

## 2. SYMMETRY ASPECTS OF EXCITATIONS

## 2.2.9.4. Relativistic effects

If a solid contains only light elements, non-relativistic calculations are well justified, but as soon as heavier elements are present in the system of interest relativistic effects can no longer be neglected. In the medium range of atomic numbers (up to about 54), so-called scalar relativistic schemes are often used (Koelling & Harmon, 1977), which describe the main contraction or expansion of various orbitals (due to the Darwin  $s$ -shift or the mass-velocity term) but omit spin-orbit splitting. Unfortunately, the spin-orbit term couples spin-up and spin-down wavefunctions. If one has  $n$  basis functions without spin-orbit coupling, then including spin-orbit coupling in the Hamiltonian would lead to a  $2n \times 2n$  matrix equation, which requires about eight times as much computer time to solve it (due to the  $n^3$  scaling). Since the spin-orbit effect is generally small (at least for the valence states), one can simplify the procedure by diagonalizing the Hamiltonian including spin-orbit coupling in the space of the low-lying bands as obtained in a scalar relativistic step. This version is called second variational method (see *e.g.* Singh, 1994). For very heavy elements it may be necessary to solve Dirac's equation, which has all these terms (Darwin  $s$ -shift, mass-velocity and spin-orbit) included. Additional aspects are illustrated in Section 2.2.14 in connection with the uranium atom.

## 2.2.10. Density functional theory

The most widely used scheme for calculating the electronic structure of solids is based on density functional theory (DFT). It is described in many excellent books, for example that by Dreizler & Gross (1990), which contains many useful definitions, explanations and references. Hohenberg & Kohn (1964) have shown that for determining the ground-state properties of a system all one needs to know is the electron density  $\rho(\mathbf{r})$ . This is a tremendous simplification considering the complicated wavefunction of a crystal with (in principle infinitely) many electrons. This means that the total energy of a system (a solid in the present case) is a functional of the density  $E[\rho(r)]$ , which is independent of the external potential provided by all nuclei. At first it was just proved that such a functional exists, but in order to make this fundamental theorem of practical use Kohn & Sham (1965) introduced orbitals and suggested the following procedure.

In the universal approach of DFT to the quantum-mechanical many-body problem, the interacting system is mapped in a unique manner onto an effective non-interacting system of quasi-electrons with the same total density. Therefore the electron density plays the key role in this formalism. The non-interacting particles of this auxiliary system move in an effective local one-particle potential, which consists of a mean-field (Hartree) part and an exchange-correlation part that, in principle, incorporates all correlation effects exactly. However, the functional form of this potential is not known and thus one needs to make approximations.

Magnetic systems (with collinear spin alignments) require a generalization, namely a different treatment for spin-up and spin-down electrons. In this generalized form the key quantities are the spin densities  $\rho_\sigma(r)$ , in terms of which the total energy  $E_{\text{tot}}$  is

$$E_{\text{tot}}(\rho_\uparrow, \rho_\downarrow) = T_s(\rho_\uparrow, \rho_\downarrow) + E_{ee}(\rho_\uparrow, \rho_\downarrow) + E_{Ne}(\rho_\uparrow, \rho_\downarrow) + E_{xc}(\rho_\uparrow, \rho_\downarrow) + E_{NN}, \quad (2.2.10.1)$$

with the electronic contributions, labelled conventionally as, respectively, the kinetic energy (of the non-interacting particles), the electron-electron repulsion, the nuclear-electron attraction and the exchange-correlation energies. The last term  $E_{NN}$  is the repulsive Coulomb energy of the fixed nuclei. This expression is still exact but has the advantage that all terms but one can be calculated very accurately and are the dominating (large) quantities.

The exception is the exchange-correlation energy  $E_{xc}$ , which is defined by (2.2.10.1) but must be approximated. The first important methods for this were the local density approximation (LDA) or its spin-polarized generalization, the local spin density approximation (LSDA). The latter comprises two assumptions:

(i) That  $E_{xc}$  can be written in terms of a local exchange-correlation energy density  $\varepsilon_{xc}$  times the total (spin-up plus spin-down) electron density as

$$E_{xc} = \int \varepsilon_{xc}(\rho_\uparrow, \rho_\downarrow) * [\rho_\uparrow + \rho_\downarrow] dr. \quad (2.2.10.2)$$

(ii) The particular form chosen for  $\varepsilon_{xc}$ . For a homogeneous electron gas  $\varepsilon_{xc}$  is known from quantum Monte Carlo simulations, *e.g.* by Ceperley & Alder (1984). The LDA can be described in the following way. At each point  $\mathbf{r}$  in space we know the electron density  $\rho(\mathbf{r})$ . If we locally replace the system by a homogeneous electron gas of the same density, then we know its exchange-correlation energy. By integrating over all space we can calculate  $E_{xc}$ .

The most effective way known to minimize  $E_{\text{tot}}$  by means of the variational principle is to introduce (spin) orbitals  $\chi_{jk}^\sigma$  constrained to construct the spin densities [see (2.2.10.7) below]. According to Kohn and Sham (KS), the variation of  $E_{\text{tot}}$  gives the following effective one-particle Schrödinger equations, the so-called Kohn-Sham equations (Kohn & Sham, 1965) (written for an atom in Rydberg atomic units with the obvious generalization to solids):

$$[-\nabla^2 + V_{Ne} + V_{ee} + V_{xc}^\sigma] \chi_{jk}^\sigma(r) = \epsilon_{jk}^\sigma(r) \chi_{jk}^\sigma(r), \quad (2.2.10.3)$$

with the external potential (the attractive interaction of the electrons by the nucleus) given by

$$V_{Ne}(r) = \frac{2Z}{r}, \quad (2.2.10.4)$$

the Coulomb potential (the electrostatic interaction between the electrons) given by

$$V_{ee}(\mathbf{r}) = V_C(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad (2.2.10.5)$$

and the exchange-correlation potential (due to quantum mechanics) given by the functional derivative

$$V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(r)]}{\delta \rho}. \quad (2.2.10.6)$$

In the KS scheme, the (spin) electron densities are obtained by summing over all occupied states, *i.e.* by filling the KS orbitals (with increasing energy) according to the *Aufbau* principle.

$$\rho_\sigma(r) = \sum_{j,k} \rho_{jk}^\sigma |\chi_{jk}^\sigma(r)|^2. \quad (2.2.10.7)$$

Here  $\rho_{jk}^\sigma$  are occupation numbers such that  $0 \leq \rho_{jk}^\sigma \leq 1/w_k$ , where  $w_k$  is the symmetry-required weight of point  $\mathbf{k}$ . These KS equations (2.2.10.3) must be solved self-consistently in an iterative process, since finding the KS orbitals requires the knowledge of the potentials, which themselves depend on the (spin) density and thus on the orbitals again. Note the similarity to (and difference from) the Hartree-Fock equation (2.2.9.1). This version of the DFT leads to a (spin) density that is close to the exact density provided that the DFT functional is sufficiently accurate.

In early applications, the local density approximation (LDA) was frequently used and several forms of functionals exist in the literature, for example by Hedin & Lundqvist (1971), von Barth & Hedin (1972), Gunnarsson & Lundqvist (1976), Vosko *et al.* (1980) or accurate fits of the Monte Carlo simulations of Ceperley & Alder (1984). The LDA has some shortcomings, mostly due to the tendency of overbinding, which causes, for example, too-small lattice constants. Recent progress has been

## 2.2. ELECTRONS

made going beyond the LSDA by adding gradient terms or higher derivatives ( $\nabla\rho$  and  $\nabla^2\rho$ ) of the electron density to the exchange–correlation energy or its corresponding potential. In this context several physical constraints can be formulated, which an exact theory should obey. Most approximations, however, satisfy only part of them. For example, the exchange density (needed in the construction of these two quantities) should integrate to  $-1$  according to the Fermi exclusion principle (Fermi hole). Such considerations led to the generalized gradient approximation (GGA), which exists in various parameterizations, *e.g.* in the one by Perdew *et al.* (1996). This is an active field of research and thus new functionals are being developed and their accuracy tested in various applications.

The Coulomb potential  $V_c(\mathbf{r})$  in (2.2.10.5) is that of all  $N$  electrons. That is, any electron is also moving in its own field, which is physically unrealistic but may be mathematically convenient. Within the HF method (and related schemes) this self-interaction is cancelled exactly by an equivalent term in the exchange interaction [see (2.2.9.1)]. For the currently used approximate density functionals, the self-interaction cancellation is not complete and thus an error remains that may be significant, at least for states (*e.g.*  $4f$  or  $5f$ ) for which the respective orbital is not delocalized. Note that delocalized states have a negligibly small self-interaction. This problem has led to the proposal of self-interaction corrections (SICs), which remove most of this error and have impacts on both the single-particle eigenvalues and the total energy (Parr *et al.*, 1978).

The Hohenberg–Kohn theorems state that the total energy (of the ground state) is a functional of the density, but the introduction of the KS orbitals (describing quasi-electrons) are only a tool in arriving at this density and consequently the total energy. Rigorously, the Kohn–Sham orbitals are *not electronic orbitals* and the KS eigenvalues  $\varepsilon_i$  (which correspond to  $E_k$  in a solids) are *not* directly related to electronic *excitation energies*. From a formal (mathematical) point of view, the  $\varepsilon_i$  are just Lagrange multipliers without a physical meaning.

Nevertheless, it is often a good approximation (and common practice) to partly ignore these formal inconsistencies and use the orbitals and their energies in discussing electronic properties. The gross features of the eigenvalue sequence depend only to a smaller extent on the details of the potential, whether it is orbital-based as in the HF method or density-based as in DFT. In this sense, the eigenvalues are mainly determined by orthogonality conditions and by the strong nuclear potential, common to DFT and the HF method.

In processes in which one removes (ionization) or adds (electron affinity) an electron, one compares the  $N$  electron system with one with  $N - 1$  or  $N + 1$  electrons. Here another conceptual difference occurs between the HF method and DFT. In the HF method one may use Koopmans’ theorem, which states that the  $\varepsilon_i^{\text{HF}}$  agree with the ionization energies from state  $i$  assuming that the corresponding orbitals do not change in the ionization process. In DFT, the  $\varepsilon_i$  can be interpreted according to Janak’s theorem (Janak, 1978) as the partial derivative with respect to the occupation number  $n_i$ ,

$$\varepsilon_i = \frac{\partial E}{\partial n_i}. \quad (2.2.10.8)$$

Thus in the HF method  $\varepsilon_i$  is the total energy difference for  $\Delta n = 1$ , in contrast to DFT where a differential change in the occupation number defines  $\varepsilon_i$ , the proper quantity for describing metallic systems. It has been proven that for the exact density functional the eigenvalue of the highest occupied orbital is the first ionization potential (Perdew & Levy, 1983). Roughly, one can state that the further an orbital energy is away from the highest occupied state, the poorer becomes the approximation to use  $\varepsilon_i$  as excitation energy. For core energies the deviation can be significant, but one may use Slater’s transition state (Slater,

1974), in which half an electron is removed from the corresponding orbital, and then use the  $\varepsilon_i^{\text{TS}}$  to represent the ionization from that orbital.

Another excitation from the valence to the conduction band is given by the energy gap, separating the occupied from the unoccupied single-particle levels. It is well known that the gap is not given well by taking  $\Delta\varepsilon_i$  as excitation energy. Current DFT methods significantly underestimate the gap (half the experimental value), whereas the HF method usually overestimates gaps (by a factor of about two). A trivial solution, applying the ‘scissor operator’, is to shift the DFT bands to agree with the experimental gap. An improved but much more elaborate approach for obtaining electronic excitation energies within DFT is the GW method in which quasi-particle energies are calculated (Hybertsen & Louie, 1984; Godby *et al.*, 1986; Perdew, 1986). This scheme is based on calculating the dielectric matrix, which contains information on the response of the system to an external perturbation, such as the excitation of an electron.

In some cases, one can rely on the total energy of the states involved. The original Hohenberg–Kohn theorems (Hohenberg & Kohn, 1964) apply only to the ground state. The theorems may, however, be generalized to the energetically lowest state of any symmetry representation for which any property is a functional of the corresponding density. This allows (in cases where applicable) the calculation of excitation energies by taking total energy differences.

Many aspects of DFT from formalism to applications are discussed and many references are given in the book by Springborg (1997).

### 2.2.11. Band-theory methods

There are several methods for calculating the electronic structure of solids. They have advantages and disadvantages, different accuracies and computational requirements (speed or memory), and are based on different approximations. Some of these aspects have been discussed in Section 2.2.9. This is a rapidly changing field and thus only the basic concepts of a few approaches in current use are outlined below.

#### 2.2.11.1. LCAO (linear combination of atomic orbitals)

For the description of crystalline wavefunctions (Bloch functions), one often starts with a simple concept of placing atomic orbitals (AOs) at each site in a crystal denoted by  $|m\rangle$ , from which one forms Bloch sums in order to have proper translational symmetry:

$$\chi_{\mathbf{k}}(\mathbf{r}) = \sum_m \exp(i\mathbf{k}\mathbf{T}_m)|m\rangle. \quad (2.2.11.1)$$

Then Bloch functions can be constructed by taking a linear combination of such Bloch sums, where the linear-combination coefficients are determined by the variational principle in which a secular equation must be solved. The LCAO can be used in combination with both the Hartree–Fock method and DFT.

#### 2.2.11.2. TB (tight binding)

A simple version of the LCAO is found by parameterizing the matrix elements  $\langle m'|\mathbb{H}|m\rangle$  and  $\langle m'|m\rangle$  in a way similar to the Hückel molecular orbital (HMO) method, where the only non-vanishing matrix elements are the on-site integrals and the nearest-neighbour interactions (hopping integrals). For a particular class of solids the parameters can be adjusted to fit experimental values. With these parameters, the electronic structures of rather complicated solids can be described and yield quite satisfactory results, but only for the class of materials for which such a parametrization is available. Chemical bonding and symmetry aspects can be well described with such schemes, as Hoffmann has illustrated in many applications (Hoffmann, 1988).