Cambridge University Press
0521815916 - Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods
Jorge Kohanoff
Excerpt
More information

Part I

Theory

Cambridge University Press
0521815916 - Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods
Jorge Kohanoff
Excerpt
More information

1

The problem of the structure of matter

The description of the physical and chemical properties of matter is a central issue that has occupied the minds of scientists since the age of the ancient Greeks. In their route to dissect matter down to what cannot be divided any further, they coined the term *atom*, the *indivisible*. Matter became then a collection of atoms. More than twenty centuries had to pass until the development of a more precise concept of atom, thanks, amongst others, to the systematic studies of Mendeleyev and the establishment in 1869 of the periodic table of the elements (Mendeleyev, 1869). The discovery of the electron in 1897 and the first modern model of the atomic structure by Sir Joseph Thomson were soon refined by his student, Sir Ernest Rutherford, who in 1910 showed that an atom was made of a positively charged small nucleus and a number of negatively charged electrons that neutralize the nuclear charge. Much in the spirit of planetary systems, and drawing from the analogy between gravitational and electrostatic interactions, scientists in the beginning of the twentieth century built an image of the atom that consisted of a number Z of electrons – of elementary charge $-e$ – orbiting around the nucleus of charge Ze.

A number of experimental observations, though, were incompatible with this idea of orbiting electrons. In particular, according to the successful electromagnetic theory, charged electrons in orbital (radially accelerated) motion should radiate energy, thus decelerating and eventually collapsing onto the nucleus. Clearly, such a picture would imply that matter is essentially unstable, in flagrant contradiction with our everyday experience of the very existence of matter. It was this kind of incompatibility that motivated the idea that matter at such small scale does not obey the laws of classical mechanics and electromagnetism, but a different set of laws, whose body became known as *quantum mechanics* (Bohr, 1913). To solve the problem of electron radiation, Bohr postulated the existence of certain peculiar orbits for which the electron would not radiate. These orbits correspond to specific energies and radii, and the promotion from one orbit to another could only happen through the absorption or emission of a *quantum* of energy. One of the appealing aspects of this theory was its conceptual compatibility with Planck's theory of black body (electromagnetic) radiation, which also required discrete energy levels. The emission of light of very precise frequency by atoms could then be explained as the decay of an electron from a higher to a lower energy level, by emitting a quantum of radiation whose frequency was proportional to the energy difference between these levels. These ebullient new ideas gave a tremendous momentum to the field, and, in a few decades, the mathematical apparatus, the language of quantum mechanics, was largely developed. Schrödinger's equation was published in 1926 (Schrödinger, 1926), and it was soon applied to multi-electronic atoms and to polyatomic systems such as molecules (Heitler and London, 1927) and solids (Bloch, 1928). It is the main goal of this book to describe the different approaches to the description of matter at the atomic scale, i.e. in terms of atomic nuclei and electrons.

In general terms, we can imagine a piece of matter as a collection of interacting atoms, sometimes under the influence of an external field. This ensemble of particles may be in the gas phase (molecules, clusters) or in a condensed phase (bulk solids, surfaces, wires). It could be in a solid, liquid or amorphous phase, either homogeneous or heterogeneous (molecules in solution, interfaces, adsorbates on surfaces). However, at this scale, we can unambiguously describe all these systems as a set of atomic nuclei and electrons interacting via coulombic, electrostatic forces. Formally, we can write the Hamiltonian of such a system in the following general form:

$$\begin{aligned}\hat{\mathcal{H}} = &-\sum_{I=1}^{P}\frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_{i=1}^{N}\frac{\hbar^2}{2m}\nabla_i^2 + \frac{e^2}{2}\sum_{I=1}^{P}\sum_{J\neq I}^{P}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|}\\ &+\frac{e^2}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} - e^2\sum_{I=1}^{P}\sum_{i=1}^{N}\frac{Z_I}{|\mathbf{R}_I-\mathbf{r}_i|},\end{aligned} \tag{1.1}$$

where $\mathbf{R} = \{\mathbf{R}_I, \quad I = 1, \ldots, P\}$ is a set of P nuclear coordinates, and $\mathbf{r} = \{\mathbf{r}_i, \quad i = 1, \ldots, N\}$ is a set of N electronic coordinates. Z_I and M_I are the nuclear charges and masses, respectively. Since the electrons are fermions, the total electronic wave function must be antisymmetric, i.e. it should change sign whenever the coordinates of any two electrons are exchanged. Different nuclear species are distinguishable, but nuclei of the same species also obey a specific statistics according to the nuclear spin. They are fermions for half-integer nuclear spin (e.g. H, ^{3}He) and bosons for integer spin (e.g. D, ^{4}He, H_2). All the ingredients are well known and, in principle, all the properties can be derived by solving the time-independent Schrödinger equation:

$$\hat{\mathcal{H}}\ \Psi_n(\mathbf{R},\mathbf{r}) = \mathcal{E}_n\Psi_n(\mathbf{R},\mathbf{r}), \tag{1.2}$$

where $\mathcal{E}_n$ are the energy eigenvalues and $\Psi_n(\mathbf{r}, \mathbf{R})$ are the corresponding eigenstates, or wave functions, which must be antisymmetric with respect to exchange of electronic coordinates in $\mathbf{r}$, and symmetric or antisymmetric with respect to exchange of nuclear variables in $\mathbf{R}$.

In practice, this problem is almost impossible to treat within a full quantum mechanical framework. Only in a few cases, such as hydrogenoid atoms or the H_2^+ molecule, a complete analytic solution is available. Exact numerical solutions are also limited to a few cases, mostly atoms and very small molecules. There are several features that contribute to this difficulty, but the most important is that this is a multi-component many-body system, and the two-body nature of the Coulomb interaction makes the above Schrödinger equation not separable.

In order to fix ideas, let us confine ourselves to the case of an atom with Z electrons, and focus on the electronic wave function. First of all, to respect the antisymmetry of the wave function against electron exchange, we can, in principle, write such a wave function as an antisymmetrized product of one-electron wave functions (a so-called Slater determinant). This assumes, however, some kind of separability of the Schrödinger equation, implying that the probability of finding an electron at some point in space is essentially independent of where the other electrons are located. The repulsive electron–electron interaction is quite at odds with this picture, because an electron located at point $\mathbf{r}$ in space precludes other electrons from approaching this location, much in the spirit of an *exclusion zone*. Hence, the probability of finding an electron at $\mathbf{r}$ depends on the location of the other $Z-1$ electrons. This phenomenon is known as *correlation*, and it implies that the exact many-body wave function should contain factors depending on two electronic coordinates. Therefore, the image in terms of one-electron wave functions can be somewhat crude in many cases.

This means that the full Schrödinger equation cannot be easily decoupled into a set of equations, so that, in general, we have to deal with $3(P+N)$ coupled degrees of freedom. The usual choice is to resort to a few reasonable and well-controlled approximations, which encompass a wide variety of problems of interest. Of course, there are systems where the hypotheses leading to these approximations are violated, and these are by no means uninteresting. They require, however, a much larger theoretical and computational effort, and thus, historically, precedence has been given to the *easier* systems.

In the first part of this book we shall develop the theory starting from two major approximations: the adiabatic separation of the nuclear and electronic degrees of freedom, and the classical treatment of atomic nuclei. We shall then discuss in detail the different approaches to tackling the electronic problem, as emerged from two, often contrasting, but mostly cooperating, communities: *chemists* and *physicists*, trying to establish a common language between them. This will be done in the spirit of facilitating the understanding of chemical approaches by physicists, and vice versa. In fact, similar approaches are sometimes adopted by the two communities using different names, as is the case of tight-binding and semiempirical methods. In the second part, we shall concentrate on the different computational methods proposed to cast the problem of the structure of matter in a form amenable to numerical treatment, devoting our attention also to specific algorithms.

1.1 Adiabatic approximation

A first observation is that the time scale associated with the motion of nuclei is usually much slower than that associated with electrons. In fact, the most unfavorable case of a single proton already corresponds to a mass ratio of 1:1836, i.e. less than 1%. Within a classical picture we could say that, under typical conditions, the velocity of the electron is much larger than that of the heavy particle (the proton). In this spirit, Born and Oppenheimer (1927) proposed a scheme for separating the motion of nuclei from that of the electrons. The original work studied the time-independent Hamiltonian (1.1) perturbatively in the mass ratio $\kappa = (m/M)^{1/4}$. The influence of the nuclei on the electronic wave functions becomes apparent when considering a new set of nuclear variables defined by $\mathbf{R} = \mathbf{R}_0 + \kappa\mathbf{u}$, where $\mathbf{u}$ represents the displacement of the nuclei with respect to their equilibrium positions $\mathbf{R}_0$. Using this change of variables and expanding the $\mathbf{R}$-dependent terms in the potential in powers of $\mathbf{u}$, they obtained an expansion of the Hamiltonian in powers of κ. By keeping terms up to fourth order in κ, they showed that no mixing of different electronic stationary states happened due to the interaction with the nuclei. Therefore, under appropriate conditions that are discussed below, *the electrons do not undergo transitions between stationary states*. This is called the *adiabatic approximation*. The reason for the name is based more on dynamical, rather than stationary, arguments. We will first present a semiclassical picture which serves to fix the ideas, and then the quantum mechanical derivation from time-dependent perturbation theory. This will allow us to inspect the limits of validity of the adiabatic approximation.

Let us first analyze the case of a molecular system, where the electronic spectrum is discrete. The arguments here have to be adapted to the case of infinite systems such as solids or liquids, which exhibit a continuum spectrum of excitations. In a molecule there are basically three types of motion: electronic, nuclear vibrations, and nuclear rotations, each one corresponding to a typical time (or energy) scale. The energy scale of the electronic motion is given by the separation between successive electronic eigenstates, which is of the same order of magnitude as the ground state energy. This quantity is of the order of

$E_e \approx \hbar^2/(ma^2)$, where a is a typical interatomic distance in the molecule. This distance a gives an indication of the order of magnitude of the amplitude of the electronic motion. Typical distances are of the order of a few Å, and the associated energies are in the order of 1 eV. The energy scale of nuclear vibrations is given, in the harmonic approximation, by the quantum $E_v = \hbar\omega$, where ω is the frequency of the vibrational motion. The order of magnitude of such a frequency can be obtained by thinking that the energy of a classical oscillatory motion of amplitude a is $M\omega^2a^2$. With a being a typical interatomic distance, a motion of this amplitude would cause the extraction of an atom from the molecule. Since atoms in molecules bind mostly via shared electrons (chemical bonding), this energy can be identified with that of breaking a chemical bond, which is of the same order of magnitude as the electronic energy E_e. Hence, $M\omega^2a^2 \approx E_e$, and replacing a in terms of E_e from the first expression leads to $E_v = \hbar\omega \approx (m/M)^{1/2}E_e$. Rotational energies are the smallest of them all. They are related to the angular momentum of the molecule, which is quantized in levels separated by $\approx \hbar$. The energy is $E_r = L^2/I$, where I is the moment of inertia of the molecule, which is of the order of Ma^2. Therefore, $E_r \approx \hbar^2/Ma^2 = (m/M)\hbar^2/ma^2 = (m/M)E_e$. In summary, there is the following relation between these three energy scales:

$$E_r \approx \kappa^2 E_v \approx \kappa^4 E_e. \tag{1.3}$$

Therefore, if we take a value of $\kappa \approx 0.066$ (corresponding to the N_2 molecule), we have that rotational energies are about two orders of magnitude smaller than vibrational energies, and these in turn are two orders of magnitude smaller than electronic energies. For example, the first electronic excitation energy in the N_2 molecule is 7.5 eV, the vibrational excitation energy is about 300 meV, and the rotational level separation is around 0.5 meV.

In extended systems the electronic spectrum is continuous and, in principle, so is the excitation spectrum. Here we have to distinguish two cases. For insulating and semiconducting systems the smallest electronic excitation is given by the energy gap, which, as in the case of molecules, falls in the region of a few eV ($E_g \approx 1.1$ eV for silicon and ≈ 4 eV for diamond). For metallic systems the electronic excitations form a continuum, so that, formally, $E_e = 0$ and, strictly speaking, the adiabatic approximation should not be applicable. We shall discuss this at the end of this section, but, for the moment being, let us analyze the situation in which, for some reason, the electronic energy levels are more narrowly spaced than what has been assumed above. In that case the electronic energy scale becomes comparable to that of the nuclear vibrational motion, the adiabatic separation breaks down, and nuclei and electrons have to be treated in a unified framework. Examples of this behavior are when the electronic gap closes due to some external factor, e.g. pressure, doping, or temperature. States can be formed where electrons are intimately coupled to a nuclear vibration, thus giving rise to exotic entities like *polarons*. Thermal agitation also modifies the spacing between electronic levels but in a dynamical way. Occasionally, two or more levels may become so close that the nuclear motion promotes an electronic transition from one to another level, thus violating the hypotheses of the adiabatic approximation. If these crossings involve electron–nuclear interactions that occur during a limited period of time, then they can be studied using asymptotic techniques that view the non-adiabatic region in terms of transition probabilities, like the Landau–Zener approach (Messiah, 1961). The same phenomenon happens in atomic and molecular collisions. A more general approach requires the dynamical treatment of the electronic motion coupled to the nuclear motion, but described at the quantum-mechanical level. Practical schemes to achieve this goal within a harmonic description of the nuclear motion (phonons) have only very recently become available (see, e.g., Horsfield *et al.,* 2004).

The electrons can then be thought of as *instantaneously following the motion of the nuclei, while remaining always in the same stationary state of the electronic Hamiltonian*. This stationary state will vary in time because of the electrostatic coupling of the two sets of degrees of freedom, but, if the electrons were, e.g., in the (many-electron) ground state, they would remain there forever. In other words, as the nuclei follow their dynamics, the electrons instantaneously adjust their wave function according to the nuclear wave function. This approximation ignores the possibility of having non-radiative transitions between different electronic eigenstates. Transitions can only arise through the coupling with an external electromagnetic field, but this issue will not be addressed in the following.

All this can be cast in a formal mathematical framework by proposing a solution to Eq. (1.2) of the following form:

$$\Psi(\mathbf{R}, \mathbf{r}, t) = \sum_n \Theta_n(\mathbf{R}, t)\Phi_n(\mathbf{R}, \mathbf{r}), \tag{1.4}$$

where $\Theta_n(\mathbf{R}, t)$ are wave functions describing the evolution of the nuclear subsystem in each one of the *adiabatic* electronic eigenstates $\Phi_n(\mathbf{R}, \mathbf{r})$. These satisfy the time-independent Schrödinger equation

$$\hat{h}_{\mathrm{e}}\Phi_n(\mathbf{R}, \mathbf{r}) = E_n(\mathbf{R})\Phi_n(\mathbf{R}, \mathbf{r}), \tag{1.5}$$

where the electronic Hamiltonian is:

$$\hat{h}_{\mathrm{e}} = \hat{T} + \hat{U}_{\mathrm{ee}} + \hat{V}_{\mathrm{ne}} = \hat{\mathcal{H}} - \hat{T}_{\mathrm{n}} - \hat{V}_{\mathrm{nn}}. \tag{1.6}$$

Here, $\hat{T}$ is the electronic kinetic operator, $\hat{U}_{\mathrm{ee}}$ is the electron–electron interaction, $\hat{V}_{\mathrm{ne}}$ the electron–nuclear interaction, $\hat{T}_{\mathrm{n}}$ the nuclear kinetic operator, and $\hat{V}_{\mathrm{nn}}$ the inter-nuclear interaction.

In this partial differential equation on the **r** variables, the $3P$ nuclear coordinates **R** enter as parameters. This expansion, which is always mathematically possible, is called the expansion in the *adiabatic basis*, because $\Phi_n(\mathbf{R}, \mathbf{r})$ are solutions of the time-independent electronic Schrödinger equation, corresponding to a particular nuclear configuration. Equation (1.5) has to be solved for all nuclear configurations **R** where the nuclear wave function is non-vanishing.

By replacing the above ansatz into the full Schrödinger equation we obtain:

$$\begin{aligned}\left[i\hbar \frac{\partial}{\partial t} + \sum_{I=1}^{P} \frac{\hbar^2}{2M_I} \nabla_I^2 - V_{nn} - E_q(\mathbf{R}) \right] \Theta_q(\mathbf{R}, t) = \\ -\sum_n \sum_{I=1}^{P} \frac{\hbar^2}{2M_I} \langle \Phi_q | \nabla_I^2 | \Phi_n \rangle \Theta_n(\mathbf{R}, t) \\ -2 \sum_n \sum_{I=1}^{P} \frac{\hbar^2}{2M_I} \vec{\nabla}_I \Theta_n(\mathbf{R}, t) \cdot \langle \Phi_q | \vec{\nabla}_I | \Phi_n \rangle, \end{aligned} \tag{1.7}$$

which constitutes a set (infinite, in principle) of coupled partial differential equations containing off-diagonal terms. Here we have used Dirac's bra-ket notation to indicate matrix elements of the type

$$\langle \Phi_q | \vec{\nabla}_I | \Phi_n \rangle = \int \Phi_q^*(\mathbf{R}, \mathbf{r}) \vec{\nabla}_I \Phi_n(\mathbf{R}, \mathbf{r}) \, d\mathbf{r}. \tag{1.8}$$

Therefore, the reduction of the full wave function to an expression of the type

$$\Psi(\mathbf{R}, \mathbf{r}, t) = \Theta_n(\mathbf{R}, t) \Phi_n(\mathbf{R}, \mathbf{r}) \tag{1.9}$$

is not completely correct, because, even if the system was initially *prepared* in a *pure* state like the above one, the off-diagonal terms will mix (excite) the different electronic eigenstates along the temporal evolution. These are precisely the non-radiative transitions alluded to above. If this is the case, then the dynamics is said to be *non-adiabatic*. However, if the off-diagonal terms can be neglected, then an expression like (1.9) is valid because the nuclear dynamics has no means to cause electronic transitions, and the electrons remain always in the same (n) adiabatic state (ground or excited). In this case, the dynamics is said to be *adiabatic*.

The necessary condition for neglecting the non-adiabatic couplings is that

$$\left| \sum_{I=1}^{P} \frac{\hbar^2}{M_I} \langle \Theta_q | \vec{\nabla}_I | \Theta_n \rangle \cdot \langle \Phi_q | \vec{\nabla}_I | \Phi_n \rangle \right| \ll |E_q(\mathbf{R}) - E_n(\mathbf{R})| \tag{1.10}$$

or, equivalently,

$$\frac{m}{M} \left| \frac{\hbar \Omega_\text{v}}{E_q(\mathbf{R}) - E_n(\mathbf{R})} \right| \ll 1, \tag{1.11}$$

where Ω_v is the maximum frequency of rotation of the electronic wave function due to the nuclear motion, and the energies in the denominator correspond to the electronic adiabatic eigenstates (the energy gap if $q = 1$ and $n = 0$). Notice that the mass ratio m/M is always smaller than 5×10^{-4}, thus justifying the adiabatic approximation unless a very small gap occurs, as for open-shell, conical intersections or Jahn–Teller systems. The case of lighter particles such as muons would be different. Typical electronic excitations are of the order of 1 eV, while typical nuclear excitations (phonons) are of the order of 0.01 eV. This indicates that there is a clear separation of energy (and consequently time) scales. There are situations in which this approximation is not adequate, but they are rather exceptional cases.

The arguments in favor of an adiabatic treatment of metallic systems are more subtle. In principle, it can be argued that the adiabatic approximation breaks down because the energy gap is zero and electronic excitations of vanishing energy are possible. However, since typical temperatures (between room temperature and a few thousand degrees) are usually much lower than the electronic Fermi temperature, excitations are confined to a narrow region around the Fermi surface, and most properties are little affected by neglecting non-adiabatic contributions due to these few electrons. In fact, the usual treatment of electronic transport phenomena in metals begins with the adiabatic description and introduces non-adiabatic terms (in the form of electron–phonon interactions) afterwards, perturbatively. In terms of the ratio of energy scales, it can also be realized that the relevant excitations in metals at small wave numbers are not electron–hole pairs, which, besides being very few, carry a small oscillator strength (Migdal's theorem). The relevant energy scale is actually dictated by the plasmon (collective charge excitation), which is again typically of the order of a few eV.

1.2 Classical nuclei approximation

Therefore, according to the adiabatic approximation, the total wave function can be written in the form of Expression (1.9), where $\Theta_n(\mathbf{R}, t)$ is the nuclear wave function. At room temperature the thermal wavelength is $\lambda_T = (\hbar^2/2Mk_BT)^{1/2}$, which, for hydrogen at room temperature, is of the order of 0.2 Å. Regions of space separated by more than λ_T do not exhibit quantum phase coherence. Interatomic distances are normally of the order of 1 Å, and then the total nuclear wave function can be considered as an incoherent superposition of individual nuclear wave packets:

$$\Theta_n(\mathbf{R}, t) = \prod_{I=1}^{P} \Theta_n^{(I)}(\mathbf{R}, \mathbf{R}^{(I)}(t), t), \tag{1.12}$$

where $\mathbf{R}^{(I)}(t)$ are the centers of the individual wave packets. The details of the decoherence process, i.e. the quantum-to-classical transition, have been the subject of intense debate in recent years and are now fairly well understood (see Habib *et al.*, 1998, and references therein), but they go beyond the scope of this book. The above expression for the nuclear wave function is called the time-dependent Hartree approximation, and does not include correlations between the different nuclei. Exchange effects are also absent in this expression, although they could be recovered by proposing a total wave function in the form of a Slater determinant (for odd-spin nuclei), thus leading to the so-called time-dependent Hartree–Fock approximation. However, atomic nuclei exhibit exchange effects only at very low temperatures, e.g. below 5 K in the case of hydrogen. In addition, nuclear masses are typically large enough that the individual nuclear wave packets are quite localized, provided that the curvature of the potential where they move is sufficiently large. For instance, the ground state of a proton in a typical molecular bonding environment has a width of about 0.25 Å. The combination of these two observations allows us to propose that, in most cases, atomic nuclei can be treated as *classical* particles.

The time-dependent adiabatic Schrödinger equation for the nuclear wave function is

$$i\hbar \frac{\partial \Theta_m(\mathbf{R}, t)}{\partial t} = \left(- \sum_{I=1}^{P} \frac{\hbar^2}{2M_I} \nabla_I^2 + \tilde{\varepsilon}_n(\mathbf{R}) \right) \Theta_m(\mathbf{R}, t), \tag{1.13}$$

with

$$\tilde{\varepsilon}_n(\mathbf{R}) = \varepsilon_n(\mathbf{R}) + \sum_{I=1}^{P} \frac{\hbar^2}{2M_I} \langle \Phi_q | \nabla_I^2 | \Phi_q \rangle. \tag{1.14}$$

The second term in this expression is a diagonal correction to the electronic energy levels due to the dependence of the electronic wave function on the nuclear coordinates. It can be shown by dimensional analysis that the matrix element in the second term is proportional to the electron mass, so that the correction to the energy levels is proportional to $\kappa^4 = m/M$. It is a correction that can actually be calculated, and turns out to be smaller than 0.5% in the most unfavorable cases (Handy and Lee, 1996). Therefore, in practice, it is usually neglected, as suggested by Born and Oppenheimer (1927). When this term is included, the approximation is called *adiabatic*. When it is neglected it receives the name of *Born–Oppenheimer* approximation. Very often these two terms are used indistinctly to indicate the case in which the diagonal correction is ignored. In what follows, we shall use the Born–Oppenheimer approximation and ignore this term.