
1

WHAT IS DENSITY FUNCTIONAL THEORY?

1.1 HOW TO APPROACH THIS BOOK

There are many fields within the physical sciences and engineering where the key to scientific and technological progress is understanding and controlling the properties of matter at the level of individual atoms and molecules. Density functional theory is a phenomenally successful approach to finding solutions to the fundamental equation that describes the quantum behavior of atoms and molecules, the Schrödinger equation, in settings of practical value. This approach has rapidly grown from being a specialized art practiced by a small number of physicists and chemists at the cutting edge of quantum mechanical theory to a tool that is used regularly by large numbers of researchers in chemistry, physics, materials science, chemical engineering, geology, and other disciplines. A search of the *Science Citation Index* for articles published in 1986 with the words “density functional theory” in the title or abstract yields less than 50 entries. Repeating this search for 1996 and 2006 gives more than 1100 and 5600 entries, respectively.

Our aim with this book is to provide just what the title says: an *introduction* to using density functional theory (DFT) calculations in a *practical* context. We do not assume that you have done these calculations before or that you even understand what they are. We do assume that you want to find out what is possible with these methods, either so you can perform calculations

yourself in a research setting or so you can interact knowledgeably with collaborators who use these methods.

An analogy related to cars may be useful here. Before you learned how to drive, it was presumably clear to you that you can accomplish many useful things with the aid of a car. For you to use a car, it is important to understand the basic concepts that control cars (you need to put fuel in the car regularly, you need to follow basic traffic laws, etc.) and spend time actually driving a car in a variety of road conditions. You do not, however, need to know every detail of how fuel injectors work, how to construct a radiator system that efficiently cools an engine, or any of the other myriad of details that are required if you were going to actually build a car. Many of these details may be important if you plan on undertaking some especially difficult car-related project such as, say, driving yourself across Antarctica, but you can make it across town to a friend's house and back without understanding them.

With this book, we hope you can learn to “drive across town” when doing your own calculations with a DFT package or when interpreting other people's calculations as they relate to physical questions of interest to you. If you are interested in “building a better car” by advancing the cutting edge of method development in this area, then we applaud your enthusiasm. You should continue reading this chapter to find at least one surefire project that could win you a Nobel prize, then delve into the books listed in the Further Reading at the end of the chapter.

At the end of most chapters we have given a series of exercises, most of which involve actually doing calculations using the ideas described in the chapter. Your knowledge and ability will grow most rapidly by doing rather than by simply reading, so we strongly recommend doing as many of the exercises as you can in the time available to you.

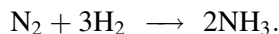
1.2 EXAMPLES OF DFT IN ACTION

Before we even define what density functional theory is, it is useful to relate a few vignettes of how it has been used in several scientific fields. We have chosen three examples from three quite different areas of science from the thousands of articles that have been published using these methods. These specific examples have been selected because they show how DFT calculations have been used to make important contributions to a diverse range of compelling scientific questions, generating information that would be essentially impossible to determine through experiments.

1.2.1 Ammonia Synthesis by Heterogeneous Catalysis

Our first example involves an industrial process of immense importance: the catalytic synthesis of ammonia (NH_3). Ammonia is a central component of

fertilizers for agriculture, and more than 100 million tons of ammonia are produced commercially each year. By some estimates, more than 1% of all energy used in the world is consumed in the production of ammonia. The core reaction in ammonia production is very simple:



To get this reaction to proceed, the reaction is performed at high temperatures ($>400^\circ\text{C}$) and high pressures (>100 atm) in the presence of metals such as iron (Fe) or ruthenium (Ru) that act as catalysts. Although these metal catalysts were identified by Haber and others almost 100 years ago, much is still not known about the mechanisms of the reactions that occur on the surfaces of these catalysts. This incomplete understanding is partly because of the structural complexity of practical catalysts. To make metal catalysts with high surface areas, tiny particles of the active metal are dispersed throughout highly porous materials. This was a widespread application of nanotechnology long before that name was applied to materials to make them sound scientifically exciting! To understand the reactivity of a metal nanoparticle, it is useful to characterize the surface atoms in terms of their local coordination since differences in this coordination can create differences in chemical reactivity; surface atoms can be classified into “types” based on their local coordination. The surfaces of nanoparticles typically include atoms of various types (based on coordination), so the overall surface reactivity is a complicated function of the shape of the nanoparticle and the reactivity of each type of atom.

The discussion above raises a fundamental question: Can a direct connection be made between the shape and size of a metal nanoparticle and its activity as a catalyst for ammonia synthesis? If detailed answers to this question can be found, then they can potentially lead to the synthesis of improved catalysts. One of the most detailed answers to this question to date has come from the DFT calculations of Honkala and co-workers,¹ who studied nanoparticles of Ru. Using DFT calculations, they showed that the net chemical reaction above proceeds via at least 12 distinct steps on a metal catalyst and that the rates of these steps depend strongly on the local coordination of the metal atoms that are involved. One of the most important reactions is the breaking of the N_2 bond on the catalyst surface. On regions of the catalyst surface that were similar to the surfaces of bulk Ru (more specifically, atomically flat regions), a great deal of energy is required for this bond-breaking reaction, implying that the reaction rate is extremely slow. Near Ru atoms that form a common kind of surface step edge on the catalyst, however, a much smaller amount of energy is needed for this reaction. Honkala and co-workers used additional DFT calculations to predict the relative stability of many different local coordinations of surface atoms in Ru nanoparticles in a way that allowed

them to predict the detailed shape of the nanoparticles as a function of particle size. This prediction makes a precise connection between the diameter of a Ru nanoparticle and the number of highly desirable reactive sites for breaking N_2 bonds on the nanoparticle. Finally, all of these calculations were used to develop an overall model that describes how the individual reaction rates for the many different kinds of metal atoms on the nanoparticle's surfaces couple together to define the overall reaction rate under realistic reaction conditions. At no stage in this process was any experimental data used to fit or adjust the model, so the final result was a truly predictive description of the reaction rate of a complex catalyst. After all this work was done, Honkala et al. compared their predictions to experimental measurements made with Ru nanoparticle catalysts under reaction conditions similar to industrial conditions. Their predictions were in stunning quantitative agreement with the experimental outcome.

1.2.2 Embrittlement of Metals by Trace Impurities

It is highly likely that as you read these words you are within 1 m of a large number of copper wires since copper is the dominant metal used for carrying electricity between components of electronic devices of all kinds. Aside from its low cost, one of the attractions of copper in practical applications is that it is a soft, ductile metal. Common pieces of copper (and other metals) are almost invariably polycrystalline, meaning that they are made up of many tiny domains called grains that are each well-oriented single crystals. Two neighboring grains have the same crystal structure and symmetry, but their orientation in space is not identical. As a result, the places where grains touch have a considerably more complicated structure than the crystal structure of the pure metal. These regions, which are present in all polycrystalline materials, are called grain boundaries.

It has been known for over 100 years that adding tiny amounts of certain impurities to copper can change the metal from being ductile to a material that will fracture in a brittle way (i.e., without plastic deformation before the fracture). This occurs, for example, when bismuth (Bi) is present in copper (Cu) at levels below 100 ppm. Similar effects have been observed with lead (Pb) or mercury (Hg) impurities. But how does this happen? Qualitatively, when the impurities cause brittle fracture, the fracture tends to occur at grain boundaries, so something about the impurities changes the properties of grain boundaries in a dramatic way. That this can happen at very low concentrations of Bi is not completely implausible because Bi is almost completely insoluble in bulk Cu. This means that it is very favorable for Bi atoms to segregate to grain boundaries rather than to exist inside grains, meaning that the

local concentration of Bi at grain boundaries can be much higher than the net concentration in the material as a whole.

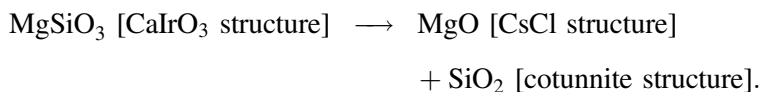
Can the changes in copper caused by Bi be explained in a detailed way? As you might expect for an interesting phenomena that has been observed over many years, several alternative explanations have been suggested. One class of explanations assigns the behavior to electronic effects. For example, a Bi atom might cause bonds between nearby Cu atoms to be stiffer than they are in pure Cu, reducing the ability of the Cu lattice to deform smoothly. A second type of electronic effect is that having an impurity atom next to a grain boundary could weaken the bonds that exist across a boundary by changing the electronic structure of the atoms, which would make fracture at the boundary more likely. A third explanation assigns the blame to size effects, noting that Bi atoms are much larger than Cu atoms. If a Bi atom is present at a grain boundary, then it might physically separate Cu atoms on the other side of the boundary from their natural spacing. This stretching of bond distances would weaken the bonds between atoms and make fracture of the grain boundary more likely. Both the second and third explanations involve weakening of bonds near grain boundaries, but they propose different root causes for this behavior. Distinguishing between these proposed mechanisms would be very difficult using direct experiments.

Recently, Schweinest, Paxton, and Finnis used DFT calculations to offer a definitive description of how Bi embrittles copper; the title of their study gives away the conclusion.² They first used DFT to predict stress–strain relationships for pure Cu and Cu containing Bi atoms as impurities. If the bond stiffness argument outlined above was correct, the elastic moduli of the metal should be increased by adding Bi. In fact, the calculations give the opposite result, immediately showing the bond-stiffening explanation to be incorrect. In a separate and much more challenging series of calculations, they explicitly calculated the cohesion energy of a particular grain boundary that is known experimentally to be embrittled by Bi. In qualitative consistency with experimental observations, the calculations predicted that the cohesive energy of the grain boundary is greatly reduced by the presence of Bi. Crucially, the DFT results allow the electronic structure of the grain boundary atoms to be examined directly. The result is that the grain boundary electronic effect outlined above was found to not be the cause of embrittlement. Instead, the large change in the properties of the grain boundary could be understood almost entirely in terms of the excess volume introduced by the Bi atoms, that is, by a size effect. This reasoning suggests that Cu should be embrittled by any impurity that has a much larger atomic size than Cu and that strongly segregates to grain boundaries. This description in fact correctly describes the properties of both Pb and Hg as impurities in Cu, and, as mentioned above, these impurities are known to embrittle Cu.

1.2.3 Materials Properties for Modeling Planetary Formation

To develop detailed models of how planets of various sizes have formed, it is necessary to know (among many other things) what minerals exist inside planets and how effective these minerals are at conducting heat. The extreme conditions that exist inside planets pose some obvious challenges to probing these topics in laboratory experiments. For example, the center of Jupiter has pressures exceeding 40 Mbar and temperatures well above 15,000 K. DFT calculations can play a useful role in probing material properties at these extreme conditions, as shown in the work of Umemoto, Wentzcovitch, and Allen.³ This work centered on the properties of bulk MgSiO_3 , a silicate mineral that is important in planet formation. At ambient conditions, MgSiO_3 forms a relatively common crystal structure known as a perovskite. Prior to Umemoto et al.'s calculations, it was known that if MgSiO_3 was placed under conditions similar to those in the core–mantle boundary of Earth, it transforms into a different crystal structure known as the CaIrO_3 structure. (It is conventional to name crystal structures after the first compound discovered with that particular structure, and the naming of this structure is an example of this convention.)

Umemoto et al. wanted to understand what happens to the structure of MgSiO_3 at conditions much more extreme than those found in Earth's core–mantle boundary. They used DFT calculations to construct a phase diagram that compared the stability of multiple possible crystal structures of solid MgSiO_3 . All of these calculations dealt with bulk materials. They also considered the possibility that MgSiO_3 might dissociate into other compounds. These calculations predicted that at pressures of ~ 11 Mbar, MgSiO_3 dissociates in the following way:



In this reaction, the crystal structure of each compound has been noted in the square brackets. An interesting feature of the compounds on the right-hand side is that neither of them is in the crystal structure that is the stable structure at ambient conditions. MgO , for example, prefers the NaCl structure at ambient conditions (i.e., the same crystal structure as everyday table salt). The behavior of SiO_2 is similar but more complicated; this compound goes through several intermediate structures between ambient conditions and the conditions relevant for MgSiO_3 dissociation. These transformations in the structures of MgO and SiO_2 allow an important connection to be made between DFT calculations and experiments since these transformations occur at conditions that can be directly probed in laboratory experiments. The transition pressures

predicted using DFT and observed experimentally are in good agreement, giving a strong indication of the accuracy of these calculations.

The dissociation reaction predicted by Umemoto et al.'s calculations has important implications for creating good models of planetary formation. At the simplest level, it gives new information about what materials exist inside large planets. The calculations predict, for example, that the center of Uranus or Neptune can contain MgSiO_3 , but that the cores of Jupiter or Saturn will not. At a more detailed level, the thermodynamic properties of the materials can be used to model phenomena such as convection inside planets. Umemoto et al. speculated that the dissociation reaction above might severely limit convection inside "dense-Saturn," a Saturn-like planet that has been discovered outside the solar system with a mass of ~ 67 Earth masses.

A legitimate concern about theoretical predictions like the reaction above is that it is difficult to envision how they can be validated against experimental data. Fortunately, DFT calculations can also be used to search for similar types of reactions that occur at pressures that are accessible experimentally. By using this approach, it has been predicted that NaMgF_3 goes through a series of transformations similar to MgSiO_3 ; namely, a perovskite to postperovskite transition at some pressure above ambient and then dissociation in NaF and MgF_2 at higher pressures.⁴ This dissociation is predicted to occur for pressures around 0.4 Mbar, far lower than the equivalent pressure for MgSiO_3 . These predictions suggest an avenue for direct experimental tests of the transformation mechanism that DFT calculations have suggested plays a role in planetary formation.

We could fill many more pages with research vignettes showing how DFT calculations have had an impact in many areas of science. Hopefully, these three examples give some flavor of the ways in which DFT calculations can have an impact on scientific understanding. It is useful to think about the common features between these three examples. All of them involve materials in their solid state, although the first example was principally concerned with the interface between a solid and a gas. Each example generated information about a physical problem that is controlled by the properties of materials on atomic length scales that would be (at best) extraordinarily challenging to probe experimentally. In each case, the calculations were used to give information not just about some theoretically ideal state, but instead to understand phenomena at temperatures, pressures, and chemical compositions of direct relevance to physical applications.

1.3 THE SCHRÖDINGER EQUATION

By now we have hopefully convinced you that density functional theory is a useful and interesting topic. But what is it exactly? We begin with

the observation that one of the most profound scientific advances of the twentieth century was the development of quantum mechanics and the repeated experimental observations that confirmed that this theory of matter describes, with astonishing accuracy, the universe in which we live.

In this section, we begin a review of some key ideas from quantum mechanics that underlie DFT (and other forms of computational chemistry). Our goal here is not to present a complete derivation of the techniques used in DFT. Instead, our goal is to give a clear, brief, introductory presentation of the most basic equations important for DFT. For the full story, there are a number of excellent texts devoted to quantum mechanics listed in the Further Reading section at the end of the chapter.

Let us imagine a situation where we would like to describe the properties of some well-defined collection of atoms—you could think of an isolated molecule or the atoms defining the crystal of an interesting mineral. One of the fundamental things we would like to know about these atoms is their energy and, more importantly, how their energy changes if we move the atoms around. To define where an atom is, we need to define both where its nucleus is and where the atom's electrons are. A key observation in applying quantum mechanics to atoms is that atomic nuclei are much heavier than individual electrons; each proton or neutron in a nucleus has more than 1800 times the mass of an electron. This means, roughly speaking, that electrons respond much more rapidly to changes in their surroundings than nuclei can. As a result, we can split our physical question into two pieces. First, we solve, for fixed positions of the atomic nuclei, the equations that describe the electron motion. For a given set of electrons moving in the field of a set of nuclei, we find the lowest energy configuration, or *state*, of the electrons. The lowest energy state is known as the *ground state* of the electrons, and the separation of the nuclei and electrons into separate mathematical problems is the *Born–Oppenheimer approximation*. If we have M nuclei at positions $\mathbf{R}_1, \dots, \mathbf{R}_M$, then we can express the ground-state energy, E , as a function of the positions of these nuclei, $E(\mathbf{R}_1, \dots, \mathbf{R}_M)$. This function is known as the *adiabatic potential energy surface* of the atoms. Once we are able to calculate this potential energy surface we can tackle the original problem posed above—how does the energy of the material change as we move its atoms around?

One simple form of the Schrödinger equation—more precisely, the time-independent, nonrelativistic Schrödinger equation—you may be familiar with is $H\psi = E\psi$. This equation is in a nice form for putting on a T-shirt or a coffee mug, but to understand it better we need to define the quantities that appear in it. In this equation, H is the Hamiltonian operator and ψ is a set of solutions, or eigenstates, of the Hamiltonian. Each of these solutions,

ψ_n , has an associated eigenvalue, E_n , a real number* that satisfies the eigenvalue equation. The detailed definition of the Hamiltonian depends on the physical system being described by the Schrödinger equation. There are several well-known examples like the particle in a box or a harmonic oscillator where the Hamiltonian has a simple form and the Schrödinger equation can be solved exactly. The situation we are interested in where multiple electrons are interacting with multiple nuclei is more complicated. In this case, a more complete description of the Schrödinger is

$$\left[-\frac{\hbar^2}{2m} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^N V(\mathbf{r}_i) + \sum_{i=1}^N \sum_{j<i}^N U(\mathbf{r}_i, \mathbf{r}_j) \right] \psi = E\psi. \quad (1.1)$$

Here, m is the electron mass. The three terms in brackets in this equation define, in order, the kinetic energy of each electron, the interaction energy between each electron and the collection of atomic nuclei, and the interaction energy between different electrons. For the Hamiltonian we have chosen, ψ is the electronic wave function, which is a function of each of the spatial coordinates of each of the N electrons, so $\psi = \psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$, and E is the ground-state energy of the electrons.** The ground-state energy is independent of time, so this is the time-independent Schrödinger equation.†

Although the electron wave function is a function of each of the coordinates of all N electrons, it is possible to approximate ψ as a product of individual electron wave functions, $\psi = \psi_1(\mathbf{r})\psi_2(\mathbf{r}), \dots, \psi_N(\mathbf{r})$. This expression for the wave function is known as a Hartree product, and there are good motivations for approximating the full wave function into a product of individual one-electron wave functions in this fashion. Notice that N , the number of electrons, is considerably larger than M , the number of nuclei, simply because each atom has one nucleus and lots of electrons. If we were interested in a single molecule of CO_2 , the full wave function is a 66-dimensional function (3 dimensions for each of the 22 electrons). If we were interested in a nanocluster of 100 Pt atoms, the full wave function requires more the 23,000 dimensions! These numbers should begin to give you an idea about why solving the Schrödinger equation for practical materials has occupied many brilliant minds for a good fraction of a century.

*The value of the functions ψ_n are complex numbers, but the eigenvalues of the Schrödinger equation are real numbers.

**For clarity of presentation, we have neglected electron spin in our description. In a complete presentation, each electron is defined by three spatial variables and its spin.

†The dynamics of electrons are defined by the time-dependent Schrödinger equation, $i\hbar(\partial\psi/\partial t) = H\psi$. The appearance of $i = \sqrt{-1}$ in this equation makes it clear that the wave function is a complex-valued function, not a real-valued function.

The situation looks even worse when we look again at the Hamiltonian, H . The term in the Hamiltonian defining electron–electron interactions is the most critical one from the point of view of solving the equation. The form of this contribution means that the individual electron wave function we defined above, $\psi_i(\mathbf{r})$, cannot be found without simultaneously considering the individual electron wave functions associated with all the other electrons. In other words, the Schrödinger equation is a many-body problem.

Although solving the Schrödinger equation can be viewed as the fundamental problem of quantum mechanics, it is worth realizing that the wave function for any particular set of coordinates cannot be directly observed. The quantity that can (in principle) be measured is the probability that the N electrons are at a particular set of coordinates, $\mathbf{r}_1, \dots, \mathbf{r}_N$. This probability is equal to $\psi^*(\mathbf{r}_1, \dots, \mathbf{r}_N)\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$, where the asterisk indicates a complex conjugate. A further point to notice is that in experiments we typically do not care which electron in the material is labeled electron 1, electron 2, and so on. Moreover, even if we did care, we cannot easily assign these labels. This means that the quantity of physical interest is really the probability that a set of N electrons in any order have coordinates $\mathbf{r}_1, \dots, \mathbf{r}_N$. A closely related quantity is the density of electrons at a particular position in space, $n(\mathbf{r})$. This can be written in terms of the individual electron wave functions as

$$n(\mathbf{r}) = 2 \sum_i \psi_i^*(\mathbf{r})\psi_i(\mathbf{r}). \quad (1.2)$$

Here, the summation goes over all the individual electron wave functions that are occupied by electrons, so the term inside the summation is the probability that an electron in individual wave function $\psi_i(\mathbf{r})$ is located at position \mathbf{r} . The factor of 2 appears because electrons have spin and the Pauli exclusion principle states that each individual electron wave function can be occupied by two separate electrons provided they have different spins. This is a purely quantum mechanical effect that has no counterpart in classical physics. The point of this discussion is that the electron density, $n(\mathbf{r})$, which is a function of only three coordinates, contains a great amount of the information that is actually physically observable from the full wave function solution to the Schrödinger equation, which is a function of $3N$ coordinates.

1.4 DENSITY FUNCTIONAL THEORY—FROM WAVE FUNCTIONS TO ELECTRON DENSITY

The entire field of density functional theory rests on two fundamental mathematical theorems proved by Kohn and Hohenberg and the derivation of a

set of equations by Kohn and Sham in the mid-1960s. The first theorem, proved by Hohenberg and Kohn, is: *The ground-state energy from Schrödinger's equation is a unique functional of the electron density.*

This theorem states that there exists a one-to-one mapping between the ground-state wave function and the ground-state electron density. To appreciate the importance of this result, you first need to know what a “functional” is. As you might guess from the name, a functional is closely related to the more familiar concept of a function. A function takes a value of a variable or variables and defines a single number from those variables. A simple example of a function dependent on a single variable is $f(x) = x^2 + 1$. A *functional* is similar, but it takes a function and defines a single number from the function. For example,

$$F[f] = \int_{-1}^1 f(x) dx,$$

is a functional of the function $f(x)$. If we evaluate this functional using $f(x) = x^2 + 1$, we get $F[f] = \frac{8}{3}$. So we can restate Hohenberg and Kohn's result by saying that the ground-state energy E can be expressed as $E[n(\mathbf{r})]$, where $n(\mathbf{r})$ is the electron density. This is why this field is known as density functional theory.

Another way to restate Hohenberg and Kohn's result is that the ground-state electron density uniquely determines all properties, including the energy and wave function, of the ground state. Why is this result important? It means that we can think about solving the Schrödinger equation by finding a function of three spatial variables, the electron density, rather than a function of $3N$ variables, the wave function. Here, by “solving the Schrödinger equation” we mean, to say it more precisely, finding the ground-state energy. So for a nanocluster of 100 Pd atoms the theorem reduces the problem from something with more than 23,000 dimensions to a problem with just 3 dimensions.

Unfortunately, although the first Hohenberg–Kohn theorem rigorously proves that a functional of the electron density exists that can be used to solve the Schrödinger equation, the theorem says nothing about what the functional actually is. The second Hohenberg–Kohn theorem defines an important property of the functional: *The electron density that minimizes the energy of the overall functional is the true electron density corresponding to the full solution of the Schrödinger equation.* If the “true” functional form were known, then we could vary the electron density until the energy from the functional is minimized, giving us a prescription for finding the relevant electron density. This variational principle is used in practice with approximate forms of the functional.

A useful way to write down the functional described by the Hohenberg–Kohn theorem is in terms of the single-electron wave functions, $\psi_i(\mathbf{r})$. Remember from Eq. (1.2) that these functions collectively define the electron density, $n(\mathbf{r})$. The energy functional can be written as

$$E[\{\psi_i\}] = E_{\text{known}}[\{\psi_i\}] + E_{\text{XC}}[\{\psi_i\}], \quad (1.3)$$

where we have split the functional into a collection of terms we can write down in a simple analytical form, $E_{\text{known}}[\{\psi_i\}]$, and everything else, E_{XC} . The “known” terms include four contributions:

$$\begin{aligned} E_{\text{known}}[\{\psi_i\}] = & -\frac{\hbar^2}{m} \sum_i \int \psi_i^* \nabla^2 \psi_i d^3 r + \int V(\mathbf{r}) n(\mathbf{r}) d^3 r \\ & + \frac{e^2}{2} \int \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r d^3 r' + E_{\text{ion}}. \end{aligned} \quad (1.4)$$

The terms on the right are, in order, the electron kinetic energies, the Coulomb interactions between the electrons and the nuclei, the Coulomb interactions between pairs of electrons, and the Coulomb interactions between pairs of nuclei. The other term in the complete energy functional, $E_{\text{XC}}[\{\psi_i\}]$, is the exchange–correlation functional, and it is defined to include all the quantum mechanical effects that are not included in the “known” terms.

Let us imagine for now that we can express the as-yet-undefined exchange–correlation energy functional in some useful way. What is involved in finding minimum energy solutions of the total energy functional? Nothing we have presented so far really guarantees that this task is any easier than the formidable task of fully solving the Schrödinger equation for the wave function. This difficulty was solved by Kohn and Sham, who showed that the task of finding the right electron density can be expressed in a way that involves solving a set of equations in which each equation only involves a single electron.

The Kohn–Sham equations have the form

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{\text{XC}}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}). \quad (1.5)$$

These equations are superficially similar to Eq. (1.1). The main difference is that the Kohn–Sham equations are missing the summations that appear inside the full Schrödinger equation [Eq. (1.1)]. This is because the solution of the Kohn–Sham equations are single-electron wave functions that depend on only three spatial variables, $\psi_i(\mathbf{r})$. On the left-hand side of the Kohn–Sham equations there are three potentials, V , V_H , and V_{XC} . The first

of these also appeared in the full Schrödinger equation (Eq. (1.1)) and in the “known” part of the total energy functional given above (Eq. (1.4)). This potential defines the interaction between an electron and the collection of atomic nuclei. The second is called the Hartree potential and is defined by

$$V_H(\mathbf{r}) = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r'. \quad (1.6)$$

This potential describes the Coulomb repulsion between the electron being considered in one of the Kohn–Sham equations and the total electron density defined by all electrons in the problem. The Hartree potential includes a so-called self-interaction contribution because the electron we are describing in the Kohn–Sham equation is also part of the total electron density, so part of V_H involves a Coulomb interaction between the electron and itself. The self-interaction is unphysical, and the correction for it is one of several effects that are lumped together into the final potential in the Kohn–Sham equations, V_{XC} , which defines exchange and correlation contributions to the single-electron equations. V_{XC} can formally be defined as a “functional derivative” of the exchange–correlation energy:

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}(\mathbf{r})}{\delta n(\mathbf{r})}. \quad (1.7)$$

The strict mathematical definition of a functional derivative is slightly more subtle than the more familiar definition of a function’s derivative, but conceptually you can think of this just as a regular derivative. The functional derivative is written using δ rather than d to emphasize that it not quite identical to a normal derivative.

If you have a vague sense that there is something circular about our discussion of the Kohn–Sham equations you are exactly right. To solve the Kohn–Sham equations, we need to define the Hartree potential, and to define the Hartree potential we need to know the electron density. But to find the electron density, we must know the single-electron wave functions, and to know these wave functions we must solve the Kohn–Sham equations. To break this circle, the problem is usually treated in an iterative way as outlined in the following algorithm:

1. Define an initial, trial electron density, $n(\mathbf{r})$.
2. Solve the Kohn–Sham equations defined using the trial electron density to find the single-particle wave functions, $\psi_i(\mathbf{r})$.
3. Calculate the electron density defined by the Kohn–Sham single-particle wave functions from step 2, $n_{KS}(\mathbf{r}) = 2 \sum_i \psi_i^*(\mathbf{r})\psi_i(\mathbf{r})$.

4. Compare the calculated electron density, $n_{\text{KS}}(\mathbf{r})$, with the electron density used in solving the Kohn–Sham equations, $n(\mathbf{r})$. If the two densities are the same, then this is the ground-state electron density, and it can be used to compute the total energy. If the two densities are different, then the trial electron density must be updated in some way. Once this is done, the process begins again from step 2.

We have skipped over a whole series of important details in this process (How close do the two electron densities have to be before we consider them to be the same? What is a good way to update the trial electron density? How should we define the initial density?), but you should be able to see how this iterative method can lead to a solution of the Kohn–Sham equations that is *self-consistent*.

1.5 EXCHANGE–CORRELATION FUNCTIONAL

Let us briefly review what we have seen so far. We would like to find the ground-state energy of the Schrödinger equation, but this is extremely difficult because this is a many-body problem. The beautiful results of Kohn, Hohenberg, and Sham showed us that the ground state we seek can be found by minimizing the energy of an energy functional, and that this can be achieved by finding a self-consistent solution to a set of single-particle equations. There is just one critical complication in this otherwise beautiful formulation: to solve the Kohn–Sham equations we must specify the exchange–correlation function, $E_{\text{XC}}[\{\psi_i\}]$. As you might gather from Eqs. (1.3) and (1.4), defining $E_{\text{XC}}[\{\psi_i\}]$ is very difficult. After all, the whole point of Eq. (1.4) is that we have already explicitly written down all the “easy” parts.

In fact, the true form of the exchange–correlation functional whose existence is guaranteed by the Hohenberg–Kohn theorem is simply not known. Fortunately, there is one case where this functional can be derived exactly: the uniform electron gas. In this situation, the electron density is constant at all points in space; that is, $n(\mathbf{r}) = \text{constant}$. This situation may appear to be of limited value in any real material since it is variations in electron density that define chemical bonds and generally make materials interesting. But the uniform electron gas provides a practical way to actually use the Kohn–Sham equations. To do this, we set the exchange–correlation potential at each position to be the known exchange–correlation potential from the uniform electron gas at the electron density observed at that position:

$$V_{\text{XC}}(\mathbf{r}) = V_{\text{XC}}^{\text{electron gas}}[n(\mathbf{r})]. \quad (1.8)$$

This approximation uses only the local density to define the approximate exchange–correlation functional, so it is called the *local density approximation* (LDA). The LDA gives us a way to completely define the Kohn–Sham equations, but it is crucial to remember that the results from these equations do not exactly solve the true Schrödinger equation because we are not using the true exchange–correlation functional.

It should not surprise you that the LDA is not the only functional that has been tried within DFT calculations. The development of functionals that more faithfully represent nature remains one of the most important areas of active research in the quantum chemistry community. We promised at the beginning of the chapter to pose a problem that could win you the Nobel prize. Here it is: Develop a functional that accurately represents nature’s exact functional and implement it in a mathematical form that can be efficiently solved for large numbers of atoms. (This advice is a little like the Hohenberg–Kohn theorem—it tells you that something exists without providing any clues how to find it.)

Even though you could become a household name (at least in scientific circles) by solving this problem rigorously, there are a number of approximate functionals that have been found to give good results in a large variety of physical problems and that have been widely adopted. The primary aim of this book is to help you understand how to do calculations with these existing functionals. The best known class of functional after the LDA uses information about the local electron density and the local gradient in the electron density; this approach defines a *generalized gradient approximation* (GGA). It is tempting to think that because the GGA includes more physical information than the LDA it must be more accurate. Unfortunately, this is not always correct.

Because there are many ways in which information from the gradient of the electron density can be included in a GGA functional, there are a large number of distinct GGA functionals. Two of the most widely used functionals in calculations involving solids are the Perdew–Wang functional (PW91) and the Perdew–Burke–Ernzerhof functional (PBE). Each of these functionals are GGA functionals, and dozens of other GGA functionals have been developed and used, particularly for calculations with isolated molecules. Because different functionals will give somewhat different results for any particular configuration of atoms, it is necessary to specify what functional was used in any particular calculation rather than simply referring to “a DFT calculation.”

Our description of GGA functionals as including information from the electron density and the gradient of this density suggests that more sophisticated functionals can be constructed that use other pieces of physical information. In fact, a hierarchy of functionals can be constructed that gradually include

more and more detailed physical information. More information about this hierarchy of functionals is given in Section 10.2.

1.6 THE QUANTUM CHEMISTRY TOURIST

As you read about the approaches aside from DFT that exist for finding numerical solutions of the Schrödinger equation, it is likely that you will rapidly encounter a bewildering array of acronyms. This experience could be a little bit like visiting a sophisticated city in an unfamiliar country. You may recognize that this new city is beautiful, and you definitely wish to appreciate its merits, but you are not planning to live there permanently. You could spend years in advance of your trip studying the language, history, culture, and geography of the country before your visit, but most likely for a brief visit you are more interested in talking with some friends who have already visited there, reading a few travel guides, browsing a phrase book, and perhaps trying to identify a few good local restaurants. This section aims to present an overview of quantum chemical methods on the level of a phrase book or travel guide.

1.6.1 Localized and Spatially Extended Functions

One useful way to classify quantum chemistry calculations is according to the types of functions they use to represent their solutions. Broadly speaking, these methods use either spatially localized functions or spatially extended functions. As an example of a spatially localized function, Fig. 1.1 shows the function

$$f(x) = f_1(x) + f_2(x) + f_3(x), \quad (1.9)$$

where

$$\begin{aligned} f_1(x) &= \exp(-x^2), \\ f_2(x) &= x^2 \exp(-x^2/2), \\ f_3(x) &= \frac{1}{10}x^2(1-x)^2 \exp(-x^2/4). \end{aligned}$$

Figure 1.1 also shows f_1 , f_2 , and f_3 . All of these functions rapidly approach zero for large values of $|x|$. Functions like this are entirely appropriate for representing the wave function or electron density of an isolated atom. This example incorporates the idea that we can combine multiple individual functions with different spatial extents, symmetries, and so on to define an overall function. We could include more information in this final function by including more individual functions within its definition. Also, we could build up functions that describe multiple atoms simply by using an appropriate set of localized functions for each individual atom.

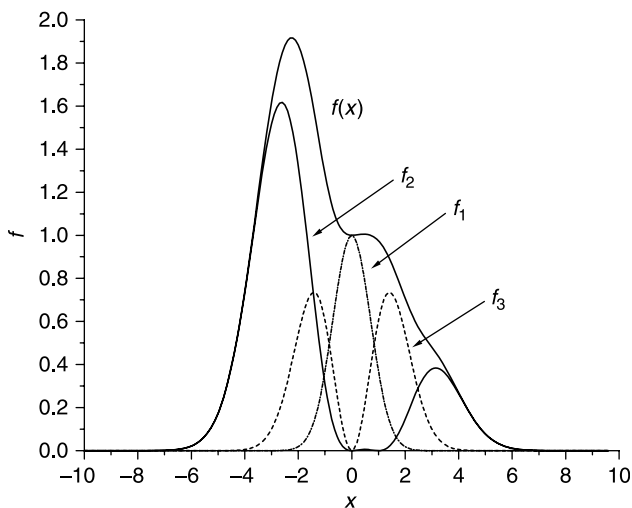


Figure 1.1 Example of spatially localized functions defined in the text.

Spatially localized functions are an extremely useful framework for thinking about the quantum chemistry of isolated molecules because the wave functions of isolated molecules really do decay to zero far away from the molecule. But what if we are interested in a bulk material such as the atoms in solid silicon or the atoms beneath the surface of a metal catalyst? We could still use spatially localized functions to describe each atom and add up these functions to describe the overall material, but this is certainly not the only way forward. A useful alternative is to use periodic functions to describe the wave functions or electron densities. Figure 1.2 shows a simple example of this idea by plotting

$$f(x) = f_1(x) + f_2(x) + f_3(x),$$

$$\begin{aligned} \text{where } f_1(x) &= \sin^2\left(\frac{\pi x}{4}\right), \\ f_2(x) &= \frac{1}{3} \cos^2\left(\frac{\pi x}{2}\right), \\ f_3(x) &= \frac{1}{10} \sin^2(\pi x). \end{aligned}$$

The resulting function is periodic; that is

$$f(x + 4n) = f(x),$$

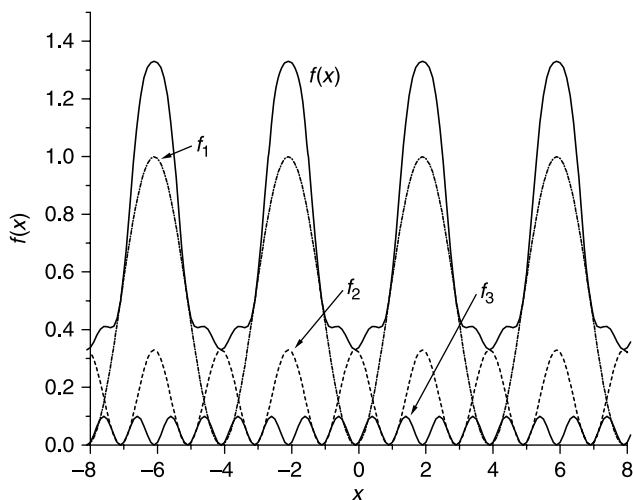


Figure 1.2 Example of spatially periodic functions defined in the text.

for any integer n . This type of function is useful for describing bulk materials since at least for defect-free materials the electron density and wave function really are spatially periodic functions.

Because spatially localized functions are the natural choice for isolated molecules, the quantum chemistry methods developed within the chemistry community are dominated by methods based on these functions. Conversely, because physicists have historically been more interested in bulk materials than in individual molecules, numerical methods for solving the Schrödinger equation developed in the physics community are dominated by spatially periodic functions. You should not view one of these approaches as “right” and the other as “wrong” as they both have advantages and disadvantages.

1.6.2 Wave-Function-Based Methods

A second fundamental classification of quantum chemistry calculations can be made according to the quantity that is being calculated. Our introduction to DFT in the previous sections has emphasized that in DFT the aim is to compute the electron *density*, not the electron *wave function*. There are many methods, however, where the object of the calculation is to compute the full electron wave function. These wave-function-based methods hold a crucial advantage over DFT calculations in that there is a well-defined hierarchy of methods that, given infinite computer time, can converge to the *exact* solution of the Schrödinger equation. We cannot do justice to the breadth of this field in just a few paragraphs, but several excellent introductory texts are available

and are listed in the Further Reading section at the end of this chapter. The strong connections between DFT and wave-function-based methods and their importance together within science was recognized in 1998 when the Nobel prize in chemistry was awarded jointly to Walter Kohn for his work developing the foundations of DFT and John Pople for his groundbreaking work on developing a quantum chemistry computer code for calculating the electronic structure of atoms and molecules. It is interesting to note that this was the first time that a Nobel prize in chemistry or physics was awarded for the development of a numerical method (or more precisely, a class of numerical methods) rather than a distinct scientific discovery. Kohn's Nobel lecture gives a very readable description of the advantages and disadvantages of wave-function-based and DFT calculations.⁵

Before giving a brief discussion of wave-function-based methods, we must first describe the common ways in which the wave function is described. We mentioned earlier that the wave function of an N -particle system is an N -dimensional function. But what, exactly, *is* a wave function? Because we want our wave functions to provide a quantum mechanical description of a system of N electrons, these wave functions must satisfy several mathematical properties exhibited by real electrons. For example, the Pauli exclusion principle prohibits two electrons with the same spin from existing at the same physical location simultaneously.[‡] We would, of course, like these properties to also exist in any approximate form of the wave function that we construct.

1.6.3 Hartree–Fock Method

Suppose we would like to approximate the wave function of N electrons. Let us assume for the moment that the electrons have *no effect* on each other. If this is true, the Hamiltonian for the electrons may be written as

$$H = \sum_{i=1}^N h_i, \quad (1.10)$$

where h_i describes the kinetic and potential energy of electron i . The full electronic Hamiltonian we wrote down in Eq. (1.1) takes this form if we simply neglect electron–electron interactions. If we write down the Schrödinger

[‡]Spin is a quantum mechanical property that does not appear in classical mechanics. An electron can have one of two distinct spins, spin up or spin down. The full specification of an electron's state must include both its location and its spin. The Pauli exclusion principle only applies to electrons with the same spin state.

equation for just one electron based on this Hamiltonian, the solutions would satisfy

$$h\chi = E\chi. \quad (1.11)$$

The eigenfunctions defined by this equation are called *spin orbitals*. For each single-electron equation there are multiple eigenfunctions, so this defines a set of spin orbitals $\chi_j(\mathbf{x}_i)$ ($j = 1, 2, \dots$) where \mathbf{x}_i is a vector of coordinates that defines the position of electron i and its spin state (up or down). We will denote the energy of spin orbital $\chi_j(\mathbf{x}_i)$ by E_j . It is useful to label the spin orbitals so that the orbital with $j = 1$ has the lowest energy, the orbital with $j = 2$ has the next highest energy, and so on. When the total Hamiltonian is simply a sum of one-electron operators, h_i , it follows that the eigenfunctions of H are products of the one-electron spin orbitals:

$$\psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \chi_{j_1}(\mathbf{x}_1)\chi_{j_2}(\mathbf{x}_2) \cdots \chi_{j_N}(\mathbf{x}_N). \quad (1.12)$$

The energy of this wave function is the sum of the spin orbital energies, $E = E_{j_1} + \cdots + E_{j_N}$. We have already seen a brief glimpse of this approximation to the N -electron wave function, the Hartree product, in Section 1.3.

Unfortunately, the Hartree product does not satisfy all the important criteria for wave functions. Because electrons are fermions, the wave function must change sign if two electrons change places with each other. This is known as the antisymmetry principle. Exchanging two electrons does not change the sign of the Hartree product, which is a serious drawback. We can obtain a better approximation to the wave function by using a Slater determinant. In a Slater determinant, the N -electron wave function is formed by combining one-electron wave functions in a way that satisfies the antisymmetry principle. This is done by expressing the overall wave function as the determinant of a matrix of single-electron wave functions. It is best to see how this works for the case of two electrons. For two electrons, the Slater determinant is

$$\begin{aligned} \psi(\mathbf{x}_1, \mathbf{x}_2) &= \frac{1}{\sqrt{2}} \det \begin{bmatrix} \chi_j(\mathbf{x}_1) & \chi_j(\mathbf{x}_2) \\ \chi_k(\mathbf{x}_1) & \chi_k(\mathbf{x}_2) \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} [\chi_j(\mathbf{x}_1)\chi_k(\mathbf{x}_2) - \chi_j(\mathbf{x}_2)\chi_k(\mathbf{x}_1)]. \end{aligned} \quad (1.13)$$

The coefficient of $(1/\sqrt{2})$ is simply a normalization factor. This expression builds in a physical description of electron exchange implicitly; it changes sign if two electrons are exchanged. This expression has other advantages. For example, it does not distinguish between electrons and it disappears if two electrons have the same coordinates or if two of the one-electron wave functions are the same. This means that the Slater determinant satisfies

the conditions of the Pauli exclusion principle. The Slater determinant may be generalized to a system of N electrons easily; it is the determinant of an $N \times N$ matrix of single-electron spin orbitals. By using a Slater determinant, we are ensuring that our method for solving the Schrödinger equation will include exchange. Unfortunately, this is not the only kind of electron correlation that we need to describe in order to arrive at good computational accuracy.

The description above may seem a little unhelpful since we know that in any interesting system the electrons interact with one another. The many different wave-function-based approaches to solving the Schrödinger equation differ in how these interactions are approximated. To understand the types of approximations that can be used, it is worth looking at the simplest approach, the Hartree–Fock method, in some detail. There are also many similarities between Hartree–Fock calculations and the DFT calculations we have described in the previous sections, so understanding this method is a useful way to view these ideas from a slightly different perspective.

In a Hartree–Fock (HF) calculation, we fix the positions of the atomic nuclei and aim to determine the wave function of N -interacting electrons. The first part of describing an HF calculation is to define what equations are solved. The Schrödinger equation for each electron is written as

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) \right] \chi_j(\mathbf{x}) = E_j \chi_j(\mathbf{x}). \quad (1.14)$$

The third term on the left-hand side is the same Hartree potential we saw in Eq. (1.5):

$$V_H(\mathbf{r}) = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r'. \quad (1.15)$$

In plain language, this means that a single electron “feels” the effect of other electrons only as an average, rather than feeling the instantaneous repulsive forces generated as electrons become close in space. If you compare Eq. (1.14) with the Kohn–Sham equations, Eq. (1.5), you will notice that the only difference between the two sets of equations is the additional exchange–correlation potential that appears in the Kohn–Sham equations.

To complete our description of the HF method, we have to define how the solutions of the single-electron equations above are expressed and how these solutions are combined to give the N -electron wave function. The HF approach assumes that the complete wave function can be approximated using a single Slater determinant. This means that the N lowest energy spin orbitals of the

single-electron equation are found, $\chi_j(\mathbf{x})$ for $j = 1, \dots, N$, and the total wave function is formed from the Slater determinant of these spin orbitals.

To actually solve the single-electron equation in a practical calculation, we have to define the spin orbitals using a finite amount of information since we cannot describe an arbitrary continuous function on a computer. To do this, we define a finite set of functions that can be added together to approximate the exact spin orbitals. If our finite set of functions is written as $\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \dots, \phi_K(\mathbf{x})$, then we can approximate the spin orbitals as

$$\chi_j(\mathbf{x}) = \sum_{i=1}^K \alpha_{j,i} \phi_i(\mathbf{x}). \quad (1.16)$$

When using this expression, we only need to find the expansion coefficients, $\alpha_{j,i}$, for $i = 1, \dots, K$ and $j = 1, \dots, N$ to fully define all the spin orbitals that are used in the HF method. The set of functions $\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \dots, \phi_K(\mathbf{x})$ is called the *basis set* for the calculation. Intuitively, you can guess that using a larger basis set (i.e., increasing K) will increase the accuracy of the calculation but also increase the amount of effort needed to find a solution. Similarly, choosing basis functions that are very similar to the types of spin orbitals that actually appear in real materials will improve the accuracy of an HF calculation. As we hinted at in Section 1.6.1, the characteristics of these functions can differ depending on the type of material that is being considered.

We now have all the pieces in place to perform an HF calculation—a basis set in which the individual spin orbitals are expanded, the equations that the spin orbitals must satisfy, and a prescription for forming the final wave function once the spin orbitals are known. But there is one crucial complication left to deal with; one that also appeared when we discussed the Kohn–Sham equations in Section 1.4. To find the spin orbitals we must solve the single-electron equations. To define the Hartree potential in the single-electron equations, we must know the electron density. But to know the electron density, we must define the electron wave function, which is found using the individual spin orbitals! To break this circle, an HF calculation is an iterative procedure that can be outlined as follows:

1. Make an initial estimate of the spin orbitals $\chi_j(\mathbf{x}) = \sum_{i=1}^K \alpha_{j,i} \phi_i(\mathbf{x})$ by specifying the expansion coefficients, $\alpha_{j,i}$.
2. From the current estimate of the spin orbitals, define the electron density, $n(\mathbf{r}')$.
3. Using the electron density from step 2, solve the single-electron equations for the spin orbitals.

4. If the spin orbitals found in step 3 are consistent with orbitals used in step 2, then these are the solutions to the HF problem we set out to calculate. If not, then a new estimate for the spin orbitals must be made and we then return to step 2.

This procedure is extremely similar to the iterative method we outlined in Section 1.4 for solving the Kohn–Sham equations within a DFT calculation. Just as in our discussion in Section 1.4, we have glossed over many details that are of great importance for actually doing an HF calculation. To identify just a few of these details: How do we decide if two sets of spin orbitals are similar enough to be called consistent? How can we update the spin orbitals in step 4 so that the overall calculation will actually converge to a solution? How large should a basis set be? How can we form a useful initial estimate of the spin orbitals? How do we efficiently find the expansion coefficients that define the solutions to the single-electron equations? Delving into the details of these issues would take us well beyond our aim in this section of giving an overview of quantum chemistry methods, but we hope that you can appreciate that reasonable answers to each of these questions can be found that allow HF calculations to be performed for physically interesting materials.

1.6.4 Beyond Hartree–Fock

The Hartree–Fock method provides an exact description of electron exchange. This means that wave functions from HF calculations have exactly the same properties when the coordinates of two or more electrons are exchanged as the true solutions of the full Schrödinger equation. If HF calculations were possible using an infinitely large basis set, the energy of N electrons that would be calculated is known as the *Hartree–Fock limit*. This energy is not the same as the energy for the true electron wave function because the HF method does not correctly describe how electrons influence other electrons. More succinctly, the HF method does not deal with electron correlations.

As we hinted at in the previous sections, writing down the physical laws that govern electron correlation is straightforward, but finding an exact description of electron correlation is intractable for any but the simplest systems. For the purposes of quantum chemistry, the energy due to electron correlation is defined in a specific way: the electron correlation energy is the difference between the Hartree–Fock limit and the true (non-relativistic) ground-state energy. Quantum chemistry approaches that are more sophisticated than the HF method for approximately solving the Schrödinger equation capture some part of the electron correlation energy by improving in some way upon one of the assumptions that were adopted in the Hartree–Fock approach.

How do more advanced quantum chemical approaches improve on the HF method? The approaches vary, but the common goal is to include a description of electron correlation. Electron correlation is often described by “mixing” into the wave function some configurations in which electrons have been excited or promoted from lower energy to higher energy orbitals. One group of methods that does this are the single-determinant methods in which a single Slater determinant is used as the reference wave function and excitations are made from that wave function. Methods based on a single reference determinant are formally known as “post–Hartree–Fock” methods. These methods include configuration interaction (CI), coupled cluster (CC), Møller–Plesset perturbation theory (MP), and the quadratic configuration interaction (QCI) approach. Each of these methods has multiple variants with names that describe salient details of the methods. For example, CCSD calculations are coupled-cluster calculations involving excitations of single electrons (S), and pairs of electrons (double—D), while CCSDT calculations further include excitations of three electrons (triples—T). Møller–Plesset perturbation theory is based on adding a small perturbation (the correlation potential) to a zero-order Hamiltonian (the HF Hamiltonian, usually). In the Møller–Plesset perturbation theory approach, a number is used to indicate the order of the perturbation theory, so MP2 is the second-order theory and so on.

Another class of methods uses more than one Slater determinant as the reference wave function. The methods used to describe electron correlation within these calculations are similar in some ways to the methods listed above. These methods include multiconfigurational self-consistent field (MCSCF), multireference single and double configuration interaction (MRDCI), and *N*-electron valence state perturbation theory (NEVPT) methods.[§]

The classification of wave-function-based methods has two distinct components: the level of theory and the basis set. The level of theory defines the approximations that are introduced to describe electron–electron interactions. This is described by the array of acronyms introduced in the preceding paragraphs that describe various levels of theory. It has been suggested, only half-jokingly, that a useful rule for assessing the accuracy of a quantum chemistry calculation is that “the longer the acronym, the better the level of theory.”⁶ The second, and equally important, component in classifying wave-function-based methods is the basis set. In the simple example we gave in Section 1.6.1 of a spatially localized function, we formed an overall function by adding together three individual functions. If we were aiming to approximate a particular function in this way, for example, the solution of the Schrödinger

[§]This may be a good time to remind yourself that this overview of quantum chemistry is meant to act something like a phrase book or travel guide for a foreign city. Details of the methods listed here may be found in the Further Reading section at the end of this chapter.

equation, we could always achieve this task more accurately by using more functions in our sum. Using a basis set with more functions allows a more accurate representation of the true solution but also requires more computational effort since the numerical coefficients defining the magnitude of each function's contribution to the net function must be calculated. Just as there are multiple levels of theory that can be used, there are many possible ways to form basis sets.

To illustrate the role of the level of theory and the basis set, we will look at two properties of a molecule of CH_4 , the C–H bond length and the ionization energy. Experimentally, the C–H bond length is 1.094 \AA ⁷ and the ionization energy for methane is 12.61 eV. First, we list these quantities calculated with four different levels of theory using the same basis set in Table 1.1. Three of the levels of theory shown in this table are wave-function-based, namely HF, MP2, and CCSD. We also list results from a DFT calculation using the most popular DFT functional for isolated molecules, that is, the B3LYP functional. (We return at the end of this section to the characteristics of this functional.) The table also shows the computational time needed for each calculation normalized by the time for the HF calculation. An important observation from this column is that the computational time for the HF and DFT calculations are approximately the same—this is a quite general result. The higher levels of theory, particularly the CCSD calculation, take considerably more computational time than the HF (or DFT) calculations.

All of the levels of theory listed in Table 1.1 predict the C–H bond length with accuracy within 1%. One piece of cheering information from Table 1.1 is that the DFT method predicts this bond length as accurately as the much more computationally expensive CCSD approach. The error in the ionization energy predicted by HF is substantial, but all three of the other methods give better predictions. The higher levels of theory (MP2 and CCSD) give considerably more accurate results for this quantity than DFT.

Now we look at the properties of CH_4 predicted by a set of calculations in which the level of theory is fixed and the size of the basis set is varied.

TABLE 1.1 Computed Properties of CH_4 Molecule for Four Levels of Theory Using pVTZ Basis Set^a

Level of Theory	C–H (Å)	Percent Error	Ionization (eV)	Percent Error	Relative Time
HF	1.085	–0.8	11.49	–8.9	1
DFT (B3LYP)	1.088	–0.5	12.46	–1.2	1
MP2	1.085	–0.8	12.58	–0.2	2
CCSD	1.088	–0.5	12.54	–0.5	18

^aErrors are defined relative to the experimental value.

TABLE 1.2 Properties of CH₄ Calculated Using DFT (B3LYP) with Four Different Basis Sets^a

Basis Set	Number of Basis Functions	C–H (Å)	Percent Error	Ionization (eV)	Percent Error	Relative Time
STO-3G	27	1.097	0.3	12.08	−4.2	1
cc-pVDZ	61	1.100	0.6	12.34	−2.2	1
cc-pVTZ	121	1.088	−0.5	12.46	−1.2	2
cc-pVQZ	240	1.088	−0.5	12.46	−1.2	13

^aErrors are defined relative to the experimental value. Time is defined relative to the STO-3G calculation.

Table 1.2 contains results of this kind using DFT calculations with the B3LYP functional in each case. There is a complicated series of names associated with different basis sets. Without going into the details, let us just say that STO-3G is a very common “minimal” basis set while cc-pVDZ, cc-pVTZ, and cc-pVQZ (D stands for double, T for triple, etc.) is a popular series of basis sets that have been carefully developed to be numerically efficient for molecular calculations. The table lists the number of basis functions used in each calculation and also the computational time relative to the most rapid calculation. All of the basis sets listed in Table 1.2 give C–H bond lengths that are within 1% of the experimental value. The ionization energy, however, becomes significantly more accurate as the size of the basis set becomes larger.

One other interesting observation from Table 1.2 is that the results for the two largest basis sets, pVTZ and pVQZ, are identical (at least to the numerical precision we listed in the table). This occurs when the basis sets include enough functions to accurately describe the solution of the Schrödinger equation, and when it occurs the results are said to be “converged with respect to basis set.” When it happens, this is a good thing! An unfortunate fact of nature is that a basis set that is large enough for one level of theory, say DFT, is not necessarily large enough for higher levels of theory. So the results in Table 1.2 do not imply that the pVTZ basis set used for the CCSD calculations in Table 1.1 were converged with respect to basis set.

In order to use wave-function-based methods to converge to the true solution of the Schrödinger equation, it is necessary to simultaneously use a high level of theory *and* a large basis set. Unfortunately, this approach is only feasible for calculations involving relatively small numbers of atoms because the computational expense associated with these calculations increases rapidly with the level of theory and the number of basis functions. For a basis set with N functions, for example, the computational expense of a conventional HF calculation typically requires $\sim N^4$ operations, while a conventional coupled-cluster calculation requires $\sim N^7$ operations. Advances have been made that improve the scaling of both HF and post-HF calculations. Even with these improvements, however you can appreciate the problem with

scaling if you notice from Table 1.2 that a reasonable basis set for even a tiny molecule like CH_4 includes hundreds of basis functions. The computational expense of high-level wave-function-based methods means that these calculations are feasible for individual organic molecules containing 10–20 atoms, but physical systems larger than this fall into either the “very challenging” or “computationally infeasible” categories.

This brings our brief tour of quantum chemistry almost to an end. As the title of this book suggests, we are going to focus throughout the book on density functional theory calculations. Moreover, we will only consider methods based on spatially periodic functions—the so-called plane-wave methods. Plane-wave methods are the method of choice in almost all situations where the physical material of interest is an extended crystalline material rather than an isolated molecule. As we stated above, it is not appropriate to view methods based on periodic functions as “right” and methods based on spatially localized functions as “wrong” (or vice versa). In the long run, it will be a great advantage to you to understand both classes of methods since having access to a wide range of tools can only improve your chances of solving significant scientific problems. Nevertheless, if you are interested in applying computational methods to materials other than isolated molecules, then plane-wave DFT is an excellent place to start.

It is important for us to emphasize that DFT calculations can also be performed using spatially localized functions—the results in Tables 1.1 and 1.2 are examples of this kind of calculation. Perhaps the main difference between DFT calculations using periodic and spatially localized functions lies in the exchange–correlation functionals that are routinely used. In Section 1.4 we defined the exchange–correlation functional by what it does not include—it is the parts of the complete energy functional that are left once we separate out the contributions that can be written in simple ways. Our discussion of the HF method, however, indicates that it is possible to treat the exchange part of the problem in an exact way, at least in principle. The most commonly used functionals in DFT calculations based on spatially localized basis functions are “hybrid” functionals that mix the exact results for the exchange part of the functional with approximations for the correlation part. The B3LYP functional is by far the most widely used of these hybrid functionals. The B stands for Becke, who worked on the exchange part of the problem, the LYP stands for Lee, Yang, and Parr, who developed the correlation part of the functional, and the 3 describes the particular way that the results are mixed together. Unfortunately, the form of the exact exchange results mean that they can be efficiently implemented for applications based on spatially localized functions but not for applications using periodic functions! Because of this fact, the functionals that are commonly used in plane-wave DFT calculations do not include contributions from the exact exchange results.

1.7 WHAT CAN DFT NOT DO?

It is very important to come to grips with the fact that practical DFT calculations are *not* exact solutions of the full Schrödinger equation. This inexactness exists because the exact functional that the Hohenberg–Kohn theorem applies to is not known. So any time you (or anyone else) performs a DFT calculation, there is an intrinsic uncertainty that exists between the energies calculated with DFT and the true ground-state energies of the Schrödinger equation. In many situations, there is no direct way to estimate the magnitude of this uncertainty apart from careful comparisons with experimental measurements. As you read further through this book, we hope you will come to appreciate that there are many physical situations where the accuracy of DFT calculations is good enough to make powerful predictions about the properties of complex materials. The vignettes in Section 1.2 give several examples of this idea. We discuss the complicated issue of the accuracy of DFT calculations in Chapter 10.

There are some important situations for which DFT cannot be expected to be physically accurate. Below, we briefly discuss some of the most common problems that fall into this category. The first situation where DFT calculations have limited accuracy is in the calculation of electronic excited states. This can be understood in a general way by looking back at the statement of the Hohenberg–Kohn theorems in Section 1.4; these theorems only apply to the ground-state energy. It is certainly possible to make predictions about excited states from DFT calculations, but it is important to remember that these predictions are not—theoretically speaking—on the same footing as similar predictions made for ground-state properties.

A well-known inaccuracy in DFT is the underestimation of calculated band gaps in semiconducting and insulating materials. In isolated molecules, the energies that are accessible to individual electrons form a discrete set (usually described in terms of molecular orbitals). In crystalline materials, these energies must be described by continuous functions known as energy bands. The simplest definition of metals and insulators involves what energy levels are available to the electrons in the material with the highest energy once all the low-energy bands are filled in accordance with the Pauli exclusion principle. If the next available electronic state lies only at an infinitesimal energy above the highest occupied state, then the material is said to be a metal. If the next available electronic state sits a finite energy above the highest occupied state, then the material is not a metal and the energy difference between these two states is called the band gap. By convention, materials with “large” band gaps (i.e., band gaps of multiple electron volts) are called insulators while materials with “small” band gaps are called semiconductors. Standard DFT calculations with existing functionals have limited accuracy for band gaps,

with errors larger than 1 eV being common when comparing with experimental data. A subtle feature of this issue is that it has been shown that even the formally exact Kohn–Sham exchange–correlation functional would suffer from the same underlying problem.^{||}

Another situation where DFT calculations give inaccurate results is associated with the weak van der Waals (vdW) attractions that exist between atoms and molecules. To see that interactions like this exist, you only have to think about a simple molecule like CH₄ (methane). Methane becomes a liquid at sufficiently low temperatures and high enough pressures. The transportation of methane over long distances is far more economical in this liquid form than as a gas; this is the basis of the worldwide liquefied natural gas (LNG) industry. But to become a liquid, some attractive interactions between pairs of CH₄ molecules must exist. The attractive interactions are the van der Waals interactions, which, at the most fundamental level, occur because of correlations that exist between temporary fluctuations in the electron density of one molecule and the energy of the electrons in another molecule responding to these fluctuations. This description already hints at the reason that describing these interactions with DFT is challenging; van der Waals interactions are a direct result of long range electron correlation. To accurately calculate the strength of these interactions from quantum mechanics, it is necessary to use high-level wave-function-based methods that treat electron correlation in a systematic way. This has been done, for example, to calculate the very weak interactions that exist between pairs of H₂ molecules, where it is known experimentally that energy of two H₂ molecules in their most favored geometry is ~ 0.003 eV lower than the energy of the same molecules separated by a long distance.⁸

There is one more fundamental limitation of DFT that is crucial to appreciate, and it stems from the computational expense associated with solving the mathematical problem posed by DFT. It is reasonable to say that calculations that involve tens of atoms are now routine, calculations involving hundreds of atoms are feasible but are considered challenging research-level problems, and calculations involving a thousand or more atoms are possible but restricted to a small group of people developing state-of-the-art codes and using some of the world's largest computers. To keep this in a physical perspective, a droplet of water 1 μm in radius contains on the order of 10^{11} atoms. No conceivable increase in computing technology or code efficiency will allow DFT

^{||}Development of methods related to DFT that can treat this situation accurately is an active area of research where considerable progress is being made. Two representative examples of this kind of work are P. Rinke, A. Qteish, J. Neugebauer, and M. Scheffler, Exciting Prospects for Solids: Exact-Exchange Based Functional Meet Quasiparticle Energy Calculations, *Phys. Stat. Sol.* **245** (2008), 929, and J. Uddin, J. E. Peralta, and G. E. Scuseria, Density Functional Theory Study of Bulk Platinum Monoxide, *Phys. Rev. B*, **71** (2005), 155112.

calculations to directly examine collections of atoms of this size. As a result, anyone using DFT calculations must clearly understand how information from calculations with extremely small numbers of atoms can be connected with information that is physically relevant to real materials.

1.8 DENSITY FUNCTIONAL THEORY IN OTHER FIELDS

For completeness, we need to point out that the name density functional theory is not solely applied to the type of quantum mechanics calculations we have described in this chapter. The idea of casting problems using functionals of density has also been used in the classical theory of fluid thermodynamics. In this case, the density of interest is the fluid density not the electron density, and the basic equation of interest is not the Schrödinger equation. Realizing that these two distinct scientific communities use the same name for their methods may save you some confusion if you find yourself in a seminar by a researcher from the other community.

1.9 HOW TO APPROACH THIS BOOK (REVISITED)

We began this chapter with an analogy about learning to drive to describe our aims for this book. Now that we have introduced much of the terminology associated with DFT and quantum chemistry calculations, we can state the subject matter and approach of the book more precisely. The remaining chapters focus on using plane-wave DFT calculations with commonly applied functionals to physical questions involving bulk materials, surfaces, nanoparticles, and molecules. Because codes to perform these plane-wave calculations are now widely available, we aim to introduce many of the issues associated with applying these methods to interesting scientific questions in a computationally efficient way.

The book has been written with two audiences in mind. The primary audience is readers who are entering a field of research where they will perform DFT calculations (and perhaps other kinds of computational chemistry or materials modeling) on a daily basis. If this describes you, it is important that you perform as many of the exercises at the end of the chapters as possible. These exercises have been chosen to require relatively modest computational resources while exploring most of the key ideas introduced in each chapter. Simply put, if your aim is to enter a field where you will perform calculations, then you must actually do calculations of your own, not just read about other people's work. As in almost every endeavor, there are many details that are best learned by experience. For readers in this group, we recommend reading through every chapter sequentially.

The second audience is people who are unlikely to routinely perform their own calculations, but who work in a field where DFT calculations have become a “standard” approach. For this group, it is important to understand the language used to describe DFT calculations and the strengths and limitations of DFT. This situation is no different from “standard” experimental techniques such as X-ray diffraction or scanning electron microscopy, where a working knowledge of the basic methods is indispensable to a huge community of researchers, regardless of whether they personally apply these methods. If you are in this audience, we hope that this book can help you become a sophisticated consumer of DFT results in a relatively efficient way. If you have a limited amount of time (a long plane flight, for example), we recommend that you read Chapter 3, Chapter 10, and then read whichever of Chapters 4–9 appears most relevant to you. If (when?) your flight is delayed, read one of the chapters that doesn’t appear directly relevant to your specific research interests—we hope that you will learn something interesting.

We have consciously limited the length of the book in the belief that the prospect of reading and understanding an entire book of this length is more appealing than the alternative of facing (and carrying) something the size of a large city’s phone book. Inevitably, this means that our coverage of various topics is limited in scope. In particular, we do not examine the details of DFT calculations using localized basis sets beyond the cursory treatment already presented in this chapter. We also do not delve deeply into the theory of DFT and the construction of functionals. In this context, the word “introduction” appears in the title of the book deliberately. You should view this book as an entry point into the vibrant world of DFT, computational chemistry, and materials modeling. By following the resources that are listed at the end of each chapter in the Further Reading section, we hope that you will continue to expand your horizons far beyond the introduction that this book gives.

We have opted to defer the crucial issue of the accuracy of DFT calculations until chapter 10, after introducing the application of DFT to a wide variety of physical properties in the preceding chapters. The discussion in that chapter emphasizes that this topic cannot be described in a simplistic way. Chapter 10 also points to some of the areas in which rapid developments are currently being made in the application of DFT to challenging physical problems.

REFERENCES

1. K. Honkala, A. Hellman, I. N. Remediakis, A. Logadottir, A. Carlsson, S. Dahl, C. H. Christensen, and J. K. Nørskov, Ammonia Synthesis from First-Principles Calculations, *Science* **307** (2005), 555.
2. R. Schweinfest, A. T. Paxton, and M. W. Finnis, Bismuth Embrittlement of Copper is an Atomic Size Effect, *Nature* **432** (2004), 1008.

3. K. Umemoto, R. M. Wentzcovitch, and P. B. Allen, Dissociation of MgSiO_3 in the Cores of Gas Giants and Terrestrial Exoplanets, *Science* **311** (2006), 983.
4. K. Umemoto, R. M. Wentzcovitch, D. J. Weidner, and J. B. Parise, NaMgF_3 : A Low-Pressure Analog of MgSiO_3 , *Geophys. Res. Lett.* **33** (2006), L15304.
5. W. Kohn, Nobel Lecture: Electronic Structure of Matter-Wave Functions and Density Functionals, *Rev. Mod. Phys.* **71** (1999), 1253.
6. C. J. Cramer and J. T. Roberts, Computational Analysis: Y_2K , *Science* **286** (1999), 2281.
7. G. Herzberg, *Electronic Spectra and Electronic Structure of Polyatomic Molecules*, Van Nostrand, New York, 1966.
8. P. Diep and J. K. Johnson, An Accurate $\text{H}_2\text{-H}_2$ Interaction Potential from First Principles, *J. Chem. Phys.* **112** (2000), 4465.

FURTHER READING

Throughout this book, we will list resources for further reading at the end of each chapter. You should think of these lists as pointers to help you learn about topics we have mentioned or simplified in a detailed way. We have made no attempt to make these lists exhaustive in any sense (to understand why, find out how many textbooks exist dealing with “quantum mechanics” in some form or another).

Among the many books on quantum mechanics that have been written, the following are good places to start if you would like to review the basic concepts we have touched on in this chapter:

- P. W. Atkins and R. S. Friedman, *Molecular Quantum Mechanics*, Oxford University Press, Oxford, UK, 1997.
- D. A. McQuarrie, *Quantum Chemistry*, University Science Books, Mill Valley, CA, 1983.
- M. A. Ratner and G. C. Schatz, *Introduction to Quantum Mechanics in Chemistry*, Prentice Hall, Upper Saddle River, NJ, 2001.
- J. Simons and J. Nichols, *Quantum Mechanics in Chemistry*, Oxford University Press, New York, 1997.

Detailed accounts of DFT are available in:

- W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, Weinheim, 2000.
- R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, UK, 2004.
- R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, UK, 1989.

Resources for learning about the wide range of quantum chemistry calculation methods that go beyond DFT include:

- J. B. Foresman and A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian Inc., Pittsburgh, 1996.
- A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover, Mineola, NY, 1996.
- D. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, Wiley, New York, 2001.

A book that gives a relatively brief overview of band theory is:

- A. P. Sutton, *Electronic Structure of Materials*, Oxford University Press, Oxford, UK, 1993.

Two traditional sources for a more in-depth view of this topic are:

- N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College Publishing, Orlando, 1976.
- C. Kittel, *Introduction to Solid State Physics*, Wiley, New York, 1976.

A good source if you want to learn about the fluid thermodynamics version of DFT is:

- H. Ted Davis, *Statistical Mechanics of Phases, Interfaces, and Thin Films*, Wiley-VCH, 1995.

