

Chapter 8

Spin-Orbit Interaction

8.1 Origin of spin-orbit interaction

Spin-orbit interaction is a well-known phenomenon that manifests itself in lifting the degeneracy of one-electron energy levels in atoms, molecules, and solids. In solid-state physics, the nonrelativistic Schrödinger equation is frequently used as a first approximation, e.g. in electron band-structure calculations. Without relativistic corrections, it leads to doubly-degenerated bands, spin-up and spin-down, which can be split by a spin-dependent term in the Hamiltonian. In this approach, spin-orbit interaction can be included as a relativistic correction to the Schrödinger equation.

Let us briefly review the origins of this correction, following approach of Ref. [80]. For this purpose one has to consider the Dirac equation, which is the basic equation for electronic systems, including the electron spin and its relativistic behavior. One obtains the Dirac equation by linearizing the relativistic generalization of the Schrödinger equation. It is Lorentz-invariant and describes the electron spin and spin-orbit coupling from first principles. One naturally arrives at the Dirac equation when starting from the relativistic expression for the kinetic energy

$$H^2 = c^2 \mathbf{p}^2 + m^2 c^4. \quad (8.1)$$

Inclusion of the electric and magnetic potentials, ϕ and A , by substituting $\mathbf{p} - (\frac{e}{c})\mathbf{A}$ for \mathbf{p} (\mathbf{p} is the canonical momentum) and $H - e\phi$ for H (electron charge $\varepsilon = -e$) in equation 8.1 leads to:

$$(H - e\phi)^2 = (c\mathbf{p} - \varepsilon\mathbf{A})^2 + m^2 c^4. \quad (8.2)$$

By interpreting \mathbf{p} and H as operators

$$\mathbf{p} = -i\hbar\nabla \quad \text{and} \quad H = i\hbar\frac{\partial}{\partial t}, \quad (8.3)$$

one obtains the relativistic wave equation in external electric and magnetic fields. One can write the force-free form of the wave equation as

$$(H^2 - c^2 \sum_{\mu} p_{\mu}^2 - m^2 c^4) \psi = 0. \quad (8.4)$$

$p_{\mu} = p_x, p_y, p_z$ are components of the momentum operator, as follows from Eq. 8.2. A disadvantage of this differential equation is that one needs the initial values of ψ and $\frac{\partial}{\partial t} \psi$ to solve it, whereas the Schrödinger equation requires only the initial value of ψ . The equation can be expressed in the form

$$(H - c \sum_{\mu} \alpha_{\mu} p_{\mu} - \beta m c^2)(H + c \sum_{\mu} \alpha_{\mu} p_{\mu} - \beta m c^2) \psi = 0 \quad (8.5)$$

if the constant coefficients α_{μ} and β satisfy the relations

$$\begin{aligned} \alpha_{\mu} \alpha_{\mu'} + \alpha_{\mu'} \alpha_{\mu} &= 2\delta_{\mu\mu'} \\ \alpha_{\mu} \beta + \beta \alpha_{\mu} &= 0 \\ \beta^2 &= 1. \end{aligned}$$

One can linearize it by taking only the first part of Eq. 8.5. Solution of the first part will give the solution of whole equation. The linearized equation has the advantage that it is of first order in $\frac{\partial}{\partial t}$, similar to the Schrödinger equation; it is referred to as Dirac equation:

$$(H - c \sum_{\mu} \alpha_{\mu} p_{\mu} - \beta m c^2) \psi = 0, \quad (8.6)$$

with the four-component vector ψ .

Let us now compare the Dirac equation with the Schrödinger equation. For this we use the non-linearized form

$$[H - \varepsilon \phi - c \boldsymbol{\alpha} \cdot (\mathbf{p} - \frac{\varepsilon}{c} \mathbf{A}) - \beta m c^2][H - \varepsilon \phi + c \boldsymbol{\alpha} \cdot (\mathbf{p} - \frac{\varepsilon}{c} \mathbf{A}) + \beta m c^2] \psi = 0. \quad (8.7)$$

Using the approximation that the kinetic and potential energies are small compared to $m c^2$, two components of the spin function can be neglected, and equation 8.7 takes the form

$$\left(\frac{1}{2m} (\mathbf{p} - \frac{\varepsilon}{c} \mathbf{A})^2 + \varepsilon \phi - \frac{\varepsilon \hbar}{2m c} \boldsymbol{\sigma} \cdot \mathbf{B} + i \frac{\varepsilon \hbar}{4m^2 c^2} \mathbf{E} \cdot \mathbf{p} - \frac{\varepsilon \hbar}{4m^2 c^2} \boldsymbol{\sigma} \cdot (\mathbf{E} \times \mathbf{p}) \right) \psi = W \psi, \quad (8.8)$$

where $W + m c^2$ is the total energy.

The first two terms in the parentheses are equivalent to those in the Schrödinger equation for external fields. The third term corresponds to the interaction energy $-\vec{\mu} \cdot \mathbf{B}$ of the magnetic dipole, whose moment is presented by operator $\vec{\mu} = \frac{\varepsilon \hbar}{2m c} \vec{\sigma} = \frac{\varepsilon}{m c} \mathbf{s}$. The fourth term is a relativistic correction to the energy and does not have a classical analogy. The fifth term describes the spin-orbit coupling.

It can be illustrated in framework of the classical electrodynamics, where the vectors of the electromagnetic field depend on the reference system. In a reference system that moves with velocity \mathbf{v} relative to an electric field \mathbf{E} , one finds a magnetic field

$$\mathbf{B} = -\frac{1}{c}\mathbf{v} \times \mathbf{E} = \frac{1}{mc}(\mathbf{E} \times \mathbf{p}), \quad (8.9)$$

where terms of order $(\frac{v}{c})^2$ and higher order terms are neglected. In other words, the moving electron experiences a magnetic field in its rest frame that arises from the Lorentz transformation of the static (external) electric field; this field will affect the electron spin. The energy of the electron in this field, due to its magnetic moment μ , is

$$-\boldsymbol{\mu} \cdot \mathbf{B} = -\frac{\varepsilon}{mc}\mathbf{s} \cdot \mathbf{B} = -\frac{\varepsilon}{m^2c^2}\mathbf{s} \cdot (\mathbf{E} \times \mathbf{p}). \quad (8.10)$$

This additional energy term in the Hamiltonian is essentially the spin-orbit term in Eq. 8.8, except for a factor of 2. This factor (Thomas factor) is missing because we have not taken into account that changing the frame of reference also leads to a time transformation and, consequently, the precession frequency of the electron spin in the magnetic field changes.

In case of centrally symmetric electrical fields, as e.g. the orbital motion of an electron in the electric field of an atomic nucleus, one has

$$\mathbf{E} = -\frac{1}{\varepsilon} \frac{\mathbf{r}}{r} \frac{dV}{dr}, \quad (8.11)$$

and the term can be written in the form ($\mathbf{s} = \frac{\hbar}{2}\boldsymbol{\sigma}$)

$$-\frac{\varepsilon\hbar}{4m^2c^2}\boldsymbol{\sigma} \cdot (\mathbf{E} \times \mathbf{p}) = \frac{\varepsilon}{2m^2c^2}\mathbf{s} \cdot \left(-\frac{1}{\varepsilon} \frac{\mathbf{r}}{r} \frac{dV}{dr} \times \mathbf{p} \right) = \frac{1}{2m^2c^2} \frac{1}{r} \frac{dV}{dr} (\mathbf{s} \cdot \mathbf{l}). \quad (8.12)$$

This is called spin-orbit energy, as it results from an interaction of the spin with the magnetic field that is experienced by the moving electron.

8.2 Inversion symmetry and time reversal symmetry

The form of the spin-orbit interaction term in Eq. 8.2 suggests that a spin-degenerate level can be split by spin-orbit coupling (SOC) into levels with spin parallel and antiparallel to the orbit. In a free atom, spin-orbit interaction can lift the degeneracy of states with the same orbital (spatial) wave function but with opposite spins. In solids, however, such a splitting can be forbidden due to crystal symmetry.

As a fundamental principle of quantum mechanics, time reversal symmetry always preserves the Kramer's degeneracy between a function $\psi(r, s)$ and its complex conjugate $\psi^*(r, s)$; the latter function differs from ψ through a simultaneous reversal of wave vector and electron spin. Therefore, at any point of the Brillouin zone, one can write for the energy

$$E(k, \uparrow) = E(-k, \downarrow). \quad (8.13)$$

The Kramer's degeneracy remains unaffected when including the spin-orbit interaction term in the Hamiltonian. (Notice that \uparrow and \downarrow do not necessarily denote eigenfunctions of the spin operator σ_z , but correspond to the two quantum numbers denoting eigenstates for a given Bloch vector.)

If the crystal lattice has inversion symmetry (i.e. if the operation $\vec{r} \rightarrow -\vec{r}$ does not change the crystal lattice), one will obtain

$$E(k, \uparrow) = E(-k, \uparrow) \quad \text{and} \quad E(k, \downarrow) = E(-k, \downarrow). \quad (8.14)$$

From combination of the equations 8.13 and 8.14 it becomes clear that if both time reversal symmetry and inversion symmetry are present, the band structure should satisfy to the condition

$$E(k, \uparrow) = E(k, \downarrow). \quad (8.15)$$

In other words, the energy cannot depend on the electron spin. Consequently, for crystals which have inversion symmetry (like fcc, hcp), spin splitting is not allowed in the bulk, and these solids keep their spin degeneracy.

8.3 Hydrogen atom

Consideration of the electron orbital motion in the Coulomb field permits us to visualize spin-orbit interaction. We use the hydrogen atom for which the Schrödinger equation can be solved analytically. In the semiclassical model, moments are treated as vectors with z-components being already quantized. (For spin-orbit interaction, only the case of $l \geq 1$ is important, since SOC vanishes for $l = 0$.) The energy-level correction due to spin-orbit coupling takes the well-known form [81]:

$$E_{n,l,s} = E_n - \mu_s \cdot B_l = E_n + \frac{\mu_0 Z e^2}{8\pi m_e^2 r^3} (\mathbf{s} \cdot \mathbf{l}), \quad (8.16)$$

where

$$B_l = -\frac{\mu_0 Z e}{4\pi^3} (\mathbf{v} \times \mathbf{r}) = \frac{\mu_0 Z e}{4\pi^3 m} \mathbf{l}. \quad (8.17)$$

The dot product of spin and orbital momentum can be expressed through the eigenvalues of spin, orbital, and total angular momentum

$$\mathbf{s} \cdot \mathbf{l} = \frac{1}{2} [\mathbf{j}^2 - \mathbf{l}^2 - \mathbf{s}^2] = \frac{1}{2} \hbar^2 [j(j+1) - l(l+1) - s(s+1)]. \quad (8.18)$$

Therefore the energy can be written as

$$E_{n,l,s} = E_n + \frac{a}{2} [j(j+1) - l(l+1) - s(s+1)], \quad (8.19)$$

where $a = \frac{\mu_0 Z e^2 \hbar^2}{8\pi m_e^2 r^3}$, - the so-called spin-orbit coupling constant. Thus, the energy level defined by n , l splits into sublevels with different total moments, $j = l + \frac{1}{2}$ and $j = l - \frac{1}{2}$.

In quantum mechanics, the electron radius is treated using the probability density distribution. So the observable of $\frac{1}{r^3}$ should be written as its expectation value

$$\bar{a} = \frac{\mu_0 Z e^2 \hbar^2}{8\pi m_e^2} \cdot \int \psi_{n,l,m}^* \frac{1}{r^3} \psi_{n,l,m} d\tau. \quad (8.20)$$

If we substitute the Hydrogen wave function for $\psi_{n,l,m}(r, \theta, \phi)$, we obtain

$$\bar{a} = -E_n \cdot \frac{Z^2 \alpha^2}{n \cdot l(l+1/2)(l+1)}, \quad (8.21)$$

with the relativistic fine-structure constant $\alpha = \frac{\mu_0 c e^2}{4\pi \hbar} = \frac{e^2}{4\pi \epsilon_0 \hbar c} \simeq \frac{1}{137}$. Finally, the energy difference between two sublevels with opposite spin is

$$\Delta E_{l,s} = -E_n \cdot \frac{Z^2 \alpha^2}{n \cdot l(l+1)} \quad (8.22)$$

From Eq. 8.22 we immediately see two important aspects of SOC: (1) atomic spin-orbit splitting is proportional to the product of $E_n \cdot Z^2$; in case of a Coulomb field, SOC is proportional to $\frac{Z^4}{n^3}$. (2) Spin-orbit splitting decreases with increasing quantum numbers n and l .

8.4 Solids

Bulk crystals

If a crystal lattice potential has a center of inversion, spin-orbit interaction will not affect the band structure, except for certain points of the Brillouin zone (BZ), where the influence of spin-orbit interaction can be very important. As an example, we consider the center of a BZ (Γ point) that has cubic symmetry. In a tight binding model, we can build up the p bands from atomic p wave functions. In a free atom without SOC, there are three degenerate p functions, so three p bands will appear in the model. These bands will all be degenerate at the Γ point ($k = 0$, see Fig. 8.1(a)), transforming one into the other through cubic symmetry transformations at this point. Each of the three bands is also doubly degenerate in the spin, altogether a six-fold degeneracy at Γ .

When spin-orbit interaction is included, the bands will split in two subbands with fourfold $p_{3/2}$ and twofold $p_{1/2}$ degeneracy (see Fig. 8.1(b)). The bands will now behave differently as we move away from the Γ point depending on whether or not there is inversion symmetry in the lattice. If the crystal has a center of inversion, each band will preserve its spin degeneracy, as illustrated on Fig. 8.1(b). The $p_{3/2}$ bands, with $j = 3/2$, will now split into two bands with $m_j = \pm 3/2$ and $m_j = \pm 1/2$. (with projections taken e.g. in the direction of the vector k).

If a center of inversion symmetry is absent, then the $p_{3/2}$ and $p_{1/2}$ bands will additionally be split by spin-orbit interaction, removing the spin degeneracy. This situation is shown in Fig. 8.1(c).

The cases considered are really occurring in crystals, e.g. in semiconductors like Ge and InSb. Germanium has a center of inversion symmetry, and we will have the first scenario for the top valence band. For InSb, which crystallizes in a zinc-blende lattice, the spin degeneracy will be lifted as we leave the center of

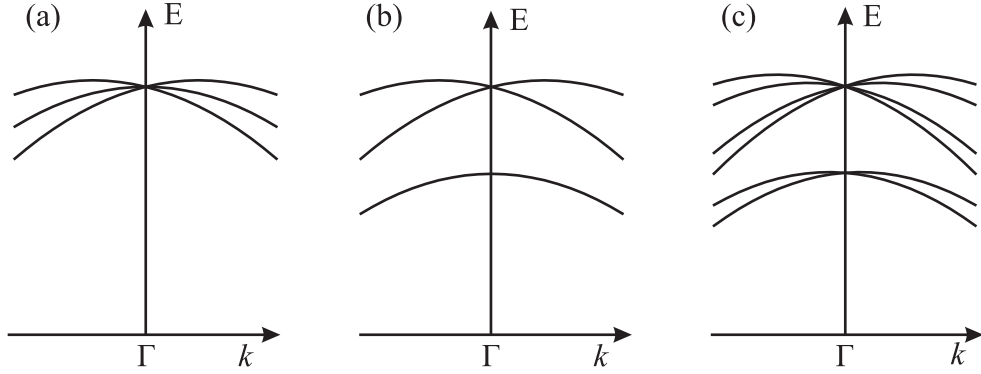


Figure 8.1: Influence of spin-orbit interaction on the p-levels at the center of the Brillouin zone: (a) Without SOC, six degenerate levels are observed at the Γ -point; (b) SOC leads to a splitting into $p_{3/2}$ and $p_{1/2}$ levels, but leaves the spin degeneracy in lattices with inversion symmetry; (c) in a lattice without inversion symmetry, the spin degeneracy is fully removed, except for the Γ -point.

the BZ. This effect of symmetry on the electronic band structure of zinc-blende crystals leads to the appearance of a spin-orbit interaction term that is cubic in k (Dresselhaus splitting).

Surfaces

Termination of the crystal by a surface breaks the 3D inversion symmetry. In this way, it permits electron surface states with the same parallel wave vector, but k_{\parallel} , opposite spins to have different energies:

$$E(k_{\parallel}, \uparrow) \neq E(k_{\parallel}, \downarrow). \quad (8.23)$$

In the absence of external or internal magnetic fields, spin-orbit splitting is the only possible cause for a splitting of spin degenerate levels at the surface, and this interaction is described by the Rashba Hamiltonian [82].