under consideration and the nuclear forces become

$$\mathbf{F}_I^{\mathrm{HF}} = -\langle \Psi_0 | \nabla_I \mathcal{H}_{\mathrm{el}} | \Psi_0 \rangle \tag{9.6}$$

This is the content of the often cited Hellmann-Feynman Theorem, which is also valid for many variational wavefunctions (e.g. Hartree-Fock or Kohn-Sham wavefunctions) .

9.2 Ehrenfest Molecular Dynamics

The more straightforward molecular dynamical scheme amounts to compute the Ehrenfest force by actually solving numerically the coupled set of nuclear and electronic propagation equations:

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \langle \Psi_0 | \mathcal{H}_{\text{el}} | \Psi_0 \rangle \tag{9.7}$$

$$i\hbar \frac{\partial \Psi}{\partial t} = \left[-\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 + \varepsilon(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) \right] \Psi \tag{9.8}$$

The Ehrenfest approach describes a real time-dependent evolution of the electronic degrees of freedom and therefore includes rigorously non-adiabatic effects like transitions between different electronic states Ψ_k and Ψ_l [2], which however are not considered in this chapter. Ehrenfest dynamics is just introduced here as a *rigorous* way of propagating both nuclear and electronic degrees of freedom in the adiabatic regime. The application of the Ehrenfest molecular dynamics approach in computational chemistry is however limited by the size of the propagation time step, which is of the order of 0.1 a.u., and therefore 10 to 100 smaller than the one used in Car-Parrinello and Born-Oppenheimer molecular dynamics, respectively.

9.3 Born-Oppenheimer Molecular Dynamics

The simplest way of including electronic structure in molecular dynamics simulations consists in straightforwardly solving the *static* electronic structure problem in each molecular dynamics step given the set of *fixed* nuclear positions at the instant of time. Thus, the electronic structure part is reduced to solving the *time-independent* quantum problem, e. g., by solving the *time-independent* Schödinger equation, concurrently to propagating the nuclei via classical molecular dynamics. In this picture, the time-dependence of the electronic structure is a consequence of the nuclear motion, and not explicit as for instance in Ehrenfest dynamics.

The resulting Born-Oppenheimer molecular dynamics method is defined by

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \min_{\Psi_0} \{ \langle \Psi_0 | \mathcal{H}_{\text{el}} | \Psi_0 \rangle \} \tag{9.9}$$

$$E_0 \Psi_0 = \mathcal{H}_{\text{el}} \Psi_0 \tag{9.10}$$

for the electronic ground state. A deep difference with respect to Ehrenfest dynamics concerning the nuclear equation of motion is that the minimum of $\langle \mathcal{H}_{\text{el}} \rangle$ has to be reached in each Born-Oppenheimer molecular dynamics step. In Ehrenfest dynamics, on the other hand, a wavefunction that minimized $\langle \mathcal{H}_{\text{el}} \rangle$ initially will also stay in its respective minimum as the nuclei move.

---

[2] see for instance: J. C. Tully, in *Modern Methods for Multidimensional Dynamics Computations in Chemistry*, ed. D.L. Thompson (World Scientific, Singapore, 1998).

9.4 Car-Parrinello Method

Since the paper by Car and Parrinello[3] first appeared in 1985, ab initio MD techniques have developed very rapidly. In their paper, Car and Parrinello proposed to integrate a Newtonian equation of motion for both the wavefunctions and the atomic coordinates. Such an approach was a completely new idea, beyond the previous common-sense idea that the electronic structure must be calculated by means of matrix diagonalization and self-consistent iterations.

Since the basic equation of motion in the Car-Parrinello method is derived from a Lagrangian, total energy conservation is automatically guaranteed in the formulation (the total energy in this method can be considered as the true total energy of the system if μ, defined below, is sufficiently small), and their approach is especially suitable for microcanonical simulations (with constant total energy). Their subsequent very active studies with various applications demonstrated the validity and effectiveness of their approach. With this technique, it is no longer necessary to treat huge eigenvalue problems and it becomes possible to reduce significantly both the computational memory and the computational time. The reason for the recent very rapid development of ab initio MD techniques is mainly due to this approach.

The basic equation of the Car-Parrinello method is given by the Lagrangian $\mathcal{L}$ as a functional of the wavefunction (=Kohn-Sham orbitals) ψ_l. (l denotes the level number of the electronic states) and the atomic position $\mathbf{R}_I$ (I stands for the atom index)

$$\mathcal{L} = \sum_l \frac{\mu}{2} \int d^3r \, |\dot{\psi}_l|^2 + \frac{1}{2} \sum_I M_I |\dot{\mathbf{R}}_I|^2 - \varepsilon \left[\{\psi_l\}, \{\mathbf{R}_I\} \right] \tag{9.11}$$

Here, μ is a fictitious electron mass which governs the motion of the wavefunctions, and $\varepsilon \left[\{\psi_l\}, \{\mathbf{R}_I\} \right]$ includes the electronic potential and the classical interaction between the nuclei. Note that this Lagrangian coincides with the true Lagrangian of the system only in the limit $\mu \to 0$, although we set μ finite. The important meaning of the existence of a Lagrangian is that the system has no energy dissipation and the total energy, including the fictitious kinetic energy (KE) $\mu_l |\dot{\psi}_l|^2$ of the electrons, is conserved. Therefore, the atomic oscillation of this system is guaranteed to be preserved permanently without damping. From this Lagrangian, one may derive the equation of motion for the wavefunctions

$$\mu \frac{d^2}{dt^2} \psi_l = -\mathcal{H} \psi_l + \sum_\nu \Lambda_{l\nu} \psi_\nu \quad , \tag{9.12}$$

where the $\Lambda_{l\nu}$ are Lagrange multipliers to ensure the orthogonality of the wavefunctions ψ_l. This multiplier is determined so as to orthonormalize the wavefunctions.

---

[3] R.Car and M. Parrinello, *Phys.Rev.Lett*, **55**, 2493 (1985)

To ensure the Born-Oppenheimer adiabatic approximation in the Car-Parrinello formalism, one has to choose the left-hand side of Eq. (9.12) to be small enough. That is, the fictitious electron mass μ should be small. Here, one should comment on the fact that even with this choice of μ, the electronic wavefunctions oscillate very rapidly around the BO surface, following the nuclear motion. This rapid oscillation of the electronic states is of course not realistic, but it is a consequence of the second-order differentiation with respect to time in Eq. (9.12), like in the case of a wave equation. However, for sufficiently small μ, the amplitude of this oscillation is usually very small.

On the other hand, the nuclear motion obeys the Newtonian equation

$$M_I \frac{d^2}{dt^2}\mathbf{R}_I = -\nabla_I E \tag{9.13}$$

with E being the sum of the Coulomb potential between the nuclei and the total energy of the electron system for nuclear positions, i.e. the physical total energy $+\sum_l (\mu/2)|\dot{\psi}_l|^2$, which is conserved. If the force on the right hand side of Eq. 9.13 is obtained, this equation can be integrated by means of, for example, the Verlet method.

9.4.1 Why does the Car-Parrinello Method Works?

In order to shed light on the title question, the dynamics generated by the Car-Parrinello Lagrangian Eq. 9.11 is analyzed in more detail invoking a "classical dynamics perspective" of a simple model system (eight silicon atoms forming a periodic diamond lattice, local density approximation to density functional theory, normconserving pseudopotentials for core electrons, plane wave basis for valence orbitals, 0.3 fs time step with $\mu = 300$ a.u., in total 20000 time steps or 6.3 ps; see G. Pastore et al., *Phys. Rev. A* 44, 6334 (1991) for full details).

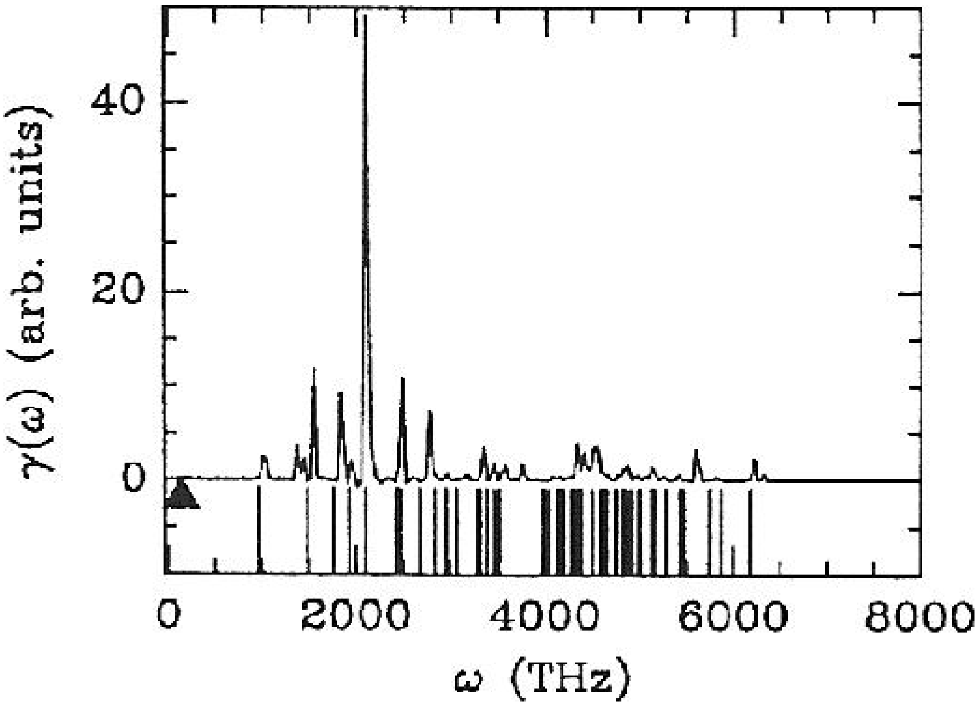


Figure 9.1: Vibrational density of states Eq. 9.14 (continuous spectrum in upper part) and harmonic approximation thereof (stick spectrum in lower part) of the electronic degrees of freedom compared to the highest frequency phonon mode $\omega_n^{\max}$ (triangle) for a model system (from wikipedia.

For this system the vibrational density of states or power spectrum of the electronic degrees of freedom, i.e. the Fourier transform of the statistically averaged velocity autocorrelation function of the classical fields

$$f(\omega) = \int_0^\infty dt\ \cos(\omega t)\ \sum_i \langle \dot{\psi}_i(r,t) | \dot{\psi}_i(r,0) \rangle \tag{9.14}$$

is compared to the highest frequency phonon mode $\omega_n^{\max}$ of the nuclear subsystem in the nex figure.

From this figure it is evident that for the chosen parameters the nuclear and electronic subsystems are dynamically separated: their power spectra do not overlap so that energy transfer from the hot to the cold subsystem is expected to be prohibitively slow.

Chapter 10

Hybrid Quantum-Mechanics/Molecular-Mechanics Approach

Quantum chemical methods are generally applicable and allow the calculation of ground and excited state properties (molecular energies and structures, energies and structures of transition states, atomic charges, reaction pathways etc.)

Molecular Mechanical methods are restricted to the classes of molecules they have been designed for and their success strongly depends on the careful calibration of a large number of parameters. In addition chemical reactions cannot be described using Molecular Mechanics because the bonds are defined once and forever by the chosen force field.

The development of the hybrid QM/MM approaches is guided by the general idea that large chemical systems may be partitioned into an electronically important region which requires a quantum chemical treatment and a remainder which only acts in a perturbative fashion and thus admits a classical description.

Using this approach it is possible to reduce the size of the quantum system, and treat the rest of the system, the environment, at a classical Molecular Mechanical level. In this way we can reduce substantially the amount of computational time and memory storage capacity required for the computation of complex systems made of thousands of atoms, like for instance a protein.

The original molecular Hamiltonian for the QM system in the Born-Oppenheimer approximation

$$\hat{H}_{tot}^{QM} = -\frac{1}{2}\sum_I \frac{1}{M_I}\nabla_I^2 - \frac{1}{2}\sum_n \nabla_n^2 + \sum_{I<J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{In}\frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_n|} + \sum_{n<m}\frac{1}{|\mathbf{r}_m - \mathbf{r}_n|} \tag{10.1}$$

is modified into

$$\hat{H}_{tot}^{QM/MM} = -\frac{1}{2}\sum_I \frac{1}{M_I}\nabla_I^2 - \frac{1}{2}\sum_n \nabla_n^2 + \sum_{I<J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{In}\frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_n|} + \sum_{n<m}\frac{1}{|\mathbf{r}_m - \mathbf{r}_n|} - \sum_{Kn}\frac{Q_K}{|\mathbf{R}_K - \mathbf{r}_n|} + \sum_{KL}\frac{Q_K Q_L}{|\mathbf{R}_K - \mathbf{R}_L|} \quad (10.2)$$

where the indices K and L run over all MM atoms with charges Q. The last two terms in Eq. 10.2 describe the interaction of the electrons of the QM sub-system with the charges of the MM atoms, and the electrostatic interaction among the MM atoms. What is still missing - but can easily be added - is the part of the MM Hamiltonian dealing with the the molecular bonded-interactions and that you already know from classical Molecular Dynamics.

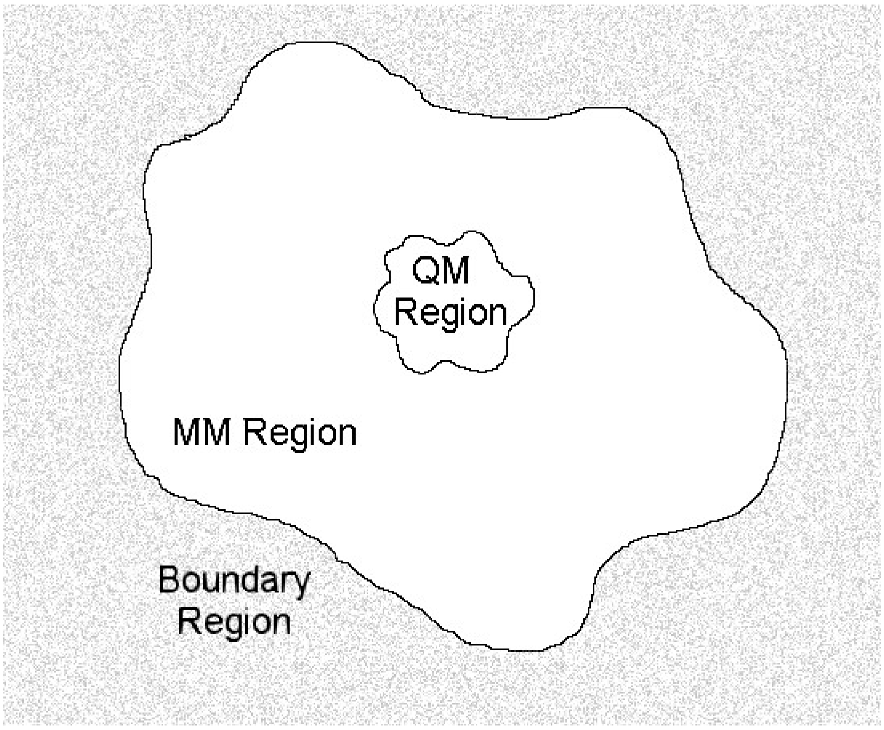


Figure 10.1: Cartoon representing a simple QM/MM setup. The central QM region can become the active site of an enzyme, the MM region the protein matrix, and the boundary region the solvent (picture taken from http://anusf.anu.edu.au.

Computational Chemistry Glossary

- **Ab initio.** The latin term "ab initio" means "from the beginning" or "from first principles". If we don't make use of any (experimentally derived) parameters in solving the Schrödinger equation for the electrons, the theory is called *ab initio.* If the nuclei are described at the level of classical mechanics and the electrons at a quantum mechanical level the approach is called *semi-classical.* In this course we will always remain within this approximation.

- **Approximations in computational quantum chemistry**
  We distinguish two main types of approximations:

  - either in the Hamiltonian (for instance by changing from a a wavefunction based to a density based description of the electronic interaction (DFT), or through the simplification of the electronic interaction term in semi-empirical, wavefunction based models).
  - or in the description of the many-electron wavefunction (for instance by describing the many-electron wavefunction with a single Slater determinant like in the Hartree-Fock method).

  In computational quantum chemistry the electronic wavefunction of a molecular system is often expanded as a sum of anti-symmetrized many electron wavefunctions (Slater determinants)

  $$\begin{aligned}\Psi_{el}(\mathbf{r}_1, s_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, s_N) = \\ = \sum_{m_1,m_2,\ldots,m_N} C_{m_1,m_2,\ldots,m_N} |\phi_{m_1}(\mathbf{r}_1, s_1)\phi_{m_2}(\mathbf{r}_2, s_2)\ldots\phi_{m_N}(\mathbf{r}_N, s_N)|\end{aligned}$$

  The components of the Slater determinant, $\phi_{m_i}(\mathbf{r}_i)$, are one-electron molecular orbitals which are usually given as an expansion in "atomic orbitals", χ_n:

  $$\phi_m(\mathbf{r}, s) = \sum_n D_{mn}\chi_n(\mathbf{r}) \otimes s \tag{10.3}$$

  (here $\mathbf{r}$ stays for the Cartesian coordinates (x, y, z) and s is the spin of the electron ($s \in \{\alpha, \beta\}$).

The collection of coefficients $D_{...}$ and $C_{...}$ fully characterizes the solution of the electronic Schrödinger equation for atoms and molecules.
The most commonly used approximate methods for the solution of the electronic molecular Schrödinger equation are:
- **Semi-empirical** (MNDO, AM1, PM3, etc.): use a single Slater determinant (only one C is 1 all the others are equal 0). Vary the coefficients D, but just use empirical estimates rather than the true integrals.
Very cheap, but only accurate for molecule similar to those used to develop the empirical estimates.
- **DFT** (B3LYP, BLYP, PW91, etc.): slightly empirical, but much more reliable than semi-empirical methods. CPU: cheap, same as HF $\mathcal{O}(N^3)$. Errors $\sim$ 4 kcal/mole (comparable accuracy to MP2 but much cheaper). Preferred method for geometries, second derivatives and transition-metal containing systems.
- **HF** (Hartree-Fock, SCF): only one many-electrons Slater determinant is used. Vary the D's, all terms in the electronic Hamiltonian calculated 'ab-initio' within the mean field approximation, no empirical parameters.
CPU: cheap $\mathcal{O}(N^3)$ *errors* ~ 15 *kcal/mol.*
- **MP2, MP4** (Moller-Plesset, MBPT): Vary the D's first, then set the C's to the values given by perturbation theory (you don't freely vary these C's, saving CPU).
MP2: medium CPU: $\mathcal{O}(N^5)$*, errors* ~ 5 *kcal/mol.*
- **CI, CISD, QCISD** (Configuration Interaction): Vary the coefficients D first, freeze them, then vary a lot of the coefficients C.
Expensive. Not used much anymore, CCSD is preferred.
- **MCSCF, CASSCF**: vary a finite set of C's and all the D's simultaneously.
Expensive. Good for understanding cases where several electronic states have comparable energies. User expertise required to select which C's to vary.
- **CAS-PT2**: Determine the D's and some C's by CASSCF, then determine more C's by perturbation theory.
Not much more expensive than CASSCF. Sometimes very good, but not reliable.
- *MRCI* (multi reference CI): Determine the D's and some C's by CASSCF or MCSCF, freeze these, then allow many of the C's to vary.
Super expensive. Very high accuracy for small systems.
- **CCSD, CCSD(T)** (Coupled Cluster): Vary the D's, fix them, then vary a lot of the C's, but constraining certain relationships between the C's. This allows you to effectively use a longer expansion without increasing the number of adjustable parameters so much. The constraints force the solution to be "size-consistent", i.e. two molecules calculated simultaneously have exactly the same energy as two molecules calculated separately.
Expensive. Often very accurate.
- **Extrapolations** ("Composite Methods"): G2, G3, CBS-q, CBS-Q, CBS-

QB3, CBS-RAD Run a series of the above calculations with different size basis sets, following some recipe. The results from all these calculations are extrapolated to an estimate of the true ground state energy (in the limit of a complete basis set and full electron correlation). These methods give excellent accuracy in less CPU time than CCSD or MRCI. However, the multiple steps involved provide many opportunities for something to go wrong.
Accuracy: usually 1-2 kcal/mol.

- **Ab initio forces.** Among the four fundamental forces in nature (electrostatic, weak and strong interactions, and gravitation) the only one that is relevant to ab initio simulations is the Coulomb interaction among nuclei and electrons.

- **Atom.** In chemistry and physics, an atom (Greek for "uncuttable") is the smallest possible particle of a chemical element that retains its chemical properties. Whereas the word atom originally denoted a particle that cannot be cut into smaller particles, the atoms of modern parlance are composed of subatomic particles:

  - electrons, which have a negative charge and are the least massive of the three;
  - protons, which have a positive charge and are about 1836 times more massive than electrons; and
  - neutrons, which have no charge and are about 1839 times more massive than electrons.

  (Protons and neutrons are not fundamental particles but are composed of quarks bound together by gluons (strong interaction, confinement.) Atoms can differ in the number of each of the subatomic particles they contain. Atoms of the same element have the same number of protons (called the atomic number). Within a single element, the number of neutrons may vary, determining the *isotope* of that element. The number of electrons associated with an atom is most easily changed, due to the lower energy of binding of electrons. The number of protons (and neutrons) in the atomic nucleus may also change, via nuclear fusion, nuclear fission or radioactive decay, in which case the atom is no longer the same element it was. This requires much larger energies than changes in the electronic composition.
  Atoms are electrically neutral if they have an equal number of protons and electrons. Atoms which have either a deficit or a surplus of electrons are called ions. Electrons that are furthest from the nucleus may be transferred to other nearby atoms or shared between atoms. By this mechanism atoms are able to bind into molecules and other types of chemical compounds like ionic and covalent network crystals.

In all the methods we discuss in this course, the nucleus is considered as a classical point charge with no extension, carrying a charge Z_I. Its position in space is defined by the position vector R_I. The electrons are described quantum mechanically. Sometimes, core electrons (the inner electrons in the shells close to the nucleus) are not considered explicitly because they are mainly chemically inert. In this case, the nuclear charge is replaced by, $Z_I - N_c$, where N_c is the number of core electrons, and the effects of the core electrons on the remaining valence electrons is replaced by the corresponding **pseudopotentials**.

- **Classical mechanics.** Classical mechanics is concerned with the set of physical laws governing and mathematically describing the motions of bodies and aggregates of bodies in the *classical* limit (neglecting quantum effects). The term classical mechanics was coined in the early 20th century to describe the system of mathematical physics developed during the 400 years since the groundbreaking works of Brahe, Kepler, and Galileo, but before the development of quantum physics and relativity. The initial stage in the development of classical mechanics is often referred to as Newtonian mechanics, and is associated with the mathematical methods invented by Newton himself, in parallel with Leibniz, and others. More abstract, and general methods include Lagrangian mechanics and Hamiltonian mechanics. While the terms classical mechanics and Newtonian mechanics are usually considered equivalent, the conventional content of classical mechanics was created in the 19th century and differs considerably (particularly in its use of analytical mathematics) from the work of Newton.

- **Computational chemistry.** Computational chemistry is a branch of chemistry that uses the results of theoretical chemistry incorporated into efficient computer programs to calculate the structures and properties of molecules and solids, applying these programs to real chemical problems. Examples of such properties are molecular structures (i.e. the expected positions of the constituent atoms), energy and interaction energy, charges, dipoles and higher multipole moments, vibrational frequencies, reactivity and spectroscopic quantities. The term computational chemistry is also sometimes used to cover any of the areas of science that overlap between computer science and chemistry. Electronic structure theory is a sub-discipline of computational chemistry.

- **n-body and many-body problems.** The "n-body problem" is the problem of finding, given the initial positions, masses, and velocities of n bodies, their subsequent motions as determined by classical mechanics, i.e., Newton's laws of motion. The equivalent problem in quantum mechanics is named "many-body problem".

  In classical mechanics (n-body problem) an exact solution is only known for

the case $n = 2$.The three-body problem ($n = 3$) is much more complicated; its solution can be chaotic.
In quantum mechanics the "many-body problem" is usually posed as the question of solving more complex problems than the hydrogen atom. Depending on the complexity of the molecule, different models are used. **The kind of methods and approximations that can be used is the main subject of this course**.

- **Force field.** In Force Field Methods (also called Molecular Mechanics or Classical Molecular Dynamics Methods) the energy of a molecular system is written as a parametric function of the nuclear coordinates. The parameters that enter the function are fitted to experimental or higher level computational data. Molecules are described as atoms held together by potentials that model chemical bonds. Three and four atom energy terms are also included to describe angles and dihedrals. All atoms are represented by point charges with fractional charges that interact through a Coulomb potential. The ensemble of all parameters is called a *Force Field.* For a given molecule, the assignment of force field parameters to the differend atoms, bonds, angles, and dihedrals defines its *topology.* Once the topology is defined the constraints to the dynamics of the system are set once forever. No chemical reactions can occur.
  Forces on the atoms are computed as derivative of the total energy and the system evolves according to Newton's law.

- **Fundamental forces.** Traditionally, modern physics describes all natural phenomena using four fundamental interactions: gravitation, electromagnetism, the weak interaction, and the strong interaction.
  In this course we consider exclusively electrostatic interactions between nuclei and electrons.

- **Hamiltonian operator.** In quantum mechanics, the quantum Hamiltonian $\hat{H}$ is the operator that determines to the total energy of the system, i.e. the total energy is the observable associated with $\hat{H}$. As with all observables, the spectrum of the Hamiltonian are the possible outcomes when one measures the total energy of a system.

- **Molecular dynamics (MD) and ab-initio molecular dynamics.** Once the Hamiltonian is given (classical or semiclassical) one can compute its derivative with respect to the nuclear positions to calculate the forces that act on the nuclei. These are used together with the Newton equation to compute the dynamics of the nuclei. Nuclear dynamics can be carried out in different thermodynamic ensembles, microcanonical (NVE, constant energy), canonical (NVT or NPT, constant temperature or grand canonical (μVT, variable number of particles).

The *electronic dynamics* can be carried out explicitly by solving the time-dependent Schrödinger equation, or - due to the much smaller mass of the electrons - can be considered as an instantaneous relaxation of the electrons into the new energy minimum provided by the new atomic positions. In the second case, named Born-Oppenheimer dynamics, there is no real electron dynamics , but the only dynamical variables are the nuclear positions.

$$M_I \ddot{R}_I(t) = \nabla_I \min_{\Phi_0} \{ \langle \Phi_0 | \mathcal{H}(R_I) | \Phi_0 \rangle \}$$

Born-Oppenheimer MD can be accelerated using the Car-Parrinello MD scheme.

- **Molecular mechanics.** The term molecular mechanics refers to the use of Newtonian mechanics to model molecular systems. Molecular mechanics approaches are widely applied in molecular structure refinement, molecular dynamics simulations, Monte Carlo simulations and ligand docking simulations. Molecular mechanics can be used to study small molecules as well as large biological systems or material assemblies with many thousands to millions of atoms. All-atomistic molecular mechanics methods have the following properties:
  - Each atom is simulated as a single hard spherical particle.
  - To each such particle a radius (typically the van der Waals radius) and a constant net charge (generally derived from high-level quantum calculations and/or experiment) is assigned.
  - Bonded interactions are treated as "springs" with equilibrium distances and harmonic force constants equal to the experimental or calculated bond lengths and vibrational frequencies.

- **Optimization of wavefunctions and geometries.** By wavefunction optimization we refer to the proccess of finding numerical approximations of the system's wavefunction by successive iterations. This process requires a variational principle that guarantees convergence by minimization of the total energy.

  $$E \leq \langle \Phi | \mathcal{H} | \Phi \rangle \, .$$

  During a geometry optimization, nuclear forces are computed at the end of each wavefunction optimization process. The nuclei are then shifted in direction of the computed forces and a new wavefunction is computed for the new positions. The process is iterated until convergence. The final nuclear coordinates correspond to the (global) minimum of the potential energy surface.

- **Pseudopotentials.** In quantum mechanics, the pseudopotential formalism is an attempt to replace the complicated effects of the motion of the core (i.e. non-valence) electrons of an atom or ion and its nucleus with an effective potential, or pseudopotential, so that the Schrödinger equation contains a modified potential term instead of e.g. the Coulombic potential term normally found in the Schrödinger equation and the electronic Schrödinger equation can be solved for the valence electrons only.

- **Quantum chemistry.** Quantum chemistry is a branch of theoretical chemistry, which applies quantum mechanics to address issues and problems in chemistry. The description of the electronic behavior of atoms and molecules to describe their reactivity is one of the applications of quantum chemistry. Quantum chemistry lies on the border between chemistry and physics, and significant contributions have been made by scientists from both fields. It has a strong and active overlap with the fields of atomic physics and molecular physics, as well as physical chemistry.

- **Quantum mechanics.** Quantum mechanics is a fundamental branch of theoretical physics that replaces classical mechanics and classical electromagnetism at the atomic and subatomic levels. It is the underlying mathematical framework of many fields of physics and chemistry, including condensed matter physics, atomic physics, molecular physics, computational chemistry, quantum chemistry, particle physics, and nuclear physics. Along with general relativity, quantum mechanics is one of the pillars of modern physics.

- **Relativistic quantum mechanics.** Relativistic effects in quantum mechanics are described by the Dirac equations. The notion of wavefunction and spin is replaced by the four-components spinor (solution of the Dirac equation). In the non-relativistic limit

  $$\varepsilon \sim \frac{p^2}{2m} \ll pc \tag{10.4}$$

  the Dirac equations reduces to the Schrödinger equation. Here p is the particle momentum, m is its mass and c is the speed of light.
  In this course we do not consider relativistic effects.

- **Schrödinger equation.** In physics, the Schrödinger equation, proposed by the Austrian physicist Erwin Schrödinger in 1925, describes the space- and time-dependence of quantum mechanical systems. It is of central importance to the theory of quantum mechanics, playing a role analogous to Newton's law in classical mechanics.
  In the mathematical formulation of quantum mechanics, each system is associated with a complex Hilbert space such that each instantaneous state of the system is described by a unit vector in that space. This state vector encodes the probabilities for the outcomes of all possible measurements applied to the system. As the state of a system generally changes over time,

the state vector is a function of time. The Schrödinger equation provides a quantitative description of the rate of change of the state vector.

- **Statistical Mechanics.** Statistical mechanics is the application of probability theory, which includes mathematical tools for dealing with large populations, to the field of mechanics, which is concerned with the motion of particles or objects when subjected to a force. It provides a framework for relating the microscopic properties of individual atoms and molecules to the macroscopic or bulk properties of materials that can be observed in everyday life, therefore explaining thermodynamics as a natural result of statistics and mechanics (classical and quantum) at the microscopic level. In particular, it can be used to calculate the thermodynamic properties of bulk materials from the spectroscopic data of individual molecules.