

Electrons and Phonons in Periodic Lattice/ Oded Agam



These lecture notes are inspired by a course given by Prof. Igor Aleiner from Columbia University in New York. They are intended only for the use of the students of the course: “Condensed Matter A: Electrons and Phonons in Periodic Lattice” of the Hebrew University. It is forbidden to distribute these notes in any form and use them for any commercial purpose.

© 2021 Oded Agam. All rights reserved.

Table of Contents

1 Introduction

1.1 The conductivity of Metals – Drude Model	7
1.2 Bloch's theorem	10
1.3 Effective mass, electrons, and holes.....	16
1.4 Fermi surfaces, metals, and insulators.....	18
1.5 Exercises.....	23

2 Lattices and spatial symmetries

2.1 Translation vectors.....	24
2.2 The translation group.....	25
2.3 Bravais Lattice	26
2.4 Point groups.....	28
2.5 Space groups.....	32
2.6 Decorated lattices.....	33
2.7 The Wigner-Seitz cell.....	35
2.8 Three-dimensional Bravais lattices.....	35
2.9 Directions and planes in a crystal (Miller indices).....	42
2.10 The reciprocal lattice.....	43
2.11 Exercises.....	46

3 Time-reversal symmetry

3.1 Implications of time-reversal symmetry on the electron spectrum.....	47
3.2 The time-reversal operation for spin $\frac{1}{2}$ systems.....	49
3.3 Spin-orbit interaction and Kramer's degeneracy.....	51
3.4 Rashba term.....	53
3.5 Exercises	56

4 Group Theory: Basic concepts

4.1 Definitions.....	57
4.2 Symmetry operations of point groups.....	60
4.3 Schoenflies notation of point groups.....	60
4.4 Conjugate elements and conjugacy classes.....	63
4.5 Representations of groups.....	65
4.6 Characters of irreducible representations.....	71
4.7 Character tables.....	73
4.8 Basis functions.....	75
4.9 Mulliken symbols of irreducible representations.....	81
4.10 Molecular vibrations.....	82
4.11 Irreducible representations in quantum mechanics.....	87
4.12 Exercises.....	90

5 Graphene and Dichalcogenides

5.1 The graphene lattice.....	93
5.2 The “little group” and the spectrum near the K -point.....	94
5.3 Schur’s lemma and multiplication of irreducible representations.....	101
5.4 Dichalcogenides.....	105
5.5 Exercises.....	109

6 Extended groups and double groups

6.1 Extended groups.....	111
6.2 Double groups.....	115
6.3 Spin-orbit interaction in graphene – the Kane Mele term.....	118
6.4 Topological insulators.....	120
6.5 Exercises.....	122

7 Methods for calculating band structure

7.1 The nearly free electron approximation.....	123
7.2 The tight-binding approximation.....	129
7.3 The $k \cdot p$ approximation, Kane’s model, heavy holes and light holes.....	139
7.4 Exercises.....	145

8 Topological metals

8.1 Energy bands in mercury telluride – a qualitative discussion.....	148
8.2 Weyl points.....	151
8.3 Curvature, parallel transport, and topological numbers.....	151
8.4 Berry’s curvature.....	157
8.5 Chern numbers and chiral charge.....	162
8.6 Dirac points.....	163
8.7 Fermi arcs.....	163
8.8 Exercises.....	166

9 Crystals in a constant electric field

9.1 Physical response in dielectric crystals.....	168
9.2 The effective mass approximation.....	170
9.3 Stark ladder and Wannier-Stark oscillations.....	174
9.4 Beyond the effective mass approximation -the $k \cdot p$ approximation.....	176
9.5 Dielectric breakdown (Landau-Zener tunneling)	179
9.6 Exercises.....	188

10 Crystals in a constant magnetic field

10.1	Gauge invariance and charge conservation.....	190
10.2	The effective mass approximation.....	193
10.3	Perturbative corrections to the effective mass approximation.....	196
10.4	Charged particles in a crystal subjected to magnetic field.....	199
10.5	Bohr-Sommerfeld quantization and Landau levels.....	204
10.6	Magnetic breakdown.....	207
10.7	Strong magnetic field – Preliminary discussion (Moiré patterns).....	214
10.8	Magnetic translations.....	217
10.9	Hofstadter’s butterfly.....	221
10.10	Exercises.....	222

11 Elastic deformations, sound waves, and phonons

11.1	The strain tensor.....	224
11.2	The energy of elastic deformations.....	227
11.3	Sound waves in crystals.....	231
11.4	Phonons.....	238
11.5	Optical vibrations in crystals (optical phonons).....	240
11.6	Symmetry approach to optical modes in crystals.....	243
11.7	Exercises.....	247

12 Crystals in an electromagnetic field

12.1	Infrared activity.....	250
12.2	Raman scattering.....	253
12.3	Electromagnetic waves in polarizable crystals and polaritons	256
12.4	Exercises.....	260

13 Piezoelectric and polar crystals

13.1	Piezoelectric crystals	262
13.2	Pyroelectric crystals.....	264
13.3	Ferroelectric crystals.....	266
13.4	Exercises.....	272

14 Electrons in deformed crystals

14.1	The effect of lattice deformation on the electron’s energy spectrum.....	273
14.2	Deformation of the Fermi surface of metals due to lattice deformations.....	276
14.3	Dislocations.....	280
14.4	Peierls instability.....	285
14.5	Exercises.....	290

15 Electron-phonon interaction

15.1 Deformation interaction.....	292
15.2 Transition rate between electronic states.....	297
15.3 The kinetic equations for electrons and phonons distribution functions.....	299
15.4 Thermodynamic equilibrium.....	301
15.5 The thermodynamic equilibrium from entropy considerations.....	306
15.6 Piezoelectric interaction.....	308
15.7 Fröhlich polaron.....	309
15.8 Exercises.....	313

16 Umklapp processes

16.1 Qualitative discussion.....	318
16.2 The umklapp Hamiltonian: Group theory approach.....	320
16.3 The umklapp Hamiltonian from the modulated hopping approach.....	321
16.4 Exercises.....	326

17 Semiconductors and disordered crystals

17.1 Impurities and defects in crystals.....	327
17.2 The Jahn-Teller effect of an impurity in a crystal.....	335
17.3 Elastic scattering and the optical theorem.....	341
17.4 The kinetic equations for elastic scattering.....	349
17.5 Exercises.....	352

18 Transport coefficients & thermoelectric effects

18.1 Diffusion.....	354
18.2 The electric conductivity and Einstein's relation.....	357
18.3 The second law of thermodynamics in electronic systems.....	360
18.4 Thermoelectric effects.....	367
18.5 Relations between thermoelectric/kinetic coefficients.....	369
18.6 Exercises.....	372

1 Introduction

The matter around us appears in three main phases, gas, liquid, and solid. When the temperature is sufficiently low, atoms in solids may rearrange themselves in a crystal characterized by periodic spatial structure. This course aims at understanding how this periodic structure manifests itself in the behavior of the electrons and the phonons in the system. We shall see how this periodicity dictates the structure of the energy spectrum of the electrons in the crystal, its optical properties, and the amount of electrical current induced by an electric field or temperature gradient. The scope of this course is limited to cases where the interactions among electrons do not play an important role.

The precise periodic structure of a crystal depends on the type of atoms from which it is composed and the type of chemical bonding among them. In general, there are 230 different possible periodic structures in three dimensions that are characterized by different sets of symmetry operations. However, here we shall consider only simple periodic structures (with an emphasis on two-dimensional crystals that constitutes the state-of-the-art systems in current research), but the ideas that will be developed along the course provide the basis for understanding more complicated structures.

In this introductory chapter, I will present the Drude model for conductivity which, as we shall see, motivates the study of systems with periodic structures. Following, I will discuss several basic ideas of the field: Bloch's theorem (in one dimension), Fermi surfaces, effective mass, and charge carriers, i.e. electrons and holes.

1.1 The conductivity of metals – the Drude model

The Drude model for conductivity assumes that electrons may be treated as free particles scattered from impurities. This scattering results in relaxation of the electron momentum, $\mathbf{p}$, thus Newton's second law takes the form:

$$\frac{d\langle \mathbf{p} \rangle}{dt} = -e\mathbf{E} - \frac{\langle \mathbf{p} \rangle}{\tau}, \quad (1.1)$$

where $\mathbf{E}$ is the electric field, $-e$ is the electron charge, and τ is the typical relaxation time of the momentum. The angular brackets, $\langle \dots \rangle$, denote thermodynamic averaging over the ensemble of electrons. Notice that the relaxation of momentum is generally due to a change of the momentum direction rather than a reduction of its absolute value.

The steady-state solution of the above equation yields:

$$\langle \mathbf{p} \rangle = -\tau e\mathbf{E}, \quad (1.2)$$

and if one assumes that, in the absence of collisions and external fields, electrons behave as free particles with the Hamiltonian, $H = p^2/2m$, where m is the electron mass, then the average velocity of the electron - also known as the *drift velocity* - is:

$$\langle \mathbf{v} \rangle = \frac{\langle \mathbf{p} \rangle}{m} = -\frac{\tau e}{m} \mathbf{E} . \quad (1.3)$$

The absolute value of the ratio of the drift velocity to the electric field, $\mu = e\tau/m$, is called *electron mobility*. From the solution (1.3), it follows that the electric current density is given by

$$\mathbf{j} = -en\langle \mathbf{v} \rangle = \frac{ne^2\tau}{m} \mathbf{E} , \quad (1.4)$$

where n is the electron density. Thus, we obtain the *Drude conductivity* as the ratio of the current density to the electric field,

$$\sigma = \frac{ne^2\tau}{m} . \quad (1.5)$$

Since the electron density, its charge, and mass can be measured independently, knowing the conductivity of the sample allows one to deduce the value of the momentum relaxation time, τ . This empirical value can be compared to the theoretical prediction for τ , which we turn to estimate now.

Let us assume that an electron can scatter from any ion in the lattice and that in each scattering event, the electron completely loses memory about its initial direction. In other words, the scattering time is also the momentum relaxation time.

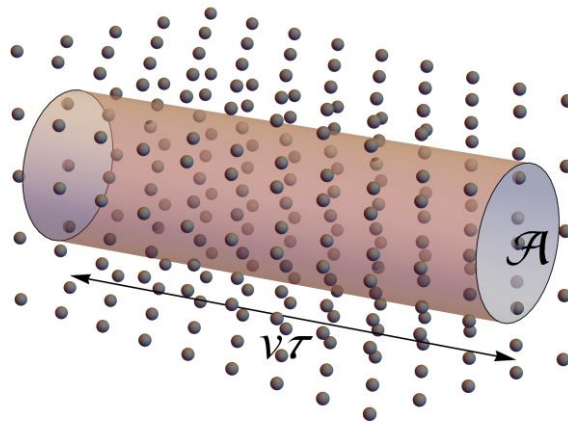


Fig. 1-1 Classical scattering of electrons in a periodic lattice

Now, consider a cylinder of length $v\tau$ and basis area $\mathcal{A}$ as demonstrated in Fig. 1-1. Let n_{ion} be the density of ions in the cylinder and σ_{scat} the scattering cross-section of each ion. An electron that enters the cylinder will be scattered with a probability of order one if the

sum of scattering cross-sections of all atoms is of the order of the area, $\mathcal{A}$, of the cylinder basis, i.e., when $\mathcal{A}v\tau n_{\text{ion}}\sigma_{\text{scat}} \sim \mathcal{A}$, thus

$$\tau \sim \frac{1}{vn_{\text{ion}}\sigma_{\text{scat}}} . \quad (1.6)$$

Let a be the spatial period of the lattice, known as the *lattice constant*, then $n_{\text{ion}} \sim 1/a^3$. Assuming also that the typical electron wavelength, λ_F , equals a , then $\sigma_{\text{scat}} \sim a^2$, and therefore $1/\tau \sim v/a \sim vp_F/\hbar \sim \varepsilon_F/\hbar$, where ε_F is the Fermi energy. To obtain this estimation, the relation between the momentum and the wavelength of the electron, $p_F \sim \hbar/\lambda_F \sim \hbar/a$, has been used. However, the scattering rate measured in experiments that were conducted during the 20th century, was $1/\tau \sim (10^{-4} \div 10^{-5})\varepsilon_F/\hbar$, i.e. smaller by 4 to 5 orders of magnitude compared to that predicted by Drude theory.

This discrepancy between the theoretical prediction and the empirical results has been resolved by Sommerfeld and Bloch (1928). They showed that an electron moving in a perfect periodic lattice does not scatter from ions at all! Qualitatively, this result can be understood by the Huygens-Fresnel principle: Imagine an electron entering the lattice with a wave function in the form of a plane wave. Each ion scatters this plane wave and behaves as a point source of circular waves. By the Huygens-Fresnel principle, we know that the constructive interference from all these circular waves reconstructs the plane wave (albeit with a different wave velocity – similar to the change of the speed of light moving in dielectric materials), meaning that the electron does not scatter.

The Huygens-Fresnel principle applies in free space where the point sources on a wavefront can be arbitrarily close to each other. When the distances between nearest neighbors' point sources are finite and fixed, constructive interference occurs only for particular values of wavelengths. These values dictate the electron momentum and form bands of allowed energies. The scattering rate of an electron moving in a perfect lattice, with energy situated in one of these bands, is infinite $\tau = \infty$.

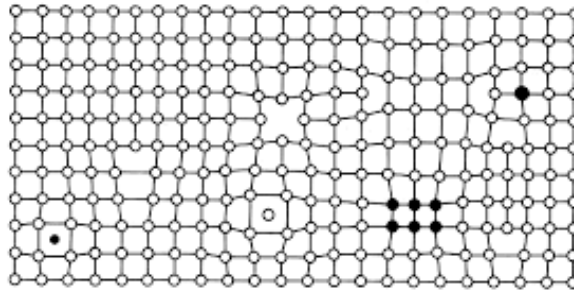


Figure 1-2 An illustration of lattice defects

At finite temperature, vibrations destroy the perfect periodicity of the lattice and lead to the scattering of the electron. Similarly, the existence of various defects in the lattice such

as interstitials, substitutional atoms, vacancies, dislocations, and disclinations, see illustration in Fig. 1-2, also destroy the periodic structure and generate scattering. This scattering produces a finite conductivity even at zero temperature.

1.2 Bloch's theorem

How does the periodicity of a potential manifest itself in the wave functions and energy spectrum of a particle moving in a lattice? This issue is one of the central questions that will be discussed in this course, and as we shall see, it has many interesting aspects. Here, as part of the introduction, we discuss the simple case of a particle moving in a one-dimensional periodic potential and prove Bloch's theorem stating that all wave functions of the particle can be written in the form:

$$\psi(x) = \exp(ikx)\phi_k(x), \quad (1.7)$$

where k is a real number called *Bloch's wavenumber*, while $\phi_k(x)$ is a periodic function with the periodicity of the potential.

To prove Bloch's theorem, consider the following Hamiltonian:

$$H(x) = -\frac{\hbar}{2m} \frac{\partial^2}{\partial x^2} + u(x). \quad (1.8)$$

Here x is the particle position, m is its mass, and $u(x)$ is a periodic potential with the periodicity of the lattice constant, a , i.e.

$$u(x+a) = u(x) \quad (1.9)$$

for any value of x . The (time-independent) Schrödinger equation of this system is

$$H(x)\psi(x) = \varepsilon\psi(x), \quad (1.10)$$

and symmetry to translation by the lattice constant implies that

$$H(x+a)\psi(x+a) = \varepsilon\psi(x+a). \quad (1.11)$$

But $H(x+a) = H(x)$, therefore $\psi(x+a)$ is also a solution of the Schrödinger equation (1.10) with the same energy. Moreover, translation operation does not reverse the direction of the momentum, and since the energy spectrum in a one-dimensional system is nondegenerate (we'll prove it later), $\psi(x+a)$ should be the same as $\psi(x)$ up to a phase factor, namely $\psi(x+a) = A\psi(x)$ with $|A| = 1$. In particular, one can set the constant A to be $\exp(ika)$, where k is a real free parameter, thus

$$\psi(x+a) = \exp(ika)\psi(x). \quad (1.12)$$

We turn to show that a function that satisfies this condition for any x should have the structure of a Bloch wave function (1.7). Multiplying Eq. (1.12) by $\exp[-ik(x+a)]$ we obtain

$$\exp[-ik(x+a)]\psi(x+a) = \exp(-ikx)\psi(x). \quad (1.13)$$

This equation completes our proof because it implies that $\phi_k(x) = \exp(-ikx)\psi(x)$ is a periodic function of x with period a .

Example: The Kronig-Penney model (1931)

The Kronig-Penney model describes a particle that moves in a one-dimensional periodic potential made from square well potentials of width b , depth $-u_0$, and distance a between neighboring wells. It is shown in the following figure:

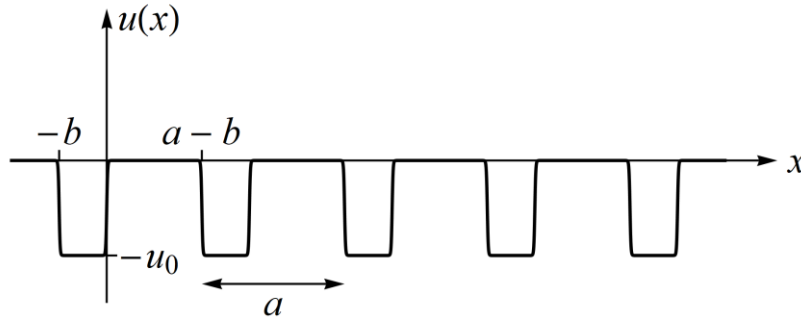


Figure 1-3 The potential of the Kronig-Penney model

This potential is piecewise constant. In each interval where the potential is constant, the wave function is a superposition of exponential functions (either real or imaginary) describing particle that moves in opposite directions. Thus, the wave function in the unit cell, $-b < x \leq a-b$, takes the form:

$$\psi(x) = \begin{cases} A\exp(i\alpha x) + A'\exp(-i\alpha x) & 0 < x \leq a-b \\ B\exp(i\beta x) + B'\exp(-i\beta x) & -b < x \leq 0 \end{cases}, \quad (1.14)$$

where A, A', B, B', α , and β are constants that should be calculated. Substituting $\psi(x)$ in Schrödinger equation (1.10) shows that the energy ε satisfies the following relations:

$$\varepsilon = \frac{\hbar^2 \alpha^2}{2m} \quad \text{and} \quad \varepsilon + u_0 = \frac{\hbar^2 \beta^2}{2m}. \quad (1.15)$$

Thus, the value of α determines that of β and vice versa. Notice that if $\varepsilon < 0$, the constant α is purely imaginary.

The continuity of the wave function (1.14) and its derivative at $x = 0$ yields the following conditions:

$$\begin{aligned} A + A' &= B + B', \\ i\alpha(A - A') &= i\beta(B - B'). \end{aligned} \quad (1.16)$$

To obtain two additional conditions, we use Bloch's theorem, implying that

$$\begin{aligned} \phi_k(x) &= \exp(-ikx)\psi(x) \\ &= \begin{cases} A \exp[i(\alpha - k)x] + A' \exp[-i(\alpha + k)x] & 0 < x \leq a - b \\ B \exp[i(\beta - k)x] + B' \exp[-i(\beta + k)x] & -b < x \leq 0 \end{cases} \end{aligned} \quad (1.17)$$

is a periodic function of x with period a . The periodicity of this function and its derivative,

$$\phi_k(-b) = \phi_k(a - b) \quad \text{and} \quad \phi'_k(-b) = \phi'_k(a - b), \quad (1.18)$$

leads to the following equations:

$$\begin{aligned} B \exp[-i(\beta - k)b] + B' \exp[i(\beta + k)b] \\ = A \exp[i(\alpha - k)(a - b)] + A' \exp[-i(\alpha + k)(a - b)], \end{aligned} \quad (1.19)$$

and

$$\begin{aligned} i(\beta - k)B \exp[-i(\beta - k)b] - i(\beta + k)B' \exp[i(\beta + k)b] \\ = i(\alpha - k)A \exp[i(\alpha - k)(a - b)] - i(\alpha + k)A' \exp[-i(\alpha + k)(a - b)]. \end{aligned} \quad (1.20)$$

The four equations that we obtained for the constants A, A', B , and B' , can be written as a matrix equation:

$$\begin{pmatrix} 1 & 1 & -1 & -1 \\ \alpha & -\alpha & -\beta & \beta \\ e^{i(\alpha - k)(a - b)} & e^{-i(\alpha + k)(a - b)} & -e^{-i(\beta - k)b} & -e^{i(\beta + k)b} \\ (\alpha - k)e^{i(\alpha - k)(a - b)} & -(\alpha + k)e^{-i(\alpha + k)(a - b)} & -(\beta - k)e^{-i(\beta - k)b} & (\beta + k)e^{i(\beta + k)b} \end{pmatrix} \begin{pmatrix} A \\ A' \\ B \\ B' \end{pmatrix} = 0. \quad (1.21)$$

A nontrivial solution of this equation exists only if the determinant of the above matrix vanishes. This condition yields (see Ex. 1) an equation that determines the possible values of Bloch's wavenumber, k , and their relations to the constant α :

$$\cos(ka) = \cos(\beta b) \cos[\alpha(a-b)] - \frac{\alpha^2 + \beta^2}{2\alpha\beta} \sin(\beta b) \sin[\alpha(a-b)]. \quad (1.22)$$

To simplify this equation, we focus our attention on the limit of very deep and narrow wells, $u_0 \rightarrow \infty$ and $b \rightarrow 0$, keeping the product $u_0 b$ constant. From the second relation of Eq. (1.15) it follows that in this limit, $\beta \rightarrow \infty$, but $\beta^2 b$ is a constant, hence $\beta b \rightarrow 0$. Thus Eq. (1.22) reduces to

$$\cos(ka) = \cos(\alpha a) - \mu \frac{\sin(\alpha a)}{\alpha a} \quad (1.23)$$

with

$$\mu = \frac{\beta^2 b a}{2} \quad (1.24)$$

as a free dimensionless parameter that characterizes the potential strength. In particular, if $\mu = 0$, then $\alpha = k$ and $\varepsilon = \hbar^2 k^2 / (2m)$ as expected for a free particle.

If $\mu \neq 0$ one should consider two cases: One is when α is real, and the second is when it is purely imaginary. The first case is associated with positive energy $\varepsilon = \hbar^2 \alpha^2 / (2m)$ and describes solutions in which the particle moves above the potential barriers that separate the wells. A graphical solution of equation (1.23) for real α is depicted in Fig. 1-4.

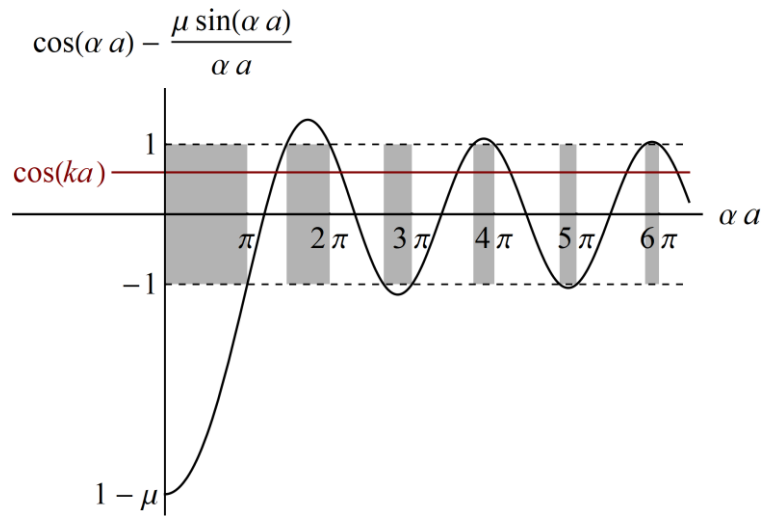


Figure 1-4 A graphical solution of Eq. (1.23) for real α

Here the right-hand side of Eq. (1.23) is drawn by solid black line as a function of αa , while the left-hand side of the equation, $\cos(ka)$, is drawn by the red line. As this constant resides within the interval between -1 and 1 (depending on k), there are values of αa for which there is no solution to the secular equation (1.23). The gray regions in the figure designate these values. Recalling that the energy is given by $\varepsilon = \hbar^2 \alpha^2 / (2m)$, this behavior of the solution implies that the energy spectrum of the system, $\varepsilon(k)$, consists of a set of separated “energy bands”.

Consider now the second case where α is purely imaginary. Substituting $\alpha = i\alpha'$ (where α' is real) in Eq. (1.23) we obtain

$$\cos(ka) = \cosh(\alpha' a) - \mu \frac{\sinh(\alpha' a)}{\alpha' a}. \quad (1.25)$$

The graphical solution of this equation is shown in Fig. 1-5.

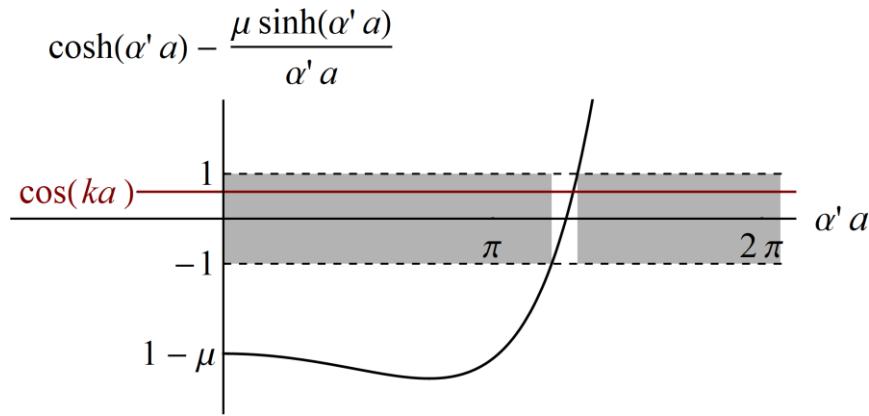


Figure 1-5 The graphical solution of Eq. (1.25)

Now there is only a single domain of $\alpha' a$ for which there is a solution of the equation. The particle energy associated with this solution is negative, $\varepsilon = -\hbar^2 \alpha'^2 / (2m)$, therefore, it forms the lowest energy band of the system. To obtain an approximate analytic expression for this band, consider the limit $\mu \gg 1$. In this case one expects $\alpha' a$ to be close to μ , therefore both functions, $\cosh(\alpha' a)$ and $\sinh(\alpha' a)$, can be approximated by $\exp(\alpha' a)/2$. With this approximation, linearization of Eq. (25) in the vicinity of $\alpha' a = \mu$ gives

$$\cos(ka) = \cosh(\alpha' a) - \mu \frac{\sinh(\alpha' a)}{\alpha' a} \simeq \frac{\exp(\mu)}{2\mu} (\alpha' a - \mu), \quad (1.26)$$

and the lowest energy band (which we denote by the subscript 0) is given by the approximate formula:

$$\varepsilon_0(k) = -\frac{\hbar^2 \alpha'^2}{2m} \simeq -\frac{\hbar^2 \mu^2}{2ma^2} [1 + 4\exp(-\mu)\cos(ka)]. \quad (1.27)$$

These negative energy states are associated with the tunneling of the particle from one well to the other. On the other hand, the solutions of Eq. (1.23) for real values of α yield a set of positive energy bands, $\varepsilon_n(k)$, associated with states where the particle travels above the energy barriers between the wells. A diagram of the energy levels obtained from a numerical solution of Eq. (1.23) is shown in Fig. 1-6 below.

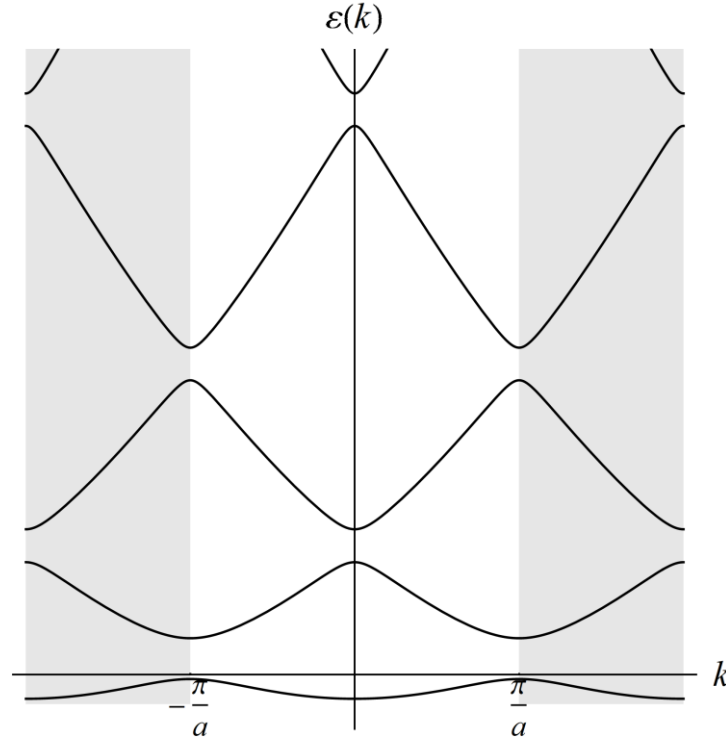


Figure 1-6 The energy levels of the Kronig-Penney model

The main conclusions drawn from the above discussion are:

- (a) The energy spectrum of a particle moving in a (one-dimensional) periodic potential comprises a set of allowed energy bands described by some functions, $\varepsilon_n(k)$. Thus, a state of the particle is characterized by two numbers: the band index, n , and Bloch's wave number, k .
- (b) The energy spectrum, as a function of Bloch's wavenumber, k , is periodic with a period $2\pi/a$. This is because the energy levels are functions of $\cos(ka)$ (see Eq. (1.22)). Therefore, to avoid double counting of the energy levels, the domain of allowed values of k should be restricted to one period. It is customary to choose this restricted domain to be $-\pi/a < k \leq \pi/a$. This domain is called the *first Brillouin zone* and is displayed by the white stripe in Fig. 1-6.

1.3 Effective mass, electrons, and holes.

Unlike free particles, the energy spectrum of a particle moving in a periodic lattice is no longer a parabolic function of the wavenumber. To appreciate the importance of this property, consider an electron moving in a one-dimensional periodic potential and subjected to a weak electric field E (a comprehensive discussion of this issue will be given in Chapter 9). Furthermore, let us adopt a semiclassical approach according to which Hamilton's equations can describe the dynamics of the electron using the Hamiltonian:

$$H = \varepsilon(p) + eEx, \quad (1.28)$$

where $\varepsilon(p)$ is the energy spectrum of the electron in the lattice which is some function of the momentum, p , and $-e$ is the electron charge. Hamilton's equations, in this case, take the form:

$$\begin{aligned} \frac{dx}{dt} = v &= \frac{\partial \varepsilon(p)}{\partial p} \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial x} = -eE \end{aligned} \quad (1.29)$$

For a free electron, $\varepsilon(p) = p^2/(2m)$, the first equation implies that the electron velocity is $v = p/m$. Taking the derivative of the latter equation and using the second equation of (1.29) we obtain Newton's second law for the acceleration of the electron:

$$\frac{dv}{dt} = \frac{-e}{m} E. \quad (1.30)$$

Now consider the case of an electron near the bottom of the lowest energy band (see, for example, Fig. 1-6). In this region, the spectrum can be approximated by parabolic spectrum, which may be presented in the form

$$\varepsilon(p) = \frac{p^2}{2m_{\text{eff}}}, \quad (1.31)$$

where we assume that the momentum satisfies the relation $p = \hbar k$, with k as Bloch's wavenumber (the justification of this assumption is deferred to Chapter 9). Repeating the calculation that leads to formula (1.30) yields the same expression but with m replaced by m_{eff} . Thus, the electron behaves as a free particle, albeit with a mass determined by the lattice properties that may be very different from that of a free electron. This mass is known as the *effective mass* of the electron.

We turn to discuss a situation where the electron is still in the lowest energy band but close to the edge of the Brillouin zone, $k \sim k_0 = \pi/a$. Now $\varepsilon(p)$ is near a maximum and can be approximated by inverted parabola:

$$\varepsilon(p) = \varepsilon_0 - \frac{(p - \hbar k_0)^2}{2m'_{\text{eff}}}. \quad (1.32)$$

Substituting this formula in Hamilton's equations (1.29) we obtain:

$$\begin{aligned} v &= \frac{\partial \varepsilon(p)}{\partial p} = -\frac{p - \hbar k_0}{m'_{\text{eff}}}, \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial x} = -eE, \end{aligned} \quad (1.33)$$

and once again, taking the time derivative of the first equation and substituting it in the second one, we obtain a formula for the acceleration like Eq. (1.30) but with a different sign and different mass:

$$\frac{dv}{dt} = \frac{+e}{m'_{\text{eff}}} E. \quad (1.34)$$

A possible interpretation of the above result is that the electron has a negative mass. However, in reality, we do not measure the mass directly but only as a response to force; hence an alternative interpretation of the above result is that an electron near the upper edge of the band behaves like a particle with a positive charge. Such a particle is called a "hole".

To clarify the reason for this nomenclature, we consider two cases. In the first, we look at a single electron (near the bottom of the band) that moves on a lattice where all sites are unoccupied, as demonstrated below:

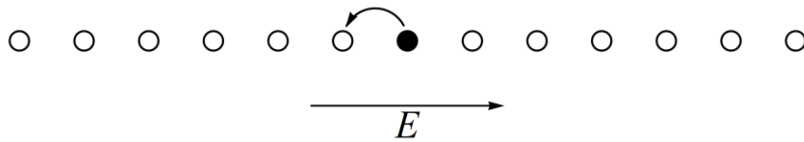


Figure 1-7 An electron in a lattice near the bottom of the lowest band

Here the electric field, E , induces transitions of the electron from one lattice site to another unoccupied site.

Consider, now, the second case where all lattice sites are occupied except for one as illustrated in Fig. 1-8 on the next page. Ignoring the spin degree of freedom (assuming, e.g., that all spins of the electrons point in the same direction), all electrons are frozen at their position due to Pauli's exclusion principle, except for the electrons near the vacant lattice point. The application of an electric field induces the transition of the electron on the right side of the vacant lattice point to move to the left. However, this process is

equivalent to a right jump of the vacant lattice point – i.e. the hole – as illustrated in Fig. 1-8.

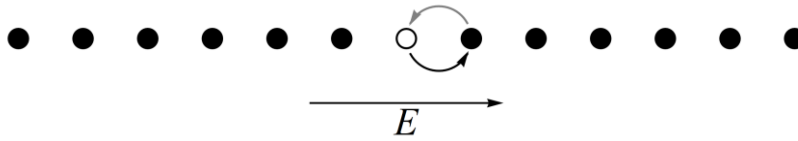


Figure 1-8 Electrons in a lattice near at the top of the lowest band

While the above picture motivates the notion of a “hole” as a particle, it may be misleading. The behavior manifested in Eq. (1.34) refers to a single particle near the edge of the Brillouin zone without reference to the occupation of states by other electrons. The electron’s acceleration changes sign because of constructive interference of the reflected waves from the atoms, which becomes stronger as the electron’s energy increases. This is, essentially, Bragg’s reflection phenomenon: Near the edge of the Brillouin zone, the electron’s wavelength is close to being twice the lattice constant; hence the waves reflected from neighboring lattice cells interfere constructively.

1.4 Fermi surfaces, metals, and insulators

At sufficiently low temperatures, electronic systems become degenerate. The electrons (which are Fermions) essentially occupy all states up to energy, ε_F , known as the *Fermi energy*, excluding a narrow band near the Fermi level - whose width is determined by the temperature. In this band, the occupation is partial. The Fermi-Dirac distribution of the electrons accounts for this property:

$$f(\varepsilon) = \frac{1}{1 + \exp\left(\frac{\varepsilon - \varepsilon_F}{k_B T}\right)}. \quad (1.35)$$

Here T is the temperature, and k_B is Boltzmann’s constant. Only those electrons with energies located within the narrow stripe near the Fermi level react to weak external perturbations because electrons with energies deep below the Fermi level are frozen due to Pauli’s exclusion principle. Therefore, properties of the system, such as the heat capacitance or the electric conductivity, are determined only by electrons near the Fermi level. Hence, the characterization of the physical properties of crystals requires information about the position and structure of the Fermi level in $\mathbf{k}$ space.

The *Fermi surface* is a surface in the wavenumber space, $\mathbf{k}$, that is obtained from the solution of the equation:

$$\varepsilon(\mathbf{k}) = \varepsilon_F. \quad (1.36)$$

In one-dimensional systems, the Fermi surface is a set of points. In two dimensions, Fermi surfaces are curves, while in three dimensions, they are surfaces in $\mathbf{k}$ space.

In general, a crystal becomes a metal or an insulator depending on the location of the Fermi level. When the latter is within a band, the electrons fill all the states below the Fermi level, while those above the Fermi level remain empty. An electric field applied to the system induces transitions of electrons from the occupied states to vacant energy states with an energy difference that can be as small as we wish (assuming the system is large enough). These transitions of the electrons can produce electric current; hence, the system is metallic. Such metals, known as band metals, are obtained in two primary manners as illustrated in the following figure (for one-dimensional systems):

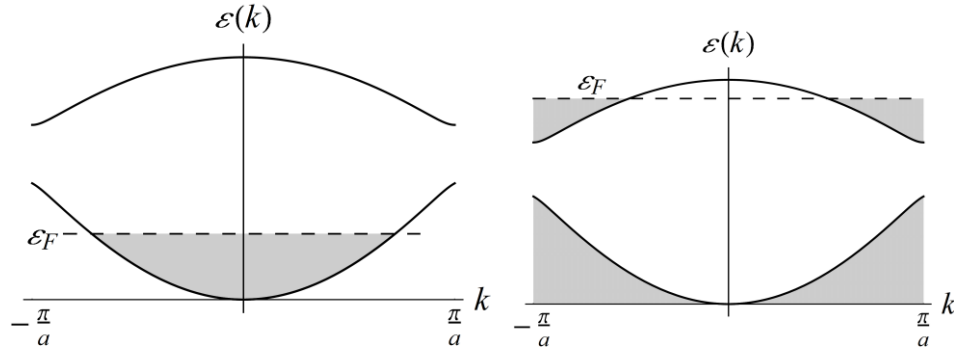


Figure 1-9 An electron-like (left) and a hole-like (right) Fermi surfaces

The real space illustrations of one-dimensional systems with electron-like and hole-like Fermi surfaces are shown in Figs. 1-7 and 1-8, respectively.

In contrast, when the Fermi level is located between two bands, the electrons fill all the states of the lower band, and the only possibility to change the occupation is by moving an electron to a higher band. However, usually, this process requires a large amount of energy that a weak electric field cannot provide. In this case, the system is an insulator because the electrons do not have close energy levels into which they can move in order to produce an electric current. This type of insulator is called a *band insulator*. It is illustrated in Fig. 1-10. (There are also different kinds of insulators called *Anderson insulators*. The mechanism that suppresses the electric current in these insulators is interference. We shall not discuss Anderson insulators in this course.)

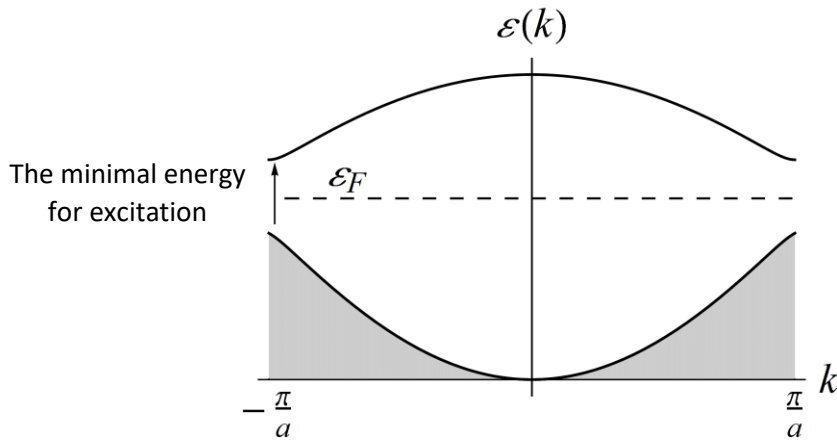


Figure 1-10 A band insulator

The question that remains is what determines whether the Fermi level is within a band or between two bands. To answer this question, we should first show how to count the number of states in a band. For a one dimensional system, assuming its length is L (with, say, periodic boundary conditions), each state occupies a range $\Delta k = 2\pi/L$ in k space; therefore, the number of states in the band is:

$$N_{\text{band}} = 2 \int_{-\frac{\pi}{a}}^{\frac{\pi}{a}} \frac{Ldk}{2\pi} = 2 \frac{L}{a} = 2N_{\text{uc}}, \quad (1.37)$$

where the factor 2 is due to the spin degree of freedom, a is the lattice constant, and N_{uc} is the number of unit cells in a system of length L .

These states are occupied by electrons that come from the atoms in each unit cell. Now, there are two possibilities: The first is when the number of electrons coming from the atoms, in a single unit cell, is an even number, and the second possibility is that this number is odd. In the first case, the electrons fill all the states in the bands and we obtain an insulator. In the second case, the last band will be only half-filled because it contains N_{uc} electrons, whereas the band contains $2N_{\text{uc}}$ states. In this case, the lattice is a metal.

The above argument is independent of the dimensionality of the system. However, there are, of course, exceptions. But before presenting examples of these exceptions, let us discuss the typical structures of Fermi surfaces in systems of higher dimensions. For simplicity, we consider two-dimensional systems for which the Fermi surfaces are curves in the first Brillouin zone of $\mathbf{k}$ space.

An *electron-like Fermi surface* is illustrated in Fig. 1-11 on the next page. Here the shadowed region represents a region of occupied states in the first Brillouin zone in $\mathbf{k}$ space.

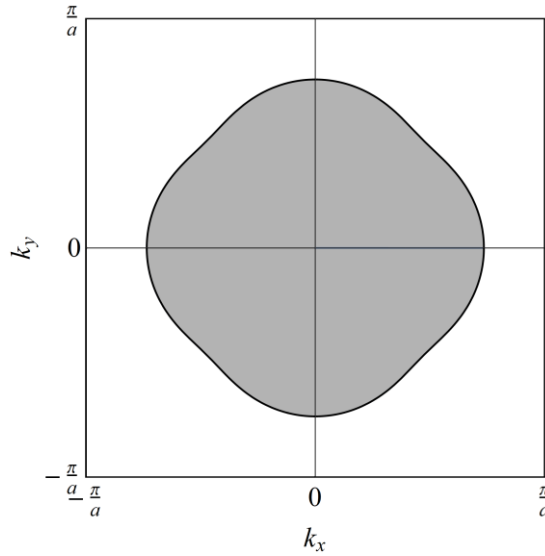


Figure 1-11 An electron-like Fermi surface in a two-dimensional system

A *hole-like Fermi surface* is illustrated in Fig. 1-12. Choosing to describe the systems using holes instead of electrons, the Fermi surface of the holes is depicted in Fig. 1-13. Notice that, in this case, the Fermi surface of the holes is approximately circular because there are periodic boundary conditions on the Brillouin zone. (Alternatively, one may shift the Brillouin zone to a position such that the hole Fermi surface is at the center. This is justified because there is an arbitrariness in choosing the location of the first Brillouin zone.).

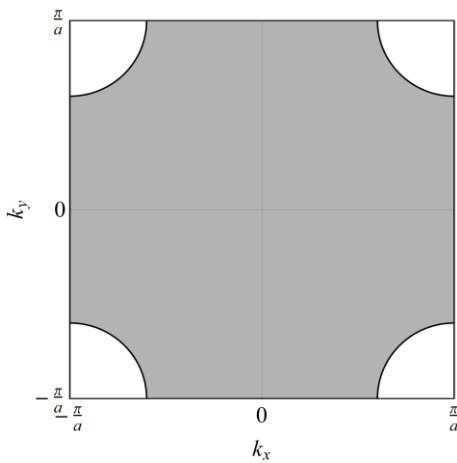


Figure 1-12 A hole-like Fermi surface

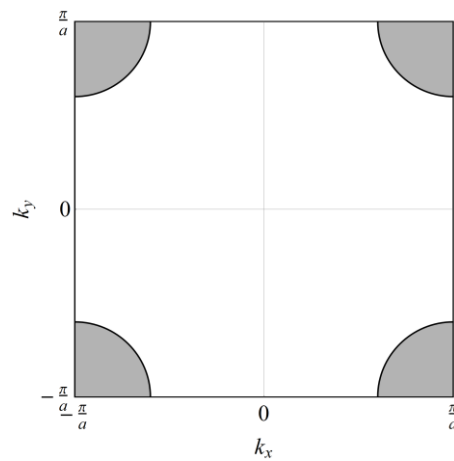


Figure 1-13 The Fermi surface of holes for the system shown in Fig. 1-12

Another classification of Fermi surfaces refers to *closed* and *open surfaces*. All Fermi surfaces described above are closed Fermi surfaces because, by proper choice of the Brillouin zone, closed curves describe them. An example of an open Fermi surface is shown in the left panel of Fig. 1-14. Here the Fermi surface does not form a closed curve for any

choice of the Brillouin zone. In such cases, however, the distinction between the type of charge carriers is meaningless, as demonstrated in the right panel of Fig. 1-14.

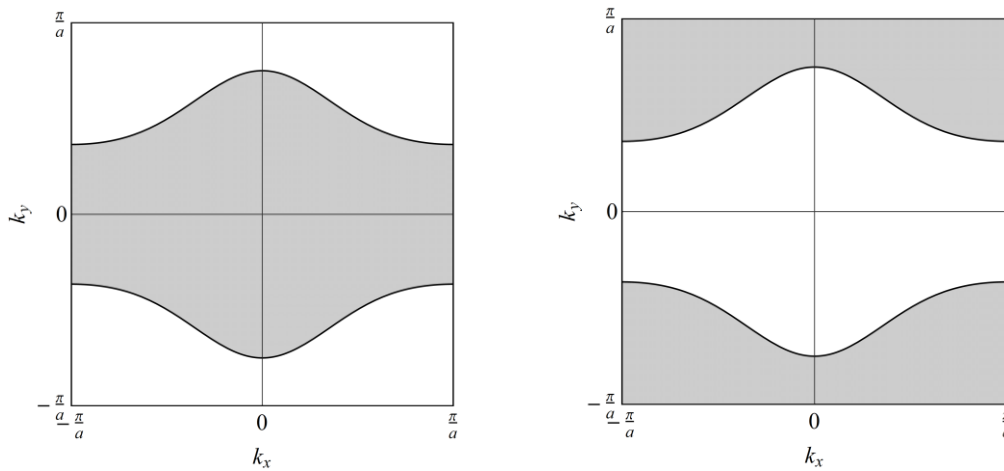


Figure 1-14 An example of an open Fermi surface in a two-dimensional system. Lefts panel for electrons and right panel for holes.

More complicated Fermi surfaces may simultaneously contain both pockets of electrons and pockets of holes, as illustrated in the figure below.

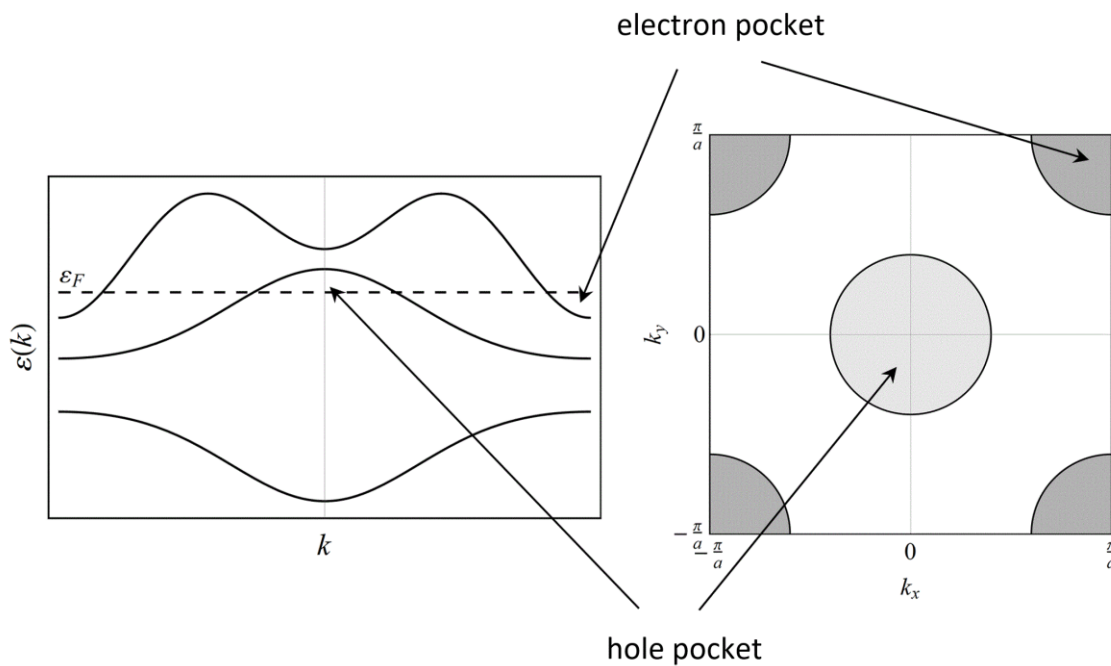


Figure 1-14 An illustration of the Fermi surface in semi-metals

The left panel of this figure shows a diagonal cross-section of the spectrum in the first Brillouin zone depicted on the right panel. Systems of this type are called *semi-metals*. They contain an even number of electrons in each unit cell and, as such, should be insulators. However, the spatial structure of the energy bands in k space, where the

minimum of one band is below the maximum of the lower band, creates pockets of charge carriers, making the system conductive.

Another possibility of having a conductive system, although each unit cell contributes an even number of electrons, is when having band-touching points. In this case, the valance band is wholly filled by electrons, but the energy gap to the conductive band is zero. This type of degeneracy points results from symmetries of the lattice, as we shall discuss later in this course.

1.5 Exercises

1. Prove Eqs. (1.22) and (1.23).
2. Consider the Kronig-Penney model with potential made from a periodic set of δ -barriers:

$$u(x) = u_0 \sum_n \delta(x - na), \quad (1.38)$$

where a is the lattice constant, and $u_0 > 0$. To solve this model, assume that between each neighboring pair of δ -barriers the wave function is a linear combination of $\cos(\alpha x)$ and $\sin(\alpha x)$, and construct the secular equation for the coefficients of these functions. Draw a graphical solution of the equation that you obtained.

3. Show that the energy gaps between neighboring bands in the Kronig-Penney model approach a constant value in the limit $\varepsilon \rightarrow \infty$.
4. Show that the ground state energy of the Kronig-Penney model, obtained in Eq. (1.27) in the limit $\mu \rightarrow \infty$, is that of a particle in a delta potential:

$$H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} - \frac{\hbar^2 \mu}{ma} \delta(x). \quad (1.39)$$

2 Lattices and spatial symmetries

Spatial periodic structures are associated with symmetry operations such as translations in certain directions, reflections, and rotations in specific angles. The set of symmetry operations that characterizes a crystal is of paramount importance. As we shall see in the coming chapters, it dictates the band structure of the energy levels of the electrons as well as the crystal vibrational properties. In this chapter, we present the basic concepts and terminology of this issue.

2.1 Translation vectors

Let $u(\mathbf{r})$ be the periodic potential that acts on an electron moving in a lattice (without impurities or other defects). In a one-dimensional system, the periodicity of the potential implies that

$$u(x+a) = u(x) \quad \text{for any } x, \quad (2.1)$$

where a is the lattice constant as illustrated in Fig. 2-1. Clearly, the potential is also periodic in any multiple of the lattice constant, $a_x^{(j)} = ja$, where j is an integer.



Figure 2-1 A one-dimensional lattice with lattice constant a

In two spatial dimensions, a periodic potential satisfies the condition:

$$u(\mathbf{r} + \mathbf{a}) = u(\mathbf{r}) \quad \text{with the translation vector } \mathbf{a} = j_1 \mathbf{a}_1 + j_2 \mathbf{a}_2, \quad (2.2)$$

where j_1 and j_2 are integers, while $\mathbf{a}_1$ and $\mathbf{a}_2$ are two linearly independent vectors that represent the lattice's shortest translation vectors, as illustrated in Fig. 2-2. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are called the *primitive basis of the translation vectors* or the *primitive lattice vectors*.

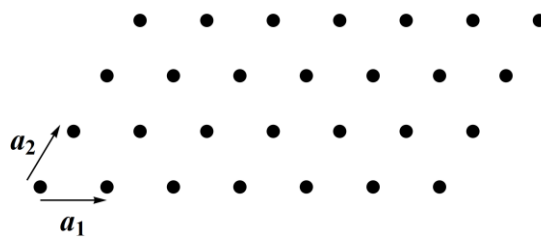


Figure 2-2 A two-dimensional lattice with primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$

The natural generalization of the translation vectors to three-dimensional lattices is

$$\mathbf{a} = j_1 \mathbf{a}_1 + j_2 \mathbf{a}_2 + j_3 \mathbf{a}_3, \quad (2.3)$$

where j_i ($i=1,2,3$) are integers, while $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are three independent (three dimensional) vectors. Finally, we comment that each dot in the above figures represents a whole unit cell of the potential, which may have a complicated internal structure.

2.2 The translation group

The spatial periodicity of the potential $u(\mathbf{r})$ is a manifestation of symmetry of the system: Translation of the (infinite) system by any linear combination of the primitive basis vectors, with integer coefficients, leaves the system unchanged. The mathematical framework for treating symmetries is *group theory*, and here we introduce it by using the simple example of the *translation group* in crystals.

Let us define T_a to be the operator that translates any function, $f(\mathbf{r})$, by the vector $\mathbf{a}$, i.e.:

$$T_a f(\mathbf{r}) = f(\mathbf{r} + \mathbf{a}). \quad (2.4)$$

Operating on this formula with an additional translation operator, $T_{a'}$, yields

$$T_{a'} T_a f(\mathbf{r}) = T_{a'} f(\mathbf{r} + \mathbf{a}) = f(\mathbf{r} + \mathbf{a} + \mathbf{a'}). \quad (2.5)$$

Thus, the translation operators satisfy the following property:

$$(a) \quad T_{a'} T_a = T_{a+a'}. \quad (2.6)$$

I.e., the “multiplication” of any two translation operators is also a translation operator. This property is called “closure”. In addition, choosing $\mathbf{a}' = -\mathbf{a}$ implies that

$$(b) \quad T_{-\mathbf{a}} T_a = E, \quad (2.7)$$

where E is the identity operator whose action on any function leaves it intact, $E f(\mathbf{r}) = f(\mathbf{r})$. The identity operator also satisfies the property

$$(c) \quad T_a E = E T_a = T_a. \quad (2.8)$$

Finally, it is easy to see that the translation operators also satisfy associative property:

$$(d) \quad (T_{a''} T_{a'}) T_a = T_{a''} (T_{a'} T_a). \quad (2.9)$$

A set of operations (or operators) that meet properties (a-d), namely, closure, the existence of the identity operation, the presence of an inverse operator for any operator, and associativity of multiplication of operators, is called *group*. In the example discussed above, the set of translation operators constitute the group of translations.

A group is called *abelian* if the order of operations can be changed, namely

$$T_{a'}T_a = T_aT_{a'}, \quad (2.10)$$

for any pair of operators in the group. The translation group is abelian.

2.3 Bravais Lattices

Bravais Lattice is a lattice of points obtained by acting on a single point by the whole set of translation operators. This lattice is the simplest one that describes the system because each point represents a unit cell whose internal (possibly complicated) structure is ignored. In this lattice, the “view” seen from any lattice point is precisely the same.

In one dimension, the Bravais lattice is the set of an infinite number of points on a straight line, such that the distance between any neighboring points is fixed, as illustrated in Fig. 2-1.

In two dimensions, a general Bravais lattice looks as in Fig.2-2. However, it is instructive to classify these lattices according to their symmetry level (with respect to symmetry operations such as rotations and reflections). The primitive basis vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ determine this symmetry level. The least symmetric two-dimensional lattice is the *oblique lattice*, for which $|\mathbf{a}_1| \neq |\mathbf{a}_2|$ and the angle between these two vectors is different from 90° as shown in Fig. 2-3. The only symmetry operation of this lattice (apart from translations) is a rotation by 180° around any lattice point.

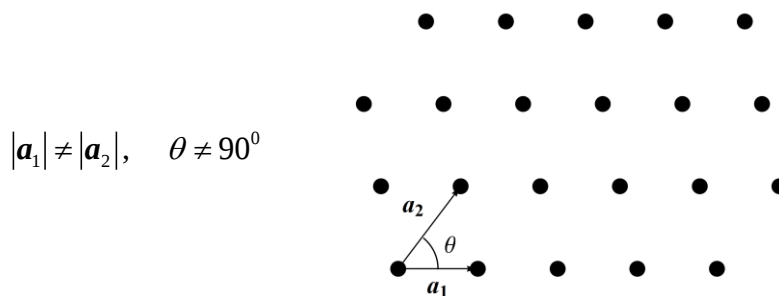


Figure 2-3 Oblique lattice

A rectangular Bravais lattice can be realized in two manners, as demonstrated in Fig. 2-4. The lattice shown on the left panel is the *primitive rectangular lattice*, while that on the right panel is the *centered rectangular lattice*. Both versions of the rectangular lattice are symmetric for rotation in 180° , and reflections through horizontal and vertical axes.

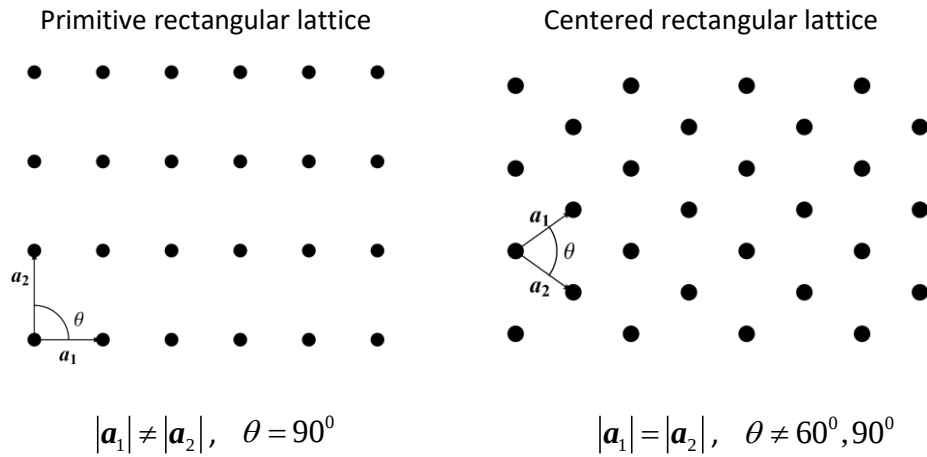


Figure 2-4 Rectangular Bravais lattices

The next Bravais lattice is the square lattice with even higher symmetry (i.e. with a larger set of symmetry operations other than translations). It is obtained when the primitive basis vectors have the same length, and the angle between them is 90° as shown in Fig. 2-5. This lattice possesses symmetry to rotations in 90° , 180° and 270° as well as reflections through 4 axes: horizontal, vertical, and two diagonals.

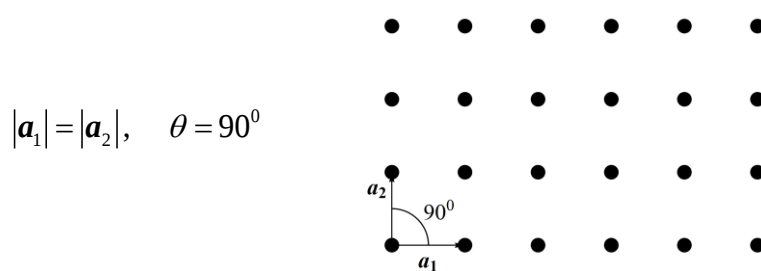


Figure 2-5 Square lattice

Finally, the Bravais lattice with the highest symmetry in two dimensions is the hexagonal (or triangular) lattice, obtained when the primitive basis vectors are of the same length, $|a_1| = |a_2|$, while the angle between them is 60° , as shown in Fig. 2-6. The symmetry operations of this lattice include five rotations in multiples of 60° , and six reflections through 6 axes that can be obtained by 0° , 60° and 120° rotations of the horizontal and the vertical axes (see also Fig. 2-11 below).

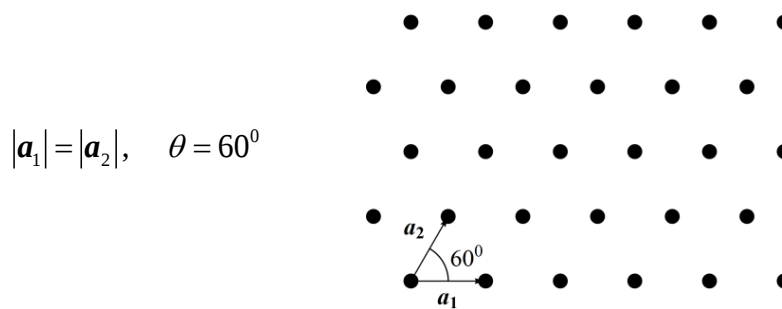


Figure 2-6 Hexagonal lattice

2.4 Point groups

As we have seen, lattices are characterized by a set of symmetry operations, such as rotations and reflections, in addition to the translation operations. These symmetry elements form a group which is called “*point group*”. This name reflects the property that (at least) one point in space is unaffected by all symmetry operations of the group. Point groups play a central role in the characterization of lattices in two and three dimensions. Here, as a preliminary exposition of the subject, we discuss point groups in the context of two-dimensional lattices. A more rigorous discussion is given in chapter 4.

Consider the shape:

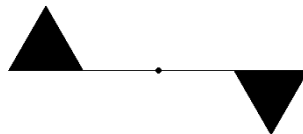


Figure 2-7 A shape symmetric for 180° rotation

Rotation by 180° around an axis that perpendicular pierces the page at the central point leaves this shape unchanged. Therefore this rotation is a symmetry of the system. We denote such a rotation by c_2 . More generally, c_n represents a rotation by $360^\circ/n$, where n is an integer.

The symmetry operation, c_2 , together with the identity operation, E , form a group. Operating twice with c_2 yields a 360° rotation, which is the identity operator, $c_2^2 = E$. Hence, these two operations satisfy the four conditions that define a group: closure, the presence of the identity operation, the existence of an inverse operator for any operator (which here is the operator itself), and associativity.

Take a look now at the following shape:

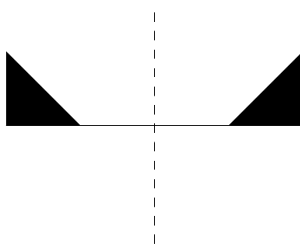


Figure 2-8 A shape symmetric to reflection

It does not have any rotation symmetry (in angle different from 360°), but it is symmetric to reflection through the dashed line. The symbol for reflection operation is σ . It is customary to add a subscript in order to define the axis of reflection - e.g., σ_y is a reflection through the y -axis). As in the previous example, σ and E form a group because operating twice with σ yields the identity operator, $\sigma^2 = E$.

Next, consider the shape:

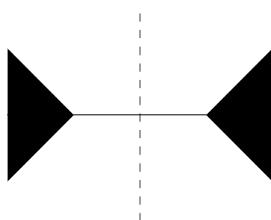


Figure 2-9 A shape symmetric to two reflections and rotation in 180°

It is symmetric with respect to the following operations: c_2 - rotation in 180° ; σ_x - reflection through the horizontal axis; and σ_y - reflection through the vertical axis. If we locate the origin of the coordinate system at the center of the shape, then σ_x transforms the point (x, y) to $(x, -y)$; σ_y takes (x, y) to $(-x, y)$; and c_2 transforms (x, y) to $(-x, -y)$. From these relations it follows that $c_2\sigma_x = \sigma_x c_2 = \sigma_y$, and that the set of operation, E , c_2 , σ_x and σ_y forms a group. This group is denoted by C_{2v} , and it is an abelian group. The multiplication table of its elements is presented on the next page.

The multiplication table of C_{2v}

C_{2v}	E	c_2	σ_x	σ_y
E	E	c_2	σ_x	σ_y
c_2	c_2	E	σ_y	σ_x
σ_x	σ_x	σ_y	E	c_2
σ_y	σ_y	σ_x	c_2	E

Comment: The group C_{2v} is abelian; therefore, the order of operations is not important. However, in what follows, we shall also consider non-abelian groups, where the order of operations is important. The convention we use in these cases is that the element ab in a specific cell of the table is obtained from the product (from the left) of the element a , that appears on the leftmost cell of the same row, by the element b that appears on the top cell of the same column.

Consider now the symmetry group of a square. The square is symmetric to reflections through 4 axes: σ_x , σ_y , σ_{xy} , and $\sigma_{x\bar{y}}$, as illustrated in Fig. 2-10. It is also symmetric to rotations by 90° , 180° and 270° denoted by c_4 , c_2 , and c_4^3 , respectively.

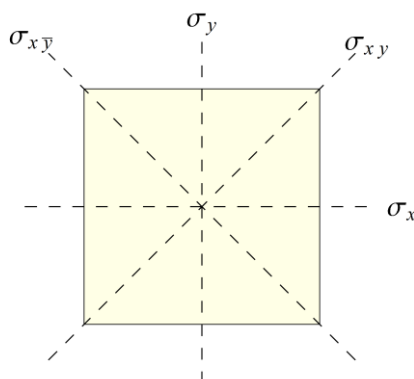


Figure 2-10 the reflection axes of a square

When placing the origin of the coordinate system at the center of the square, it is easy to verify that the symmetry operations acting on some arbitrary point (x, y) satisfy:

$$c_4 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}; \quad c_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ -y \end{pmatrix}; \quad c_4^3 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ -x \end{pmatrix}; \quad (2.11)$$

$$\sigma_x \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}; \quad \sigma_y \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}; \quad \sigma_{xy} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}; \quad \sigma_{x\bar{y}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ -x \end{pmatrix} \quad (2.12)$$

Using these equations, one can construct the multiplication table of the group (see table below) and verify that it is not abelian. This symmetry group is C_{4v} .

The multiplication table of C_{4v}

C_{4v}	E	c_4	c_2	c_4^3	σ_x	σ_y	σ_{xy}	$\sigma_{x\bar{y}}$
E	E	c_4	c_2	c_4^3	σ_x	σ_y	σ_{xy}	$\sigma_{x\bar{y}}$
c_4	c_4	c_2	c_4^3	E	σ_{xy}	$\sigma_{x\bar{y}}$	σ_y	σ_x
c_2	c_2	c_4^3	E	c_4	σ_y	σ_x	$\sigma_{x\bar{y}}$	σ_{xy}
c_4^3	c_4^3	E	c_4	c_2	$\sigma_{x\bar{y}}$	σ_{xy}	σ_x	σ_y
σ_x	σ_x	$\sigma_{x\bar{y}}$	σ_y	σ_{xy}	E	c_2	c_4^3	c_4
σ_y	σ_y	σ_{xy}	σ_x	$\sigma_{x\bar{y}}$	c_2	E	c_4	c_4^3
σ_{xy}	σ_{xy}	σ_x	$\sigma_{x\bar{y}}$	σ_y	c_4	c_4^3	E	c_2
$\sigma_{x\bar{y}}$	$\sigma_{x\bar{y}}$	σ_y	σ_{xy}	σ_x	c_4^3	c_4	c_2	E

We conclude this section by presenting the symmetry group of a regular hexagon, C_{6v} . This point group contains 12 symmetry operations: The identity operation E , five rotations c_6 , c_3 , c_2 , c_3^2 , and c_6^5 ; and six reflections, σ_1 , σ_2 , σ_3 , $\sigma_{\bar{1}}$, $\sigma_{\bar{2}}$, and $\sigma_{\bar{3}}$, as demonstrated in Fig. 2-11.

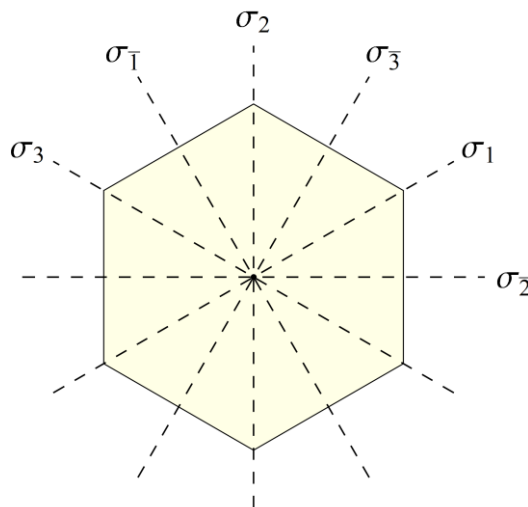


Figure 2-11 The reflection axes of a regular hexagon

2.5 Space groups

Combining the operations of the translation group with all other symmetry operations of the lattice, such as rotations and reflections (as well other symmetry operations that will be described later), gives the complete symmetry group of the lattice, called *space group*. For instance, the space group of one-dimensional Bravais lattice is the collection of all operations obtained from the multiplication of the elements in the point group C_2 (identity and rotation by 180°) by the elements of the translation group T :

$$C_2 \otimes T = \{T_a C_2, T_a\} . \quad (2.13)$$

To see that this space group is not abelian, notice that in one dimension, $c_2 x = -x$ where x denotes a lattice point. Therefore

$$, c_2 T_a c_2 x = c_2 T_a (-x) = c_2 (a - x) = x - a = T_{-a} x . \quad (2.14)$$

From here it follows that $c_2 T_a c_2 = T_{-a}$ or $c_2 T_a = T_{-a} c_2$ which is different from $T_a c_2$, in general. Thus, these symmetry operations are not commutative.

Similar considerations apply to Bravais lattices in two dimensions. In the table below, we summarize the space group of each lattice.

Space groups of Bravais lattices in two dimensions

oblique lattice	$C_2 \otimes T$
(centered) rectangular lattice	$C_{2v} \otimes T$
square lattice	$C_{4v} \otimes T$
hexagonal lattice	$C_{6v} \otimes T$

Space groups obtained from the multiplication of translation group by a point group are called *symmorphic groups*. There are, however, lattices whose space groups are not symmorphic. In the next section, we present an example of such a lattice.

2.6 Decorated lattices

Each point in a Bravais lattice represents a unit cell of the lattice. The latter may have a structure that affects the symmetry of the lattice. Consider, for instance, a one-dimensional lattice made of two types of atoms that form two sublattices, A and B as demonstrated in the following figure:

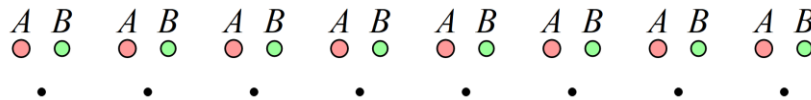


Figure 2-12 A lattice made of two types of atoms that form two sublattices.

In this figure, the row of black points below the colored lattice points represents the corresponding Bravais lattice. As explained above, the space group of Bravais lattice in one dimension is $C_2 \otimes T$. However, it is evident that the lattice shown in the figure does not have the symmetry of C_2 ; therefore, its symmetry is reduced.

A similar way of reducing the symmetry of a lattice is by decoration. Here each lattice cell has a structure that may reduce the symmetry, as shown in the following figure:

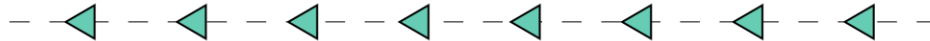


Figure 2-13 Decorated Bravais lattice in one dimension

As in the previous example, the above lattice decoration breaks the C_2 symmetry of the corresponding Bravais lattice.

A more interesting example is the one-dimensional lattice shown in the following figure:

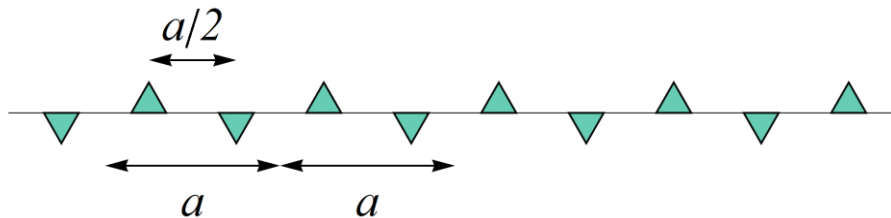


Figure 2-14 Decorated Bravais lattice with two sublattices

Here each unit cell contains two triangles pointing in opposite directions. The distance between two adjacent triangles is half the lattice constant. The C_2 symmetry of the lattice is preserved, but there is an additional symmetry operation called *glide reflection*. It is obtained by reflection through the horizontal axis followed by translation along the same

axis by half the lattice constant. (In three dimensions, the analogous operation is reflection through some plane followed by translation along the same plane.) This glide reflection operation is denoted by $\left\{ \sigma_x \mid \frac{a}{2} \right\}$, and it is easy to see that $\left\{ \sigma_x \mid \frac{a}{2} \right\}^2 = T_a$. The symmetry group of this lattice is obtained from all translations, glide translations, and rotations by 180° . This group is *non-symmorphic* because it is not a product of a point group by translation group.

In two dimensions, it is possible to construct 17 different lattices (which are not necessarily Bravais lattices) such that each one of them has a different space group. An example of such a lattice (that we will discuss extensively in this course) is graphene. It is a two-dimensional lattice of carbon atoms in a honeycomb structure, as shown in Fig 2-15. Here, each atom is bonded to three other atoms such that the angle between any two bonds is 120° . One can view graphene as built from two hexagonal sublattices, *A* and *B*, such that the nearest neighbors of an atom in sublattice *A* are only atoms in sublattice *B*, and vice versa.

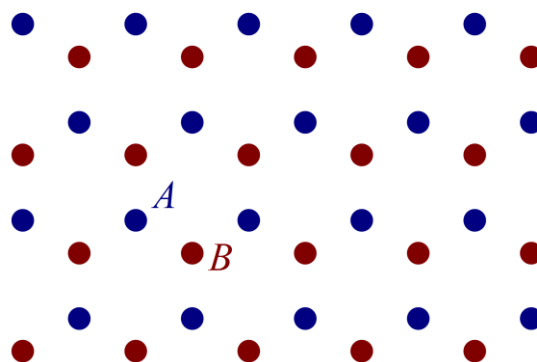


Figure 2-15 Graphene

Thus, a unit cell of graphene contains two atoms (one from each sublattice), and the corresponding Bravais lattice is hexagonal, as illustrated by the green hexagons in Fig. 2-16. Notice, however, that there is an arbitrariness in the way we choose the unit cell. One may also choose it to contain $1/3$ of an atom from 6 different atoms, as shown by the red hexagon in Fig. 2-16. The point group of graphene is C_{6v} .

Another lattice having a similar structure is the two-dimensional lattice of boron nitride, BN. From a symmetry viewpoint, this lattice has the same honeycomb structure but with sublattices made from two different atoms (namely, the red and the blue disks in Fig. 2-15 represent two types of atoms). The point group symmetry of this lattice is C_{3v} . It contains rotations by $\pm 120^\circ$, and reflections through three axes defined by the lines connecting an atom with its nearest neighbors. It will be discussed extensively in Chapter 5.

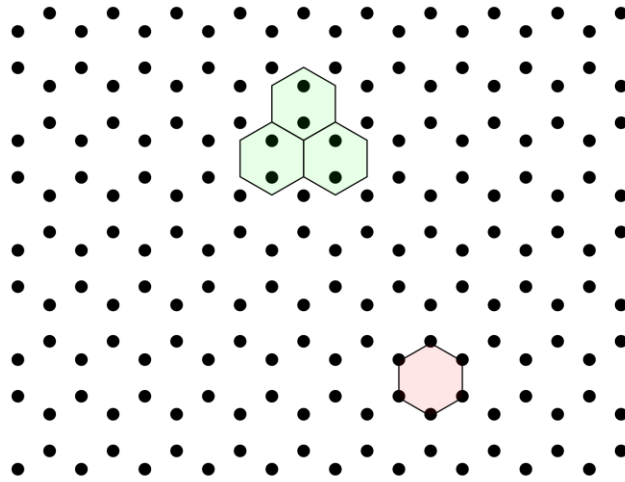


Figure 2-16 Unit cells in graphene

2.7 The Wigner-Seitz cell

The choice of a primitive unit cell in Bravais lattice is not unique. Choosing it arbitrarily may conceal the symmetry of the lattice. To avoid this problem, one constructs the unit cell according to the accepted procedure proposed by Wigner and Seitz. The Wigner-Seitz cell is unique and contains a single lattice point. The cell is constructed in the following manner: First, we choose a lattice point and connect it by segments to all its neighboring lattice points. Next, the segments are bisected by perpendicular lines. The smallest convex region enclosed by these lines is the Wigner-Seitz cell. This construction is demonstrated in Fig. 2-17.

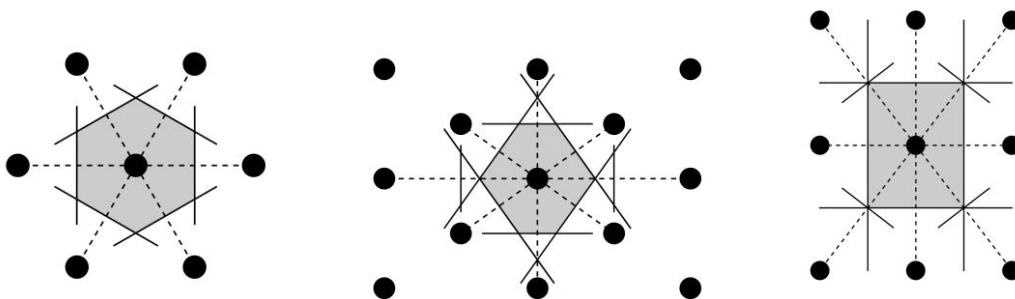


Figure 2-17 The construction Wigner-Seitz cells for various two-dimensional lattices.

2.8 Three-dimensional Bravais lattices

There are 14 different Bravais lattices in three dimensions. Each one of them is uniquely defined by the three primitive basis vectors, $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. The relation between their lengths and the angles between them determine the type of lattice and its symmetry. To construct the Bravais lattices in three dimensions, we shall use the two-dimensional

lattices described above as a basis on which we add lattice points in the third dimension. Usually (although not always), we choose the primitive basis vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ to be in the xy plane, and then add a third vector, $\mathbf{a}_3$, that has a component in the z direction.

1. The lowest symmetry lattice is the *triclinic* lattice. It is constructed by taking an oblique lattice in the xy plane (with $|\mathbf{a}_1| \neq |\mathbf{a}_2|$ and the angle between them $\theta \neq 90^\circ$) and adding a non-perpendicular basis vector, $\mathbf{a}_3$, whose length is different from the two other vectors $|\mathbf{a}_1| \neq |\mathbf{a}_2| \neq |\mathbf{a}_3|$, as illustrated in Fig. 2-18. The only symmetry operation the triclinic lattice possesses (apart from translations) is inversion, c_i . When setting the origin of a coordinate system at one of the lattice points, the action of the inversion operator, on a general vector, is:

$$c_i \begin{pmatrix} x \\ y \\ z \end{pmatrix} = - \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (2.15)$$

Thus, the point group associated with this lattice contains only two elements, the identity operator, E , and the inversion operator that satisfies the relation $c_i^2 = E$. This point group is denoted by C_i .

2. Next is the *primitive monoclinic* lattice. Here we also start from the two-dimensional oblique lattice, but now the third vector, $\mathbf{a}_3$, is set to be perpendicular to the xy plane, as shown in Fig. 2-19. This lattice has the following symmetry elements: c_2 rotation around the z axis, inversion c_i , and reflection through the xy plane. The latter operation is denoted by σ_h , and its action on a general vector is

$$\sigma_h \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ -z \end{pmatrix}. \quad (2.16)$$

One can quickly check that $c_2 = \sigma_h c_i$ and that the four symmetry operations, E , c_2 , c_i and σ_h form a group. This group is denoted by C_{2h} (an explanation about the symbols that designate the point groups can be found in Chapter 4).

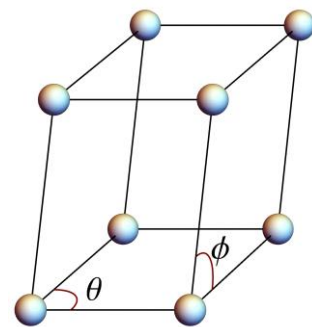


Figure 2-18 Triclinic lattice

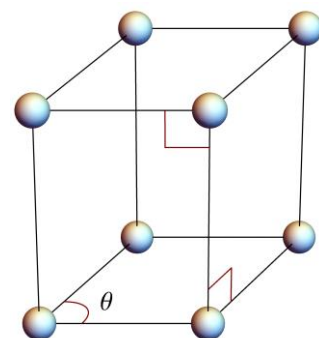


Figure 2-19 Primitive monoclinic lattice

3. Another Bravais lattice characterized by the symmetry group C_{2h} is the *base-centered monoclinic* lattice shown in Fig. 2-20. Starting from a centered rectangular lattice in the xy plane (with primitive basis vectors satisfying $|\mathbf{a}_1| = |\mathbf{a}_2|$ and angle between them $\theta \neq 90^\circ$), this lattice is obtained by adding a third primitive vector $\mathbf{a}_3$ in the plane defined by the z axis and the vector $\mathbf{a}_1 + \mathbf{a}_2$, such that the angle it creates with the xy plane is $\gamma \neq 90^\circ$, see figure.

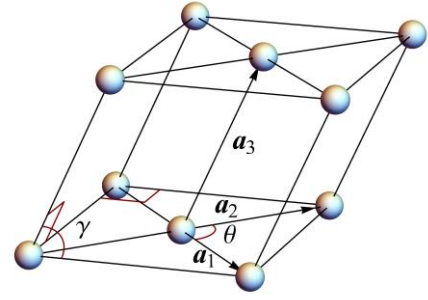


Figure 2-20 Base centered monoclinic lattice

Assuming $\mathbf{a}_1 + \mathbf{a}_2$ is parallel to the x axis, the reflection symmetry is through the xz plane, while the c_2 rotation is around the y axis. This lattice is also invariant to inversion.

4. Consider now a lattice whose base is the primitive rectangle lattice (in the xy plane), and a third primitive vector, $\mathbf{a}_3$, is in the z direction. When the lengths of all primitive basis vectors are different, $|\mathbf{a}_1| \neq |\mathbf{a}_2| \neq |\mathbf{a}_3|$, the resulting lattice is the *primitive orthorhombic* lattice shown in Fig. 2-21. Apart from translations, it is invariant under the following symmetry operations: three rotations by 180° around the axes x , y and z ; three reflections through the planes that are perpendicular to these axes; inversion; and the identity operation. For instance, the rotation around the x axis, $c_2^{(x)}$, and the reflection through the plane which is perpendicular to the same axis, σ_x , acting on a general vector in space, give:

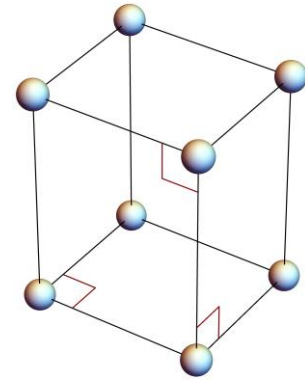


Figure 2-21 Primitive orthorhombic lattice

$$c_2^{(x)} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ -y \\ -z \end{pmatrix} \quad \text{and} \quad \sigma_x \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -x \\ y \\ z \end{pmatrix}. \quad (2.17)$$

The point group that describes the symmetry of the primitive orthorhombic lattice is denoted by D_{2h} .

5. Similarly, starting from a basis of centered rectangular lattice in a plane, and choosing $\mathbf{a}_3$ to be perpendicular to that plane, we obtain *the base centered orthorhombic* lattice shown in Fig. 2-22. The point group associated with this lattice is the same as that of the primitive orthorhombic lattice, D_{2h} .
6. An additional lattice that belongs to the orthorhombic system (i.e., characterized by the same point group as the two previous lattices) is the *body-centered orthorhombic* lattice depicted in Fig. 2-23. Here, at the center of each cuboid, there is an additional lattice point. The lattice points on each diagonal plane that passes through opposite edges is a centered rectangular lattice. Notice that in Fig. 2-23 none of the primitive basis vectors lie in the xy plane.
7. Finally, the last lattice that belongs to the orthorhombic system is the *face-centered orthorhombic* lattice presented in Fig. 2-24. Here at the center of each cuboid's face, there is an additional lattice point. Choosing our coordinate system to coincide with the edges of the cuboid, and the lengths of these edges to be a , b , and c , the primitive basis vectors are given by:

$$\begin{aligned} \mathbf{a}_1 &= \frac{1}{2}(a, b, 0), \quad \mathbf{a}_2 = \frac{1}{2}(a, 0, c), \\ \text{and } \mathbf{a}_3 &= \frac{1}{2}(0, b, c) \end{aligned} \quad (2.18)$$

so that

$$\begin{aligned} a\hat{\mathbf{x}} &= \mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3 \\ b\hat{\mathbf{y}} &= \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3, \\ b\hat{\mathbf{z}} &= -\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3 \end{aligned} \quad (2.19)$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are unit vectors in the directions of the axes of the coordinate system. One can verify that this choice of primitive basis vectors indeed gives the lattice.

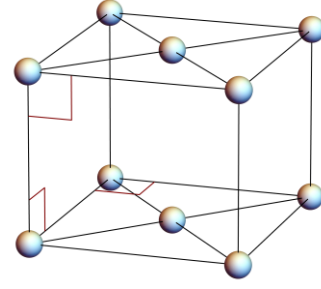


Figure 2-22 Base centered orthorhombic lattice

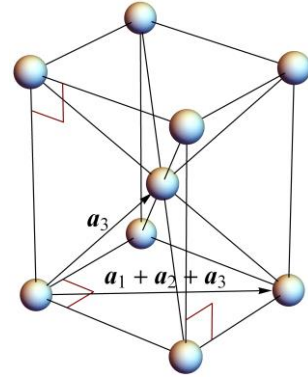


Figure 2-23 Body-centered orthorhombic lattice

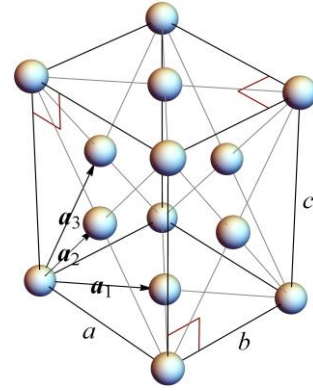


Figure 2-24 Face-centered orthorhombic lattice

8. The next lattice system is the tetragonal system. Here the starting point is a two-dimensional square lattice in the xy plane, with square sides of length a . The simplest lattice of this system is obtained by choosing the vector $\mathbf{a}_3$ to be in the z direction, and $|\mathbf{a}_3| \neq a$. The resulting lattice, called the *primitive tetragonal* lattice, is presented in Fig. 2-25.

Its point group contains 16 symmetry operations: the identity E ; inversion c_i ; rotations in multiples of 90° around the z axis c_4 , c_2 , and c_4^3 ; three reflections through planes that are perpendicular to the axes σ_x , σ_y , and σ_z ; two reflections through diagonal planes that contain the z axis and pass through opposite vertices of the square σ_{xy} , and $\sigma_{x\bar{y}}$; four rotations in 180° around axes that are perpendicular to the z axis and either the parallel to the diagonals of the square or parallel to its sides $c_2^{(x)}$, $c_2^{(y)}$, $c_2^{(xy)}$ and $c_2^{(x\bar{y})}$; and two additional symmetry operations, S_4 and S_4^3 called *improper rotations*.

An improper rotation, S_n , is a rotation by $360^\circ/n$ degrees followed by reflection through the plane, which is perpendicular to the rotation axis. In our case $S_4 = \sigma_z c_4$, and its action on a vector in space is:

$$S_4 \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -y \\ x \\ -z \end{pmatrix}. \quad (2.20)$$

An illustration of a shape that possesses an improper rotation symmetry, S_4 , but does not have c_4 symmetry or inversion symmetry is shown in Fig. 2-26. The operation S_4^3 is a rotation by 270° degrees followed by reflection. Notice that $S_2 = c_i$ and $S_4^2 = c_2$. The point group of the tetragonal lattice is denoted by D_{4h} .

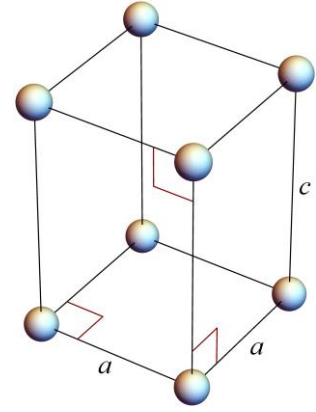


Figure 2-25 Primitive tetragonal lattice

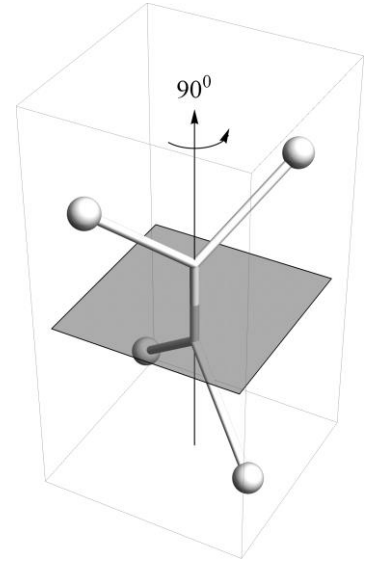


Figure 2-26 A shape with an improper rotation symmetry, S_4

9. The *body-centered tetragonal* lattice is the primitive tetragonal lattice with an extra lattice point at the center of each cuboid, see Fig. 2-27. This lattice can also be viewed as a face-centered tetragonal lattice. It is seen by choosing the vertical sides of the cuboid as the planes containing the z axis and running parallel to the diagonals of the square lattice (in the xy plane).
10. The repeating element of the *rhombohedral* lattice is obtained by deforming a cube along its diagonal, keeping all the length of its sides intact, see Fig. 2-28. In this figure, we have colored the lattice points with different colors to highlight the lattice symmetry to rotations by $\pm 120^\circ$ degrees around the principal symmetry axis. The latter passes through the two gray points shown in the figure. This symmetry is easier to perceive from the “top view” of the lattice depicted in Fig. 2-29.

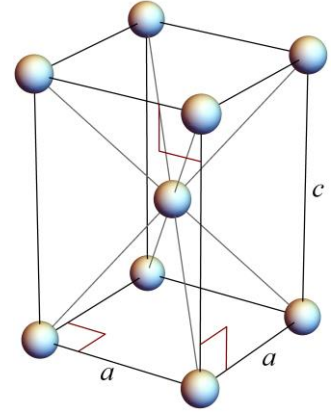


Figure 2-27 Body centered tetragonal lattice

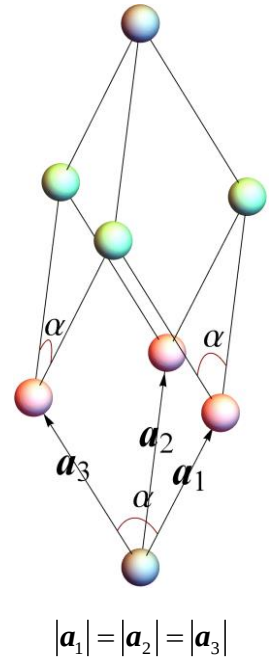


Figure 2-28
Rhombohedral lattice

A possible choice of the primitive basis vector for this lattice is:

$$\mathbf{a}_{1,3} = \left(\pm \frac{a}{2}, -\frac{a}{2\sqrt{3}}, \frac{c}{3} \right), \quad \mathbf{a}_2 = \left(0, \frac{a}{\sqrt{3}}, \frac{c}{3} \right), \quad (2.21)$$

where a and c are arbitrary. One can check that these basis vectors have equal lengths,

$$|\mathbf{a}_1| = |\mathbf{a}_2| = |\mathbf{a}_3| = \sqrt{\frac{a^2}{3} + \frac{c^2}{9}}, \quad (2.22)$$

and that the angle between any pair of them is

$$\alpha = \cos^{-1} \left(\frac{2c^2 - 3a^2}{2(c^2 + 3a^2)} \right). \quad (2.23)$$

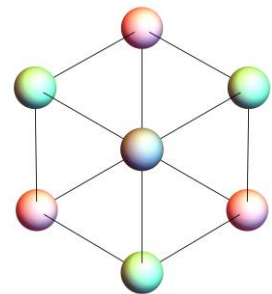


Figure 2-29
“Top view” of the
rhombohedral lattice

11. Next is the *hexagonal lattice* shown in Fig. 2-30. It is obtained by setting a two-dimensional hexagonal lattice as a basis in the xy plane and choosing the third basis vector, $\mathbf{a}_3$, to be in the z direction. The point group associated with this lattice is C_{6v} . It contains 24 elements: the identity; inversion; five rotations around the principal symmetry (z) axis in multiples of 60° ; reflection through the xy plane; six reflections through planes perpendicular to the xy plane; six rotations by 180° around axes perpendicular to the principal axis; and four improper rotations, S_6 , S_3 , S_6^5 , and S_3^5 .

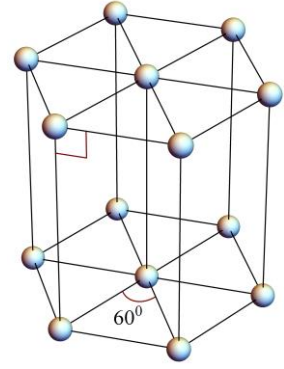


Figure 2-30
Hexagonal lattice

12. The last lattice system is the cubic system. The simplest one is the *primitive cubic lattice* shown in Fig. 2-31. The point group of this lattice, denoted by O_h , is the group symmetry of a cube. It contains 48 symmetry operations: the identity; inversion; reflections through three planes; eight rotations by 120° (two around each diagonal of the cube); eight improper rotations by 60° (around the same axes mentioned above); six rotations by 90° ; six improper rotations by 90° ; nine rotations by 180° (three around the axes that pierce the centers of opposite faces and six that bisect opposite sides of the cube); and six reflections through planes that contain two opposite sides of the cube.

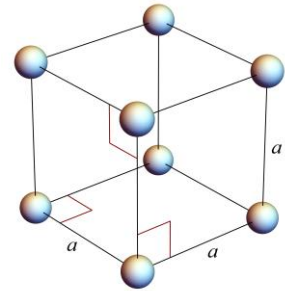


Figure 2-31
Primitive cubic lattice

13. The *body-centered cubic lattice*, shown in Fig. 2-32, belongs to the same lattice system. The primitive basis vectors of this lattice can be selected to be:

$$\mathbf{a}_1 = \frac{a}{2}(-1, 1, 1), \mathbf{a}_2 = \frac{a}{2}(1, -1, 1), \mathbf{a}_3 = \frac{a}{2}(1, 1, -1), \quad (2.24)$$

and the volume of the primitive unit cell is $a^3/2$.

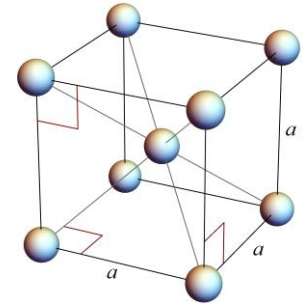


Figure 2-32 Body-centered

14. Finally, the *face-centered cubic lattice* shown in Fig. 3-23 is obtained (for example) from the following basis vectors:

$$\mathbf{a}_1 = \frac{a}{2}(0, 1, 1), \mathbf{a}_2 = \frac{a}{2}(1, 0, 1), \mathbf{a}_3 = \frac{a}{2}(1, 1, 0). \quad (2.25)$$

The volume of its unit cell is $a^3/4$ (recall that a unit cell contains a single lattice point).

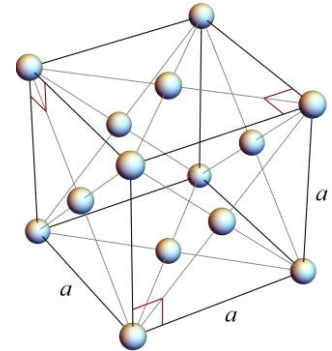


Figure 2-33 Face-centered cubic (fcc) lattice

Altogether we have listed 14 different Bravais lattices (in three dimensions) classified by seven lattice systems: triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, and cubic. Adding sublattices and decorations, one obtains lattices described by 230 different space groups. Seventy-three of them are symmorphic groups. The rest are non-symmorphic groups characterized by symmetry operations such as glide reflection $\{\sigma | \mathbf{a}_i/2\}$ (explained above), and *screw displacement* (also called screw operation or *rotary displacement*), $\{c_n | \mathbf{a}/2\}$. The latter is a rotation by $360^\circ/n$ degrees followed by a translation along the rotation axis by half the lattice constant.

2.9 Directions and planes in a crystal (Miller indices)

The primitive basis vectors of a lattice, $\mathbf{a}_i$ ($i=1,2,3$), can be used to define lattice directions (i.e., the directions of vectors connecting pairs of lattice points) by $h\mathbf{a}_1 + k\mathbf{a}_2 + l\mathbf{a}_3$, where h , k and l are coprime numbers, namely integers whose common factor is only one. The *Miller index* denoting such a direction is $[h,k,l]$. It is customary to replace negative values of these integers with an overbar, e.g., instead of writing $[1,2,-3]$ we use $[1,2,\bar{3}]$.

Similarly, planes in the lattice are defined by three points, $\mathbf{a}_1/h$, $\mathbf{a}_2/k$ and $\mathbf{a}_3/l$, see Fig. 2-34, and using Miller indices denoted by (h,k,l) . This symbol refers to the whole family of planes that are parallel to each other. Notice that the direction $[h,k,l]$ is generally not perpendicular to the plane (h,k,l) unless we are dealing with a cubic lattice.

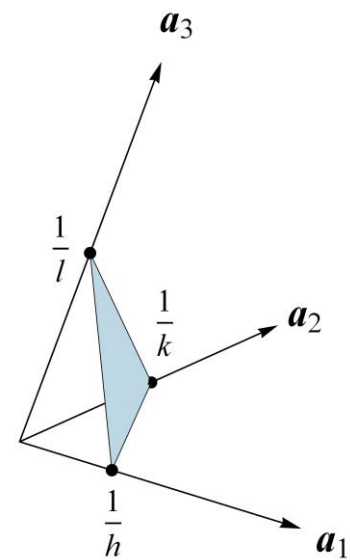


Figure 2-34 Definition of lattice planes using Miller indices

The above definition of the lattice planes ensures that the lattice points of a three-dimensional Bravais lattice that reside on such a plane form a two-dimensional Bravais lattice. For example, for a primitive cubic lattice, the lattice points on the $(1,0,0)$ plane form a square lattice; those of the $(1,1,0)$ plane are ordered in a simple rectangular lattice; while the points on the $(1,1,1)$ plane create a hexagonal lattice.

2.10 The reciprocal lattice and Brillouin zones

The *reciprocal lattice* of a Bravais lattice whose lattice points are $\mathbf{a} = j_1 \mathbf{a}_1 + j_2 \mathbf{a}_2 + j_3 \mathbf{a}_3$ (with $\mathbf{a}_i$ as the primitive basis vectors and j_i integers) is the set of points $\mathbf{b}$ that satisfy the condition:

$$\mathbf{b} \cdot \mathbf{a} = 2\pi m, \quad (2.26)$$

where m is an integer. The reciprocal lattice is a Bravais lattice by itself, and as such, its lattice points are given by a linear combination of primitive basis vectors,

$$\mathbf{b} = n_1 \mathbf{b}_1 + n_2 \mathbf{b}_2 + n_3 \mathbf{b}_3, \quad (2.27)$$

where n_i are integers. In two dimensions, the primitive basis vectors of the reciprocal lattice are given by

$$\mathbf{b}_1 = 2\pi \frac{\hat{\mathbf{z}} \times \mathbf{a}_2}{|\mathbf{a}_1 \times \mathbf{a}_2|} \quad \text{and} \quad \mathbf{b}_2 = 2\pi \frac{\hat{\mathbf{z}} \times \mathbf{a}_1}{|\mathbf{a}_1 \times \mathbf{a}_2|}. \quad (2.28)$$

Here we assume the lattice to reside in the xy plane, and $\hat{\mathbf{z}}$ is a unit vector perpendicular to that plane. The denominator in these formulas, $|\mathbf{a}_1 \times \mathbf{a}_2|$, is the area of a unit cell of the lattice (notice that the vector product of two-dimensional vectors is a scalar).

In three dimensions

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{V_{uc}}, \quad \mathbf{b}_2 = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{V_{uc}}, \quad \text{and} \quad \mathbf{b}_3 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{V_{uc}}, \quad (2.29)$$

where

$$V_{uc} = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) \quad (2.30)$$

is the volume of a unit cell (one can always choose the order of these primitive vectors such that this volume is positive). The proof of these formulas is given as an exercise.

Comment: The symmetries of a lattice and its reciprocal are the same. Namely, both belong to the same lattice system. However, the reciprocal lattice is not necessarily the same as the original one. For example, the reciprocal lattice of fcc is bcc and vice versa (the proof of this property is given as an exercise).

To gain a better understanding of the meaning of the reciprocal lattice, consider the problem of an electron moving in a periodic potential,

$$u(\mathbf{r} + \mathbf{a}) = u(\mathbf{r}). \quad (2.31)$$

From Fourier theory, we know that any periodic function can be represented as a Fourier series,

$$u(\mathbf{r}) = \sum_{\mathbf{b}} u_{\mathbf{b}} \exp(i\mathbf{b} \cdot \mathbf{r}), \quad (2.32)$$

where $u_{\mathbf{b}}$ are the Fourier expansion coefficients. The periodicity of the potential implies that

$$\begin{aligned} u(\mathbf{r} + \mathbf{a}) &= \sum_{\mathbf{b}} u_{\mathbf{b}} \exp[i\mathbf{b} \cdot (\mathbf{r} + \mathbf{a})] = \sum_{\mathbf{b}} u_{\mathbf{b}} \exp(i\mathbf{b} \cdot \mathbf{r} + i\mathbf{b} \cdot \mathbf{a}) \\ &= u(\mathbf{r}) = \sum_{\mathbf{b}} u_{\mathbf{b}} \exp(i\mathbf{b} \cdot \mathbf{r}). \end{aligned} \quad (2.33)$$

Therefore $\mathbf{b} \cdot \mathbf{a}$ must be an integer multiple of 2π . Thus the vectors $\mathbf{b}$ satisfy Eq. (2.26) and therefore belong to the reciprocal lattice. Hence, the reciprocal lattice represents the momentum space of the problem. Thus, using Wigner-Seitz procedure for the reciprocal lattice yields the Brillouin zone of the system.

Example – The Brillouin zone of a two-dimensional hexagonal lattice.

In this example, we identify the Brillouin zone of a system whose Bravais lattice is a two-dimensional hexagonal lattice. It will be obtained by constructing the Wigner-Seitz cell of the reciprocal lattice. First, we choose the primitive basis vectors of the hexagonal lattice to be

$$\mathbf{a}_1 = a(1, 0) \quad \text{and} \quad \mathbf{a}_2 = \frac{a}{2}(1, \sqrt{3}), \quad (2.34)$$

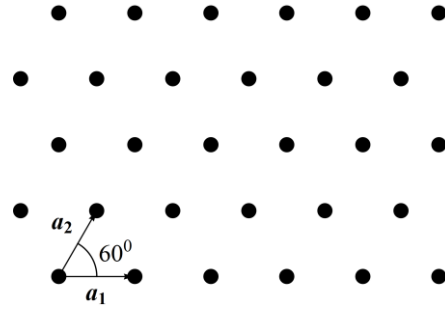


Figure 2-35 The primitive basis vectors of hexagonal lattice

as illustrated in Fig. 2-35. With this choice, we obtain that the area of the lattice unit cell is

$\mathbf{a}_1 \times \mathbf{a}_2 = a^2 \sqrt{3}/2$, and from formulas (2.28) for the primitive basis vectors of the reciprocal lattice we obtain:

$$\begin{aligned} \mathbf{b}_1 &= 2\pi \frac{\hat{\mathbf{z}} \times \mathbf{a}_2}{|\mathbf{a}_1 \times \mathbf{a}_2|} = 2\pi \frac{\frac{a}{2}(-\sqrt{3}, 1)}{\frac{a^2 \sqrt{3}}{2}} = \frac{2\pi}{a} \left(-1, \frac{1}{\sqrt{3}} \right) \\ \mathbf{b}_2 &= 2\pi \frac{\hat{\mathbf{z}} \times \mathbf{a}_1}{|\mathbf{a}_1 \times \mathbf{a}_2|} = 2\pi \frac{a(0, 1)}{\frac{a^2 \sqrt{3}}{2}} = \frac{2\pi}{a} \left(0, \frac{2}{\sqrt{3}} \right) \end{aligned} \quad (2.35)$$

These primitive lattice vectors define the reciprocal lattice points, $\mathbf{b} = n_1 \mathbf{b}_1 + n_2 \mathbf{b}_2$, as shown in Fig. 2-36. It is evident that the reciprocal lattice is also hexagonal, yet it is rotated by 90° degrees with respect to the original lattice. The Wigner-Seitz cell of the reciprocal lattice is constructed as described in section 2.7 and is shown in Fig. 2-37.

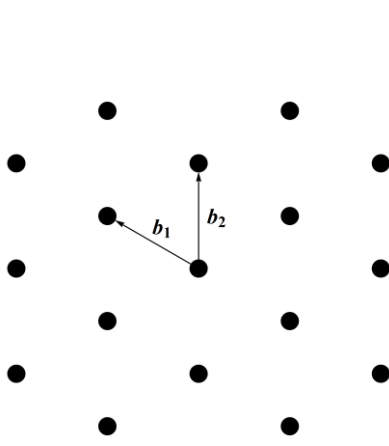


Figure 2-36 The primitive basis vectors of reciprocal lattice

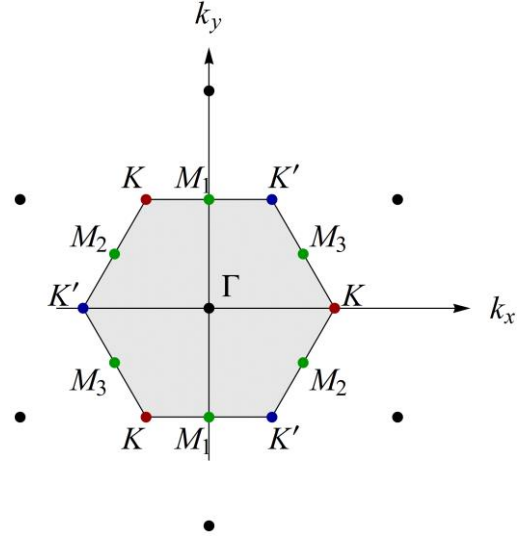


Figure 2-37 The Brillouin zone of hexagonal lattice

It is customary to give names to the special symmetry points of the Brillouin zone. The zero momentum point, $\mathbf{k} = (0, 0)$, is the Γ -point; the vertices of the hexagonal of the Brillouin zone are the K points; and the central points on each side of the hexagon are the M -points. Taking into account the periodic structure of the reciprocal lattice, one sees that the three K points in Fig. 2-37 are, in fact, the same point, and so are the three K' points. Also, since opposite sides of the hexagonal are identical, there are only three inequivalent M -points. These are denoted by M_1 , M_2 , and M_3 in the figure.

To conclude this chapter, we present the Brillouin zones of bcc and fcc lattices obtained by the same procedure described in the example:

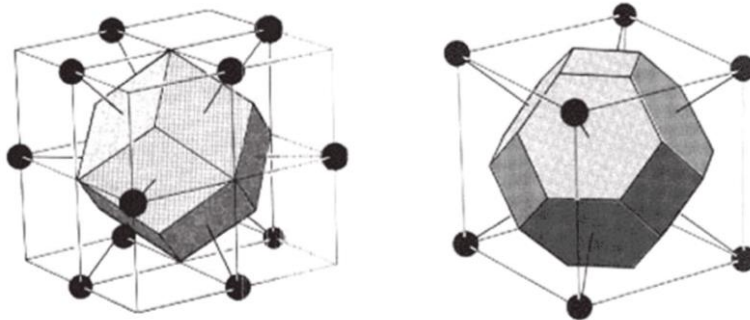


Figure 2-38 The Brillouin zones of fcc (left panel) and bcc (right panel) lattices.

2.11 Exercises

1. Prove that the only possible rotation symmetry operations, c_n , in a Bravais lattice is with $n = 1, 2, 3, 4$ and 6 .

Hint: Consider two rotations in angles $\phi = 360^\circ/n$ around two adjacent lattice points, A and B as illustrated in Fig. 2-39. Find a simple condition ensuring that the two rotated points, A' and B' , are located on the same Bravais lattice. Notice that the segments $A'B'$ and AB are parallel.

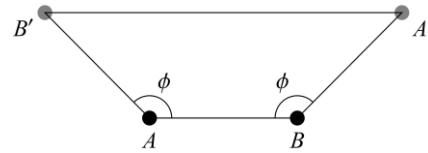


Figure 2-39

2. Identify the symmetry operations of the group C_{3v} (the symmetry group of an equilateral triangle) and construct the multiplication table of the group.

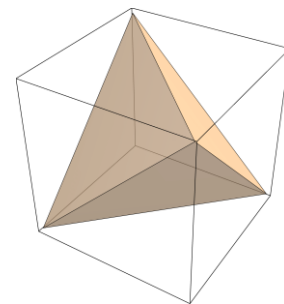


Figure 2-40 A regular tetrahedron

3. Identify all 24 symmetry operations of the point group, T_d , of a regular tetrahedron. A regular tetrahedron, also known as a triangular pyramid, is a polyhedron composed of four equilateral triangular faces. It has four equivalent vertex corners, as shown in Fig. 2-40.
4. Identify the two-dimensional Bravais lattices on the $(1,1,1)$ and $(1,1,\bar{1})$ planes of a bcc lattice.
5. Check that the primitive translation vectors of a lattice and its reciprocal satisfy the condition $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, and show that the reciprocal lattice of bcc is fcc and vice versa.
6. Identify the reciprocal lattices of the following lattices: primitive orthorhombic, base-centered orthorhombic, face-centered orthorhombic, and body-centered orthorhombic.
7. Verify that formulas (2.28-2.30) satisfy Eq. (2.26).

3 Time-reversal symmetry

In the previous chapter, we have presented various lattices and identified their spatial symmetries. The next step is to harness tools from group theory to understand the implications of these symmetries on various physical aspects of the system, such as the electron energy spectrum and the phonon vibrations. However, before turning to this task, let us consider an important and simpler symmetry – the time-reversal symmetry:

$$t \rightarrow -t. \quad (3.1)$$

This symmetry applies to the fundamental laws of nature and describes the microscopic behavior of matter (in the absence of an external magnetic field). Namely, there is nothing in the fundamental physical laws that distinguishes the direction of time. In other words, if we could make a film of the microscopic movements of all particles in a closed system, we would not be able to say whether the film is running forward or backward. Of course, in the macroscopic world, this symmetry is broken (for example, we can see that when a glass of water falls, it breaks into small pieces, but we never see the reverse process). However, this subject is for a different course. Here we focus on understanding the implication of the time-reversal symmetry (3.1) on the spectrum of electrons in a lattice. We start by discussing systems where the spin degree of freedom can be ignored. Next, we define the time-reversal operator for spin $\frac{1}{2}$ particles, and explain the implication of time-reversal symmetry for such systems.

3.1 Implications of time-reversal symmetry on the electronic spectrum

In the systems that we shall consider, time-reversal symmetry implies invariance to reversing the momentum direction, $\mathbf{k} \rightarrow -\mathbf{k}$. To find out the significance of this symmetry on the electronic spectrum,

$$\varepsilon(\mathbf{k}) = \varepsilon(-\mathbf{k}), \quad (3.2)$$

it is instructive to analyze a specific example: Consider a system of an electron moving in a two-dimensional hexagonal lattice. The Brillouin zone of this system is depicted in Fig. 2-37. Our goal is to understand what can we learn from time-reversal symmetry (3.2) about the local behavior of the spectrum in the vicinity of the special points of the Brillouin zone, Γ , K , and M .

To begin with, let us take a look at the Taylor expansion of $\varepsilon(\mathbf{k})$ near the Γ point ($\mathbf{k} = 0$). Time reversal symmetry (3.2) implies that the linear term of this expansion must vanish, therefore

$$\varepsilon(\mathbf{k}) = \sum_{\alpha\beta} k_{\alpha} \frac{\hbar^2}{2m_{\alpha\beta}} k_{\beta} + O(k^4), \quad (3.3)$$

where $m_{\alpha\beta}$ is the effective mass tensor. Thus, time-reversal symmetry dictates that the spectrum near this point is quadratic. In a cubic lattice (or a square lattice in two dimensions), similar considerations lead to the same result, but one can also show that spatial symmetry is sufficiently high to ensure that $1/m_{\alpha\beta} = \delta_{\alpha\beta}/m_{\text{eff}}$, hence

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2 k^2}{2m_{\text{eff}}} + O(k^4), \quad (3.4)$$

where m_{eff} is the effective mass of the electrons near $\mathbf{k} = 0$.

Consider now the energy spectrum near one of the M points of the Brillouin zone, and let $\delta\mathbf{k}$ be the deviation of the wavenumber vector from that point. I.e., $\mathbf{k} = \mathbf{k}_M + \delta\mathbf{k}$, where $\mathbf{k}_M$ is the value of the wavenumber at the M point. With this definition, time-reversal symmetry yields

$$\varepsilon(\mathbf{k}_M + \delta\mathbf{k}) = \varepsilon(-\mathbf{k}_M - \delta\mathbf{k}) = \varepsilon(\mathbf{k}_M - \delta\mathbf{k}), \quad (3.5)$$

where second equality follows from the fact that opposite M points on the Brillouin zone are, in fact, the same point due to the periodicity Brillouin zone in $\mathbf{k}$ space. The above equation implies that the energy spectrum near the M points is quadratic.

Finally, consider the behavior near the K point of the Brillouin zone. As before, we present the wavenumber in terms of its deviation from that point, $\mathbf{k} = \mathbf{k}_K + \delta\mathbf{k}$. In this case, we have

$$\varepsilon(\mathbf{k}_K + \delta\mathbf{k}) = \varepsilon(-\mathbf{k}_K - \delta\mathbf{k}) = \varepsilon(\mathbf{k}_{K'} - \delta\mathbf{k}). \quad (3.6)$$

Here we took into account that time-reversal symmetry exchanges between the K and the K' points of the Brillouin zone, i.e. $-\mathbf{k}_K = \mathbf{k}_{K'}$ (see Fig. 2-37). However, these points are inequivalent; hence one cannot assume that the linear term of the Taylor expansion of $\varepsilon(\mathbf{k})$ in $\delta\mathbf{k}$ vanishes. The only information obtained from time-reversal symmetry, in this case, is a relation between the slopes of the spectrum (in the same band) at the K and the K' points:

$$\left. \frac{\partial \varepsilon(\mathbf{k})}{\partial \mathbf{k}} \right|_K = - \left. \frac{\partial \varepsilon(\mathbf{k})}{\partial \mathbf{k}} \right|_{K'}. \quad (3.7)$$

3.2 The time-reversal operator for spin $\frac{1}{2}$ systems

Ignoring the electron's spin is justified when there is no coupling between the electron trajectory in space and its spin direction. In that case, the spin only implies a double degeneracy of the electron energy at each value of $\mathbf{k}$. However, in the presence of spin orbit-interaction, the situation becomes more interesting. Our goal here is to study the main manifestations of time-reversal symmetry in such cases. To this end, we must first formulate the operation of time reversal on spin $\frac{1}{2}$ particles.

Let Θ denote the time-reversal operator. The basic properties one expects from this operator are:

$$\Theta|\mathbf{r}\rangle = |\mathbf{r}\rangle, \quad \Theta|\mathbf{p}\rangle = |-\mathbf{p}\rangle \quad \text{and} \quad \Theta|\mathbf{L}\rangle = |-\mathbf{L}\rangle, \quad (3.8)$$

where $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is the angular momentum. An alternative way to present these equations is in the form of operator equations:

$$\Theta^{-1}\hat{\mathbf{r}}\Theta = \hat{\mathbf{r}}, \quad \Theta^{-1}\hat{\mathbf{p}}\Theta = -\hat{\mathbf{p}}, \quad \text{and} \quad \Theta^{-1}\hat{\mathbf{L}}\Theta = -\hat{\mathbf{L}}. \quad (3.9)$$

These relations are proved, e.g., by calculating the matrix elements of the operators on both sides of the equations in a complete set of eigenstates. Since the spin operator, $\hat{\mathbf{S}}$, behaves similar to angular momentum, time-reversal implies

$$\Theta^{-1}\hat{\mathbf{S}}\Theta = -\hat{\mathbf{S}}. \quad (3.10)$$

Consider now the action of time-reversal operation on wave functions. To begin with, let us focus our attention on a spinless particle. The action of time-reversal symmetry on the wave function gives its complex conjugate (as follows from the Schrödinger equation). Thus, for an arbitrary pair of wave functions, $|\psi_1\rangle$ and $|\psi_2\rangle$, we have

$$\langle \Theta\psi_1 | \Theta\psi_2 \rangle = \langle \psi_1 | \psi_2 \rangle^* = \langle \psi_2 | \psi_1 \rangle. \quad (3.11)$$

This relation implies that time-reversal operation does not preserve the inner product as a unitary operator. Instead, it gives the complex conjugation of the inner product. Such operator is called anti-unitary, and a mathematical theorem that we present here without a proof states that any anti-unitary operator may be represented as a product of a unitary operator, U , by an operator, K , that we shall call the “complex conjugation operator”,

$$\Theta = UK. \quad (3.12)$$

This decomposition of the time-reversal operator also applies to spinful particles, i.e., when the wavefunctions are spinors.

The easiest way to define the action of the K operator on a wave function, $|\psi\rangle$, is by using a complete set of states, $|n\rangle$. Expanding the wave function in this basis, $|\psi\rangle = \sum_n \psi_n |n\rangle$, the action of K is defined by

$$K|\psi\rangle = \sum_n \psi_n^* |n\rangle. \quad (3.13)$$

Thus K acts only on the expansion coefficients, while the basis states, $|n\rangle$, are left intact. This definition is, clearly, basis-dependent. If we choose a different basis, we may get another K . However, this is not a problem because one can always choose the unitary transformation U to account for the transformation from one basis to another. Thus there are many ways to define the operators U and K , depending on the choice of the basis. It remains to figure out how to define U for some particular set of basis functions, $|n\rangle$. Since the action of the time-reversal operator gives

$$\Theta|\psi\rangle = UK|\psi\rangle = \sum_n \psi_n^* U|n\rangle, \quad (3.14)$$

we should understand how U acts on $|n\rangle$.

In the spinless case, and for real-space basis functions, the time-reversal operation gives the complex conjugate wave function, therefore we should choose $U = 1$.

For spin $\frac{1}{2}$ particles, the wave function is a spinor containing two components. From Eq. (3.10) we expect that $\Theta|\uparrow\rangle = U|\uparrow\rangle$ gives a state that is parallel to $|\downarrow\rangle$. Similarly, $\Theta|\downarrow\rangle = U|\downarrow\rangle$ should be parallel to $|\uparrow\rangle$. Taking the spinor components in the z direction, an operation that reverses the spin direction is a π -rotation around the y axis. Since the spin, as the angular momentum, is the generator of rotations in the spin space,

$$U = \exp\left(-i \frac{\hat{S}_y}{\hbar} \pi\right) = \exp\left(-i \frac{\pi}{2} \tau_y\right) = -i\tau_y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad (3.15)$$

where τ_y is the Pauli matrix (see Eq. (3.22) below). Thus

$$\Theta = -i\tau_y K. \quad (3.16)$$

From here we see that

$$\begin{aligned} \Theta|\uparrow\rangle &= U|\uparrow\rangle = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |\downarrow\rangle \\ \Theta|\downarrow\rangle &= U|\downarrow\rangle = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} = -|\uparrow\rangle \end{aligned} \quad (3.17)$$

and therefore $\Theta^2|S\rangle = -|S\rangle$. The conclusion from this discussion is that for spin 1/2 particles

$$\Theta^2|\psi\rangle = -|\psi\rangle. \quad (3.18)$$

This property implies that $|\psi\rangle$ and $\Theta|\psi\rangle$ are orthogonal, because from the above relation and Eq. (3.11) it follows that $\langle\Theta\psi|\psi\rangle = \langle\Theta^2\psi|\Theta\psi\rangle^* = \langle\Theta\psi|\Theta^2\psi\rangle = -\langle\Theta\psi|\psi\rangle$, and this can be satisfied only if

$$\langle\Theta\psi|\psi\rangle = 0. \quad (3.19)$$

This result is valid for any particle with a non-integer spin.

Equipped with knowledge about the action of the time-reversal operator on spin 1/2 particles, we turn to study the implication of this symmetry in systems with spin-orbit interactions.

3.3 Spin-orbit interaction and Kramer's degeneracy

On a qualitative level, spin-orbit interaction can be viewed as the interaction of the electron's magnetic moment with the magnetic field, $\mathbf{B}$, seen from the electron's moving reference frame, due to the presence electric field, $\mathbf{E}$, in the laboratory frame. Assuming no magnetic field in the system, the magnetic field in the reference frame attached to an electron moving with velocity $\mathbf{v}$ is

$$\mathbf{B} = -\frac{1}{c^2}\mathbf{v} \times \mathbf{E} = \frac{1}{ec^2}\mathbf{v} \times \nabla u(\mathbf{r}), \quad (3.20)$$

where to obtain the second equality, we expressed the electric field as a gradient of the electric potential $u(\mathbf{r})/e$. The electron's magnetic dipole moment (due to spin) is

$$\boldsymbol{\mu} = -\frac{1}{2}g\mu_B\boldsymbol{\tau}, \quad (3.21)$$

where g is the g -factor, which is approximately 2 (more precisely 2.002319...), $\mu_B = e\hbar/2m$ is the Bohr magneton (m being the free electron mass), and $\boldsymbol{\tau}/2$ is the dimensionless spin operator expressed in terms of Pauli matrices, $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)$ with

$$\tau_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \tau_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.22)$$

The spin-orbit Hamiltonian comes from the interaction of the magnetic dipole moment (3.21) with the magnetic field (3.20):

$$H_{\text{so}} = -\boldsymbol{\mu} \cdot \mathbf{B} \rightarrow \frac{\hbar}{4m^2c^2} \boldsymbol{\tau} \cdot [\mathbf{p} \times \nabla u(\mathbf{r})]. \quad (3.23)$$

To obtain this formula, we replaced the velocity with the momentum divided by the mass and added a factor $1/2$. This $1/2$ factor is due to another relativistic effect called *Thomas precession* which we do not discuss here.

Taking into account spin-orbit interaction, the Schrödinger equation of an electron moving in a periodic lattice is:

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) + \frac{\hbar}{4m^2c^2} \boldsymbol{\tau} \cdot [\nabla u(\mathbf{r}) \times i\hbar \nabla] \right\} \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \varepsilon_j(\mathbf{k}) \psi_{\mathbf{k}}^{(j)}(\mathbf{r}), \quad (3.24)$$

where $\psi_{\mathbf{k}}^{(j)}(\mathbf{r})$ is a spinor containing two components associated with the spin direction:

$$\psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \begin{pmatrix} \psi_{\mathbf{k}\uparrow}^{(j)}(\mathbf{r}) \\ \psi_{\mathbf{k}\downarrow}^{(j)}(\mathbf{r}) \end{pmatrix}. \quad (3.25)$$

Here, the index j refers to the energy band and $\mathbf{k}$ is Bloch's wave number.

Notice that Bloch's theorem also holds in the presence of spin-orbit interaction because the Hamiltonian (3.23) has the same spatial periodicity as the potential energy $u(\mathbf{r})$. Moreover, as discussed earlier, time-reversal symmetry dictates the relation:

$$\varepsilon_j(\mathbf{k}) = \varepsilon_j(-\mathbf{k}). \quad (3.26)$$

We show, now, that the time-reversal symmetry in fermionic systems,

$$\Theta^{-1} H \Theta = H, \quad (3.27)$$

implies that energy levels are degenerate. Multiplying the above equation, from the right, by $\psi_{\mathbf{k}}^{(j)}(\mathbf{r})$, and from the left by Θ yields

$$H \Theta \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \Theta H \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \varepsilon_j(\mathbf{k}) \Theta \psi_{\mathbf{k}}^{(j)}(\mathbf{r}). \quad (3.28)$$

Thus, if $\psi_{\mathbf{k}}^{(j)}(\mathbf{r})$ is an eigenstate of the Hamiltonian with energy $\varepsilon_j(\mathbf{k})$, then $\Theta \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = -i\tau_y \psi_{\mathbf{k}}^{(j)*}(\mathbf{r})$ is also an eigenstate with the same energy. Moreover, we have proved that $\Theta \psi_{\mathbf{k}}^{(j)}(\mathbf{r})$ is orthogonal to $\psi_{\mathbf{k}}^{(j)}(\mathbf{r})$ (see Eq. (3.19)); hence energy levels are, at least, doubly degenerate. This degeneracy is called *Kramer's degeneracy*. The two degenerate eigenstates have opposite momenta and opposite spins.

When transitions between bands can be neglected, the Hamiltonian that takes into account the spin-orbit interaction of an electron moving in a periodic lattice (and restricted

to the lowest band) can be approximated by

$$H(\mathbf{k}) = \varepsilon_0(\mathbf{k})I + \boldsymbol{\tau} \cdot \mathbf{w}(\mathbf{k}), \quad (3.29)$$

where $\varepsilon_0(\mathbf{k})$ is the spectrum in the absence of spin-orbit interaction, I is the identity matrix in the spinor space, and $\mathbf{w}(\mathbf{k}) = \langle \psi_k^{(0)} | H_{\text{so}} | \psi_k^{(0)} \rangle$, $\psi_k^{(0)}$ being the Bloch wave function when spin-orbit interaction is absent. Time reversal symmetry, $\Theta^{-1}H\Theta = H$, implies:

$$\varepsilon_0(\mathbf{k}) = \varepsilon_0(-\mathbf{k}), \text{ and } \mathbf{w}(\mathbf{k}) = -\mathbf{w}(-\mathbf{k}). \quad (3.30)$$

The second relation follows from $\boldsymbol{\tau} \cdot \mathbf{w}(\mathbf{k}) = \Theta^{-1} \boldsymbol{\tau} \cdot \mathbf{w}(\mathbf{k}) \Theta = \Theta^{-1} \boldsymbol{\tau} \Theta \cdot \mathbf{w}(-\mathbf{k}) = -\boldsymbol{\tau} \cdot \mathbf{w}(-\mathbf{k})$. Thus, in particular, $\mathbf{w}(0) = 0$. Diagonalization of the Hamiltonian (3.29) gives $\varepsilon(\mathbf{k}) = \varepsilon_0(\mathbf{k}) \pm |\mathbf{w}(\mathbf{k})|$; hence the energy at $k = 0$ is (at least) doubly degenerate.

3.4 The Rashba term

In many situations, the periodic potential of the lattice is too weak to contribute effectively to the spin-orbit interaction. However, the situation may be different near the system's boundary because the potential energy changes rapidly in space. Near the system's edge, the potential energy gradient is large, and the contribution to the spin-orbit interaction can be significant - see Eq. (2.23). In this section, we calculate the spectrum of electrons moving near the surface of a crystal, taking into account this effect.

For simplicity, we consider the surface of a cubic lattice, say the $(0,0,1)$ plane, and denote by $\hat{\mathbf{n}}$ a unit vector perpendicular to the surface and pointing outwards. The approximate Hamiltonian that describes an electron moving on the surface is:

$$H = \frac{\hbar^2 \mathbf{k}^2}{2m_{\text{eff}}} + \alpha_R \boldsymbol{\tau} \cdot (\hat{\mathbf{n}} \times \mathbf{k}), \quad (3.31)$$

where m_{eff} is the effective mass of the electron. The second term of the Hamiltonian (3.31) is due to spin-orbit interaction. It is called *Rashba term*. Here, α_R is the *Rashba parameter* that characterizes the strength of the spin-orbit interaction. The structure of Rashba term can be deduced from the following argument: The system contains only three vectors - the normal to the surface $\hat{\mathbf{n}}$, the electron's momentum $\hbar \mathbf{k}$, and the electron's spin $\hbar \boldsymbol{\tau}/2$. The only scalar that can be constructed from these vectors, which is the lowest order in $\mathbf{k}$ and have the property (3.30), is $\boldsymbol{\tau} \cdot (\hat{\mathbf{n}} \times \mathbf{k}) = \mathbf{k} \cdot (\boldsymbol{\tau} \times \hat{\mathbf{n}}) = \hat{\mathbf{n}} \cdot (\mathbf{k} \times \boldsymbol{\tau})$. Therefore, at low enough energy, one can describe the system by the Hamiltonian (3.31).

Now, let us choose the surface on which the electron moves to be the xy plane. The normal to the surface is $\hat{n} = \hat{z}$, so that the Rashba term is:

$$\boldsymbol{\tau} \cdot (\hat{z} \times \mathbf{k}) = \begin{vmatrix} \tau_x & \tau_y & \tau_z \\ 0 & 0 & 1 \\ k_x & k_y & k_z \end{vmatrix} = k_x \tau_y - k_y \tau_x. \quad (3.32)$$

Assuming that the electrons move only parallel to the plane, the Hamiltonian (3.31) reduces to:

$$H = \begin{pmatrix} \frac{\hbar^2}{2m_{\text{eff}}} (k_x^2 + k_y^2) & -\alpha_R (k_y + ik_x) \\ -\alpha_R (k_y - ik_x) & \frac{\hbar^2}{2m_{\text{eff}}} (k_x^2 + k_y^2) \end{pmatrix}. \quad (3.33)$$

Diagonalization of this Hamiltonian gives the energy spectrum of the system:

$$\varepsilon_{\pm}(k) = \frac{\hbar^2 k^2}{2m_{\text{eff}}} \pm |\alpha_R| k = \frac{\hbar^2 \left(k \pm \frac{m_{\text{eff}}}{\hbar^2} |\alpha_R| \right)^2}{2m_{\text{eff}}} - \frac{m_{\text{eff}} \alpha_R^2}{2\hbar^2}, \quad (3.34)$$

with $k = \sqrt{k_x^2 + k_y^2}$. The energy surfaces described by this formula are depicted in Fig. 3-1.

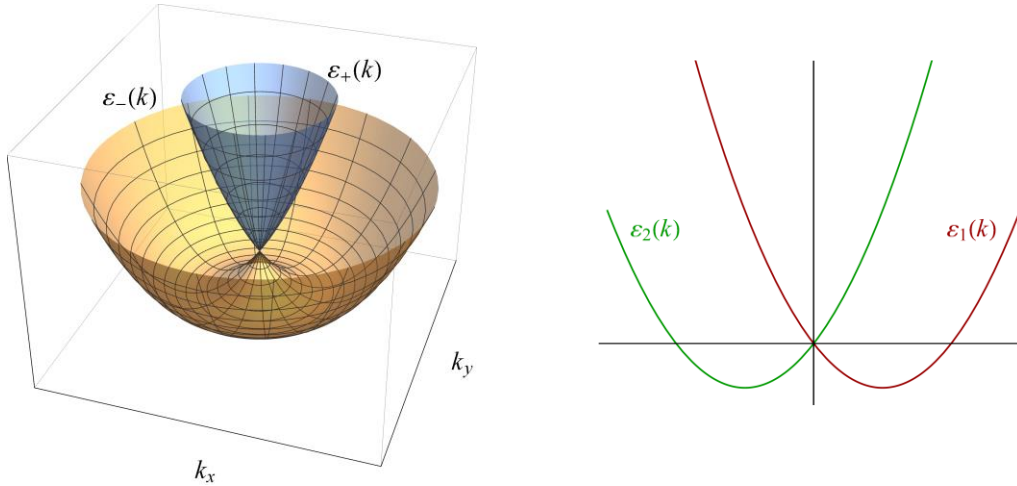


Figure 3-1 The energy spectrum described by Eq. (3.34)

On the left panel of the figure, the orange and blue surfaces represent the spectrum's two branches, $\varepsilon_-(k)$ and $\varepsilon_+(k)$, respectively. These branches touch at a single point, $k = 0$. The primary influence of the weak spin-orbit interaction is near this point, where the dispersion becomes linear. The right panel of Fig. 3-1 shows a cross-section of the spectrum in the direction of the k_x -axis. Along this line (and any other axis that passes

through the origin), the spectrum looks like two parabolas shifted away from each other. This illustration highlights the time-reversal symmetry property of the system because, here, it is clear that the two branches of the spectrum satisfy the relation $\varepsilon_1(\mathbf{k}) = \varepsilon_2(-\mathbf{k})$.

We turn now to discuss the spin configuration of the wave functions in the problem. The eigenfunctions of the Hamiltonian (3.33) (see Exercise 2) are

$$\psi_{\pm} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm \text{sign}(\alpha_R) \exp(i\tilde{\phi}) \end{pmatrix} \quad \text{with} \quad \tilde{\phi} = \arctan\left(\frac{k_x}{k_y}\right). \quad (3.35)$$

These functions are particular cases of the general spinor wave function,

$$|\hat{\mathbf{d}}\rangle = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right) \exp(i\phi) \end{pmatrix} \quad (3.36)$$

that describes a spin pointing in the direction

$$\hat{\mathbf{d}} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta). \quad (3.37)$$

Here θ and ϕ are the polar angles defined in Fig. 3-2. From the comparison of Eqs. (3.35) and (3.36), it follows that the spin is perpendicular to $\mathbf{k}$ and lies in the xy plane. Thus, the electron momentum, the electron spin, and the normal to the surface are perpendicular to each other.

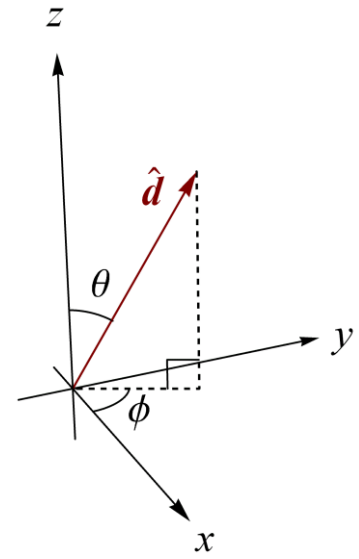


Figure 3-2 The polar angles of the spinor $|\hat{\mathbf{d}}\rangle$

In Fig. 3-3 we present the Fermi surfaces of the system near the band touching point with arrows showing the spin's direction for any momentum. The arrows change direction when α_R changes sign. Here they are plotted for $\alpha_R > 0$, and assuming that k_x , k_y , and z form a right-handed triple.

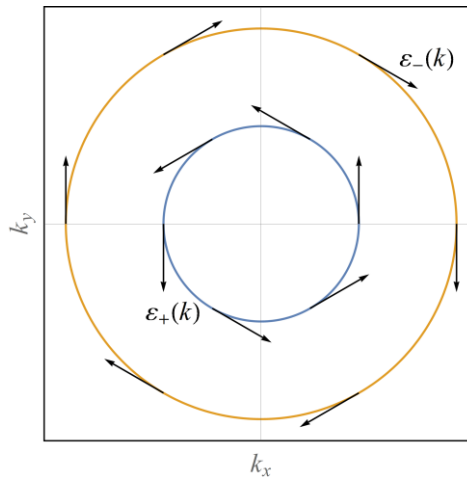


Figure 3-3 The spin direction on the Fermi surface of the spectrum shown in Fig. 3-1

3.5 Exercises

1. Diagonalize the Hamiltonian (3.33) and prove Eqs. (3.34) and (3.35).
2. Prove that the spinor function in Eq. (3.36) represents a spin that points in the direction of the vector (3.37). Namely, show that

$$\hat{\mathbf{d}} \cdot \boldsymbol{\tau} |\hat{\mathbf{d}}\rangle = |\hat{\mathbf{d}}\rangle. \quad (3.38)$$

3. Calculate the energy spectrum of Luttinger Hamiltonian:

$$H = Ak^2 + B(\mathbf{k} \cdot \hat{\mathbf{L}})^2 + \Delta_{\text{so}} \hat{\mathbf{s}} \cdot \hat{\mathbf{L}}, \quad (3.39)$$

where A , B , and Δ_{so} are constants, $\hat{\mathbf{L}}$ is the orbital angular momentum with $l=1$, and $\hat{\mathbf{s}}$ is the spin operator. The last term in this Hamiltonian represents the spin-orbit interaction. To shorten the formulas, assume $\hbar=1$.

Advice: Choose the z axis in the direction of the wave number vector, and rewrite the Hamiltonian using $\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{s}}$ and L_z . Notice that L_z commutes with J_z but not with $\hat{\mathbf{J}}^2$. Now consider the following cases $J_z = \pm 3/2$ (where the Hamiltonian is diagonal), and $J_z = \pm 1/2$ where for each sign, the Hamiltonian is a 2×2 matrix. Show that the spectrum you got is doubly degenerate.

4 Group theory: Basic concepts

From time immemorial, scientists used symmetry considerations to analyze physical systems. However, with time, symmetry evolved from being only a tool to the stage that constitutes the theory's cornerstone. In many situations, one can deduce the structure of the energy spectrum, or the system's response to some external perturbations, solely from symmetry considerations even without knowing the precise form of the Hamiltonian. We shall encounter a few such examples in the coming chapters.

The mathematical framework for the study of symmetry is *group theory*, and in this chapter, we shall present the main ideas and tools needed to analyze crystals. We focus our attention on point groups that are sufficient for our purposes. We skip the mathematical proofs of the theorems as they can be easily found in the literature. This chapter is meant to be self-contained and therefore has some overlap with chapter 2.

4.1 Definitions

A set of elements $G = \{a, b, c, d, \dots\}$ is called a *group* if there exists an operation - we call "multiplication" - between any pair of elements that satisfies the following requirements:

- (a) Closure: The multiplication of any pair of elements is an element of the group.
- (b) The existence of the identity operation: One of the set elements, denoted by E , is the identity element that satisfies the condition $aE = Ea = a$ for any a .
- (c) The presence of inverse elements: To each element, a , there is an inverse element, a^{-1} , (in the group) that satisfies the condition $aa^{-1} = a^{-1}a = E$.
- (d) Associativity: $a(bc) = (ab)c$.

Examples

- The simplest group is a group that contains a single element $\{E\}$.
- The only group that contains two elements is $C_2 = \{E, a\}$ where a is the inverse of itself, $a^2 = E$.
- There is also only one group that contains three elements $C_3 = \{E, a, b\}$. The multiplication table of this group is presented on the next page. This group can also be represented in the form $C_3 = \{E, a, a^2\}$ where $a^3 = E$.

C_3	E	a	b
E	E	a	b
a	a	b	E
b	b	E	a

D_2	E	σ	σ'	c_2
E	E	σ	σ'	c_2
σ	σ	E	c_2	σ'
σ'	σ'	c_2	E	σ
c_2	c_2	σ'	σ	E

- All three examples presented above are particular examples of the cyclic group $C_n = \{E, a, a^2 \dots a^{n-1}; a^n = E\}$, generated by repeated multiplication of a single element.
- Only two groups contain four elements: The cyclic group C_4 , and the dihedral group D_2 (which is the simplest non-cyclic group). The latter is the symmetry group of a rectangle, see Fig. 4-1. Its symmetry elements are: rotation in 180° , c_2 ; reflection through the vertical axis, σ ; reflection through the horizontal axis σ' ; and the identity operator, E . The multiplication table of this group is shown above.

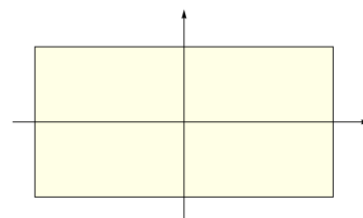


Figure 4-1 The dihedral group D_2 as the symmetry group of a rectangle

Definition: A group is called *abelian* if the multiplication of any pair of elements in the group is commutative $ab = ba$.

All groups that have been presented so far are abelian. This property is manifested by a symmetric matrix structure of the multiplication table.

Definition: The *group order* is the number of elements in the group.

- The simplest non-abelian group (i.e., the non-abelian group of lowest order) contains six elements. It is the group associated with the symmetry of an equilateral triangle, see Fig. 4-2. Its elements are: the identity E ; two rotations by $\pm 120^\circ$, c_3 and c_3^2 ; and three reflections

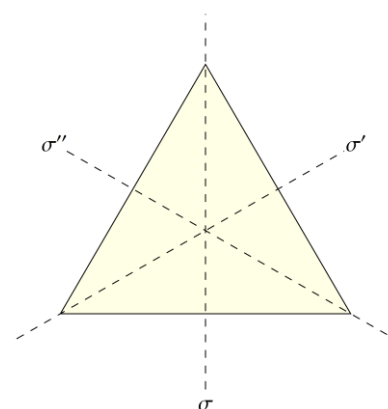


Figure 4-2 The equilateral triangle whose symmetry group is C_{3v}

σ , σ' and σ'' through the axes shown in the figure. This group is denoted by C_{3v} , and its multiplication table is given below¹.

C_{3v}	E	σ	σ'	σ''	c_3	c_3^2
E	E	σ	σ'	σ''	c_3	c_3^2
σ	σ	E	c_3	c_3^2	σ'	σ''
σ'	σ'	c_3^2	E	c_3	σ''	σ
σ''	σ''	c_3	c_3^2	E	σ	σ'
c_3	c_3	σ''	σ	σ'	c_3^2	E
c_3^2	c_3^2	σ'	σ''	σ	E	c_3

This table is not symmetric; hence some of the products are not commutative. An example of a non-commutative product is illustrated in Fig. 4-3.

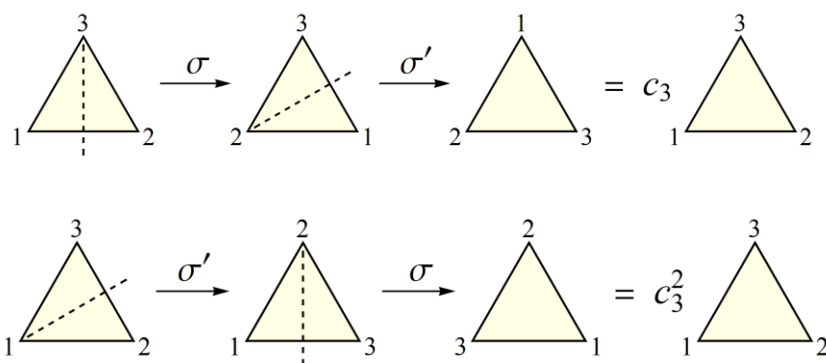


Figure 4-3 An illustration of a non-commutative product of reflections in the group C_{3v}

The group C_{3v} has the same multiplication table as the dihedral group D_3 . The latter is obtained when replacing the reflection axes in Fig 4-2 by rotations axes in 180° . Two point groups are said to be *isomorphic* if their multiplication tables are identical (up to reordering of rows or columns).

The rearrangement theorem: Each row and each column in the group multiplication table lists each of the group elements once and only once.

¹ The product order convention for group multiplication tables is that the element, ab , which appears in a given cell, is the product (from the left) of the element a that appears on the leftmost cell of the same row, by the element b that appears on the top cell of the same column.

4.2 Symmetry operations of point groups

Point groups are groups whose symmetry operations (acting on a set of points) do not affect (at least) one point. The set of symmetry operations and their symbols are listed below:

E - the identity operation that leaves the system unchanged.

c_n - rotation in $360^\circ/n$ around symmetry axis.

i - inversion, the operation $\mathbf{r} \rightarrow -\mathbf{r}$.

σ_h - reflection through a plane perpendicular to the principal symmetry axis (of rotation).

σ_v - reflection through a plane that contains the principal symmetry axis.

σ_d - reflection through a diagonal plane that contains the principal symmetry axis. This plane also contains the bisector between two secondary rotation axes, c_2 , which is perpendicular to the principal axis (see Fig. 4-8 below).

S_n - improper rotation. Rotation in $360^\circ/n$ around the principal symmetry axis followed by reflection through a plane perpendicular to this axis (see Fig. 2-26 and the explanation on the same page).

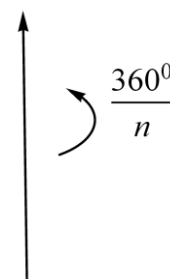


Figure 4-4 The rotation axis associated with cyclic groups

4.3 Schoenflies notation of point groups

Point groups are denoted by letters and subscripts. Below we list the main notations of point groups.

C_n - (The cyclic group) describes systems with a single symmetry axis around which the system is symmetric to rotations in $360^\circ/n$ as shown in Fig. 4-4. This group contains n elements.

C_{nv} - This group contains the following elements: n -fold rotation around a single axis (as in the cyclic group), and n reflections through planes that contain this axis. The angle between nearby planes is $180^\circ/n$ as demonstrated in Fig. 4-5 for C_{2v} . This group includes $2n$ elements.

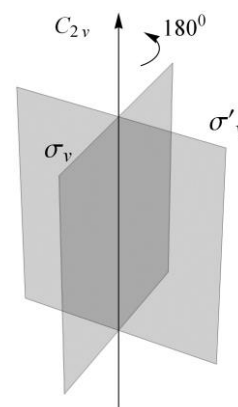


Figure 4-5 The symmetry operations of C_{2v}

C_{nh} - This group's elements are obtained from n -fold rotations and reflection through a plane perpendicular to the rotation axis - see Fig. 4-6. The group has $2n$ elements.

D_n - (The dihedral group) The symmetry associated with this group is that of n -fold rotation around the principle symmetry axis and additional n twofold rotations around secondary axes perpendicular to the principal axis. The angle between nearby secondary axes is $180^\circ/n$. This group contains $2n$ elements. An illustration showing the symmetry operations of D_2 is depicted in Fig. 4-7.

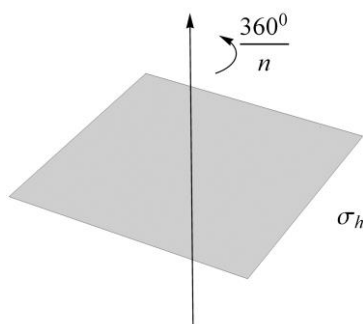


Figure 4-6 The symmetry operations of C_{nh}

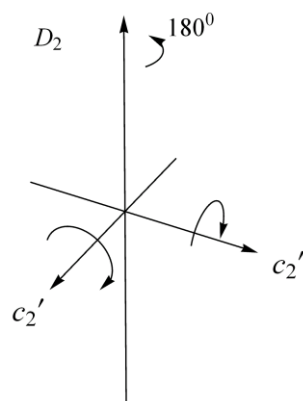


Figure 4-7 The symmetry operations of D_2

D_{nd} - This group contains all symmetry operations of D_n plus reflections through n planes that contain the principal symmetry axis and one of the bisectors between nearby secondary axes. An illustration of these symmetry operations for D_{2d} is shown in Fig. 4-8 (on the next page). The group contains $4n$ elements.

D_{nh} - This group contains all symmetry operations of D_n plus reflection through a plane perpendicular to the principal symmetry axis. This reflection symmetry implies there must be additional n reflection planes. These planes contain the principal symmetry axis and one of the secondary axes, as illustrated in Fig. 4-9 (on the next page) for D_{2h} . The group has $4n$ elements.

T_d - The tetrahedral group is the symmetry group of a regular tetrahedron, see Fig. 2-40. It contains 24 symmetry operations: the identity operator; eight rotations in 120° ; three rotations in 180° ; six improper rotations S_4 ; and six reflections.

O_h - The octahedral group is the symmetry group of a cube (as well as an octahedron). It contains 48 elements that include the identity; eight rotations in 120° ; six rotations in 90° ; nine rotations in 180° ; inversion; six improper rotations S_4 ; eight improper rotations S_6 ; and nine reflections.

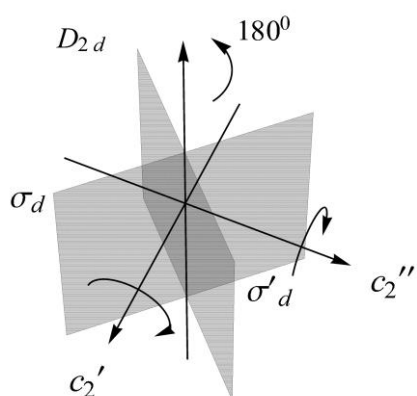


Figure 4-8 The symmetry operations of D_{2d}

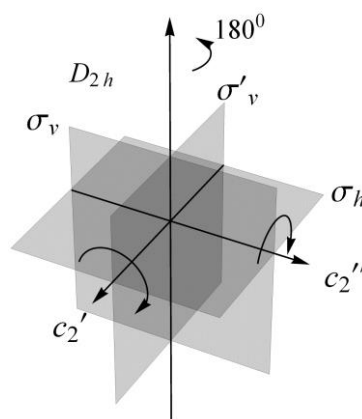


Figure 4-9 The symmetry operations of D_{2h}

Example

Let us identify the point groups that describe the symmetry of the following molecules:

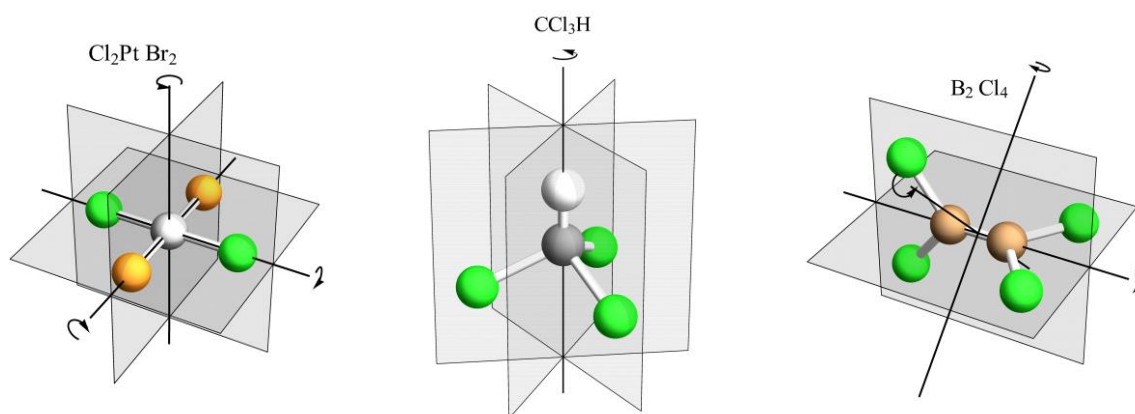


Figure 4-10 The symmetry operations of a few molecules

Consider, first, the molecule Cl_2PtBr_2 that appears on the left panel of the figure. It has two-fold rotation symmetry around an axis perpendicular to the molecule's plane and two additional twofold rotations around axes that pass through identical atoms. It also has three reflection symmetries through the planes shown in the figure. The group associated with these symmetry operations is, therefore, D_{2h} (see Fig. 4-9).

Let's Look, now, at the molecule CCl_3H shown in the middle panel of the figure. Here there is an axis of three-fold rotations and three reflection planes that contain the principal symmetry axis and one of the chlorine atoms. The point group is, therefore, C_{3v} .

Finally, consider the molecule B_2Cl_4 shown on the right panel of the figure. Here, one can quickly identify the two-fold rotation symmetry around the principal axis passing through the Boron atoms. Two reflection planes contain the principal axis and pairs of chlorine atoms. Finally, it is a bit more challenging to see the two additional two-fold rotations. These rotations are around axes perpendicular to the principal symmetry axis and parallel to the bisectors of the reflection planes. Thus, the point group associated with B_2Cl_4 is D_{2d} .

4.4 Conjugate elements and conjugacy classes

Definition: An element $b \in G$ is said to be *conjugate* to $a \in G$, if there is an element p in the group such that:

$$b = pap^{-1} \quad (4.1)$$

Conjugacy relation between two elements, a , and b , is customarily denoted by $a \sim b$. It is easy to prove the following properties:

- Each element is conjugate to itself, $a \sim a$.
- If $a \sim b$, then $b \sim a$.
- If $a \sim b$ and $b \sim c$, then $a \sim c$.

Definition: All the elements of a group that are conjugate to each other form a *conjugacy class* of the group.

It is easy to prove that:

- Each element of the group belongs to one and only one conjugacy class.
- The identity operation has its own conjugacy class (that contains a single element).

From a physical viewpoint, conjugate elements are associated with similar symmetry operations. For instance, in a system with four-fold rotation symmetry around the z axis, reflections through the yz plane and the xz plane are of the same nature. It is because reflection through the yz plane can be obtained by rotating the system by -90° around the z axis, reflecting through the xz plane, and finally rotating back by 90° around the same axis. This set of operations is precisely that of Eq. (4.1), where p is the rotation operation, while a and b are the two reflections.

Example: The conjugacy classes of C_{2v}

This point group (see group multiplication table on page 30) is abelian; therefore, each element creates its own conjugacy class, because $x = yxy^{-1}$ where y can be any element of the group. Looking at Fig. 4-5, it is clear that all symmetry operations are of different nature (there is no 90° rotation that allows conjugation of the two reflection symmetries). A diagram of the conjugacy classes of this group is shown in Fig. 4-11.

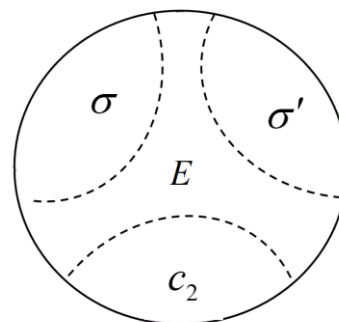


Figure 4-11 The conjugacy classes of C_{2v}

Example: The conjugacy classes of C_{3v}

This is a nonabelian group. Here one may expect the conjugacy classes to contain more than one element. In particular, the three reflection operations are of the same nature because the rotation operation conjugates them; see the middle panel of Fig. 4-10. For instance, from the multiplication table presented on page 59, one sees that $\sigma' = c_3\sigma c_3^2$. Similarly, the rotation operations are conjugated by reflections, $c_3 = \sigma c_3^2 \sigma$. The conjugacy classes of this group are shown in Fig. 4-12.

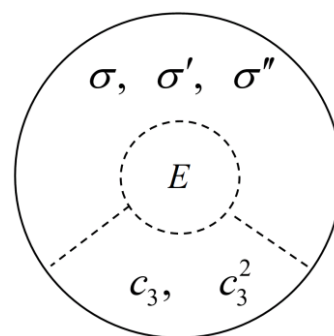


Figure 4-12 The conjugacy classes of C_{3v}

Example: The conjugacy classes of D_4

The dihedral group D_4 contains eight symmetry operations: The identity, four-fold rotations, and four two-fold rotations around axes perpendicular to the principal axis, as illustrated in Fig. 4-13. From this figure, one quickly sees that c_2 and c_2'' are conjugated by 90° rotation, and so are c_2' and c_2''' . However, these two conjugacy classes are different because there is no 45° rotation to conjugate them. c_4 and c_4^3 are also conjugated because $c_4^3 = c_2' c_4 c_2'$. On the other hand, c_4^2 conjugates only to itself, because $c_4^2 = p^{-1} c_4^2 p$, where $p = c_2, c_2', c_2'',$ or c_2''' . Thus, there are four conjugacy classes, as shown in Fig. 4-14 on the next page.

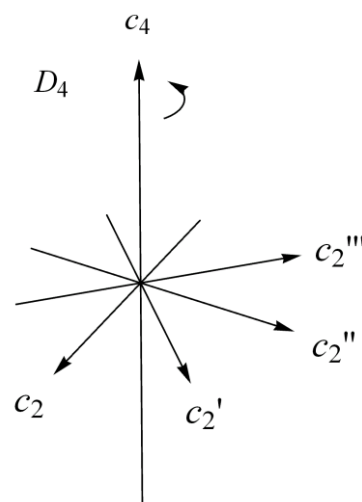


Figure 4-13 The symmetry operations of D_4

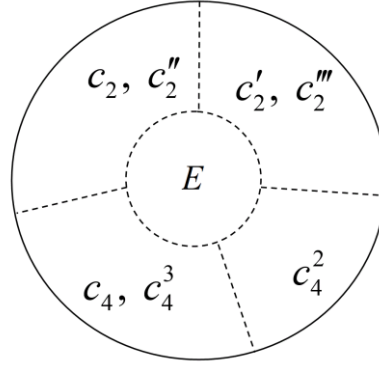


Figure 4-14 The conjugacy classes of D_4

4.5 Representations of groups

A *group representation* is a group of mathematical objects that describe the symmetry operations of the group. Here we shall confine our discussion only to those cases where these mathematical objects are square invertible matrices, and multiplication of group operations corresponds to matrix multiplication. Thus, the group element a , will be represented by a matrix $\Gamma(a)$, such that for any two elements of the group:

$$\Gamma(a)\Gamma(b) = \Gamma(ab). \quad (4.2)$$

The size of the matrix is called the *dimension of the representation*.

Example: C_2

A one-dimensional representation of this group can be, for example, $\Gamma(E) = \Gamma(a) = 1$, because it trivially satisfies the group multiplication table that appears on the right. Another one-dimensional representation is $\Gamma(E) = 1$ and $\Gamma(a) = -1$. As one can easily see, it satisfies the multiplication table.

C_2	E	a
E	E	a
a	a	E

A two-dimensional representation of the group can be constructed from these one-dimensional representations. For example:

$$\Gamma(E) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ and } \Gamma(a) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (4.3)$$

One can also complicate this representation by rotating the matrices. The identity matrix is left unchanged by rotation:

$$\Gamma(E) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (4.4)$$

but the matrix associated with the a element takes a different form:

$$\Gamma(a) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} \cos^2 \theta - \sin^2 \theta & -2 \sin \theta \cos \theta \\ -2 \sin \theta \cos \theta & \sin^2 \theta - \cos^2 \theta \end{pmatrix}. \quad (4.5)$$

It is clear, by construction, that the above matrices satisfy the group multiplication table.

This example shows that there is an infinite number of representations for any group with more than one element (in the above example, the rotation angle, θ , can be arbitrary).

Definition: Two group representations, Γ and Γ' , are said to be *equivalent representations* if they are related by similarity transformation:

$$\Gamma' = S \Gamma S^{-1}, \quad (4.6)$$

where S is some matrix (but the same one for all group elements).

Definition: A group representation is *faithful* if each group element is represented by a different matrix.

Example: C_4

Let us construct representations of the cyclic group $C_4 = \{E, a, a^2, a^3\}$ with $a^4 = E$. A trivial representation is the identity representation where all elements are represented by 1, $\Gamma(e) = \Gamma(a) = \Gamma(a^2) = \Gamma(a^3) = 1$. Another possibility for one-dimensional representation is:

$$\Gamma(a^n) = \exp\left(i \frac{2\pi}{4} n\right), \text{ where we define } a^0 = E. \quad (4.7)$$

More generally, we can choose representation in the form $\Gamma(a^n) = \exp(i 2\pi mn/4)$ with $m=0,1,2,3$ (larger values of m give the same representations). The identity representation is associated with $m=0$. This representation, as well as the $m=2$ representation, are not faithful. The other two are faithful representations.

Example: C_{2v}

The symmetry operations of this group are illustrated in Fig. 4-5. One way of constructing representation is to identify the symmetry operation with the matrices that transform a general vector in space:

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (4.8)$$

Choosing σ to be a reflection through the xz plane; σ' to be a reflection through the yz plane; and c_2 a rotation around the z axis, we obtain the following matrices:

The identity operator:
$$\Gamma(E) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (4.9)$$

A reflection through xz plane:
$$\Gamma(\sigma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (4.10)$$

A reflection through yz plane:
$$\Gamma(\sigma') = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (4.11)$$

A Rotation in 180° around the z axis:
$$\Gamma(c_2) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.12)$$

More generally, one can construct group representations using matrices that act on a set of coordinates, such as the coordinates of the atoms in a molecule with the corresponding symmetry. In the case of C_{2v} point group, such a molecule is, for example, the water molecule shown in Fig. 4-15.

Here, reflection through the plane perpendicular to the molecule's plane (the reflection operation, σ') transforms the coordinates of the atoms in the following manner:

$$\begin{aligned} x_1 &\rightarrow -x_1, \quad x_2 \rightarrow -x_3, \quad x_3 \rightarrow -x_2, \\ y_2 &\leftrightarrow y_3, \quad z_2 \leftrightarrow z_3, \end{aligned} \quad (4.13)$$

while y_1 and z_1 are left unchanged. This transformation can be expressed as a matrix 9×9 multiplying the vector built from the coordinates of the atoms:

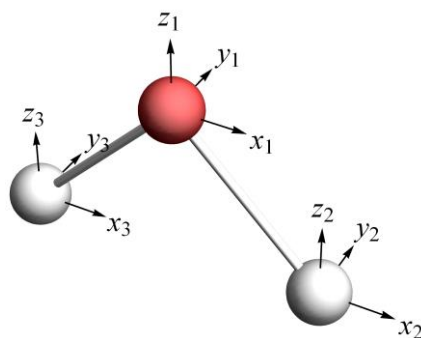


Figure 4-15 The coordinates of the atoms of water molecule as a basis for representation of the C_{2v} group

$$\begin{pmatrix}
 -1 & 0 & 0 & & & \\
 0 & 1 & 0 & & 0 & \\
 0 & 0 & 1 & & & \\
 & & & -1 & 0 & 0 \\
 & 0 & & 0 & 1 & 0 \\
 & & & 0 & 0 & 1 \\
 & & -1 & 0 & 0 & \\
 0 & 0 & 1 & 0 & & 0 \\
 & 0 & 0 & 1 & &
 \end{pmatrix}
 \begin{pmatrix}
 x_1 \\
 y_1 \\
 z_1 \\
 x_2 \\
 y_2 \\
 z_2 \\
 x_3 \\
 y_3 \\
 z_3
 \end{pmatrix}
 =
 \begin{pmatrix}
 -x_1 \\
 y_1 \\
 z_1 \\
 -x_3 \\
 y_3 \\
 z_3 \\
 -x_2 \\
 y_2 \\
 z_2
 \end{pmatrix} . \quad (4.14)$$

The above matrix represents the reflection operation σ' . Similar matrices can be constructed for the other symmetry elements of the group. The set of these matrices is a representation of the group.

It is evident that one can build representations with dimensions as high as desired. However, in general, it will be possible to decompose them into smaller representations using similarity transform, $\Gamma' = S\Gamma S^{-1}$, that brings each matrix into a block diagonal form. Representations for which such a process can be executed are called *reducible representations* since each block in the matrix constitutes a representation by itself.

Definition: Group representations of the lowest possible dimension, i.e. representations that cannot be further reduced into block diagonal form, are called *irreducible representations*.

The great orthogonality theorem of irreducible representations

Let $\Gamma_{mn}^{(\alpha)}(g)$ be the matrix element of an irreducible representation, α , corresponding to the group element g . We shall denote the dimension of this representation by ℓ_α , and the order of the group (i.e. the number of elements in the group) by $|G|$. The *great orthogonality theorem of irreducible representations* states that

$$\sum_{g \in G} \Gamma_{mn}^{(\alpha)*}(g) \Gamma_{m'n'}^{(\beta)}(g) = \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta} \delta_{mm'} \delta_{nn'} , \quad (4.15)$$

where the sum is over all the elements in the group G . The interpretation of this equation is that the matrix elements of irreducible representations behave like orthogonal vectors in a $|G|$ dimensional space.

Corollary: From the orthogonality theorem of irreducible representations it follows that the number of irreducible representations of a group cannot exceed $|G|$. It is simply because in N dimensional space there are, at most, N orthogonal vectors.

Example: C_2

On page 65, we have presented two one-dimensional irreducible representations of the group C_2 . Since the group order is 2, these representations are the only possible irreducible representations of the group:

$$\Gamma^{(1)}(E) = \Gamma^{(1)}(a) = 1 \quad \text{and} \quad \Gamma^{(2)}(E) = 1, \quad \Gamma^{(2)}(a) = -1. \quad (4.16)$$

Let us illustrate the orthogonality theorem using this example. For the same representations ($\alpha = \beta$ in Eq. (4.15)) we obtain:

$$\Gamma^{(\alpha)*}(E)\Gamma^{(\alpha)}(E) + \Gamma^{(\alpha)*}(a)\Gamma^{(\alpha)}(a) = 2 \quad \text{for } \alpha = 1, 2, \quad (4.17)$$

while for different representations, $\alpha \neq \beta$, we see that

$$\Gamma^{(1)*}(E)\Gamma^{(2)}(E) + \Gamma^{(1)*}(a)\Gamma^{(2)}(a) = 1 \cdot 1 + 1 \cdot (-1) = 0. \quad (4.18)$$

Example: C_{3v}

The irreducible representations of C_{3v} (the symmetry group of an equilateral triangle) are listed in the following table:

C_{3v}	E	σ	σ'	σ''	c_3	c_3^2
$\Gamma^{(1)}$	1	1	1	1	1	1
$\Gamma^{(2)}$	1	-1	-1	-1	1	1
$\Gamma^{(3)}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$

From here, we see that the group has three irreducible representations, two one-dimensional representations, and one representation that is two-dimensional. To illustrate the orthogonality theorem, we rewrite this table such that elements in each row correspond to one of the matrix elements of the representations:

C_{3v}	E	σ	σ'	σ''	C_3	C_3^2
$\Gamma^{(1)}$	1	1	1	1	1	1
$\Gamma^{(2)}$	1	-1	-1	-1	1	1
$\Gamma_{11}^{(3)}$	1	-1	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$
$\Gamma_{12}^{(3)}$	0	0	$-\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$
$\Gamma_{21}^{(3)}$	0	0	$-\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$
$\Gamma_{22}^{(3)}$	1	1	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$

Each row in this table is a six-dimensional vector. According to the great orthogonality theorem of the irreducible representations, these vectors are orthogonal. For example, the five lower rows are orthogonal to the first one because the sum of their elements vanishes. Another example of the orthogonality property is of the second and third rows:

$$\Gamma_{11}^{(3)*} \cdot \Gamma^{(2)} = 1 \cdot 1 + (-1) \cdot (-1) + \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot (-1) - \frac{1}{2} \cdot 1 - \frac{1}{2} \cdot 1 = 0. \quad (4.19)$$

Similarly, one can check that all vectors are orthogonal to each other. The norms of the one-dimensional representations also satisfy the condition set by the theorem:

$$\Gamma^{(1)*} \cdot \Gamma^{(1)} = \Gamma^{(2)*} \cdot \Gamma^{(2)} = 6 = |G|, \quad (4.19)$$

and so are the norms of the vectors of the two-dimensional representation:

$$\Gamma_{11}^{(3)*} \cdot \Gamma_{11}^{(3)} = \Gamma_{12}^{(3)*} \cdot \Gamma_{12}^{(3)} = \Gamma_{21}^{(3)*} \cdot \Gamma_{21}^{(3)} = \Gamma_{22}^{(3)*} \cdot \Gamma_{22}^{(3)} = 3 = \frac{6}{2} = \frac{|G|}{\ell_3}. \quad (4.20)$$

The examples that we brought here illustrate the following theorem:

Theorem: The sum of squares of the dimensions of the irreducible representations equals the group order:

$$\sum_{\alpha} \ell_{\alpha}^2 = |G|. \quad (4.21)$$

In many situations, this equation is sufficient for determining the dimensions of the irreducible representations of a group. We illustrate this with two examples.

Example: The dimensions of the irreducible representations of D_2

The dihedral group, D_2 , contains four elements. Since any group has the identity representation (where all elements are represented by one), Eq. (4.21) reduces to

$$1 + \sum'_{\alpha} \ell_{\alpha}^2 = 4, \quad (4.22)$$

where Σ' denotes a sum over the irreducible representations that do not include the identity representation. The only way to satisfy the above equation is by choosing all representations to be one-dimensional, so that $1^2 + 1^2 + 1^2 + 1^2 = 4$.

Example: The dimensions of the irreducible representations of D_3

The dihedral group D_3 contains six elements. In this case, Eq. (4.21) reads

$$1 + \sum'_{\alpha} \ell_{\alpha}^2 = 6. \quad (4.23)$$

This equation has two solutions. One is when all irreducible representations are one-dimensional. However, the only group of order six that has six one-dimensional representations is the cyclic group C_6 . This group is not isomorphic to D_3 ; hence one should look for another solution of (4.23). There is only one additional solution which is $1^2 + 1^2 + 2^2 = 6$. Thus D_3 has two one-dimensional representations and one irreducible representation, which is two-dimensional.

4.6 Characters of irreducible representations

As we have seen, there is an arbitrariness in the choice of irreducible representation having a dimension larger than one. Any two representations related by a similarity transformation, $\Gamma' = S\Gamma S^{-1}$, are equivalent. It is desirable to develop tools that are free of this problem. The most natural candidate is the trace of the matrix associated with the representation because traces are invariant under similarity transform:

$$\text{Tr}(\Gamma') = \text{Tr}(S\Gamma S^{-1}) = \text{Tr}(S^{-1}S\Gamma) = \text{Tr}(\Gamma). \quad (4.24)$$

Accordingly, the *character of an irreducible representation, α* , of the group element g is defined to be:

$$\chi^{(\alpha)}(g) = \text{Tr}[\Gamma^{(\alpha)}(g)] = \sum_{m=1}^{\ell_{\alpha}} \Gamma_{mm}^{(\alpha)}(g). \quad (4.25)$$

From this definition it follows that all the elements in the same conjugacy class have the same character. It is because if $a \sim b$, then $b = pap^{-1}$ and the representations of a and b

are related by the similarity transformation, $\Gamma(b) = \Gamma(p)\Gamma(a)\Gamma^{-1}(p)$; hence $\text{Tr}\Gamma(b) = \text{Tr}\Gamma(a)$.

Let us show that the characters of the irreducible representations behave like orthogonal vectors:

$$\sum_{g \in G} \chi^{(\alpha)*}(g) \chi^{(\beta)}(g) = |G| \delta_{\alpha\beta}. \quad (4.26)$$

This property follows directly from the orthogonality theorem of irreducible representations because from Eq. (4.15) we have

$$\sum_{g \in G} \Gamma_{mm}^{(\alpha)*}(g) \Gamma_{m'm'}^{(\beta)}(g) = \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta} \delta_{mm'}, \quad (4.27)$$

and summing over m' gives

$$\sum_{g \in G} \Gamma_{mm}^{(\alpha)*}(g) \chi^{(\beta)}(g) = \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta}. \quad (4.28)$$

Finally, summing over m leads to Eq. (4.26).

Eq. (4.26) can be rewritten as a sum over conjugacy classes because characters of elements in the same conjugacy class are equal. Thus

$$\sum_{k \in \text{classes}} N_k \chi^{(\alpha)*}(g_k) \chi^{(\beta)}(g_k) = |G| \delta_{\alpha\beta}, \quad (4.29)$$

where N_k is the number of elements in the k -th conjugacy class, and g_k is a representative element of that class. This equation implies that the characters of irreducible representations form an orthogonal set of vectors in the space of conjugacy classes. Therefore, the number of irreducible representations must be smaller or equal to the number of conjugacy classes. One can prove that it precisely equals the number of conjugacy classes, but we skip this proof.

The conclusion from this discussion is that the irreducible representations of a group satisfy the following properties:

(a) The number of irreducible representations equals the number of conjugacy classes.

(b) $\sum_{\alpha} \ell_{\alpha}^2 = |G|.$

In many cases, these properties allow us to uniquely determine the dimensions of the irreducible representations of the group.

Example: The dimensions of the irreducible representations of D_4

We have seen that the dihedral group D_4 contains eight elements and five conjugacy classes, see Fig. 4-14. This information, together with the properties mentioned above, implies that

$$1^2 + \ell_1^2 + \ell_2^2 + \ell_3^2 + \ell_4^2 = 8, \quad (4.30)$$

where we took into account that any group has the one-dimensional identity representation. From this equation it is evident that D_4 cannot have a three-dimensional irreducible representation. One can neither choose all its representations to be one-dimensional. The only possible way of satisfying Eq. (4.30) is by:

$$1^2 + 1^2 + 1^2 + 1^2 + 2^2 = 8 \quad (4.31)$$

Hence, D_4 has a single two-dimensional irreducible representation and four one-dimensional representations.

4.7 Character tables

One of the main tools for implementing group theory in physics is the character tables. A character table is a list of all characters associated with the irreducible representations of a group. An example of such a table, for the C_{3v} group, is shown here:

C_{3v}	E	3σ	$2c_3$
A_1	1	1	1
A_2	1	-1	1
E	2	0	-1

In this table, the columns and the rows classify the group's conjugacy classes and irreducible representations, respectively. The top row lists the conjugacy classes (with the number of elements in each class), while the leftmost column lists the Mulliken symbols (to be explained later) of the irreducible representations. The number in each cell is the character associated with an element in the conjugacy class of the corresponding irreducible representation. The table is a square matrix because the number of conjugacy classes equals the number of irreducible representations.

The character tables of all point groups can be found in the literature (or the internet). Yet, in many cases, they can be quickly constructed using the following rules (that follow from the above discussion):

- (1) The size of the table is the number of the conjugacy classes of the group.
- (2) The dimensions of the irreducible representations are constrained by $\sum_{\alpha} \ell_{\alpha}^2 = |G|$.
- (3) The rows of the table are orthogonal to each other.
- (4) The sum of squares of (the absolute value of) all characters in a row equals $|G|$.
- (5) All character tables contain the identity representation.
- (6) An additional property that we present here without proof is that the columns of the table are also orthogonal to each other; namely, they satisfy the equation:

$$\sum_{\alpha} \chi^{(\alpha)*}(g_k) \chi^{(\alpha)}(g_n) = \frac{|G|}{N_k} \delta_{kn}, \quad (4.32)$$

where N_k is the number of elements in the k -th conjugacy class.

- (7) From Eq. (4.32), and the fact that the identity element of the group forms a conjugacy class of its own, it follows that the sum of squares of the characters of the identity operator equals the group order $\sum_{\alpha} |\chi^{(\alpha)}(E)|^2 = |G|$.

Example: Construction of the character table of D_4

As we already know, D_4 has five conjugacy classes and five irreducible representations: one two-dimensional representation and the rest are one-dimensional. The first row of the table corresponds to the identity representation where all entries are one. The first column of the table, corresponding to the identity operation, is also easy to fill because the identity operation is represented by identity matrices, whose traces equal the representation's dimension. Thus the first row and the first column of the table are known.

D_4	E	$2c_4$	c_4^2	$2c_2$	$2c'_2$
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
B_1	1	-1	1	1	-1
B_2	1	-1	1	-1	1
E	2	0	-2	0	0

Let us now identify the entries of the cells in the red rectangle of the above table. These are obtained from the orthogonality of the rows (recall that one should take into account the number of elements in each conjugacy class). Here, the entries are either $+1$ or -1 , because it is the only possibility to satisfy the orthogonality condition given that c_4^2 and E form conjugacy classes of their own, while all other conjugacy classes contain two elements. Finally, to determine the characters of the two-dimensional representation (i.e., the last row of the above table), we use the property that all columns are orthogonal to the first one, see Eq. (4.32).

4.8 Basis functions

Character tables are usually supplemented by additional columns that list functions associated with each one of the irreducible representations. The functions are called *basis functions*, and in this section, we explain their meaning and show how to construct them.

The basis functions associated with an irreducible representation α of dimension ℓ_α are defined as the sets of functions, $\mathbf{f}^{(\alpha)} = (f_1^{(\alpha)}, f_2^{(\alpha)}, \dots, f_{\ell_\alpha}^{(\alpha)})$, that satisfy the condition:

$$g\mathbf{f}^{(\alpha)} = \Gamma^{(\alpha)}(g)\mathbf{f}^{(\alpha)}, \text{ or in components } gf_j^{(\alpha)} = \sum_{i=1}^{\ell_\alpha} \Gamma_{ji}^{(\alpha)}(g) f_i^{(\alpha)}. \quad (4.33)$$

Namely, the behavior of these functions under the symmetry operations of the group reflects the nature of the irreducible representations. It is instructive to start with a simple example. Consider the group C_{2v} whose all irreducible representations are one-dimensional, and therefore characters are the representation matrices. The character table of this group, together with the basis functions, is given below.

C_{2v}	E	c_2	σ_v	σ'_v		
A_1	1	1	1	1	z	x^2, y^2, z^2
A_2	1	1	-1	-1	R_z	xy
B_1	1	-1	1	-1	x, R_y	xz
B_2	1	-1	-1	1	y, R_x	yz

Here, the rotation is around the z axis, while the reflections, σ_v and σ'_v , are through the molecule's plane (xz -plane) and the perpendicular plane (yz -plane), respectively.

The second column from the right shows linear functions of the position vector. Observe first the coordinate z . It is not affected by any symmetry operation, and therefore reflects

the nature of the identity representation, A_1 . On the other hand, the coordinate x changes sign under the action of c_2 and σ'_v , while unaffected by σ_v . It is associated with the B_1 irreducible representation because the characters of both c_2 and σ'_v are -1 while that of σ_v is $+1$. Similarly, one can check that the y coordinate is a basis function of the B_2 irreducible representation.

A rotation around the z axis, denoted by R_z , is unaffected by rotation around the same axis but reverses direction when reflected through any plane containing the z axis. Hence R_z transforms as the A_2 representation. Rotation around the y axis, R_y , changes sign when rotated around the z axis and by reflection through the yz plane. However, the rotation direction is not reversed by reflection through the xz plane, as demonstrated in Fig. 4-16. Thus R_y is a basis function of the B_1 representation. Similar considerations show that R_x is a basis function of B_2 .

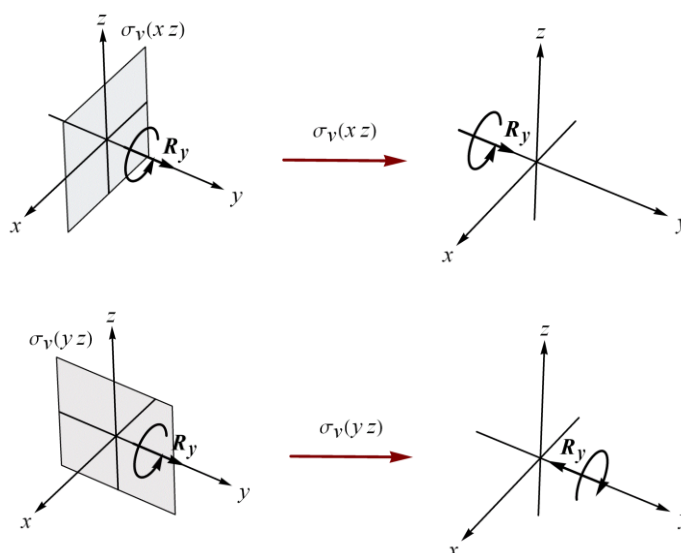


Figure 4-16 The action of reflection through a plane on the rotation basis function

The quadratic basis functions of C_{2v} are listed in the rightmost column of the character table shown in the previous page. The basis function xy belongs to the A_2 representation because the c_2 rotation transforms $x \rightarrow -x$ and $y \rightarrow -y$, and leaves their product unchanged. On the other hand, reflections change the sign of only one of these coordinates, hence reversing the sign of the product xy .

As a second example, consider the group C_{3v} whose character table that includes the basis function is given on the next page. Here, the rightmost column lists the cubic basis functions. Notice, also, that each basis function of the two-dimensional representation contains two components. This is because two-dimensional representations are 2×2 matrices that act on vectors (or spinors).

C_{3v}	E	$2C_3$	3σ			
A_1	1	1	1	z	$x^2 + y^2; z^2$	$z^3; \text{Re}(x + iy)^3; z(x^2 + y^2)$
A_2	1	1	-1	R_z		$\text{Im}(x + iy)^3$
E	2	-1	0	(x, y) (R_x, R_y)	$(x^2 - y^2, 2xy)$ (xz, yz)	$(xz^2, yz^2); (2xyz, zx^2 - zy^2)$ $(x^3 + xy^2, y^3 + yx^2)$

How to construct basis functions of irreducible representations?

The idea is to project some general functions down to the subspace of functions that belong to the irreducible representation. To execute this program, one should first identify the projection operator associated with a given irreducible representation.

Starting from definition (4.33) of the basis function, we multiply this equation from the left by $\Gamma_{kl}^{(\beta)*}(g)$. Then summing over the group elements leads to:

$$\begin{aligned}
 \sum_g \Gamma_{kl}^{(\beta)*}(g) g f_j^{(\alpha)} &= \sum_{i=1}^{\ell_\alpha} \sum_g \Gamma_{kl}^{(\beta)*}(g) \Gamma_{ji}^{(\alpha)}(g) f_i^{(\alpha)} \\
 &= \sum_{i=1}^{\ell_\alpha} \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta} \delta_{kj} \delta_{li} f_i^{(\alpha)} = \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta} \delta_{kj} f_l^{(\alpha)}.
 \end{aligned} \tag{4.34}$$

To obtain the second line in this equation, we employed the great orthogonality theorem of irreducible representations (4.15). By defining the operator

$$P_{kl}^{(\beta)} = \sum_g \Gamma_{kl}^{(\beta)*}(g) g, \tag{4.35}$$

one may rewrite Eq. (4.34) in the form

$$P_{kl}^{(\beta)} f_j^{(\alpha)} = \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta} \delta_{kj} f_l^{(\alpha)}. \tag{4.36}$$

Finally, setting $k = l$ and summing over l gives:

$$P^{(\beta)} f_j^{(\alpha)} = \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta} f_j^{(\alpha)}, \tag{4.37}$$

where

$$P^{(\beta)} = \sum_g \chi^{(\beta)*}(g) g. \tag{4.38}$$

Equation (4.37) shows that when acting with $P^{(\beta)}$ on any function that does not belong to the basis functions of the β irreducible representation, the result is zero. Therefore, $P^{(\beta)}$

$$P^{(\text{B}_1)} = E - c_4 - c_4^3 + c_2 + \sigma_v + \sigma_v' - \sigma_d - \sigma_d', \quad (4.41)$$

$$P^{(B_2)} = E - c_4 - c_4^3 + c_2 - \sigma_v - \sigma'_v + \sigma_d + \sigma'_d, \quad (4.42)$$

$$P^{(E)} = 2E - 2c_2. \quad (4.43)$$

To construct the linear basis functions, we apply the above operators on each component of the position vector, (x, y, z) , where x and y coordinates are shown in Fig. 4-17, while the z coordinate is perpendicular to the plane of the square (and therefore invariant under all group operations). Let us first list the action of the group elements on the position vector:

$$\begin{aligned} E \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} x \\ y \\ z \end{pmatrix}, & c_4 \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} -y \\ x \\ z \end{pmatrix}, & c_4^3 \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} y \\ -x \\ z \end{pmatrix}, & c_2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} -x \\ -y \\ z \end{pmatrix} \\ \sigma_v \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} -x \\ y \\ z \end{pmatrix}, & \sigma'_v \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} x \\ -y \\ z \end{pmatrix}, & \sigma_d \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} x \\ y \\ z \end{pmatrix}, & \sigma'_d \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} -y \\ -x \\ z \end{pmatrix} \end{aligned} \quad (4.44)$$

Using the above formulas, it is easy to see that projection of the position vector on the identity representation gives

$$P^{(A_1)} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 8 \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}, \quad (4.45)$$

therefore z is the basis function of the A_1 irreducible representation. Similarly, applying the projection operator of the two-dimensional representation, E , gives

$$P^{(E)} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 4 \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}. \quad (4.46)$$

Hence the pair (x, y) is a basis function of the E irreducible representation. One can check that all other projection operators nullify the vector (x, y, z) because we already have all the linear basis functions, and there cannot be additional ones.

To construct quadratic basis functions, let us define a matrix whose components include all possible combinations of products of pairs of coordinates. It is obtained from the external product of the position vector by itself:

$$Q = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \begin{pmatrix} x & y & z \end{pmatrix} = \begin{pmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{pmatrix} \quad (4.47)$$

Using Eq. (4.44) we deduce the action of the group elements on this matrix:

$$\begin{aligned}
 EQ &= \begin{pmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{pmatrix}, \quad c_4Q = \begin{pmatrix} y^2 & -yx & -yz \\ -yx & x^2 & xz \\ -yz & xz & z^2 \end{pmatrix}, \quad c_4^3Q = \begin{pmatrix} y^2 & -yx & yz \\ -yx & x^2 & -xz \\ yz & -xz & z^2 \end{pmatrix}, \\
 c_2Q &= \begin{pmatrix} x^2 & xy & -xz \\ xy & y^2 & -yz \\ -xz & -yz & z^2 \end{pmatrix}, \quad \sigma_vQ = \begin{pmatrix} x^2 & -xy & -xz \\ -xy & y^2 & yz \\ -xz & yz & z^2 \end{pmatrix}, \quad \sigma'_vQ = \begin{pmatrix} x^2 & -xy & xz \\ -xy & y^2 & -yz \\ xz & -yz & z^2 \end{pmatrix}, \quad (4.48) \\
 \sigma_dQ &= \begin{pmatrix} y^2 & xy & yz \\ xy & x^2 & xz \\ yz & xz & z^2 \end{pmatrix}, \quad \sigma'_dQ = \begin{pmatrix} y^2 & xy & -yz \\ xy & x^2 & -xz \\ -yz & -xz & z^2 \end{pmatrix}.
 \end{aligned}$$

With the help of these formulas, we can calculate the projection of Q on irreducible representations. In particular, projection on the identity representation gives:

$$P^{(A_1)}Q = 4 \begin{pmatrix} x^2 + y^2 & 0 & 0 \\ 0 & x^2 + y^2 & 0 \\ 0 & 0 & 2z^2 \end{pmatrix}, \quad (4.49)$$

therefore $x^2 + y^2$ and z^2 are quadratic basis functions of A_1 . Similarly,

$$P^{(B_1)}Q = 4 \begin{pmatrix} x^2 - y^2 & 0 & 0 \\ 0 & x^2 - y^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (4.50)$$

hence $x^2 - y^2$ is the quadratic basis function of the B_1 representation. Projection on B_2 gives

$$P^{(B_2)}Q = 8 \begin{pmatrix} 0 & xy & 0 \\ xy & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (4.51)$$

thus xy is the corresponding basis function. Finally,

$$P^{(E)}Q = 4 \begin{pmatrix} 0 & 0 & xz \\ 0 & 0 & yz \\ xz & yz & 0 \end{pmatrix}, \quad (4.52)$$

Hence $(zx, zy) = z(x, y)$ is the quadratic basis function of the two-dimensional irreducible representation E. One can also check that $P^{(A_2)}Q = 0$; therefore, the A_2 irreducible representation does not have quadratic basis functions.

4.9 Mulliken symbols of irreducible representations

The symbols on the leftmost column of a character table denote the irreducible representations of the group. These symbols are called Mulliken symbols, and now we turn to explain them.

1. The dimensionality of the representation is denoted by a capital letter in the following manner:

Dimension of the irreducible representation	Mulliken symbol
1	A or B
2	E
3	F or T
4	G
5	H

2. Choosing between the A and the B letters (of the one-dimensional representations) depends on the sign of the character of the n-fold rotation, c_n , around the principal axis:

$\chi(c_n)$	Mulliken symbol
+1	A
-1	B

3. Lower indices of the Mulliken symbols represent classification according to a possible sign change of the basis functions of the irreducible representation, ψ , under the following symmetry operations: If there is a c_2 axis perpendicular to the principal axis or a reflection through a plane that contains the principal axis, then the lower index is determined by the following rule:

ψ	Index
No sign change	1
A sign change	2

4. An additional classification is provided when the symmetry group contains an inversion symmetry. Depending on the character of the inversion operation, the Mulliken symbol is supplemented by a lower index:

$\chi(i)$	Index
+1	<i>g</i>
-1	<i>u</i>

5. Prime or double prime are added to the Mulliken symbol depending on the character of the reflection operator through a plane perpendicular to the principal rotation axis, σ_h :

$\chi(\sigma_h)$	Primes
+1	'
-1	"

4.10 Molecular vibrations

Irreducible representations and basis functions are valuable tools for the analysis of physical systems. To begin understanding their importance, we demonstrate how to use them to identify and calculate the normal vibrational modes of simple molecules. As we shall see, these normal modes are, in fact, basis functions of the irreducible representations of the symmetry group of the molecule.

Moreover, recall that the normal modes of a molecule account for all possible shape deformations from its symmetric (ground state) configuration. Thus, the irreducible representations classify all possible channels by which a molecule (and more generally a system) moves from its symmetric state.

Identification of the normal vibrational modes of a molecule can be obtained in the following steps: First, find the irreducible representations associated with the vibrational modes of the molecule. Next, construct the projection operators associated with these representations. Finally, choose coordinates that describe a general deformation of the molecule and apply the projection operators on these coordinates.

Let us demonstrate this procedure with the example of a water molecule H_2O (see Fig. 4-15), whose symmetry is expressed by the point group C_{2v} . This group has four symmetry elements and four irreducible representations (all one-dimensional). The character table of the group can be found on page 75.

The water molecule has three atoms and nine degrees of freedom (3 for each atom). However, six of them describe translations of the center of mass and rotations. Only three degrees of freedom account for the vibrations of the molecule. Our first task is to find the irreducible representations associated with them.

Consider the reducible representation obtained from the matrices that operate on a 9-component vector made from the coordinates of all three atoms as in Eq. (4.14). This representation is called the *translation vectors representation* and is denoted by $\Gamma_{\text{trans.vec}}$. It includes the irreducible representations of the nine degrees of freedom of the molecule: translations, rotations, and vibrational modes.

We wish to decompose $\Gamma_{\text{trans.vec}}$ into its irreducible representations:

$$\Gamma_{\text{trans.vec}} = \bigoplus_{\alpha} n_{\alpha} \Gamma^{(\alpha)}. \quad (4.53)$$

The symbol $\bigoplus_{\alpha}$ (called “direct sum”) on the right-hand side of this equation should be understood as the collection of the irreducible representations that appear in the block diagonal form of the translation vector representation. n_{α} is the number of times that the irreducible representation α appears in this collection.

From (4.53) it follows that any symmetry operation, g , satisfies the relation:

$$\chi_{\text{trans.vec}}(g) = \sum_{\alpha} n_{\alpha} \chi^{(\alpha)}(g), \quad (4.54)$$

where $\chi_{\text{trans.vec}}(g)$ is the character of the translation vector representation, while $\chi^{(\alpha)}(g)$ is the character of the α irreducible representation.

Multiplying Eq. (4.54) by $\chi^{(\beta)*}(g)$ and summing over the group elements, we obtain:

$$\sum_g \chi^{(\beta)*}(g) \chi(g) = \sum_{\alpha} n_{\alpha} \sum_g \chi^{(\beta)*}(g) \chi^{(\alpha)}(g) = \sum_{\alpha} n_{\alpha} |G| \delta_{\alpha\beta} = n_{\beta} |G|, \quad (4.55)$$

where $|G|$ is the group order. To obtain this result, we employed the orthogonality of the characters of irreducible representations expressed by Eq. (4.26). Solving the above equation for n_β we get the number of times that β irreducible representation appears in $\Gamma_{\text{trans.vec}}$:

$$n_\beta = \frac{1}{|G|} \sum_g \chi^{(\beta)*}(g) \chi_{\text{trans.vec}}(g). \quad (4.56)$$

Thus to identify the decomposition (4.53) of the translation vector representation, we should first calculate the characters, $\chi_{\text{trans.vec}}(g)$. For this purpose, we do not need the full matrix structure associated with the symmetry element g . It is sufficient to identify only its diagonal elements because they determine the trace. Hence, the character of g is determined only by those atoms that stay in their positions after applying this symmetry operation. Each component of the vector of such an atom contributes ± 1 depending on whether it reversed its direction or not. The character is then the sum of these numbers. In particular, for the water molecule shown in Fig. 4-15, the rotation c_2 leaves only the oxygen atom in its place with $z_1 \rightarrow z_1$, $x_1 \rightarrow -x_1$, and $y_2 \rightarrow -y_2$, hence $\chi(c_2) = 1 - 1 - 1 = -1$. Reflection through the molecular plane (xz plane) leaves all atoms at their positions. One component of their translation vectors reverses direction while the other two remain intact. Therefore $\chi(\sigma_v) = 3(1 + 1 - 1) = 3$. Reflection through the yz plane leaves only one atom at its position and reverses a single component of its translation vector, thus, $\chi(\sigma'_v) = 1$. Finally, the character of the identity operation equals the number of degrees of freedom, $\chi(E) = 9$. These characters are summarized in the following table:

C_{2v}	E	c_2	$\sigma_v(xz)$	$\sigma'_v(yz)$
$\Gamma_{\text{trans.vec}}$	9	-1	3	1

Now, using Eq. (4.56) with this result and the characters of the irreducible representations of C_{2v} , that appears on page 75, we obtain that the number of times that the identity representation appears in $\Gamma_{\text{trans.vec}}$ is

$$n_{A_1} = \frac{1}{4} \sum_g \chi^{(A_1)*}(g) \chi_{\text{trans.vec}}(g) = \frac{1}{4} (1 \cdot 9 + 1 \cdot (-1) + 1 \cdot 3 + 1 \cdot 1) = 3. \quad (4.57)$$

Similarly, for the other irreducible representations, we have:

$$n_{A_2} = \frac{1}{4} \sum_g \chi^{(A_2)*}(g) \chi_{\text{trans.vec}}(g) = \frac{1}{4} (1 \cdot 9 + 1 \cdot (-1) + (-1) \cdot 3 + (-1) \cdot 1) = 1, \quad (4.58)$$

$$n_{B_1} = \frac{1}{4} \sum_g \chi^{(B_1)*}(g) \chi_{\text{trans.vec}}(g) = \frac{1}{4} (1 \cdot 9 + (-1) \cdot (-1) + 1 \cdot 3 + (-1) \cdot 1) = 3, \quad (4.59)$$

and

$$n_{B_2} = \frac{1}{4} \sum_g \chi^{(B_2)*}(g) \chi_{\text{trans.vec}}(g) = \frac{1}{4} (1 \cdot 9 + (-1) \cdot (-1) + (-1) \cdot 3 + 1 \cdot 1) = 2. \quad (4.60)$$

Thus the translation vectors representation includes three A_1 representations, one A_2 representation, three B_1 representations, and twice the B_2 representation:

$$\Gamma_{\text{trans.vec}} = 3A_1 \oplus A_2 \oplus 3B_1 \oplus 2B_2. \quad (4.61)$$

Altogether we have nine (one-dimensional) irreducible representations that account for all nine degrees of freedom of the molecule. From the character table of C_{2v} , (on page 75) we see that translations of the molecules, described by the linear basis functions, x , y , and z , are associated with the, B_1 , B_2 , and A_1 irreducible representations, respectively. Similarly, rotations of the molecule, corresponding to the basis functions, R_x , R_y and R_z , are associated with the B_2 , B_1 , and A_2 representations, respectively. Subtracting these irreducible representations from $\Gamma_{\text{trans.vec}}$ we are left with a representation that includes only the vibrational modes of the molecule:

$$\Gamma_{\text{vib}} = 2A_1 \oplus B_1. \quad (4.62)$$

Thus, two normal mods of the molecule are basis functions of the identity representation, A_1 , while the third is a basis function of the B_1 representation.

To reveal the spatial structure of the molecule's normal modes, it is convenient to use the *internal coordinates representation*. The internal coordinates of a molecule are the bonds lengths and the angles between bonds. These coordinates account only for the deformations of the molecule because they are unaffected by translations and rotations. Thus their number equals the number of the normal modes of the molecule. Clearly, there is some arbitrariness in their choice. For the water molecule, it is convenient to choose them to be the bond lengths between the oxygen and the hydrogen atoms, and the angle between these bonds, as shown in Fig. 4-18.

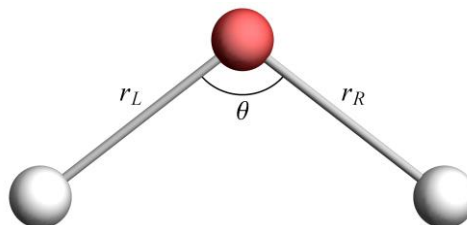


Figure 4-18 The internal coordinates of a water molecule H_2O

The action of the symmetry operations on these coordinates is:

$$E \begin{pmatrix} r_R \\ r_L \\ \theta \end{pmatrix} = \begin{pmatrix} r_R \\ r_L \\ \theta \end{pmatrix}, \quad c_2 \begin{pmatrix} r_R \\ r_L \\ \theta \end{pmatrix} = \begin{pmatrix} r_L \\ r_R \\ \theta \end{pmatrix}, \quad \sigma_v \begin{pmatrix} r_R \\ r_L \\ \theta \end{pmatrix} = \begin{pmatrix} r_R \\ r_L \\ \theta \end{pmatrix}, \quad \text{and} \quad \sigma'_v \begin{pmatrix} r_R \\ r_L \\ \theta \end{pmatrix} = \begin{pmatrix} r_L \\ r_R \\ \theta \end{pmatrix}, \quad (4.63)$$

hence the characters of this representation are: $\chi(E)=3$, $\chi(c_2)=1$, $\chi(\sigma_v)=3$, and $\chi(\sigma'_v)=1$. From here, it is easy to check that the internal coordinate representation decomposes into the three irreducible representations, as expressed in Eq. (4.62).

Our goal now is to construct the linear basis functions for each of the irreducible representations of the molecule H_2O using its internal coordinates. Since the normal vibrational modes represent small deviations from the equilibrium state, they are given by these linear basis functions.

To this end, we apply the projection operators, defined in Eq. (4.38), on each one of the internal coordinates. Consider first the application of the projection, associated with the identity representation, on the coordinate r_R . Using Eqs. (4.63), and the fact that all characters of the identity representation equal one, we have

$$\begin{aligned} P^{(A_1)} r_R &= \left[\chi^{(A_1)}(E)E + \chi^{(A_1)}(c_2)c_2 + \chi^{(A_1)}(\sigma_v)\sigma_v + \chi^{(A_1)}(\sigma'_v)\sigma'_v \right] r_R \\ &= [E + c_2 + \sigma_v + \sigma'_v] r_R = r_R + r_L + r_R + r_L = 2(r_R + r_L) \end{aligned} \quad (4.64)$$

and

$$P^{(A_1)} r_L = 2(r_R + r_L), \quad \text{while} \quad P^{(A_1)} \theta = 4\theta. \quad (4.65)$$

Similarly, projections on the B_1 representation (obtained with the help of the character table on page 75) give:

$$\begin{aligned} P^{(B_1)} r_R &= \left[\chi^{(B_1)}(E)E + \chi^{(B_1)}(c_2)c_2 + \chi^{(B_1)}(\sigma_v)\sigma_v + \chi^{(B_1)}(\sigma'_v)\sigma'_v \right] r_R \\ &= [E - c_2 + \sigma_v - \sigma'_v] r_R = r_R - r_L + r_R - r_L = 2(r_R - r_L), \end{aligned} \quad (4.66)$$

and

$$P^{(B_1)} r_L = 2(r_L - r_R), \quad P^{(B_1)} \theta = 0. \quad (4.67)$$

Thus $r_R + r_L$ and θ are two basis functions of the A_1 irreducible representation, while $r_R - r_L$ is the basis function of the B_1 representation.

These linear basis functions are the normal coordinates associated with the vibrational modes of the molecule shown in Fig. 4-19 on the next page. The normal coordinates change sinusoidally in time with the frequency of the corresponding vibrational mode. Thus the periodic change in time of the basis function, $r_R + r_L$, associated with the A_1 irreducible representation, describes the stretch and compression mode shown in the left panel of Fig. 4-19. The second normal mode of the same irreducible representation is

described by a sinusoidal change, in time, of the angle θ . This is the bending mode shown in the middle panel of the figure. Finally, the vibrational mode associated with the B_1 representation is obtained when $r_R - r_L$ changes periodically in time. This mode describes vibrations in which one bond stretches while the other compresses and vice versa. It is illustrated in the right panel of the figure.

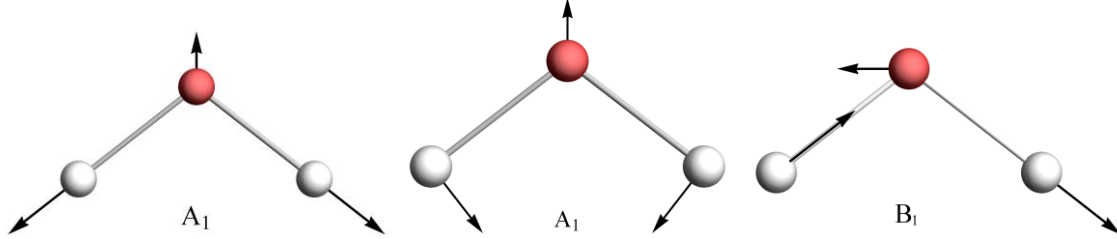


Figure 4-19 The normal modes of a water molecule

Notice that the normal modes shown in the above figure are drawn for the case where the molecule's center of mass stays put. This condition implies that, in general, all atoms move from their equilibrium position. The internal coordinates, however, behave as described above.

4.10 Irreducible representations in quantum mechanics

In this section, we show that the energy levels of a quantum system can be classified according to the irreducible representations of the symmetry group². In particular, the energy-levels degeneracy is determined by the dimension of the irreducible representation, and the corresponding wave functions are basis functions of the representation.

Let $G = \{g_i\}$ be the group of symmetry operations that leave the Hamiltonian, H , invariant, i.e. $[H, g_i] = 0$ or $H = g_i^{-1} H g_i$ for any $g_i \in G$. Consider the time-independent Schrödinger equation,

$$H\psi = \varepsilon\psi. \quad (4.68)$$

Applying one of the symmetry operations on this equation gives

$$g_j H \psi = \varepsilon g_j \psi, \quad (4.69)$$

and since the Hamiltonian commutes with g_j , $H g_j \psi = \varepsilon g_j \psi$. Thus $g_j \psi$ is a solution of the Schrödinger equation with the same energy ε . If we know that the energy level is not

²Similar considerations apply to the eigenfrequencies of the molecular vibrational modes, discussed above.

degenerate, then $g_j\psi$ can be different from ψ , at most, by a phase factor, $g_j\psi = \exp(i\varphi_j)\psi$. Applying an additional symmetry operation gives $g_i g_j\psi = \exp(i\varphi_j)g_i\psi = \exp(i\varphi_j)\exp(i\varphi_i)\psi$. However, $g_i g_j$ is also a symmetry element of G , therefore $g_i g_j\psi = \exp(i\varphi_{ij})\psi$. Thus $\exp(i\varphi_{ij}) = \exp(i\varphi_i)\exp(i\varphi_j)$; hence the phase factors constitute a one-dimensional irreducible representation of the group.

Consider now a different situation where the application of some symmetry operator on ψ generates a linear combination of ℓ orthogonal wave functions of the Hamiltonian. In this case, the energy level ε must be degenerate, at least, ℓ times. We call this degeneracy a *normal degeneracy* if there are no other wave functions with the same energy. A situation where there are other wave functions with the same energy that cannot be obtained by applying one of the symmetry elements on ψ is called an *accidental degeneracy*. The reason for accidental degeneracies is not symmetry. Usually, they are associated with fine-tuning the Hamiltonian's parameters; therefore, accidental degeneracies are not generic.

In what follows, we focus our attention on the case of a normal degeneracy. We denote by ψ_n , where $n=1,2,\dots,\ell$, the orthogonal set of wavefunctions associated with the degenerate energy subspace:

$$H\psi_n = \varepsilon\psi_n. \quad (4.70)$$

Application of a symmetry element on one of these wave functions, generally, yields a linear combination of all wave functions of the subspace. It is convenient to define a vector of the wave functions $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_\ell)^T$ so that

$$g_j\boldsymbol{\psi} = \Gamma(g_j)\boldsymbol{\psi}, \quad (4.69)$$

where $\Gamma(g_j)$ is a $\ell \times \ell$ matrix. From here we obtain

$$g_i(g_j\boldsymbol{\psi}) = \Gamma(g_i g_j)\boldsymbol{\psi} = \Gamma(g_i)[\Gamma(g_j)\boldsymbol{\psi}] = \Gamma(g_i)\Gamma(g_j)\boldsymbol{\psi}. \quad (4.70)$$

Thus, the set of matrices $\{\Gamma(g_j)\}$ constitutes a representation of the symmetry group with dimension ℓ :

$$\Gamma(g_i g_j) = \Gamma(g_i)\Gamma(g_j). \quad (4.71)$$

Take note that one can always choose the set of wave functions, $\{\psi_n\}$, to be an orthonormal set, and for this choice, $\Gamma(g_j)$, are unitary matrices.

We turn now to show that this representation is irreducible. For this purpose, we first consider a change of basis in the subspace of degenerate wave functions, $\{\psi_n\}$. A unitary transformation, U , describes the transition from one basis to another:

$$\psi' = U\psi. \quad (4.72)$$

Multiplying this equation by U^{-1} , from the left, and applying the symmetry operation, g_j , we obtain

$$g_j U^{-1} \psi' = g_j \psi = \Gamma(g_j) \psi = \Gamma(g_j) U^{-1} \psi'. \quad (4.73)$$

However, U and g_j act on two different spaces: U acts on the wave functions space while g_j acts on the coordinate space, therefore

$$g_j U^{-1} \psi' = U^{-1} g_j \psi', \quad (4.74)$$

and from the last two equations, we conclude that

$$g_j \psi' = U \Gamma(g_j) U^{-1} \psi'. \quad (4.75)$$

Thus, a change of basis is nothing but a similarity transformation:

$$\Gamma(g_j) \rightarrow U \Gamma(g_j) U^{-1}. \quad (4.76)$$

Now suppose that the representation, Γ , is reducible. Then there must be a basis in which all matrices $\Gamma(g_j)$ are block diagonal. But this property implies that there are at least two different groups of wave functions that are not mixed by any symmetry operation. However, this contradicts our assumption of normal degeneracy because it is the case of accidental degeneracy. Thus, the representation is irreducible, and the eigenfunctions are basis functions of this representation.

To conclude:

- To each eigenenergy of H corresponds one irreducible representation of the symmetry group of the Hamiltonian. The eigenstates associated with this energy are basis functions of the representation.
- The degeneracy of the eigenenergies is the dimension of the irreducible representation.
- Group theory provides “good quantum numbers”: it associates an irreducible representation to each eigenenergy of the Hamiltonian.

4.11 Exercises

1. Identify the conjugacy classes of the point group C_{6v} and the dimensions of its irreducible representations. Build the character table of the group.

Hint: Notice that C_{6v} is a direct product of $C_{3v}(\{E, 2C_3, 3\sigma\})$ by $C_2(\{E, c_2\})$. By direct product, we mean the multiplication of each element in one group by the elements of the other group.

2. Prove the convolution theorem of irreducible representations:

$$\sum_g \Gamma_{ij}^{(\alpha)}(xg^{-1}) \Gamma_{kl}^{(\beta)}(g) = \frac{|G|}{\ell_\alpha} \delta_{\alpha\beta} \delta_{jk} \Gamma_{il}^{(\alpha)}(x), \quad (4.77)$$

where x and g are elements of the same group and ℓ_α is the dimension of the irreducible representation α .

3. In continuation to exercise 2 of Chapter 2, identify the conjugacy classes of T_d , and the dimensions of its irreducible representations.
4. Identify the irreducible representations of the vibrational modes of Methane, CH_4 . This molecule has the shape of a tetrahedron, as illustrated in Fig. 4-20. Its symmetry group is T_d , and its character table is given below.

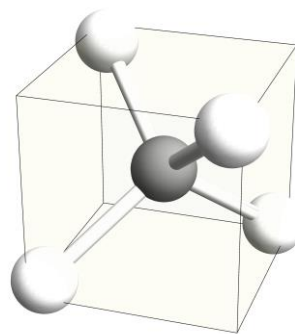


Figure 4-20 The shape of Methane molecule

T_d	E	$8C_3$	$3C_2$	$6S_4$	$6\sigma_d$		
A_1	1	1	1	1	1	-	$x^2 + y^2 + z^2$
A_2	1	1	1	-1	-1	-	-
E	2	-1	2	0	0	-	$(2z^2 - x^2 - y^2, \sqrt{3}x^2 - \sqrt{3}y^2)$
F_1	3	0	-1	1	-1	(R_x, R_y, R_z)	-
F_2	3	0	-1	-1	1	(x, y, z)	(yz, xz, xy)

5. Find the vibrational modes of a molecule having the shape of an equilateral triangle, as shown in Fig. 4-21.
6. Point groups are subgroups of the orthogonal group in three dimensions, $O(3)$. The latter contains rotations (at any angle) and reflections through any plane containing the origin. The irreducible representations of the rotation group $SO(3)$ (a subgroup of $O(3)$) are the angular momentum states. For angular momentum J , the dimension of the representation is $2J+1$. If we choose θ to be the rotation angle around the z axis, the representation of this rotation operation is described by the diagonal matrix:

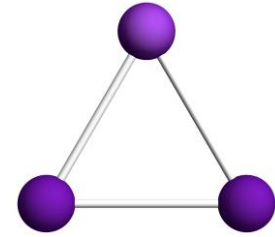


Figure 4-21 A molecule with the shape of equilateral triangle

$$\Gamma^{(J)}(\theta) = \langle J, M | \exp(iL_z \theta) | J, M' \rangle = \delta_{MM'} \exp(i\theta M), \quad (4.78)$$

where $-J \leq M \leq J$ is the projection of the angular momentum on the z axis, and L_z is the z component of the angular momentum operator. Taking the trace of this matrix, we obtain the character of the rotation operation:

$$\chi^{(J)}(\theta) = \frac{\sin\left[\left(J + \frac{1}{2}\right)\theta\right]}{\sin\left(\frac{\theta}{2}\right)} \quad (4.79)$$

Clearly this character is independent of the direction of the rotation axis.

Consider the octahedral group, O , consisting of (only) rotations of an octahedron - an object built from 8 equilateral triangles, as shown in Fig. 4-22. The character table of this group is given below³.

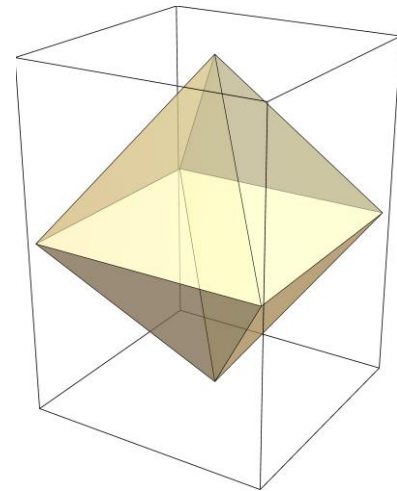


Figure 4-22 Octahedron

O	E	$8C_3$	$6C'_2$	$6C_4$	$3C_2 = C_4^2$
T_d	E	$8C_3$	$6\sigma_d$	$6S_4$	$3C_2$
A_1	1	1	1	1	1
A_2	1	1	-1	-1	1
E	2	-1	0	0	2
F_1	3	0	-1	1	-1
F_2	3	0	1	-1	-1

³ This group is isomorphic to the tetrahedral group, T_d , as shown by the character table.

Taking into account that the symmetry operations of the octahedral group are only rotations, use considerations of character orthogonality to identify the composition of the J -th representation of the rotation group in terms of the irreducible representations of the octahedral group. In particular, prove the following table:

J	Composition of O representations
0	A_1
1	F_1
2	$E \oplus F_2$
3	$A_2 \oplus F_1 \oplus F_2$

7. Use the results of the previous exercise to build the basis functions of second order for the E and the F_2 irreducible representations of the group O .

Advice: The basis functions of the rotation group $SO(3)$ are the spherical harmonic functions $Y_l^m(\theta, \varphi)$. For $J = 2$ they are given by

$$\begin{aligned}
 Y_2^2(\theta, \varphi) &= \frac{1}{4} \sqrt{\frac{15}{2\pi}} \sin^2 \theta \exp(2i\varphi), \\
 Y_2^{-2}(\theta, \varphi) &= \frac{1}{4} \sqrt{\frac{15}{2\pi}} \sin^2 \theta \exp(-2i\varphi), \\
 Y_2^1(\theta, \varphi) &= -\frac{1}{2} \sqrt{\frac{15}{2\pi}} \sin \theta \cos \theta \exp(i\varphi), \\
 Y_2^{-1}(\theta, \varphi) &= \frac{1}{2} \sqrt{\frac{15}{2\pi}} \sin \theta \cos \theta \exp(-i\varphi), \\
 Y_2^0(\theta, \varphi) &= \frac{1}{4} \sqrt{\frac{5}{\pi}} (3\cos^2 \theta - 1).
 \end{aligned} \tag{4.80}$$

From the table above, we deduce that these five wave functions should compose the basis functions of the F_2 representation (a 3-component vector) and of the E representation (a 2-component vector). From the above wave functions, construct five real functions normalized to unity (notice that the spherical harmonic functions in Eq. (4-80) are also normalized to unity), multiply them by r^2 , and rewrite the result in terms of the coordinates $x = r \sin \theta \cos \varphi$, $y = r \sin \theta \sin \varphi$, and $z = r \cos \theta$. Now identify the basis functions that belong to each one of the irreducible representations.

8. Prove the statement of the footnote on page 87.

5 Graphene and Dichalcogenides

In this chapter, we show how to employ symmetry considerations exclusively to deduce the structure of the energy bands of graphene and dichalcogenides (i.e., without knowing what the Hamiltonians that describe these systems are.)

5.1 The graphene lattice

Graphene is a monolayer of carbon atoms arranged in a honeycomb lattice. Each atom is connected to three neighboring atoms by σ - bonds and contributes one electron to the conduction band. In the graphene lattice, one can identify two hexagonal sublattices, A and B as demonstrated in Fig. 5-1

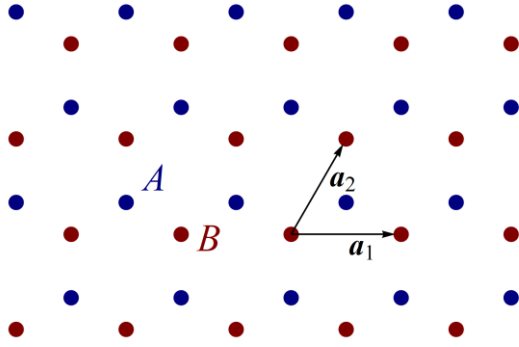


Figure 5-1 The two sublattices of graphene

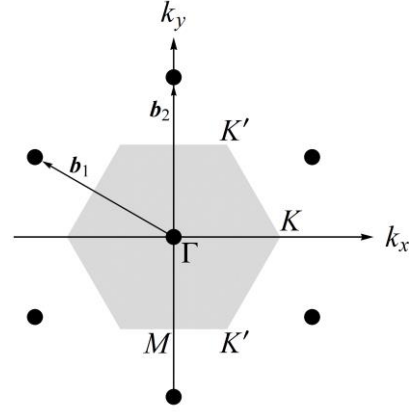


Figure 5-2 The reciprocal lattice and the first Brillouin zone of graphene

The Bravais lattice of graphene is a two-dimensional hexagonal lattice, and each unit cell contains two atoms – one of each sublattice. If we denote by a the distance between nearest neighbors sites on the same sublattice (say the red sublattice), then one can choose the two primitive basis vectors of the Bravais lattice to be:

$$\mathbf{a}_1 = a(1, 0) \quad \text{and} \quad \mathbf{a}_2 = \frac{a}{2}(1, \sqrt{3}), \quad (5.1)$$

as shown in Fig. 5-1. The primitive basis vectors of the reciprocal lattice (see Eq. 2.28) are, therefore:

$$\mathbf{b}_1 = \frac{2\pi}{a} \left(-1, \frac{1}{\sqrt{3}} \right), \quad \text{and} \quad \mathbf{b}_2 = \frac{4\pi}{a\sqrt{3}} (0, 1). \quad (5.2)$$

The reciprocal lattice and the first Brillouin zone with its special points are displayed in Fig. 5-2.

5.2 The “little group” and the graphene spectrum near the K - point

The graphene point group is C_{6v} . By choosing the origin of the coordinates to be at the center of one of the lattice’s hexagons, it is easy to see that the system has a six-fold rotation symmetry and a reflection symmetry through six planes similar to those shown in Fig. 2-11 for the hexagonal lattice¹. The character table of the C_{6v} group is presented below. This table shows that the graphene’s symmetry group contains two-dimensional irreducible representations. The C_{6v} symmetry of the reciprocal lattice, suggests that the graphene’s spectrum is characterized by degeneracy points in k -space.

The character table of C_{6v}								
C_{6v}	E	$2C_6$	$2C_3$	C_2	$3\sigma_v$	$3\sigma_d$		
A_1	1	1	1	1	1	1	-	$x^2 + y^2$
A_2	1	1	1	1	-1	-1	R_z	-
B_1	1	-1	1	-1	1	-1	-	-
B_2	1	-1	1	-1	-1	1	-	-
E_1	2	1	-1	-2	0	0	$(x, y), (R_x, R_y)$	-
E_2	2	-1	-1	2	0	0	-	$(x^2 - y^2, 2xy)$

The degeneracy points that we seek to describe here are points where two bands meet. Naturally, one expects them to be the high symmetry points in the Brillouin zone, such as the special points, Γ , K and M (see Fig. 5-2). To analyze the spectrum in the vicinity of these points, we need to introduce the notion of a “little group”. The little group is a subgroup of the crystal symmetry group that acts on Bloch vectors. It is the group of symmetry operations that returns a Bloch wave vector to itself (up to a reciprocal lattice vector). Clearly, this group depends on the choice of the wave vector, $\mathbf{k}$; therefore, special symmetry points of the Brillouin zone play an important role. Consider, for example, the

¹ The full symmetry group of graphene point is, in fact, D_{6h} where in addition to the elements of C_{6v} there are six C_2 -rotations around axes that are perpendicular to the principle axis, and an additional reflection through the horizontal plane where the atoms reside. However, being interested only in the two-dimensional properties of the system, it is sufficient to consider only those symmetry operations that do not involve the third dimension of the problem. These symmetry operations constitute the subgroup C_{6v} of D_{6h} .

Γ point. This point is of the highest symmetry in the Brillouin zone because any symmetry operation of C_{6v} returns it to itself. The K -point, on the other hand, has a lower symmetry. It is invariant only to rotations in $\pm 120^\circ$ (recall that K and K' are inequivalent), and to reflections through three lines: One is the horizontal axis and two other lines obtained from its rotations by $\pm 120^\circ$. Thus, the little group of the K -point is C_{3v} . The character table of this group is listed below (for the moment, ignore the basis functions – they are presented here for future use).

C_{3v}	E	$2C_3$	$3\sigma_v$			
A_1	1	1	1	z	$x^2 + y^2, z^2$	I
A_2	1	1	-1			τ_z^{AB}
E	2	-1	0	$\begin{pmatrix} x+iy \\ x-iy \end{pmatrix}$	$\begin{pmatrix} (x-iy)^2 \\ (x+iy)^2 \end{pmatrix}$	$\begin{pmatrix} \tau_x^{AB} + i\tau_y^{AB} \\ \tau_x^{AB} - i\tau_y^{AB} \end{pmatrix}$

How does this property of the K -point manifest itself ?

In the previous chapter, we have learned that the wave functions of the system must belong to one of the irreducible representations of the symmetry group of the system. Now we consider only those wave functions that belong to the K -point, and have lower symmetry described by the C_{3v} group. Nevertheless, the character table shows that C_{3v} contains a two-dimensional representation. Hence the K -point might be a degenerate energy point at which two bands touch. In other words, the energy spectrum of wave functions associated with this irreducible representation feature band crossing at the K -point, as illustrated schematically in Fig. 5-3.

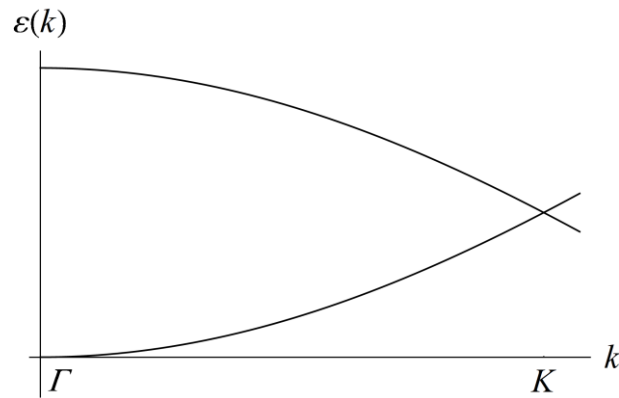


Figure 5-3 An illustration of the behavior of the energy levels along the line connecting Γ and K points of the Brillouin zone (Time reversal symmetry implies that near Γ the spectrum is quadratic – see Chap 3.)

Our goal is to construct the local Hamiltonian near the K -point from pure symmetry considerations. But before we turn to this task, let us identify the irreducible representations of the C_{3v} group. Knowing that the wave functions of the problem are the basis functions of the irreducible representations and that these also include a two-dimensional representation, we look for representations that act on a two-component wave function, i.e., a pseudospinor. In graphene, the natural choice for this pseudospinor as

$$\psi = \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix}, \quad (5.3)$$

where ψ_A and ψ_B are the wave functions on sublattices A and B , respectively.

First, Let's identify the representations of the symmetry operations (of the little group) associated with this function. Consider, first, reflection through the (horizontal) x -axis. From Fig 5-4, we see that this reflection swaps the two sublattices; therefore, the matrix representation of this operation is:

$$\Gamma(\sigma_x) \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \tau_x^{AB} \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix}, \quad (5.4)$$

where τ_x^{AB} denotes a Pauli matrix that acts on the sublattice space.

The character of this operation, $\chi(\sigma) = 0$, may be associated with two options: One is that the matrix τ_x^{AB} belongs to the two-dimensional representation, E , of C_{3v} . The second option is that τ_x^{AB} belongs to a reducible representation containing a pair of one-dimensional representations, $A_1 \oplus A_2$, whose characters ± 1 sum up is zero (see character table above)

To distinguish between the two possibilities, consider the representation of the c_3 rotation. This rotation does not mix the two sublattices. Therefore, let us look, first, at its effect on one of the two sublattices - say, sublattice A . The group which describes rotations on this sublattice is the cyclic group C_3 . All representations of this group are one-dimensional. Choosing the identity representation implies that $\Gamma(c_3)\psi_A = \psi_A$. But c_3 does not mix the sublattices; hence its representation must be a diagonal matrix. The only sensible choice for the action of this matrix on the second component of the wave function is $\Gamma(c_3)\psi_B = \psi_B$ which implies that $\chi(c_3) = 2$. Therefore, this choice corresponds to the reducible representation $A_1 \oplus A_2$.

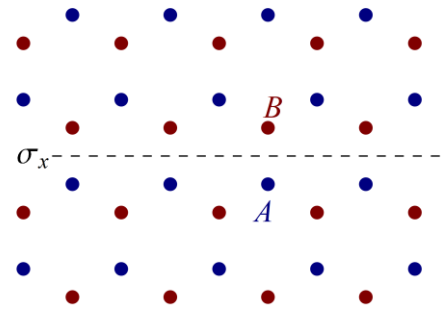


Figure 5-4 Reflection σ_x in graphene

Consider now a faithful representation of the C_3 group that describes the rotation on the A - sublattice. Let us denote the points on this sublattice by 1 , $\exp(i2\pi/3)$, and $\exp(i4\pi/3) = \exp(-i2\pi/3)$ as shown in Fig. 5-5. A clockwise rotation of this sublattice amounts for multiplication of the lattice points by $\exp(-i2\pi/3)$, thus $\Gamma(c_3)\psi_A = \exp(-i2\pi/3)\psi_A$.

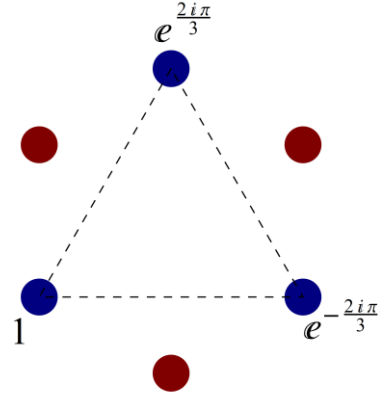


Figure 5-5 Representation of C_3 rotation on sublattice A

Let us now deduce the form of the 2×2 diagonal matrix that acts on the vector wave function (5.3). The only option that complies with the character table on page 95 is $\Gamma(c_3)\psi_B = \exp(i2\pi/3)\psi_B$. With this choice, the matrix that describes the action of c_3 rotation on the spinor is:

$$\Gamma(c_3) \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \begin{pmatrix} \exp(-i\frac{2\pi}{3}) & 0 \\ 0 & \exp(i\frac{2\pi}{3}) \end{pmatrix} \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \exp\left(-i\frac{2\pi}{3}\tau_z^{AB}\right) \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix}, \quad (5.5)$$

and the corresponding character is $\chi(c_3) = 2\cos(2\pi/3) = -1$. Alternatively, one may deduce this result by demanding that each unit cell be assigned with the same phase as demonstrated in Fig. 5-6. This condition implies that a clockwise rotation of the sublattice B is represented by $\exp(i2\pi/3)$ multiplication, i.e. $\Gamma(c_3)\psi_B = \exp(i2\pi/3)\psi_B$.

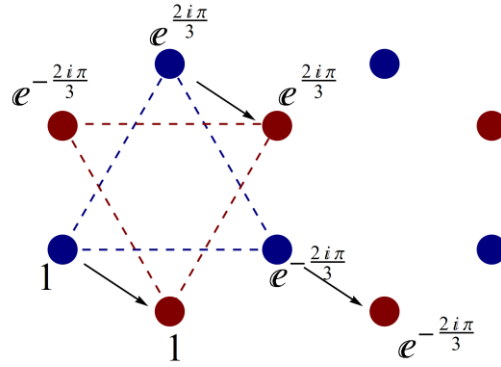


Figure 5-6 Representation of C_3 rotation on sublattice B

One can verify that the matrix representing the c_3^2 rotation is $\Gamma^*(c_3)$ and that reflections through the two other axes can be obtained by rotating $\Gamma(\sigma_x)$ appropriately, i.e., $\Gamma(c_3)\Gamma(\sigma_x)\Gamma(c_3^2)$ and $\Gamma(c_3^2)\Gamma(\sigma_x)\Gamma(c_3)$. Together with the identity operation, these matrices constitute the two-dimensional irreducible representation of C_{3v} point group.

We turn now to construct the local Hamiltonian near the K -point. For this purpose, it is convenient to define $\delta\mathbf{k}$ to be a wavenumber vector measured from the K -point of the Brillouin zone. Then the most general form of a 2×2 Hamiltonian is

$$\hat{\varepsilon}(\delta\mathbf{k}) = R_0(\delta\mathbf{k})I + \mathbf{R}(\delta\mathbf{k}) \cdot \boldsymbol{\tau}^{AB}, \quad (5.6)$$

where I is the identity matrix, $\boldsymbol{\tau}^{AB} = (\tau_x^{AB}, \tau_y^{AB}, \tau_z^{AB})$ are Pauli matrices acting on the sublattice space. The functions $R_0(\delta\mathbf{k})$ and $\mathbf{R}(\delta\mathbf{k}) = (R_x(\delta\mathbf{k}), R_y(\delta\mathbf{k}), R_z(\delta\mathbf{k}))$ are, for the time being, arbitrary functions of $\delta\mathbf{k}$. Without loss of generality, we set the energy at the K -point to be zero, $\hat{\varepsilon}(0) = 0$; hence these functions vanish at $\delta\mathbf{k} = 0$, i.e., $R_{0,x,y,z}(0) = 0$.

The time-independent Schrödinger equation is

$$\hat{\varepsilon}(\delta\mathbf{k}) \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \varepsilon \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix}. \quad (5.7)$$

Application of the symmetry operation g on both sides of this equation yields

$$\hat{\varepsilon}(g\delta\mathbf{k})\Gamma(g) \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \varepsilon\Gamma(g) \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix}, \quad \text{i.e.} \quad \Gamma^{-1}(g)\hat{\varepsilon}(g\delta\mathbf{k})\Gamma(g) \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} = \varepsilon \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix}. \quad (5.8)$$

Notice that g acts on $\delta\mathbf{k}$, but in the sublattice space is represented by $\Gamma(g)$. The wave functions are also functions of $\delta\mathbf{k}$, but to avoid cumbersome formulas, we suppressed this dependence.

It follows that invariance of the local Hamiltonian under all symmetry operations of the little group dictates the relations:

$$\hat{\varepsilon}(\delta\mathbf{k}) = \Gamma^{-1}(g)\hat{\varepsilon}(g\delta\mathbf{k})\Gamma(g) \quad \text{for any } g \in C_{3v}. \quad (5.9)$$

This equation imposes constraints on the form of the Hamiltonian (5.6). Consider, for instance, the matrix τ_z^{AB} . It is invariant under rotation:

$$\Gamma^{-1}(c_3)\tau_z^{AB}\Gamma(c_3) = \exp\left(i\frac{2\pi}{3}\tau_z^{AB}\right)\tau_z^{AB}\exp\left(-i\frac{2\pi}{3}\tau_z^{AB}\right) = \tau_z^{AB}, \quad (5.10)$$

but changes sign under reflection,

$$\Gamma^{-1}(\sigma)\tau_z^{AB}\Gamma(\sigma) = -\tau_z^{AB}. \quad (5.11)$$

Therefore, it cannot be part of the two-dimensional representation associated with band touching (i.e. degeneracy) at the K -point. Comparing the sign changes under these

operations, with those of the character table on page 95, shows that τ_z^{AB} is a basis function of the A_2 representation.

Consider now the other Pauli matrices. Guided by our expectation that they belong to the two-dimensional irreducible representation of C_{3v} , it makes sense to define the following linear combinations of these matrices, $\tau_{\pm}^{AB} = (\tau_x^{AB} \pm i\tau_y^{AB})/2$. For these matrices, we have:

$$\Gamma^{-1}(\sigma)\tau_{\pm}^{AB}\Gamma(\sigma) = \tau_x^{AB}\tau_{\pm}^{AB}\tau_x^{AB} = \tau_{\mp}^{AB}, \quad (5.12)$$

while

$$\Gamma^{-1}(c_3)\tau_{\pm}^{AB}\Gamma(c_3) = \exp\left(i\frac{2\pi}{3}\tau_z^{AB}\right)\tau_{\pm}^{AB}\exp\left(-i\frac{2\pi}{3}\tau_z^{AB}\right) = \exp\left(\mp i\frac{2\pi}{3}\right)\tau_{\pm}^{AB}. \quad (5.13)$$

(The last formula can be quickly proved using the matrix form of $\Gamma(c_3)$ shown in Eq. (5.5).)

A similar calculation gives

$$\Gamma^{-1}(c_3^2)\tau_{\pm}^{AB}\Gamma(c_3^2) = \exp\left(-i\frac{2\pi}{3}\tau_z^{AB}\right)\tau_{\pm}^{AB}\exp\left(i\frac{2\pi}{3}\tau_z^{AB}\right) = \exp\left(\pm i\frac{2\pi}{3}\right)\tau_{\pm}^{AB}. \quad (5.14)$$

From here it follows that the vector of matrices, $(\tau_+^{AB}, \tau_-^{AB})$, is a basis function of the two-dimensional representation of C_{3v} . This basis function is listed in the rightmost column of the character table on page 95. The other columns show the linear and quadratic basis functions of the group.

The basis functions of the two-dimensional representations are not invariant under the group operation. For instance, a clockwise rotation of the linear basis functions, $\delta k_x \pm i\delta k_y$ gives a phase factor², $(\delta k_x \pm i\delta k_y)\exp(\mp i2\pi/3)$, similar to the rotation of $\tau_{\pm}^{AB}$, see Eq. (5.13). Nevertheless, one may construct combinations of these basis functions that are invariant under all group operations, as required by Eq. (5.9). Consider, for example, the product $(\delta k_x - i\delta k_y)(\tau_x^{AB} + i\tau_y^{AB})$. The two factors in this product collect opposite phases; therefore the product is invariant under c_3 rotation. However, it is not invariant under reflection through the x axis. To see why, notice that this reflection transforms $\delta k_x - i\delta k_y$ to its complex conjugate, while $\tau_x^{AB} + i\tau_y^{AB}$ transforms to $\tau_x^{AB} - i\tau_y^{AB}$, see Eq. (5.12). Thus $(\delta k_x - i\delta k_y)(\tau_x^{AB} + i\tau_y^{AB}) \rightarrow (\delta k_x + i\delta k_y)(\tau_x^{AB} - i\tau_y^{AB})$. However, one can see that the combination $(\delta k_x - i\delta k_y)(\tau_x^{AB} + i\tau_y^{AB}) + (\delta k_x + i\delta k_y)(\tau_x^{AB} - i\tau_y^{AB})$ is invariant under both symmetry operations and, in fact, under all group elements of C_{3v} . Similar

² This can be proved directly checked by rotating the vector $(\delta k_x, \delta k_y)$ in 120° clockwise and calculating the phase factor that multiplies the complex vector $k_x \pm ik_y$ as a result of this rotation.

considerations show that $(\delta k_x + i\delta k_y)^2 (\tau_x^{AB} + i\tau_y^{AB}) + (\delta k_x - i\delta k_y)^2 (\tau_x^{AB} - i\tau_y^{AB})$ is also invariant under all symmetry operations of C_{3v} . Thus, up to quadratic order in $\delta \mathbf{k}$, the local form of the Hamiltonian near the K -point is

$$\begin{aligned} \hat{\varepsilon}(\delta \mathbf{k}) = & \frac{\hbar v}{2} [(\delta k_x - i\delta k_y)(\tau_x^{AB} + i\tau_y^{AB}) + h.c.] \\ & - \frac{h_w}{2} [(\delta k_x + i\delta k_y)^2 (\tau_x^{AB} + i\tau_y^{AB}) + h.c.] \\ & + \frac{\hbar^2}{2m_{\text{eff}}} (\delta k_x^2 + \delta k_y^2) I \end{aligned} \quad (5.15)$$

where $h.c.$ stands for Hermitian conjugate. Here, v , h_w , and m_{eff} are system-dependent constants that cannot be deduced from symmetry considerations. The third term in the above equation comes from the identity representation (see table on page 95).

Simplifying and rearranging the terms in formula (5.15) result in the concise form:

$$\hat{\varepsilon}(\delta \mathbf{k}) = \hbar v \delta \mathbf{k} \cdot \boldsymbol{\tau}^{AB} - h_w [(\delta k_x^2 - \delta k_y^2) \tau_x^{AB} - 2\delta k_x \delta k_y \tau_y^{AB}] + \frac{\hbar^2 \delta k^2}{2m_{\text{eff}}} I. \quad (5.16)$$

The first term is the leading contribution that describes a Dirac-point spectrum (in two dimensions). Its diagonalization gives $\varepsilon \simeq \pm \hbar v \delta k$, where $\delta k = |\delta \mathbf{k}|$. This linear spectrum is particle-hole symmetric. The last term of the Hamiltonian breaks this symmetry; however, empirically, it is found to be negligible. The second term in Eq. (5.16) breaks the rotational symmetry of the Dirac spectrum. It deforms the Dirac cone toward a triangular shape - a property called *triangular wrapping*. A contour plot obtained from diagonalization of Hamiltonian (5.16) (without the third term) is shown in Fig. 5-7. The left panel presents the local behavior of the energy surface as a function of the distance vector $(\delta k_x, \delta k_y)$ from the K -point. The right panel shows a global picture of the energy surface obtained from the 6-fold rotation symmetry of the system.

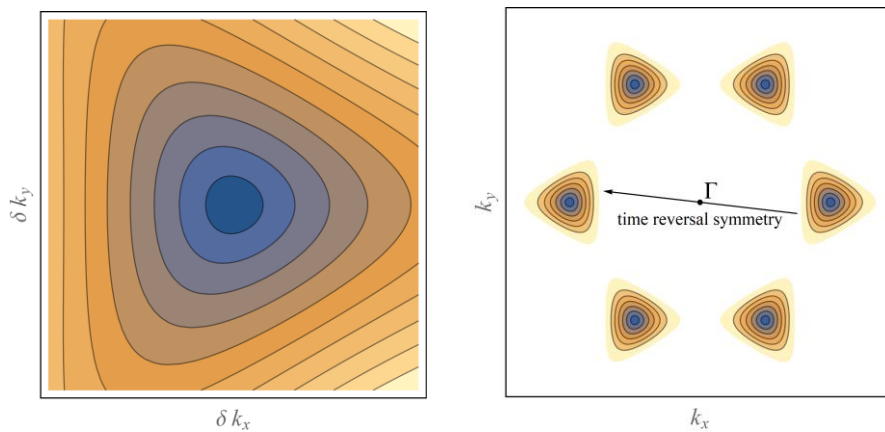


Figure 5-7 A contour plot of the energy surface of graphene obtained from symmetry considerations

We conclude this section with several comments:

- (a) When the Fermi level is low enough, the Fermi surface near the Dirac-point becomes circular and possesses an approximate symmetry $\delta\mathbf{k} \rightarrow -\delta\mathbf{k}$. This feature is called “artificial time-reversal symmetry” or “pseudo-time-reversal symmetry”.
- (b) The constants v , h_v , and m_{eff} in formula (5.16) cannot be determined from symmetry considerations. Moreover, it is possible that some of the energy levels at the K -point are associated with one-dimensional representations of the C_{3v} group A_1 and A_2 (the basis functions of the latter are cubic in the components $\delta\mathbf{k}$). In this case, the energy levels are not degenerate.
- (c) In chapter 3, we saw that time-reversal symmetry excludes the possibility of having a Dirac spectrum at the Γ point. On the other hand, the little group of this point, C_{6v} , contains two-dimensional representations; therefore, energy level may be degenerate, albeit in a quadratic manner. This behavior is demonstrated in Fig. 5-8, which presents the energy levels of graphene obtained by numerical diagonalization of a microscopic model of graphene. The plot shows the energy levels along straight lines connecting the special points in the Brillouin zone.

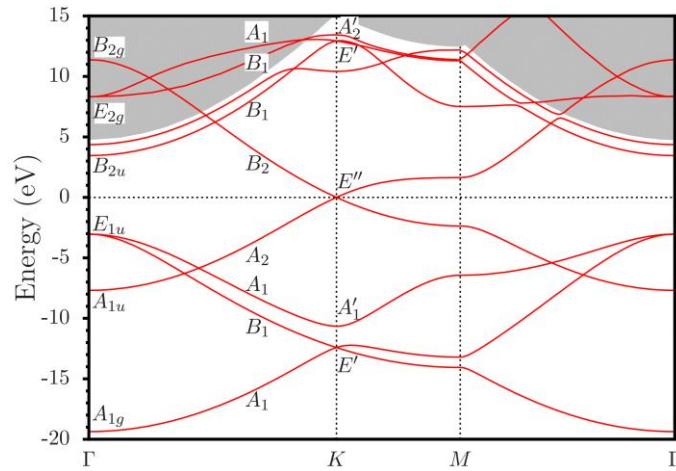


Figure 5-8 The energy levels of graphene
(taken from the paper E. Kogan and V.U. Nazarov, PRB **85**, 115418 (2012))

5.3 Schur’s lemma and multiplication of irreducible representations

The local Hamiltonian (5.16) has been constructed by an educated guess (based on properties of the basis functions) rather than an orderly procedure. In this section, we present the mathematical tools which facilitate this construction. These are based on Schur’s lemma that we give here without proof.

The (first) lemma of Schur

Let $\Gamma(G)$ be an irreducible representation of the group G acting on a vector space $\mathcal{V}$, and $\hat{C}$ some linear operator defined on the same vector space. If $[\Gamma(g), \hat{C}] = 0$ for any $g \in G$, then $\hat{C}$ is proportional to the identity matrix.

The meaning of this lemma is that if $\hat{C}$ is the Hamiltonian acting on the Hilbert space, $\mathcal{V}$, the requirement that the Hamiltonian is invariant under all symmetry operations (namely commutes with all of them) implies that it belongs to the identity representation A_1 (sometimes called “singlet”).

Consider now the general expansion of the local Hamiltonian,

$$H = \sum_{\alpha, \beta, j_1 \dots j_n} h_{\alpha, \beta, j_1 \dots j_n} k_{j_1} k_{j_2} \dots k_{j_n} |\psi_\alpha\rangle \langle \psi_\beta| \quad (5.17)$$

where $h_{\alpha, \beta, j_1 \dots j_n}$ are constants. The constituents of this expansion might belong to different irreducible representations of the symmetry group. For instance, the vector $\mathbf{k}$ may constitute a basis function of the two-dimensional representation E (as in graphene), while the wave functions, $|\psi_\alpha\rangle$ and $\langle \psi_\beta|$ might be basis functions of any irreducible representation of the symmetry group.

Thus, the above Hamiltonian can be viewed as a sum of products with factors that belong to various irreducible representations of the symmetry group. The constants $h_{\alpha, \beta, j_1 \dots j_n}$ should be chosen such that the resulting Hamiltonian is a singlet. However, in general, the product of basis functions belonging to different irreducible representations is not necessarily a basis function of the identity representation, A_1 . Our goal is to identify the condition for which this is indeed the case.

Definition: Direct products of irreducible representations

Let $\Gamma_{ij}^{(\alpha)}$ and $\Gamma_{kl}^{(\beta)}$ be the matrix elements of two irreducible representations of the group G with dimensions ℓ_α and ℓ_β , respectively. The direct product of these representations is the (generally) reducible representation of dimension $\ell_\alpha \ell_\beta$ defined by the matrix:

$$\Gamma^{(\alpha)} \otimes \Gamma^{(\beta)} = \begin{pmatrix} \Gamma_{11}^{(\alpha)} \Gamma^{(\beta)} & \Gamma_{12}^{(\alpha)} \Gamma^{(\beta)} & \dots & \Gamma_{1\ell_\alpha}^{(\alpha)} \Gamma^{(\beta)} \\ \Gamma_{21}^{(\alpha)} \Gamma^{(\beta)} & \ddots & & \vdots \\ \vdots & & & \\ \Gamma_{\ell_\alpha 1}^{(\alpha)} \Gamma^{(\beta)} & \dots & & \Gamma_{\ell_\alpha \ell_\alpha}^{(\alpha)} \Gamma^{(\beta)} \end{pmatrix}, \quad (5.18)$$

where each term in this matrix is by itself a matrix $\ell_\beta \times \ell_\beta$. In general, this representation is reducible; hence we want to obtain its decomposition into irreducible representations. This can be achieved using the character's orthogonality property, as we saw when we discussed the normal modes of molecules in section 4.9.

From Eq. (5.18) it follows that the character is a product of two irreducible representations is the product of their characters:

$$\chi[\Gamma^{(\alpha)}(g) \otimes \Gamma^{(\beta)}(g)] = \sum_i \Gamma_{ii}^{(\alpha)}(g) \sum_k \Gamma_{kk}^{(\beta)}(g) = \chi^{(\alpha)}(g) \chi^{(\beta)}(g). \quad (5.19)$$

With the help of this result, one can calculate the number of times that each irreducible representation, γ , appears in the product $\Gamma^{(\alpha)} \otimes \Gamma^{(\beta)}$ by the formula:

$$n_\gamma = \frac{1}{|G|} \sum_{g \in G} \chi^{(\gamma)*}(g) \chi^{(\alpha)}(g) \chi^{(\beta)}(g). \quad (5.20)$$

Example: The direct product $E \otimes E$ of the group C_{3v}

From the character table of C_{3v} on page 95 and Eq. (5.19) we obtain that:

$$\chi^{(E \otimes E)}(E) = 4, \quad \chi^{(E \otimes E)}(C_3) = 1 \text{ and } \chi^{(E \otimes E)}(\sigma) = 0. \quad (5.21)$$

Then using (5.20), we see that

$$n_{A_1} = \frac{1}{|G|} \sum_{g \in G} \chi^{(A_1)*}(g) \chi^{(E \otimes E)}(g) = \frac{1}{6} (1 \cdot 4 + 1 \cdot 2 \cdot 1 + 1 \cdot 3 \cdot 0) = 1, \quad (5.22)$$

$$n_{A_2} = \frac{1}{|G|} \sum_{g \in G} \chi^{(A_2)*}(g) \chi^{(E \otimes E)}(g) = \frac{1}{6} (1 \cdot 4 + 1 \cdot 2 \cdot 1 + (-1) \cdot 3 \cdot 0) = 1, \quad (5.23)$$

and

$$n_E = \frac{1}{|G|} \sum_{g \in G} \chi^{(E)*}(g) \chi^{(E \otimes E)}(g) = \frac{1}{6} (2 \cdot 4 + (-1) \cdot 2 \cdot 1 + 0 \cdot 3 \cdot 0) = 1. \quad (5.24)$$

Thus

$$E \otimes E = A_1 \oplus A_2 \oplus E. \quad (5.25)$$

Notice that the direct product $E \otimes E$ contains the identity representation A_1 . It means that one can find combinations of products of basis functions of the E representation that form a singlet.

Example for the realization of Eq. (5.25)

Consider the product $|\psi_\alpha\rangle\langle\psi_\beta|$, in which the wave functions are pseudo-spinors with two components (say, the components associated with the graphene's sublattices). One can choose a basis in the pseudo-spinor space to be, $\langle\psi_A| = (1,0)$, and $\langle\psi_B| = (0,1)$. Assuming the wave functions belong to the E representation, the product $|\psi_\alpha\rangle\langle\psi_\beta|$ is a basis function of the direct product on the left-hand side of Eq. (5.25). This product, with $\alpha = A, B$ and $\beta = A, B$, defines four independent matrices. These can be combined to form a different set of independent matrices: The identity matrix I , the Pauli matrix τ_z^{AB} , and $\tau_\pm^{AB} = \tau_x^{AB} \pm i\tau_y^{AB}$. As we have seen (see character table on page 95), the identity matrix is a basis function of A_1 ; τ_z^{AB} is a basis function of A_2 , while the pair $(\tau_+^{AB}, \tau_-^{AB})$ form a basis function of the E representation. These are the basis functions for the irreducible representations that appear on the right-hand side of Eq. (5.25).

Example: The local Hamiltonian of graphene near the K -point

Let us return to the problem of finding the local Hamiltonian of graphene near the K -point and suppose we look for a linear term in $\mathbf{k}$ (where $\mathbf{k}$ is measured from the K -point). Thus we focus our attention on a term of the form $k|\psi\rangle\langle\psi|$ in the expansion of the Hamiltonian (5.17). Now, being a vector, the irreducible representation of $\mathbf{k}$ is E . If γ is the irreducible representation of the wave function $|\psi\rangle$, then the term $k|\psi\rangle\langle\psi|$ comes from the direct product $E \otimes \gamma \otimes \gamma$. The only way of obtaining a singlet out of this product is by choosing $\gamma = E$: As we have seen in the previous example, the product $E \otimes E$ contains the irreducible representation, E , therefore $E \otimes \gamma \otimes \gamma$ includes the identity representation, A_1 . This argument shows that the local Hamiltonian of graphene may have a term linear in the wavenumber vector

Example: The local Hamiltonian of graphene near the M -point

In this example, we construct the local Hamiltonian of graphene near the M -point from symmetry considerations (time-reversal symmetry showed that it must be quadratic but did not tell anything about degeneracy). First, let us identify the little group associated with the M -point. Since opposite M -points of the Brillouin zone are identical (see Fig. 2-37), the little group is C_{2v} . The character table of the group (on page 75) shows that all

irreducible representations of the group are one-dimensional; hence energy levels are not degenerate. The basis functions associated with these representations are listed below.

Irreducible representation	Basis functions
A_1	x^2, y^2
A_2	xy
B_1	x
B_2	y

We did not include basis functions that depend on the z coordinate because they are irrelevant to our purpose.

Since the product of any one-dimensional irreducible representation by itself yields the identity representation, one cannot construct a singlet from a product of the form $B_{1,2} \otimes \gamma \otimes \gamma$. Therefore, the local Hamiltonian does not contain linear in $\mathbf{k}$ terms. Consequently, the Hamiltonian must be a sum of even powers of the wavenumber components:

$$H = h_1 k_x^2 + h_2 k_y^2 + h_3 k_x^4 + h_4 k_y^4 + h_5 k_x^2 k_y^2 + \dots, \quad (5.26)$$

where h_i are constants.

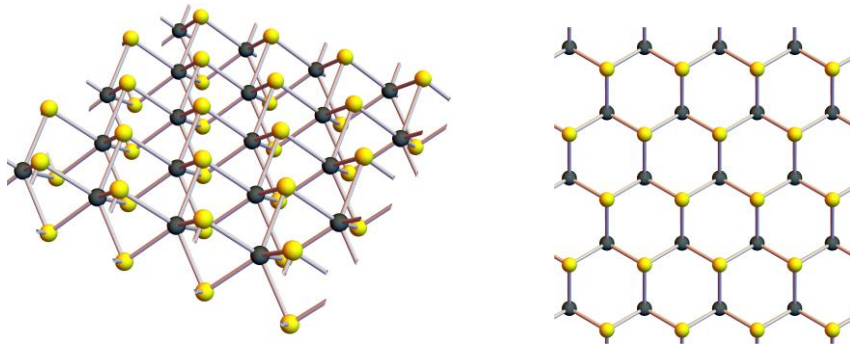


Figure 5-9 The structure of dichalcogenides: Side view (left panel) and top view (right panel)

5.4 Dichalcogenides

Dichalcogenides are a family of two-dimensional semiconductors with a chemical composition MX_2 . Here M represents a transition metal, while X is a chalcogen (i.e., an element from the same column of the Oxygen in the periodic table - such as Sulfur S , Selenium Se , and Tellurium Te). The transition metal atoms are arranged in a two-dimensional hexagonal lattice. On each side of this layer, there is a hexagonal lattice of

chalcogen atoms (with the same lattice constant); see left panel of Fig. 5-9. Here we consider the case where the two lattices of chalcogen atoms transform to each other by simple reflection symmetry. (Another option, which we shall not discuss here, is that an improper rotation, S_6 , relates them.) Thus from a top view, the system looks like a honeycomb lattice composed of two sublattices of different atoms; see right panel of Fig. 5-9.

The Bravais lattice of dichalcogenides is hexagonal as in graphene, and so is the Brillouin zone. However, the point group of the system is C_{3v} rather than C_{6v} . Namely, it contains only two rotations in $\pm 120^\circ$ and three reflections through the planes denoted by the dashed lines in Fig. 5-10.

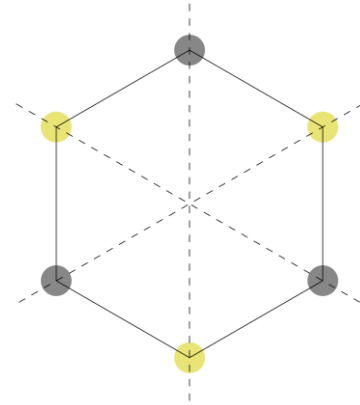


Figure 5-10 Planes of Reflection symmetry in dichalcogenides

Consider the local spectrum near the K -point of the Brillouin zone. Unlike graphene, the little group, in this case, is the cyclic group C_3 . It is not C_{3v} because the atoms on the two sublattices are different; hence the system lacks the symmetry for reflections through the horizontal axis and the two other axes obtained by $\pm 120^\circ$ rotations. However, we know that all irreducible representations of cyclic groups are one-dimensional; therefore, energy levels at the K -point are not degenerate.

The character table of C_3 is listed below:

C_3	E	c_3	c_3^2				
A_1	1	1	1		$x^2 + y^2$	$(x + iy)^3, (x - iy)^3$	τ_z^{AB}, I
A_2	1	$e^{i\frac{2\pi}{3}}$	$e^{-i\frac{2\pi}{3}}$	$x + iy$	$(x - iy)^2$		
A_3	1	$e^{-i\frac{2\pi}{3}}$	$e^{i\frac{2\pi}{3}}$	$x - iy$	$(x + iy)^2$		

We turn to construct the local Hamiltonian of dichalcogenides near the K -point. First, notice that the lack of symmetry between the two sublattices implies no combination of the matrices $\tau_\pm^{AB}$ can form a basis function, because these matrices mix the two sublattices. On the other hand, τ_z^{AB} commutes with rotations (see Eq. 5.10); hence it is a basis function of the identity representation.

Given that all irreducible representations of the little group are one dimensional, the only way of constructing the local spectrum near the K point is by combinations of the basis functions of the identity representation, thus

$$\varepsilon(\delta\mathbf{k}) = \left\{ \Delta + \frac{\hbar^2 \delta k^2}{2m_{\text{eff}}} + h_w \text{Re} \left[e^{i\phi} (\delta k_x + i\delta k_y)^3 \right] \right\} \tau_z^{AB}, \quad (5.27)$$

where $\delta\mathbf{k}$ is measured from the K -point, while Δ , m_{eff} , h_w , and, apparently, also ϕ , are arbitrary constants that cannot be determined from symmetry considerations. In the above formula, we neglected terms that break the symmetry between electrons and holes (i.e., terms proportional to the identity matrix I).

For an arbitrary value of ϕ the Fermi surface near each one of the K -points, rotates as demonstrated in Fig. 5-11. In principle, this structure is allowed by time-reversal symmetry $H(k_K + \delta k_x, \delta k_y) = H(-k_K - \delta k_x, -\delta k_y)$, as shown by the arrow in Fig. 5-11 below.

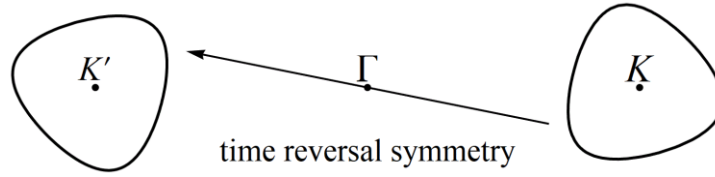


Figure 5-11 Fermi surface near K and K' points for an arbitrary value of ϕ

However, one should take into account that in addition to time-reversal symmetry, the system is also symmetric for reflection through the y -axis; see Fig. 5-10. Thus the Hamiltonian satisfies the property

$$H(k_K + \delta k_x, \delta k_y) = H(-k_K - \delta k_x, \delta k_y) = H(k_{K'} - \delta k_x, \delta k_y). \quad (5.28)$$

This constraint is not taken into account by our local analysis near the K -point. Moreover, as demonstrated in the left panel of Fig. 5-12, this constraint forces ϕ to be an integer multiple of π . Substituting $\phi = \pi$ in Eq. (5.27), we obtain that the local behavior of the energy spectrum near K -point is:

$$\varepsilon_{\pm}(\delta k_x, \delta k_y) = \pm \left[\Delta + \frac{\hbar^2 \delta k^2}{2m_{\text{eff}}} - h_w (\delta k_x^3 - 3\delta k_x \delta k_y^2) \right]. \quad (5.29)$$

Illustrations of the Fermi surface and the energy bands obtained from Eq. (5.29) are depicted in Fig. 5-12 below.

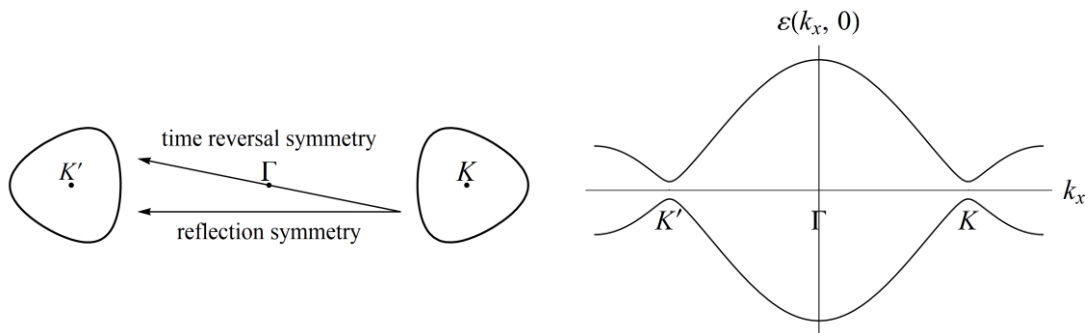


Figure 5-12 The Fermi surface (left panel) and the qualitative behavior of dichalcogenides band structure (right panel).

5.5 Exercises

1. The Brillouin zone of a two-dimensional square lattice, and its special points, Γ , X , and W , are depicted in Fig. 5-13.
- (a) Use time-reversal symmetry to characterize the spectrum of electrons moving in this lattice near the special points of the Brillouin zone.
- (b) Use spatial symmetry considerations to construct the local near the special points. Identify the possible degeneracy of the energy levels at each one of these points.

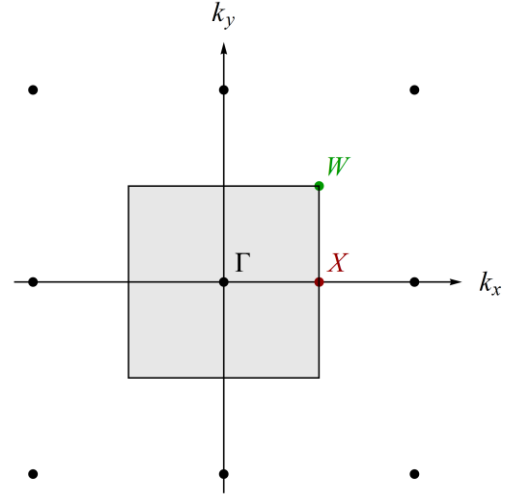


Figure 5-13 The Brillouin zone of a two-dimensional square lattice

2. Show that the local spectrum of graphene near the Γ -point cannot be a Dirac spectrum.

Advice:

- (a) First, notice that the little group of the Γ -point is C_{6v} whose character table appears on page 94.
- (b) Next, show that $(\tau_x^{AB} + i\tau_x^{AB}, \tau_x^{AB} - i\tau_x^{AB})$ is a basis function of the E_2 irreducible representation. Use the following direct products of the irreducible representations of C_{6v} :

$$\begin{aligned} E_1 \otimes E_1 &= E_2 \otimes E_2 = A_1 \oplus A_2 \oplus E_2 \\ E_1 \otimes E_2 &= B_1 \oplus B_2 \oplus E_1 \end{aligned} \tag{5.30}$$

- (c) From the above results, deduce that one cannot obtain a singlet for a term that is linear in $\mathbf{k}$.
3. Fig. 5-14 below shows a side view and a top view of a double layer of graphene. The sublattices in each layer are colored differently to ease the identification of the system symmetry. Find the group symmetry of the system and the little group associated with the K -point. Construct the local Hamiltonian near that point. Take into account the experimental evidence showing that the contribution of Dirac point to the spectrum is negligible.

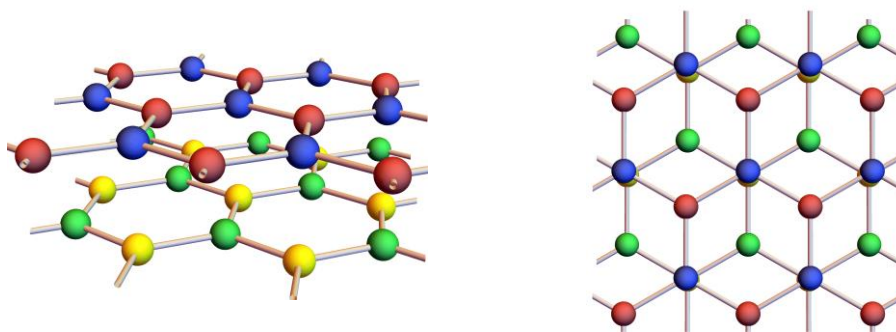


Figure 5-14 A side view and a top view of a double-layer graphene

4. For double-layer graphene, explain why applying an electric field perpendicular to layers opens a gap between the bands.
5. Explain the following rules for the subscripts in Mulliken symbols of products of irreducible representations:

$$\begin{aligned}
 g \otimes g &= g \\
 u \otimes u &= g \\
 u \otimes g &= u
 \end{aligned}
 \tag{5.31}$$

and for the subscript of A or B irreducible representations:

$$\begin{aligned}
 1 \otimes 1 &= 1 \\
 2 \otimes 2 &= 1 \\
 1 \otimes 2 &= 2
 \end{aligned}
 \tag{5.32}$$

6 Extended groups and double groups

The analysis of the spectrum of dichalcogenides near the K -point (presented in the last section of the previous chapter) demonstrated the limitation of local analysis based on the properties of the little group. As we have seen, this analysis left the phase ϕ in Eq. (5.27) undetermined, and we had to employ some global symmetry considerations, that also involved the behavior near the K' -point to set its value. This drawback raises the question of whether one can develop a group theory formalism that treats both K and K' points on equal footing. Noticing that K' point is the time-reversal counterpart of the K point (and vice versa), this generalization amounts to incorporating time-reversal symmetry into the group theoretical approach.

This chapter develops the main tools that incorporate time-reversal symmetry into group theory: *extended groups* in systems where the spin degree of freedom can be ignored and *double groups* for systems with spin-orbit interactions. Next, we use these tools to analyze the effect of spin-orbit interaction on graphene and to present a family of materials called *topological insulators*.

6.1 Extended groups

This section presents the group theoretical framework that treats the K and the K' points on equal footing. For simplicity, we begin with the example of dichalcogenides. Let us choose the primitive basis vectors of the Bravais lattice to be

$$\mathbf{a}_1 = a(1,0) \quad \text{and} \quad \mathbf{a}_2 = \frac{a}{2}(1, \sqrt{3}), \quad (6.1)$$

where a is the lattice constant. The corresponding primitive basis vectors of the reciprocal lattice are

$$\mathbf{b}_1 = \frac{2\pi}{\sqrt{3}a}(-\sqrt{3}, 1) \quad \text{and} \quad \mathbf{b}_2 = \frac{4\pi}{\sqrt{3}a}(0, 1) \quad (6.2)$$

and one can quickly check that the wavenumber vector at the K point (the one along the k_x axis) is

$$\mathbf{k}_K = \frac{4\pi}{3a}(1, 0). \quad (6.3)$$

The wave function of the system at this point can be decomposed according to Bloch's theorem:

$$\psi_K(\mathbf{r}) = d(\mathbf{r})\phi(\mathbf{r}), \quad (6.4)$$

where

$$d(\mathbf{r}) = \exp(i\mathbf{k}_K \cdot \mathbf{r}), \quad (6.5)$$

while $\phi(\mathbf{r})$ is a periodic function with a periodicity of the unit cell. This component of the wave function does not play an essential role in the following discussion and will be suppressed from now on.

Time reversal symmetry gives

$$\psi_{K'}(\mathbf{r}) = \psi_K^*(\mathbf{r}) = d^*(\mathbf{r})\phi^*(\mathbf{r}) \quad (6.6)$$

with

$$d^*(\mathbf{r}) = \exp(-i\mathbf{k}_K \cdot \mathbf{r}) = \exp(i\mathbf{k}_{K'} \cdot \mathbf{r}), \quad (6.7)$$

where we have used the relation $\mathbf{k}_{K'} = -\mathbf{k}_K$.

Let us define a vector wave function whose components are the wave function at the K -point and its time-reversed counterpart, i.e., the wave function at the K' -point:

$$\Psi(\mathbf{r}) = \begin{pmatrix} d(\mathbf{r}) \\ d^*(\mathbf{r}) \end{pmatrix}. \quad (6.8)$$

The action of the time-reversal operator, Θ , on this function is to swap K and K' points, i.e.

$$\Theta\Psi = \tau_x^{KK'}\Psi, \quad (6.9)$$

where $\tau_x^{KK'}$ is the Pauli matrix that acts in the “valleys space” of K and K' points. It follows that $\Theta^2 = I$; hence Ψ is a pseudospinor and not a spinor (for which $\Theta^2 = -I$, see Eq. (3.18)).

We turn now to identify the action of symmetry operations on the wave function Ψ . Starting with reflection through the vertical axis, σ_y , (see Fig. 5-10), we notice that this reflection $x \leftrightarrow -x$ implies that $\mathbf{k}_K \cdot \mathbf{r} \leftrightarrow -\mathbf{k}_K \cdot \mathbf{r} = \mathbf{k}_{K'} \cdot \mathbf{r}$, thus

$$\sigma_y\Psi = \begin{pmatrix} d(\sigma_y\mathbf{r}) \\ d^*(\sigma_y\mathbf{r}) \end{pmatrix} = \begin{pmatrix} d^*(\mathbf{r}) \\ d(\mathbf{r}) \end{pmatrix} = \tau_x^{KK'}\Psi. \quad (6.10)$$

Next, using arguments similar to those presented in the previous chapter for graphene, the rotation operation is described by:

$$c_3\Psi = \begin{pmatrix} d(c_3\mathbf{r}) \\ d^*(c_3\mathbf{r}) \end{pmatrix} = \exp\left(-i\frac{2\pi}{3}\tau_z^{KK'}\right)\Psi. \quad (6.11)$$

Now, let us consider the action of translation by one of the primitive basis vectors. First, notice that

$$T_{a_1} d(\mathbf{r}) = d(T_{a_1} \mathbf{r}) = \exp\left(i \frac{4\pi}{3}\right) d(\mathbf{r}) \quad (6.12)$$

because

$$\begin{aligned} d(T_{a_1} \mathbf{r}) &= \exp(i \mathbf{k}_K \cdot T_{a_1} \mathbf{r}) = \exp[i \mathbf{k}_K \cdot (\mathbf{r} + \mathbf{a}_1)] \\ &= \exp(i \mathbf{k}_K \cdot \mathbf{a}_1) \exp(i \mathbf{k}_K \cdot \mathbf{r}) = \exp\left(i \frac{4\pi}{3}\right) d(\mathbf{r}), \end{aligned} \quad (6.13)$$

where the last equality follows from Eqs. (6.1) and (6.3). Repeating the same procedure for $d^*(\mathbf{r})$ and using, $\exp(i 4\pi/3) = \exp(-i 2\pi/3)$, we obtain

$$T_{a_1} \Psi = \exp\left(-i \frac{2\pi}{3} \tau_z^{KK'}\right) \Psi. \quad (6.14)$$

A similar calculation for translation by the second primitive basis vector, $\mathbf{a}_2$, yields

$$T_{a_2} \Psi = \exp\left(i \frac{2\pi}{3} \tau_z^{KK'}\right) \Psi. \quad (6.15)$$

From the last two equations, we see that $T_{a_1} T_{a_2} = E$. It implies that the translation operators acting on Bloch wave functions, with wavenumbers $\mathbf{k}_K$ and $\mathbf{k}_{K'}$, form a group isomorphic to the cyclic group C_3 . In other words, if we set $t = T_{a_1}$, then $t^2 = T_{a_2}$ and $t^3 = E$.

The group obtained from the product of the elements of C_{3v} by these translations,

$$C_{3v}'' = C_{3v} \otimes \{E, t, t^2\}, \quad (6.16)$$

is called the *extended group* of C_{3v} . It contains 18 elements (obtained from the multiplication of the six elements of C_{3v} by the three translation operators), and six conjugacy classes. The group has two one-dimensional irreducible representations and four two-dimensional irreducible representations, such that $2 \cdot 1^2 + 4 \cdot 2^2 = 18$. To see why it is indeed so, notice that any extended group has the representations of the original group because one can always choose the identity representation for the translation operations. Therefore, knowing we have six conjugacy classes and three irreducible representations of the original group, we should have three additional irreducible representations whose sum of the square of their dimensions is 12. The only possible way of achieving that is $3 \cdot 2^2 = 12$. The character table of the extended group is listed on the next page.

C_{3v}''	E	t, t^2	$2c_3$	$c_3t, c_3^2t^2$	c_3^2t, c_3t^2	$3\sigma, 3\sigma t, 3\sigma t^2$
A_1	1	1	1	1	1	1
A_2	1	1	1	1	1	-1
E	2	2	-1	-1	-1	0
E_1''	2	-1	2	-1	-1	0
E_2''	2	-1	-1	2	-1	0
E_3''	2	-1	-1	-1	2	0

To identify which one of these irreducible representations is associated with the pseudospinor Ψ , let us calculate the characters of the extended group operations: $\chi(\sigma) = \chi(\sigma t) = \chi(\sigma t^2) = 0$, $\chi(c_3) = \chi(c_3^2) = \chi(t) = \chi(t^2) = \chi(c_3t) = \chi(c_3^2t^2) = -1$, and $\chi(E) = \chi(c_3^2t) = \chi(c_3t^2) = 2$. These characters are those of the two-dimensional irreducible representation denoted by E_3'' in the character table. It implies that the energy levels at the points of K and K' are degenerate, as illustrated in Fig. 6-1.

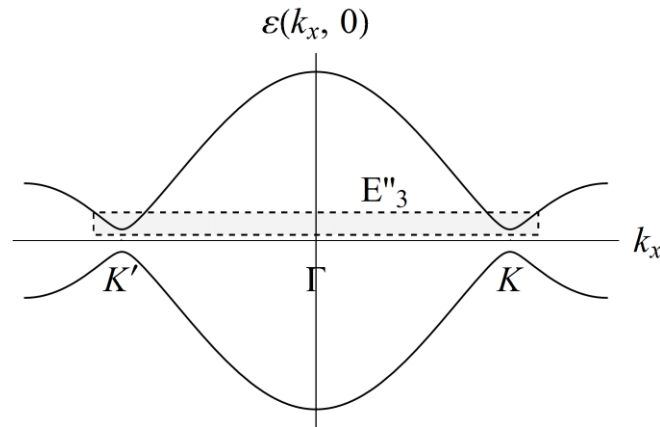


Figure 6-1 An illustration of the energy level degeneracy at the K and K' points

We now briefly repeat this procedure for graphene. Here, the wave function should include four components: two to describe the sublattices and two for the valleys. It is convenient to present the wave function in the form:

$$\Psi = \begin{pmatrix} d_A \\ d_B \\ d_A^* \\ d_B^* \end{pmatrix} \begin{array}{c} K \\ \text{---} \\ K' \end{array} \quad (6.17)$$

where the subscript A and B refer to the sublattices of the graphene. For this choice, the time-reversal operation is described by:

$$\Theta\Psi = \Psi^* = \begin{pmatrix} d_A^* \\ d_B^* \\ d_A \\ d_B \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_A \\ d_B \\ d_A^* \\ d_B^* \end{pmatrix} = \tau_x^{KK'} \otimes I^{AB} \Psi. \quad (6.18)$$

where I^{AB} is the 2×2 identity matrix acting within the sublattice space. Repeating the procedure described above, we obtain the extended group:

$$C_{6v}'' = C_{6v} \otimes \{E, t, t^2\}. \quad (6.19)$$

It contains 36 elements and nine conjugacy classes: $E, \{t, t^2\}, \{c_2, c_2t, c_2t^2\}, \{c_3, c_3^2\}, \{c_3t, c_3^2t^2\}, \{c_3t^2, c_3^2t\}, \{2c_6, 2c_6t, 2c_6t^2\}, \{3\sigma_v, 3\sigma_vt, 3\sigma_vt^2\}$, and $\{3\sigma_d, 3\sigma_dt, 3\sigma_dt^2\}$. Thus, the extended group includes the following irreducible representations: four one-dimensional, four two-dimensional, and one four-dimensional so that $4 \cdot 1^2 + 4 \cdot 2^2 + 1 \cdot 4^2 = 36$. The four-dimensional irreducible representation is the one that describes the fourfold degeneracy of the energy levels at the K and K' points.

6.2 Double groups

In the previous section, we showed that additional features of the system, such as time-reversal symmetry, can be taken into account by extending the symmetry group of the system. In this section, we adopt a similar approach in order to describe systems with spin $\frac{1}{2}$ particles. The generalized groups, in this case, are called *double groups*.

The wave function that describes a particle of spin $\frac{1}{2}$ is a spinor,

$$\Psi(\mathbf{r}) = \begin{pmatrix} \psi_{\uparrow}(\mathbf{r}) \\ \psi_{\downarrow}(\mathbf{r}) \end{pmatrix}, \quad (6.20)$$

where the subscript $s = \uparrow \downarrow$ denotes the projection of the spin state on some arbitrary direction (say the z axis). Up to now, we have considered the action of symmetry elements only on the spatial coordinate $\mathbf{r}$; however, for a spin $\frac{1}{2}$ particle, one should also take into account that these symmetry operations also act on the spin. In particular, that rotation of a spin by 360° returns the wave function (6.20) to itself with a minus sign. Thus, to construct a representation of the symmetry of a spinor, one should add a symmetry element, Q , that describes the rotation of the spin in 360° around some arbitrary axis. This element satisfies the condition $Q^2 = E$, and its relation to the rotation operations are:

$$c_n^n = Q \text{ and } c_n^{2n} = E. \quad (6.21)$$

Inversion commutes with any rotation, therefore $i^2 = E$. On the other hand, inversion can also be represented as a product of rotation in 180° followed by reflection through the plane perpendicular to the rotation axis, $i = \sigma c_2$. This relation together with $i^2 = E$ implies that

$$\sigma^2 = Q \text{ and } \sigma^4 = E \quad (6.22)$$

The double group of G denoted by G' is obtained from the direct product $G \otimes \{E, Q\}$.

The rules for building the character tables of double groups are the following:

1. If the set of elements in the original group, $\{g_i\}$, forms a conjugacy class, then $\{g_i\}$ and $\{Qg_i\}$ are two separate conjugacy classes of the double group, with two exceptions listed below.
2. The first exception refers to the case of c_2 rotation for which there is another c_2' rotation around a perpendicular axis or a reflection plane containing the c_2 rotation axis. In this case c_2 , and Qc_2 are in the same conjugacy class (see Ex. 1).
3. The second exception to rule No. 1 is for a reflection σ when there is another perpendicular reflection plane or a c_2 rotation axis within the reflection plane. Also in this case σ and $Q\sigma$ belong to the same conjugacy class (see Ex. 1).
4. Any irreducible representation of the original group is also an irreducible representation of the double group with the same characters.
5. Apart from the irreducible representations mentioned in the previous rule, there are additional irreducible representations such that the total number of irreducible representations is the number of conjugacy classes. For the spin $\frac{1}{2}$ irreducible representations of the group, the character of the element Qg is $\chi(Qg) = -\chi(g)$. When rules No. 2 and 3 apply, g and Qg are in the same conjugacy class; hence their characters should vanish because $\chi(g) = \chi(Qg) = -\chi(g)$.

Example: The character table of the double group D_2'

The dihedral group D_2 contains three c_2 -rotations around three perpendicular axes (see Figure 4-7 and the character table below). For this group, rule No. 2 applies, and the conjugacy classes of the double group are: $\{E\}$, $\{Q\}$, $\{c_2, Qc_2\}$, $\{c_2', Qc_2'\}$, and $\{c_2'', Qc_2''\}$. Thus, there are five conjugacy classes and eight elements in the double group. Since all the one-dimensional irreducible representations of D_2 appear in the double group, the additional irreducible representation of the double group must be two-dimensional to ensure that $1^2 + 1^2 + 1^2 + 1^2 + 2^2 = 8$. This irreducible representation is simply the

representation of spin $\frac{1}{2}$ particle ($\uparrow, \downarrow$). The character table, listed below, is easily constructed using rule No. 5.

The appearance of a two-dimensional irreducible representation in the double group manifests the Kramer degeneracy of spin $\frac{1}{2}$ systems with time-reversal symmetry.

D_2	E	c_2	c'_2	c''_2
A	1	1	1	1
B ₁	1	1	-1	-1
B ₂	1	-1	1	-1
B ₃	1	-1	-1	1

D'_2	E	Q	$\{c_2, Qc_2\}$	$\{c'_2, Qc'_2\}$	$\{c''_2, Qc''_2\}$
A	1	1	1	1	1
B ₁	1	1	1	-1	-1
B ₂	1	1	-1	1	-1
B ₃	1	1	-1	-1	1
E'	2	-2	0	0	0

Example: The character table of the double group C'_{3v}

The C_{3v} contains six symmetry elements; therefore, the double group has 12 elements:

$$E; \sigma_1; \sigma_2; \sigma_3; c_3; c_3^2; Q; Q\sigma_1; Q\sigma_2; Q\sigma_3; Qc_3; Qc_3^2 \quad (6.23)$$

Neither rule No. 2 nor rule No. 3 apply; hence the conjugacy classes are: $\{E\}$, $\{3\sigma\}$, $\{2c_3\}$, $\{Q\}$, $\{3Q\sigma\}$, and $\{2Qc_3\}$. Since the double group must contain the irreducible representations of the original group (by choosing the identity representation for $\{E, Q\}$), C'_{3v} must contain three additional irreducible representations, one two-dimensional and two one-dimensional representations. The character table, in this case, takes the form:

C'_{3v}	E	$2c_3$	3σ	Q	$2Qc_3$	$3Q\sigma$
A ₁	1	1	1	1	1	1
A ₂	1	1	-1	1	1	-1
E	2	-1	0	2	-1	0
E'	2	1	0	-2	-1	0
A'	$\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$	-1	i	-1	1	$-i$
A' [*]		-1	$-i$	-1	1	i

The first three irreducible representations in this table describe a spin-zero particle, while the other three describe a spin $\frac{1}{2}$ particle. Here, the characters of the additional pair of one-dimensional representations are identified using the property that characters of one-dimensional representations are representations by themselves and the relation $\sigma^2 = Q$.

Thus $\chi(\sigma) = \Gamma(\sigma) = \sqrt{\Gamma(\sigma^2)} = \sqrt{\Gamma(Q)} = \pm\sqrt{-1}$. The other entries of the table are obtained from the orthogonality condition of the rows. The pair of the one-dimensional irreducible (complex) representations of the spin are usually grouped together, $A' \oplus A'^*$, and considered as a two-dimensional representation (called *separably-degenerate representation*). It is because, by Kramer's theorem, this pair of one-dimensional representations must be associated with degenerate states.

6.3 Spin-orbit interaction in graphene – the Kane Mele term

This section aims to show that spin-orbit interaction in graphene opens a gap at the K -points. For this purpose, we should extend our description of graphene in two manners: (a) Include time-reversal symmetry by doubling the number of components of the wave function to take into account the valley space (see Eq. (6.17)); (b) Double the number of components, once again, in order to take in to account the spin degree of freedom. Thus, the wave function contains eight components classified by the sublattices A and B , the valleys K and K' , and the spin states $s = \uparrow\downarrow$.

Now let us try to identify the largest possible degeneracy of the energy levels of graphene in the presence of spin-orbit interaction. The group C_{6v} contains c_2 rotation, and all the reflection planes contain the corresponding axis of this rotation; therefore, the conjugacy classes of the double group C'_{6v} are: $E, Q, 2C_6, 2QC_6, 2C_3, 2QC_3, \{c_2, Qc_2\}, \{3\sigma_v, 3Q\sigma_v\}$, and $\{3\sigma_d, 3Q\sigma_d\}$. Altogether we have nine conjugacy classes and 24 elements. Six out of the nine irreducible representations are of the original group; therefore, the additional three irreducible representations must satisfy the condition $l_7^2 + l_8^2 + l_9^2 = 12$, which means that all spin $\frac{1}{2}$ representations are two-dimensional. The extended group obtained from this double group contains 15 conjugacy classes which can be deduced from those of C''_{6v} : $E, Q, \{t, t^2\}, \{Qt, Qt^2\}, \{c_2, Qc_2, c_2t, Qc_2t, c_2t^2, Qc_2t^2\}, \{2c_6, 2c_6t, 2c_6t^2\}, \{2QC_6, 2QC_6t, 2QC_6t^2\}, \{3\sigma_v, 3\sigma_vt, 3\sigma_vt^2, 3Q\sigma_v, 3Q\sigma_vt, 3Q\sigma_vt^2\}, \{3\sigma_d, 3\sigma_dt, 3\sigma_dt^2, 3Q\sigma_d, 3Q\sigma_dt, 3Q\sigma_dt^2\}, \{c_3, c_3^2\}, \{Qc_3, Qc_3^2\}, \{c_3t, c_3^2t^2\}, \{Qc_3t, Qc_3^2t^2\}, \{c_3t^2, c_3^2t\}$ and $\{Qc_3t^2, Qc_3^2t\}$. This extended double group contains 72 elements and 15 irreducible representations. Nine of these irreducible representations are associated with the original double group, and six more should satisfy the condition: $l_{10}^2 + l_{11}^2 + l_{12}^2 + l_{13}^2 + l_{14}^2 + l_{15}^2 = 48$. The only way of satisfying this equation (according to Kramer's degeneracy) is with $l_{10} = l_{11} = l_{12} = l_{13} = 2$ and $l_{14} = l_{15} = 4$. Thus, the largest irreducible representation of the extended double group (describing graphene with spin-orbit interactions and time-reversal symmetry) is four-dimensional. However, noticing that time-reversal symmetry implies degeneracy of $\mathbf{k} \rightarrow -\mathbf{k}$ with $s \rightarrow -s$, while the inversion symmetry of graphene implies degeneracy of $\mathbf{k} \rightarrow -\mathbf{k}$ with $s \rightarrow s$ we see

that fourfold degeneracy already accounts for Kramer's degeneracy and the valley degeneracy. Therefore, there is no possibility of having Dirac points in the spectrum as these would imply an 8-fold degeneracy of the energy levels. In other words, if we were able to tune up the strength of the spin-orbit interaction in graphene, we would have seen that it opens a gap in the Dirac spectrum.

Kane and Mele showed that the constant matrix,

$$H_{\text{so}} = \lambda_{\text{so}} \tau_z^S \otimes \tau_z^{KK'} \otimes \tau_z^{AB}, \quad (6.24)$$

describes spin-orbit interaction in graphene, where λ_{so} is a constant that determines the strength of the interaction. This Hamiltonian is called the Kane-Mele term. It does not break any spatial symmetry, and one can check that it is invariant under all required symmetry operations:

- (a) Time reversal symmetry: Taking into account that the time-reversal operator in the spin space is given by Eq. (3.16), while in the valleys and sublattices spaces by Eq. (6.18), we have $\Theta = -i\tau_y^S \hat{K} \otimes \tau_x^{KK'} \otimes I^{AB}$ (where $\hat{K}$ is the complex conjugate operator), and one can check that

$$H_{\text{so}} = \Theta^{-1} H_{\text{so}} \Theta. \quad (6.25)$$

- (b) Inversion symmetry: This symmetry exchanges the sublattices A and B , and takes K point into K' point and vice versa. Inversion does not influence the spin; thus, the inversion operator is $\hat{I} = I^S \otimes \tau_x^{KK'} \otimes \tau_x^{AB}$. One can check that the spin-orbit Hamiltonian is invariant under this symmetry operation.

$$H_{\text{so}} = \hat{I}^{-1} H_{\text{so}} \hat{I}. \quad (6.26)$$

The proof of Eqs. (6.25) and (6.26) is given as an exercise.

- (c) Finally, since the spin-orbit Hamiltonian (6.24) is independent of $\mathbf{k}$, reflections and rotation of the C_{3v} group (i.e., those operations that do not mix the sublattices), also leave the Hamiltonian invariant. All other symmetry operations are obtained from those of C_{3v} and inversion.

A derivation of the Kane-Mele term is given as an exercise (No. 4) in the next chapter. Before discussing its implication, we comment that there are, of course, additional contributions to the spin-orbit interaction which are $\mathbf{k}$ dependent. However, since these depend on the deviation, $\delta\mathbf{k}$, from the K point (or the K' point) their contribution is expected to be small compared to the Kane-Mele term.

When employing symmetry considerations to deduce the form of the Hamiltonian, the “rules of the game” are that if there is no symmetry that forbids the existence of some term in the Hamiltonian, then it generically exists. Therefore, the local Hamiltonian of graphene near the K -point should have the form

$$\hat{\varepsilon}(\delta\mathbf{k}) = \hbar v I^S \otimes \delta\mathbf{k} \cdot \boldsymbol{\tau}^{AB} + \lambda_{so} \tau_z^S \otimes \tau_z^{AB}. \quad (6.27)$$

This approximation holds only close enough to the K -point so that higher-order terms in $\delta\mathbf{k}$ can be neglected. Also, notice that the Kane-Mele term is diagonal in the spin space; therefore, spin is conserved.

Diagonalization of the above Hamiltonian gives

$$\varepsilon_{\pm, \uparrow\downarrow}(\delta\mathbf{k}) = \pm \sqrt{(\hbar v)^2 \delta k^2 + \lambda_{so}^2}. \quad (6.28)$$

From here, we see that spin-orbit interaction opens a gap in the spectrum, as shown in Fig. 6-2. This gap, in principle, turns the graphene into a band insulator. However, in reality, λ_{so} is very small - in units of Kelvin degrees, it is smaller than $1^0 mK$. For this reason, no gap has ever been observed in graphene.

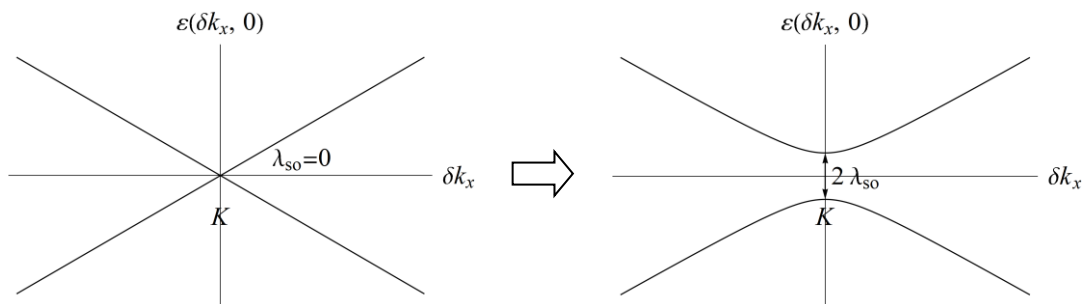


Figure 6-2 The opening of a gap in the energy spectrum of graphene due to spin-orbit interaction.

6.4 Topological insulators

Viewing graphene as a band insulator due to spin-orbit coupling (and when Fermi energy is in the middle of the gap), is to a large extent, just an academic issue. Nevertheless, it constitutes a simple example by which we can introduce the idea of *topological insulators*.

The analysis presented in the previous section holds for an infinite system. In a finite system, translation symmetry breaks - it does not apply in the direction perpendicular to the system edge - hence our result does not necessarily apply. Moreover, as we already

know, the Rashba term associated spin-orbit interaction near the system's boundary results in a Dirac point at zero energy (see Fig. 3-1). This degeneracy point is protected by time-reversal symmetry, and it is always there no matter how small the spin-orbit interaction is. The wave function associated with this Dirac point describes a particle that moves along the boundary. Its spin is perpendicular to both the particle velocity and the normal to the boundary.

How can we reconcile the behavior of the bands within the bulk with that near the boundary? In Fig. 6-3 we present a schematic picture of the energy levels in a semi-infinite graphene layer with a boundary parallel to the x axis (where the normal to the boundary is in y direction). The solid black line represents the energy level of the bulk spectrum associated with $k_y = 0$. Above this line, there are additional levels (not drawn) that correspond to $k_y \neq 0$. The red and blue lines represent the surface states of the system that describe a particle that moves along the boundary. The red curve is associated with a right-moving particle ($v_x = \partial\varepsilon/\partial\hbar k_x$) with spin pointing in the z direction, while the blue line corresponds to a left moving particle with spin pointing in the opposite direction.

In a stripe of graphene, particles moving in the same direction along the two stripe boundaries have opposite spins. In a closed sample, the particles move along the boundary in opposite directions and opposite spins (as required by time-reversal symmetry)

A *topological insulator* is a system where an electric current can flow only along the system's boundary, while the bulk is a band insulator. The surface states are protected by time-reversal symmetry; namely, they are not destroyed by defects or disorder. However, in the presence of an external magnetic field (or magnetic impurities), the system is not time-reversal symmetric anymore, a gap also opens in the surface states, and the system becomes an insulator.

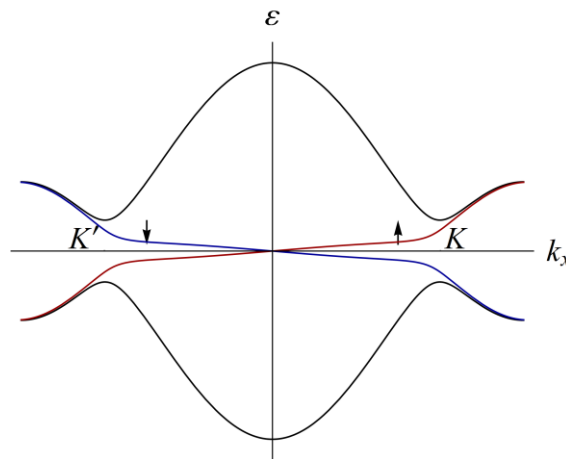


Figure 6-3 An illustration of the spectrum of a topological insulator with the boundary energy states

6.5 Exercises

1. Prove rules Nos. 2&3 for the construction of double groups (on page 116).
2. Prove Eqs. (6.25) and (6.26).
3. Construct the character tables for the double groups D'_4 , C'_{4v} , and D'_{2d} (notice these are isomorphic groups)
4. Find the dimensions of the irreducible representations of the double group O'_h .

Advice: To simplify the analysis, use the following property of the octahedral group, $O_h = O \otimes C_i$, where C_i is the group that contains the identity and the inversion elements, while O is the symmetry group of an octahedron (see Fig. 4-22, and character table on the same page). Use the five rules on page 116 to identify the conjugacy classes of the double group O' , and with this information, deduce the dimensions of its irreducible representations. Using these results, find the dimensions of the irreducible representations of the double group O'_h .

5. Rutile is a mineral made of titanium dioxide, TiO_2 . The Bravais lattice of this crystal belongs to the tetragonal system, and the structure of its unit cell is shown in Fig. 6-4, where oxygen atoms are marked in red. Find the largest possible degeneracy of the energy levels at the Γ point, taking into account spin-orbit interaction.

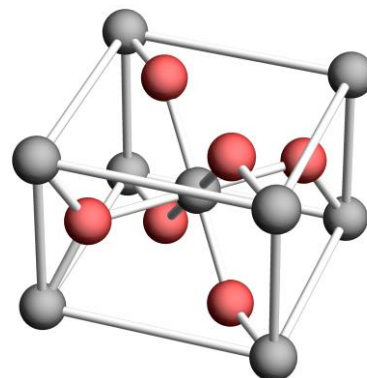


Figure 6-4 The unit cell of Rutile

Advice: Assume that the lattice is described by the D_{4h} point group (see comment below). In order to identify possible degeneracy of the energy levels, you have to find the dimensionality of the irreducible representations of the corresponding double group. To simplify the calculation, use the property $D_{nh} = D_n \otimes C_i$ which holds for even n . For D'_4 , use the results of Ex. 3.

Comment: At first sight, the point group describing the symmetries of the unit cell of Rutile is D_{2h} (which contains eight elements). However, the space group of this crystal is non-symmorphic. There are additional symmetry operations. In particular, a rotary translation $\{c_4 | \mathbf{l}\}$ with $\mathbf{l} = (\hat{x} + \hat{y})a/2 + \hat{z}c/2$, and seven additional operations: $\{c_4^3 | \mathbf{l}\}$, two $\{c_2' | \mathbf{l}\}$, $\{S_4 | \mathbf{l}\}$, $\{S_4^3 | \mathbf{l}\}$, and two $\{\sigma_v | \mathbf{l}\}$. These operations and those of D_{2h} define a space group that is isomorphic to the direct product of the translation group by D_{4h} .

7 Methods for calculating band structure

The strength of the symmetry approach that we have developed in the previous chapters is its ability to provide a qualitative picture of the band structure even without detailed information about the system. The main shortcomings of this approach are: (a) It does not account for the global behavior of the energy bands in the Brillouin zone; (b) It cannot tell us the position of the Fermi level which selects the relevant bands; and (c) It does not provide the constants that characterize the spectrum, such as the velocity v , the triangular warping constant $\hbar_w$ near the K -point of graphene, and the value of the spin-orbit coupling, λ_{so} , that opens the gap in the spectrum. To get this information, one has to diagonalize the Hamiltonian of the system. However, from the simple example of the Kronig Penney model (discussed in the first chapter), we already know that obtaining exact analytical solutions for the energy bands of a system is a difficult task. On the other hand, we usually do not need the band structure's complete information - an approximate description is sufficient.

In this chapter, we discuss three approximation methods for calculating the band structure of a system: The nearly free electron approximation, the tight-binding approximation, and the $k \cdot p$ approximation. These methods are complementary in the sense that they describe the system in different limits and under different physical assumptions.

7.1 The nearly free electron approximation

The nearly free electron approximation relies upon the assumption that the effect of the lattice is perturbative. It is instructive to present this approximation, first, for one-dimensional systems. Consider the problem of a particle moving in a one-dimensional periodic potential. The Schrödinger equation that describes this system is:

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + u(x) \right] \psi(x) = \varepsilon \psi(x), \quad (7.1)$$

with the potential

$$u(x+a) = u(x) = \sum_j u_j \exp(ijbx). \quad (7.2)$$

Here $b = 2\pi/a$ is the reciprocal lattice constant, j is an integer, and u_j are the Fourier expansion coefficients of the potential that satisfy the condition $u_{-j} = u_j^*$ (to ensure that $u(x)$ is a real function).

From Bloch's theorem it follows that the solutions of this equation have the form $\psi(x) = \exp(ikx)\phi_k(x)$, where $\phi_k(x+a) = \phi_k(x)$ is a periodic function with the same period of the potential. Therefore, $\phi_k(x)$ can also be expanded in a similar Fourier series, and the wave function may be presented in the form

$$\psi(x) = \exp(ikx) \sum_j c_k^{(j)} \exp(ijbx), \quad (7.3)$$

where $c_k^{(j)}$ are the expansion coefficients of $\phi_k(x)$. Substituting Eqs. (7.2) and (7.3) in (7.1) and using the convolution theorem, we obtain that, in Fourier's space, the Schrödinger equation is:

$$\varepsilon_0(k+jb)c_k^{(j)} + \sum_{j'} u_{j-j'} c_k^{(j')} = \varepsilon(k)c_k^{(j)}, \quad \text{with} \quad \varepsilon_0(k) = \frac{\hbar^2 k^2}{2m}. \quad (7.4)$$

This equation represents a set of an infinite number of coupled equations (i.e., a matrix equation of infinite size). Their solutions are equivalent to the exact diagonalization of the problem. Yet, usually, the coefficients u_j decay fast as a function of $|j|$ (recall that the Fourier expansion coefficients of analytic functions decay faster than any power-law function), and it might be enough to take into account only a few harmonics of the periodic potential. In what follows, we set $u_0 = 0$, without loss of generality, because this coefficient only shifts all energy levels by a constant. We also assume that the periodic potential is sufficiently weak and can be treated perturbatively.

In the first (zeroth) approximation, we set $u(x) = 0$, but its periodicity is taken into account. In this limit, the band structure is obtained by folding the spectrum of a free particle into the first Brillouin zone, as demonstrated in Fig. 7-1. Namely, the parabolic spectrum, $\varepsilon_0(k)$, is duplicated an infinite number of times, and each copy is shifted by the reciprocal lattice constant, $\varepsilon_0(k+jb)$. The energy levels within the first Brillouin zone constitute the zeroth approximation for the spectrum called the “empty lattice approximation”. This approximation is merely a

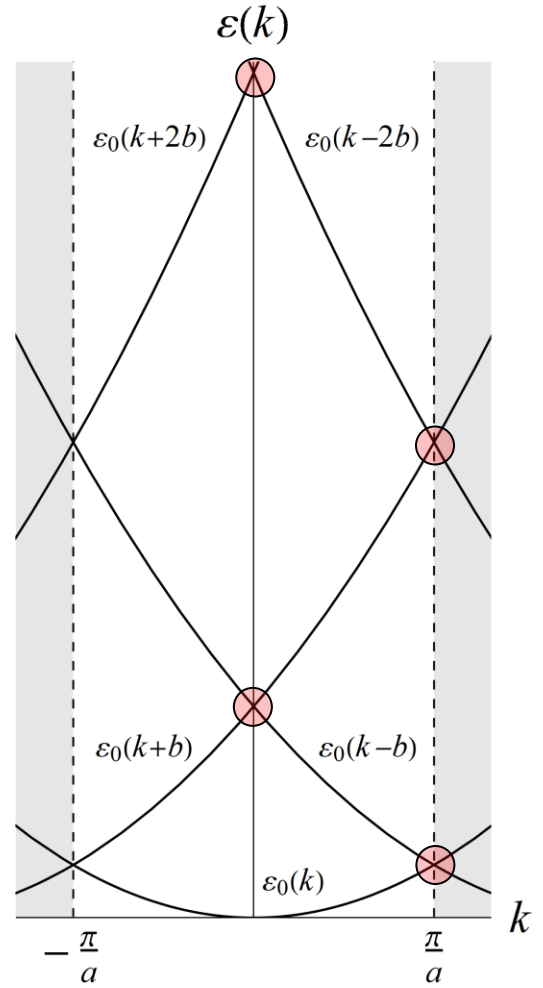


Figure 7-1 The empty lattice approximation

different way of counting the energy levels of a free particle.

Now let us include the effect of the potential. Assuming $u(x)$ is sufficiently weak, one expects it to open gaps at the degeneracy points of the energy levels obtained in the zeroth approximation, see red disks on Fig. 7-1. Consider, for example, the lowest degeneracy point, at the edge of the Brillouin zone, $k = b/2 = \pi/a$, where the functions $\varepsilon_0(k)$ and $\varepsilon_0(k-b)$ intersect. Assuming that the harmonics of the periodic potential decay fast, it is sufficient to consider only the coefficients $u_{\pm 1}$ and neglect all the others. With this assumption, one can consider the subspace of degenerate wave functions whose coefficients are $(c_k^{(0)}, c_k^{(-1)})$, hence Eq. (7.4) reduces to the 2×2 matrix equation:

$$\begin{pmatrix} \alpha + \hbar v \delta k + \beta \delta k^2 & u_1 \\ u_1^* & \alpha - \hbar v \delta k + \beta \delta k^2 \end{pmatrix} \begin{pmatrix} c_k^{(0)} \\ c_k^{(-1)} \end{pmatrix} = \varepsilon(k) \begin{pmatrix} c_k^{(0)} \\ c_k^{(-1)} \end{pmatrix}, \quad (7.5)$$

with $\delta k = k - b/2$, while

$$\alpha = \frac{\hbar^2 b^2}{8m}, \quad v = \frac{\hbar b}{2m}, \quad \text{and} \quad \beta = \frac{\hbar^2}{2m}. \quad (7.6)$$

Diagonalization of the above matrix gives the behavior of the energy levels near the lowest degeneracy point close to the edge of the Brillouin zone:

$$\varepsilon_{\pm}(\delta k) = \alpha + \beta \delta k^2 \pm \sqrt{|u_1|^2 + \hbar^2 v^2 \delta k^2}. \quad (7.7)$$

Thus, the gap that opens between the two lowest energy levels at the edge of the Brillouin zone, (i.e. when $\delta k = 0$) is $2|u_1|$.

Now, let us take a look at the intersection of the energy levels $\varepsilon_0(k \pm b)$ at $k = 0$. Considerations similar to those presented above show that in the subspace of these degenerate levels, whose coefficients are $(c_k^{(1)}, c_k^{(-1)})$, the Schrödinger equation (7.4) reduces to

$$\begin{pmatrix} 4\alpha + 2\hbar v k + \beta k^2 & u_2 \\ u_2^* & 4\alpha - 2\hbar v k + \beta k^2 \end{pmatrix} \begin{pmatrix} c_k^{(1)} \\ c_k^{(-1)} \end{pmatrix} = \varepsilon(k) \begin{pmatrix} c_k^{(1)} \\ c_k^{(-1)} \end{pmatrix}, \quad (7.8)$$

and diagonalization gives

$$\varepsilon_{\pm}(k) = 4\alpha + \beta k^2 \pm \sqrt{|u_2|^2 + 4\hbar^2 v^2 k^2}. \quad (7.9)$$

Hence, the second harmonic of the periodic potential determines the gap in the energy levels, which is $2|u_2|$.

This result raises the following question: What happens if $u_2 = 0$? Are the energy levels remain degenerate? To answer this question, consider the Schrödinger equation in a larger subspace containing three components $(c_k^{(1)}, c_k^{(0)}, c_k^{(-1)})$ (but with $u_2 = 0$):

$$\begin{pmatrix} \varepsilon_0(k-b) & u_1 & 0 \\ u_1^* & \varepsilon_0(k) & u_1 \\ 0 & u_1^* & \varepsilon_0(k+b) \end{pmatrix} \begin{pmatrix} c_k^{(1)} \\ c_k^{(0)} \\ c_k^{(-1)} \end{pmatrix} = \varepsilon(k) \begin{pmatrix} c_k^{(1)} \\ c_k^{(0)} \\ c_k^{(-1)} \end{pmatrix}. \quad (7.10)$$

Let us write the second row of this equation explicitly:

$$u_1^* c_k^{(1)} + \varepsilon_0(k) c_k^{(0)} + u_1 c_k^{(-1)} = \varepsilon(k) c_k^{(0)}. \quad (7.11)$$

As we are interested in the vicinity of $k=0$, this equation may be simplified by the following approximations: $\varepsilon_0(k) \simeq \varepsilon_0(0) = 0$, and $\varepsilon(k) \simeq \varepsilon_0(k+b) \simeq \varepsilon_0(b)$. In the second approximation, we replaced $\varepsilon(k)$ by the unperturbed energy level at $k=0$. Solving the resulting approximate equation for $c_k^{(0)}$ we obtain:

$$c_k^{(0)} = \frac{u_1^* c_k^{(1)} + u_1 c_k^{(-1)}}{\varepsilon_0(b)}. \quad (7.12)$$

Substituting this formula in the first and the third rows of Eq. (9-10) yields

$$\begin{pmatrix} \varepsilon_0(k-b) + \frac{|u_1|^2}{\varepsilon(b)} & \frac{u_1^2}{\varepsilon(b)} \\ \frac{u_1^{*2}}{\varepsilon(b)} & \varepsilon_0(k+b) + \frac{|u_1|^2}{\varepsilon(b)} \end{pmatrix} \begin{pmatrix} c_k^{(1)} \\ c_k^{(-1)} \end{pmatrix} = \varepsilon(k) \begin{pmatrix} c_k^{(1)} \\ c_k^{(-1)} \end{pmatrix}. \quad (7.13)$$

Thus, up to a constant shift of the energy levels by $|u_1|^2/\varepsilon(b)$, we have returned to the original equation (7.8) albeit with u_2 replaced by $u_1^2/\varepsilon(b)$. The conclusion drawn from this example is that even if the periodic potential does not have a second harmonic, the second-order perturbation theory lifts the degeneracy of the energy levels.

A generic one-dimensional system does not have any degenerate points in its spectrum, as illustrated in Fig. 7-2. To have a degeneracy point requires fine-tuning of the functional form of

the periodic potential. Nevertheless, since the Fourier expansion coefficients of an analytic function usually decay exponentially, the energy gaps also decay exponentially to zero in the high energy limit, $\varepsilon \rightarrow \infty$ ¹.

From the viewpoint of group theory, a one-dimensional periodic system can either have a reflection symmetry or not. In both cases, all irreducible representations of the little groups are one-dimensional. Therefore, one should not expect to have a normal degeneracy of energy levels. In other words, if there is such a degeneracy, it is accidental, namely due to a very particular and non-generic choice of the periodic potential.

In systems with higher dimensionality, the band structure calculation in the nearly free electron approximation follows a similar procedure. First, the zeroth-order (the empty lattice) approximation is constructed by folding the spectrum into the first Brillouin zone. It is obtained by duplicating and shifting the free-electron spectrum to each point of the reciprocal lattice. In Fig. 7-3, this construction is demonstrated for a two-dimensional hexagonal lattice. The dark hexagon in the figure represents the first Brillouin zone into which the free-electron spectrum is folded.

When the dimension of the system is two or three, it is customary to draw the energy levels along distinct lines within the Brillouin zone - usually, lines that connect the special points in $\mathbf{k}$ space. For hexagonal lattice, these are the lines connecting the points, Γ , K , and M as shown in the inset of Fig. 7-4. The figure shows the empty lattice spectrum along these lines. For convenience, we have indicated each energy level by a number associated with the copy of the free particle spectrum shown in Fig 7-3. For clarity, we also separated curves that fall one on top of the other.

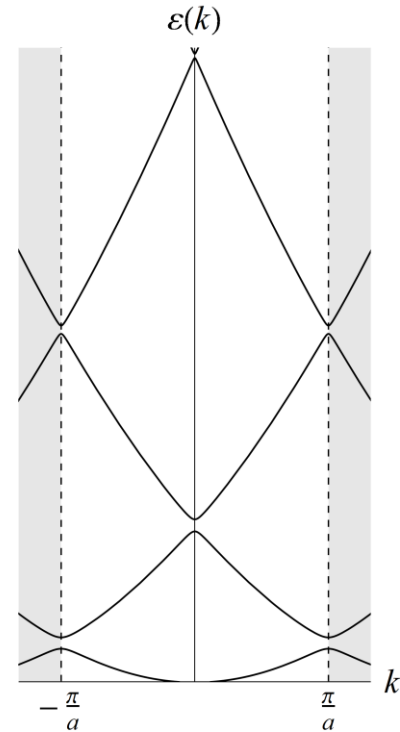


Figure 7-2 The typical spectrum of one dimensional system in the nearly free

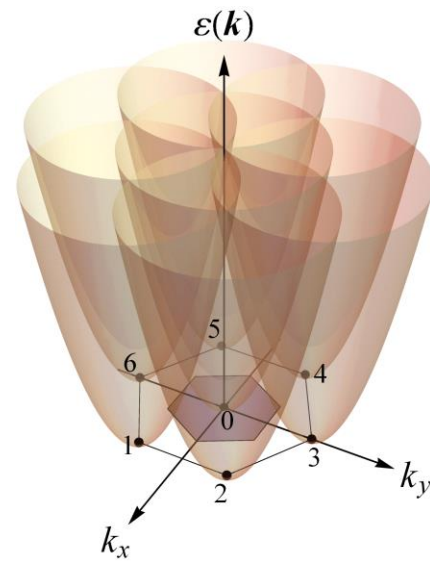


Figure 7-3 Construction of the empty lattice approximation of hexagonal lattice

¹In the Kronig-Penney model, discussed in the first chapter, the potential was non-analytic in the form of a periodic δ -function for which Fourier coefficients are constant. Hence the energy gaps remained constant also in the limit $\varepsilon \rightarrow \infty$ (see Ex. 3 on page 23).

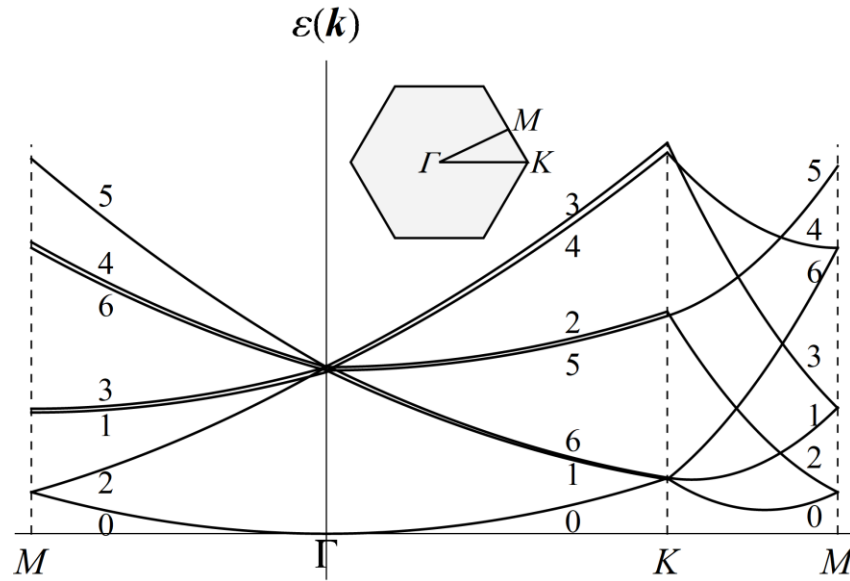


Figure 7-4 The electron spectrum of hexagonal lattice in the empty lattice approximation

From the above figure, it follows that in the framework of the empty lattice approximation, the energy spectrum has a six-fold degeneracy at the Γ -point. However, as we know, the little group at this point is C_{6v} , hence the largest possible (normal) degeneracy is two-fold. Indeed, this degeneracy is lifted when we include the effect of the periodic potential in the nearly free electron approximation - as shown in Fig. 7-5.

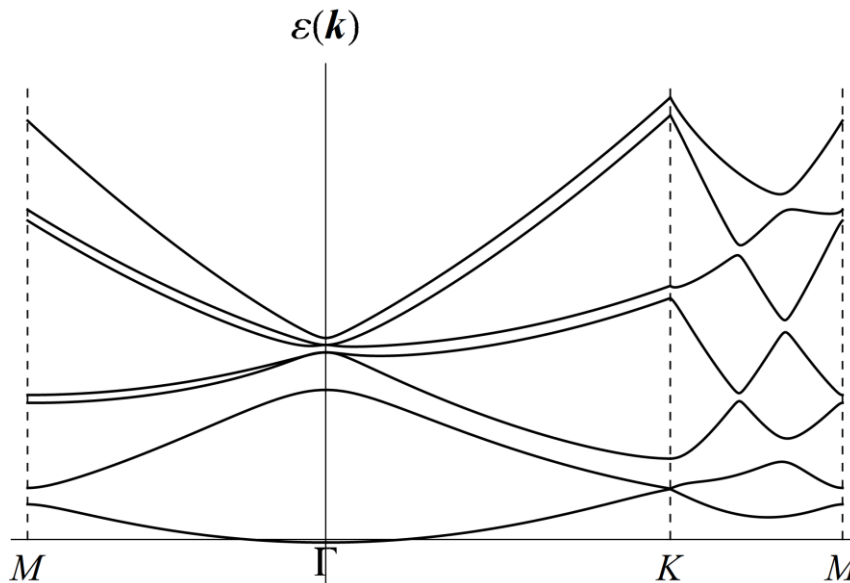


Figure 7-5 The electron spectrum of hexagonal lattice in the nearly free electron approximation

7.2 The tight-binding approximation

The tight-binding approximation applies in the opposite limit of the nearly free electron approximation. It assumes that, in the zeroth approximation, the electron's wave function is localized within one unit cell. The energy spectrum is now calculated by assuming that the transition amplitude from one cell to the other is very small and can be treated perturbatively.

At first sight, one might assume that the localized states on each unit cell are the atomic orbitals of the electrons. For instance, in the case of graphene, each carbon atom forms three σ -bonds with its neighbors, while the fourth orbital, associated with the conduction band electrons, is a $2p$ orbital perpendicular to the graphene sheet, as shown in Fig. 7-6.

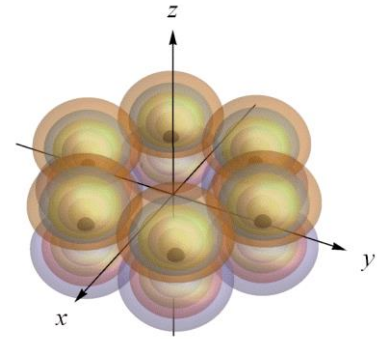


Figure 7-6 The $2p_z$ orbitals of carbon atoms in one hexagon of graphene

However, the atomic orbitals of different unit cells are not orthogonal, and in order to have a well-defined procedure for calculating the electronic spectrum, one should define an orthogonal basis of localized wave functions. The functions of this basis are called *Wannier functions*, which we turn to define now.

Let

$$\psi_k^{(j)}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) \phi_k^{(j)}(\mathbf{r}) \quad (7.14)$$

be the Bloch wave function of a particle in a lattice, where j is the band index, $\mathbf{k}$ is Bloch's wave number, and $\phi_k^{(j)}(\mathbf{r} + \mathbf{a}) = \phi_k^{(j)}(\mathbf{r})$ is a periodic function with the periodicity of the lattice. These wave functions are orthonormal; namely, they satisfy the condition

$$\int d^d r \psi_{k'}^{(j')*}(\mathbf{r}) \psi_k^{(j)}(\mathbf{r}) = \delta_{jj'} \delta_{\mathbf{k}, \mathbf{k}'}, \quad (7.15)$$

where d is the (effective) dimension of the system, and the integral is over the entire d -dimensional space of the system. To be consistent with the limit, $\mathbf{k} = \mathbf{k}'$ one should demand that

$$\int_{UC} d^d r \phi_{k'}^{(j')*}(\mathbf{r}) \phi_k^{(j)}(\mathbf{r}) = \frac{1}{N} \delta_{jj'}, \quad (7.16)$$

where the integral is over one unit cell and N is the number of unit cells in the lattice.

The Wannier functions are defined by

$$w_a^{(j)}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k} \in BZ} \exp(-i\mathbf{k} \cdot \mathbf{a}) \psi_k^{(j)}(\mathbf{r}), \quad (7.17)$$

where $\mathbf{a}$ is the Bravais lattice vector that indicates the position of the unit cell on which the Wannier function is localized, while the sum is over the wavenumber vectors in the first Brillouin zone².

The Wannier functions are orthogonal,

$$\int d^d r w_{\mathbf{a}'}^{(j')*}(\mathbf{r}) w_{\mathbf{a}}^{(j)}(\mathbf{r}) = \delta_{jj'} \delta_{\mathbf{a}\mathbf{a}'}, \quad (7.18)$$

have the same shape in any unit cell, i.e. $w_{\mathbf{a}}^{(j)}(\mathbf{r}) = w_{\mathbf{0}}^{(j)}(\mathbf{r} - \mathbf{a})$, and satisfy the closure relation:

$$\sum_{\mathbf{a}, j} w_{\mathbf{a}}^{(j)*}(\mathbf{r}) w_{\mathbf{a}}^{(j)}(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'). \quad (7.19)$$

Using Eq. (7.17), one can express the Bloch wave function as a sum over Wannier functions:

$$\psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{a}} \exp(i\mathbf{k} \cdot \mathbf{a}) w_{\mathbf{a}}^{(j)}(\mathbf{r}). \quad (7.20)$$

The proof of the last three equations is left as an exercise.

Let us now identify the Hamiltonian in $\mathbf{k}$ space using the basis of Wannier functions. This can be obtained by calculating the matrix element of the Hamiltonian in the basis of Bloch's wave functions as follows:

$$\begin{aligned} \hat{\varepsilon}_{jj'}(\mathbf{k}) &= \langle \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) | H | \psi_{\mathbf{k}}^{(j')}(\mathbf{r}) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{a}} \langle \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) | w_{\mathbf{a}}^{(j)} \rangle \langle w_{\mathbf{a}}^{(j)} | H \sum_{\mathbf{a}'} | w_{\mathbf{a}'}^{(j')} \rangle \langle w_{\mathbf{a}'}^{(j')} | \psi_{\mathbf{k}}^{(j')}(\mathbf{r}) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{a}} \sum_{\mathbf{a}'} H_{\mathbf{a}, \mathbf{a}'}^{(jj')} \exp[i\mathbf{k} \cdot (\mathbf{a}' - \mathbf{a})], \end{aligned} \quad (7.21)$$

where we have used formula (7.20) and the closure relation (7.18). In the last line, we define

$$H_{\mathbf{a}, \mathbf{a}'}^{(jj')} = \langle w_{\mathbf{a}}^{(j)} | H | w_{\mathbf{a}'}^{(j')} \rangle. \quad (7.22)$$

² In the continuum limit, i.e. when the system is large enough compared to the size of a unit cell, the sum over the wave number vectors can be replaced by an integral,

$$\sum_{\mathbf{k}} \rightarrow \frac{\text{Vol}}{(2\pi)^d} \int_{\text{BZ}} d^d k,$$

where Vol is the volume of the system. In this limit the Kronecker δ -function in Eq. (7.15) should be replaced by a δ -function according to the rule:

$$\delta_{\mathbf{k}, \mathbf{k}'} \rightarrow \frac{(2\pi)^d}{\text{Vol}} \delta(\mathbf{k} - \mathbf{k}').$$

Diagonalization of $\hat{\varepsilon}_{jj'}(\mathbf{k})$ is equivalent to the exact solution of the problem; therefore, this form of the Hamiltonian is still useless. To make progress, we shall use formula (7.21) in order to construct an efficient approximation. The tight-binding approximation is built on two main approximations: The first is to consider only those Wannier functions associated with the lowest energy band and ignore all other bands. The second approximation assumes that the significant matrix elements, given by Eq. (7.22), are only those with $\mathbf{a} = \mathbf{a}'$ and when $\mathbf{a}$ and $\mathbf{a}'$ are nearest neighbors sites, i.e., when $|\mathbf{a} - \mathbf{a}'|$ is the smallest possible distance between different lattice sites. The transition between sites separated by a larger distance is negligible because Wannier functions are localized. With these assumptions, one can rearrange the double sum in (7.21) so that one sum is over all lattice sites while the other over its nearest neighbors. The first sum gives the total number of unit cells, N , that cancels out with the factor $1/N$ in Eq. (7.21). Thus the Hamiltonian reduces to

$$\hat{\varepsilon}(\mathbf{k}) = \sum_{\mathbf{a}} H_{\mathbf{a},\mathbf{a}'} \exp[i\mathbf{k} \cdot (\mathbf{a}' - \mathbf{a})], \quad (7.23)$$

where the sum over $\mathbf{a}'$ includes the term $\mathbf{a}' = \mathbf{a}$ and the nearest neighbors of $\mathbf{a}$. Here we have suppressed the indices j and j' , as we consider only the lowest energy band.

The advantage of the tight-binding model defined by the above Hamiltonian is that one needs only a small number of parameters to characterize the system. For a given lattice, there are essentially two parameters: the onsite energy $H_{\mathbf{a},\mathbf{a}} = \langle w_{\mathbf{a}} | H | w_{\mathbf{a}} \rangle = \varepsilon_0$, and the hopping term $H_{\mathbf{a},\mathbf{a}'} = \langle w_{\mathbf{a}} | H | w_{\mathbf{a}'} \rangle = -t$, where $\mathbf{a}'$ is the nearest neighbors of $\mathbf{a}$. To calculate them, one needs the precise structure of the Wannier functions; however, these matrix elements are usually treated as phenomenological parameters.

Finally, notice that it is customary to present the tight-binding model of a system in real space by Hamiltonian of the form

$$H = \varepsilon_0 \sum_i \hat{c}_i^\dagger \hat{c}_i - t \sum_{\langle ij \rangle} \hat{c}_i^\dagger \hat{c}_j + h.c., \quad (7.24)$$

where $\hat{c}_i^\dagger$ and $\hat{c}_i$ are the creation and annihilation operators of a particle at site i , and $\langle ij \rangle$ denote sum over nearest neighbors sites.

Example: the Wannier functions for the Kronig-Penney model

In this example, the Wannier functions of the Kronig-Penney model, for δ -potential wells, are calculated. Namely, we consider a particle moving in a one dimensional periodic potential of the form:

$$u(x) = -\frac{\hbar^2 \mu}{ma} \sum_n \delta(x - na) \quad (7.25)$$

where μ is a dimensionless parameter that characterizes the potential strength (see Eq. (1.24), m is the particle's mass, and a is the lattice constant.

In order to have a reference point for comparison, let us first calculate the wave function that describes the bound state of a particle in a single δ -function well. This wave function is analogous to the atomic orbitals mentioned at the beginning of this section. The Schrödinger equation for this problem is

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \varphi(x)}{\partial x^2} - \frac{\hbar^2 \mu}{ma} \delta(x) \varphi(x) = \varepsilon \varphi(x). \quad (7.26)$$

Thus, when $x \neq 0$ this equation reduces to that of a free particle. The solution associated with a bound state, $\varepsilon < 0$, is described by a decaying exponential function and takes the form

$$\varphi(x) = B \exp\left(-\frac{1}{\hbar} \sqrt{-2m\varepsilon} |x|\right), \quad (7.27)$$

where B is the normalization constant. Next, we integrate the Schrödinger equation (7.26) from $x = 0^-$ to $x = 0^+$,

$$-\frac{\partial \varphi(x)}{\partial x} \Big|_{x=0^+} - \frac{\mu}{a} \varphi(0) = 0, \quad (7.28)$$

substitute (7.27) and solve for ε . The result is the energy of the bound state:

$$\varepsilon = -\frac{\hbar^2 \mu^2}{2ma^2}. \quad (7.29)$$

Finally, calculating the normalization constant, B , we obtain the corresponding wave function:

$$\varphi_0(x) = \sqrt{\frac{\mu}{a}} \exp\left(-\frac{\mu}{a} |x|\right). \quad (7.30)$$

Let us, now, calculate the Wannier functions of the system. The first step is to identify the Bloch wave functions. For δ -function potential, the wave function in the range $0 \leq x < a$, is a sum of decaying exponents,

$$\psi(x) = A \exp(-\alpha' x) + A' \exp(\alpha' x), \quad (7.31)$$

where A and A' are constant that we should find, and the energy

$$\varepsilon = -\frac{\hbar^2 \alpha'^2}{2m} \quad (7.32)$$

sets the value of α' . By Bloch's theorem $\psi(x) = \exp(ikx)\phi_k(x)$ where $\phi_k(x) = \phi_k(x+a)$; therefore, the periodic component of the Bloch wave function (in the range $0 \leq x < a$) is:

$$\phi_k(x) = [A \exp(-\alpha'x) + A' \exp(\alpha'x)] \exp(-ikx). \quad (7.33)$$

From the periodicity of this function, $\phi_k(0) = \phi_k(a)$, and the jump in its derivative, $-\phi'_k(0) + \phi'_k(a) - (2\mu/a)\phi_k(0) = 0$, we obtain two equations for A and A' that can be written as a matrix equation:

$$\begin{pmatrix} 1 - e^{-(\alpha' + ik)a} & 1 - e^{(\alpha' - ik)a} \\ (\alpha' + ik)(1 - e^{-(\alpha' + ik)a}) - \frac{2\mu}{a} & -(\alpha' - ik)(1 - e^{(\alpha' - ik)a}) - \frac{2\mu}{a} \end{pmatrix} \begin{pmatrix} A \\ A' \end{pmatrix} = 0. \quad (7.34)$$

A non-trivial solution of this equation is obtained when the determinant of the above matrix vanishes. This condition yields the equation

$$\cos(ka) = \cosh(\alpha'a) - \mu \frac{\sinh(\alpha'a)}{\alpha'a}. \quad (7.35)$$

The solution of this transcendental equation gives α' as a function of Bloch's wave number k . There is no closed-form solution for the equation, but it can be easily solved numerically.

Eq. (7.34) also provides a relation between A and A' . Substituting it in Eq. (7.33), we obtain:

$$\begin{aligned} \phi_k(x) &= \frac{B'}{2} \left[-\left(e^{ika} - e^{\alpha'a}\right) \exp(-\alpha'x - ikx) + \left(e^{ika} - e^{-\alpha'a}\right) \exp(\alpha'x - ikx) \right] \\ &= B' \left\{ \sinh(\alpha'x) \exp[ika] - \sinh[\alpha'(x-a)] \right\} \exp(-ikx) \end{aligned} \quad (7.36)$$

where B' is the normalization constant that we choose to be real, so that

$$\frac{1}{B'^2} = \left\{ \cos(ka) \cosh(\alpha'a) - 1 + \frac{\sinh(\alpha'a)}{a\alpha'} [\cosh(\alpha'a) + \cos(ka)] \right\} aN, \quad (7.37)$$

The above formula for $\phi_k(x)$ holds only in the range $0 \leq x < a$. In order to obtain a description over the entire range, $-\infty < x < \infty$, it should be extended periodically. Let $\phi_k^{(p)}(x) = \sum_n \phi_k(x - an)$, be this extended function, then the Bloch wave function is $\psi_k(x) = \phi_k^{(p)}(x) \exp(ikx)$. Substituting this function in Eq. (7.17), and replacing the discrete sum

over k by an integral (see footnote on page 130) we obtain that the Wannier function is expressed as the following integral over the first Brillouin zone:

$$w_0(x) = a\sqrt{N} \int_{-\frac{\pi}{a}}^{\frac{\pi}{a}} \frac{dk}{2\pi} \phi_k^{(p)}(x) \exp(ikx). \quad (7.38)$$

In the limit $\mu \gg 1$, $\alpha' \rightarrow \mu/a$ (see Eq. (1.26)), $B' \rightarrow 2\sqrt{\mu/Na} \exp(-\mu)$, and Bloch function in the range $0 \leq x < a/2$ (of the unit cell centered at the origin) reduces to

$$\phi_k(x) \exp(ikx) \rightarrow 2\sqrt{\frac{\mu}{aN}} \exp(-\mu) \left[-\sinh\left[\frac{\mu}{a}x - \mu\right] + \sinh\left(\frac{\mu}{a}x\right) \exp(ika) \right]. \quad (7.39)$$

Substituting this function in (7.38) we obtain

$$w_0(x) \rightarrow 2\sqrt{\frac{\mu}{a}} \exp(-\mu) \sinh\left[\mu - \frac{\mu}{a}x\right] \rightarrow \sqrt{\frac{\mu}{a}} \exp\left(-\frac{\mu}{a}x\right), \quad (7.40)$$

which is the same as the wave function of the bound state of a particle in a δ -potential, $\varphi_0(x)$. Thus, in the limit of deep potential wells, the difference between $w_0(x)$ and $\varphi_0(x)$ is negligible because the wave function decays to zero, essentially, within a distance of one unit cell. Thus, the overlap of two wave functions located in neighboring lattice sites is negligible, implying that they are approximately orthogonal.

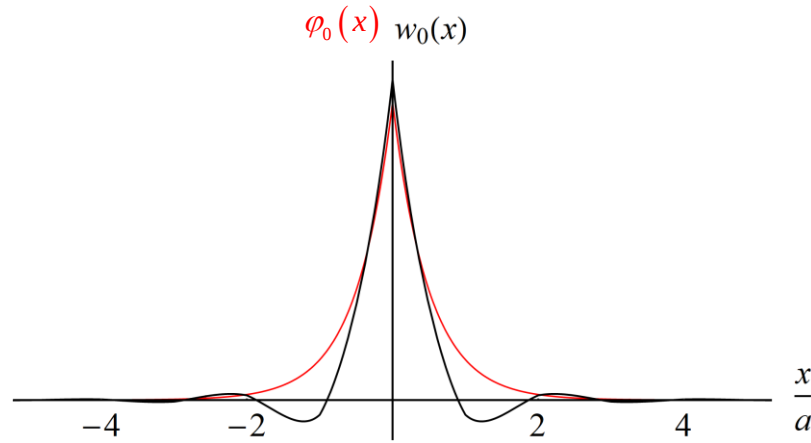


Figure 7-7 The Wannier function for the Kronig Penney model with δ -potential wells (black line), and the solution of the bound state of a particle in a single δ -potential well (red line)

However, when μ is not large, the tails of the wave function (7.30) extend over a much larger distance than that of one unit cell. The orthogonality condition of Wannier functions, in this limit, forces them to change sign. An example of this behavior is shown in Fig. 7-7, where the Wannier

function for $\mu = 2$ (obtained by numerical solution of the problem), is drawn by the solid black line, while the red line represents the wave function $\varphi_0(x)$.

Example: The tight-binding model in one dimension

In one dimension, Eq. (7.23) reduces to

$$\begin{aligned}\varepsilon(k) &= \varepsilon_0 - t \exp(ika) - t \exp(-ika) \\ &= \varepsilon_0 - 2t \cos(ka) .\end{aligned}\quad (7.41)$$

(The onsite energy, ε_0 , and the hopping term, t , are defined below Eq. (7.23).) This formula for the band energy is depicted in Fig. 7-8. It constitutes a good approximation when all other energy bands are sufficiently far such that their influence is negligible. An example of a situation where this approximation is invalid is shown in Fig. 7-2. Here the bands become very close to each other at some particular point in k space, near the edge of the Brillouin zone, for example.

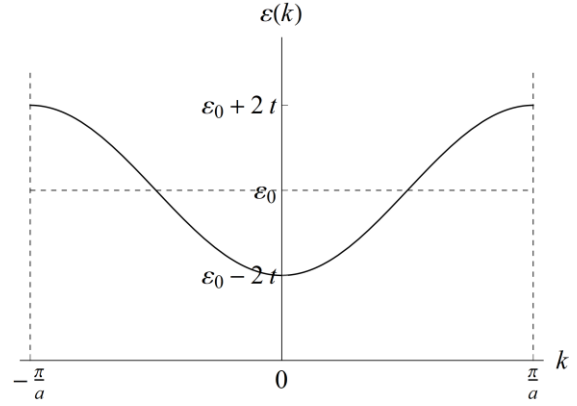


Figure 7-8 The spectrum of the tight binding model in one dimension

Example: The tight-binding model for graphene

To calculate the band structure of graphene in the framework of the tight-binding approximation, one has to take into account its two sublattices, which induce a pseudospinor structure of the wave function. In particular, the hopping matrix element between neighboring sites is from one sublattice to the other and vice versa. Therefore, in the basis of the Wannier functions, the Hamiltonian is a 2×2 matrix of the form

$$H = \begin{pmatrix} 0 & H_{AB} \\ H_{BA} & 0 \end{pmatrix}. \quad (7.42)$$

Here, without loss of generality, we set the onsite energy (i.e. the diagonal components in the above matrix) to be zero, $\varepsilon_0 = 0$. The Hamiltonian in $\mathbf{k}$ space (7.23) is therefore

$$\hat{\varepsilon}(\mathbf{k}) = \begin{pmatrix} 0 & \Delta(\mathbf{k}) \\ \Delta^*(\mathbf{k}) & 0 \end{pmatrix}, \quad (7.43)$$

where

$$\Delta(\mathbf{k}) = t \sum_{j:A \rightarrow B} \exp(i\mathbf{k} \cdot \mathbf{a}_j) \quad (7.44)$$

is the contribution of hopping from sublattice A to sublattice B . Here $\mathbf{a}_j$ are the vectors connecting a lattice point, of sublattice A , to its three nearest neighbors on sublattice B as shown in Fig. 7-9. These vectors are:

$$\begin{aligned} \mathbf{a}_1 &= \left(0, -\frac{1}{\sqrt{3}}\right)a, & \mathbf{a}_2 &= \left(\frac{1}{2}, \frac{1}{2\sqrt{3}}\right)a, \\ \mathbf{a}_3 &= \left(-\frac{1}{2}, \frac{1}{2\sqrt{3}}\right)a, \end{aligned} \quad (7.45)$$

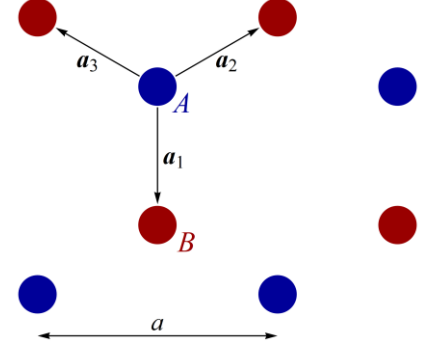


Figure 7-9 The vectors $\mathbf{a}_j$

where a is the lattice constant. Substituting Eq. (7.45) in (7.44) yields

$$\Delta(\mathbf{k}) = t \left[\exp\left(-i\frac{1}{\sqrt{3}}k_y a\right) + 2 \cos\left(\frac{1}{2}k_x a\right) \exp\left(i\frac{1}{2\sqrt{3}}k_y a\right) \right], \quad (7.46)$$

and diagonalization of the Hamiltonian (7.43) gives the energy levels:

$$\varepsilon_{\pm}(\mathbf{k}) = \pm |\Delta(\mathbf{k})| = \pm t \sqrt{1 + 4 \cos^2\left(\frac{1}{2}k_x a\right) + 4 \cos\left(\frac{1}{2}k_x a\right) \cos\left(\frac{\sqrt{3}}{2}k_y a\right)} \quad (7.47)$$

The energy surfaces described by this formula are depicted in Fig. 7-10. In the right panel of this figure, we present only the lower energy surface, $\varepsilon_{-}(\mathbf{k})$, to highlight the Dirac cone structure near the K and the K' degeneracy points in the Brillouin zone.

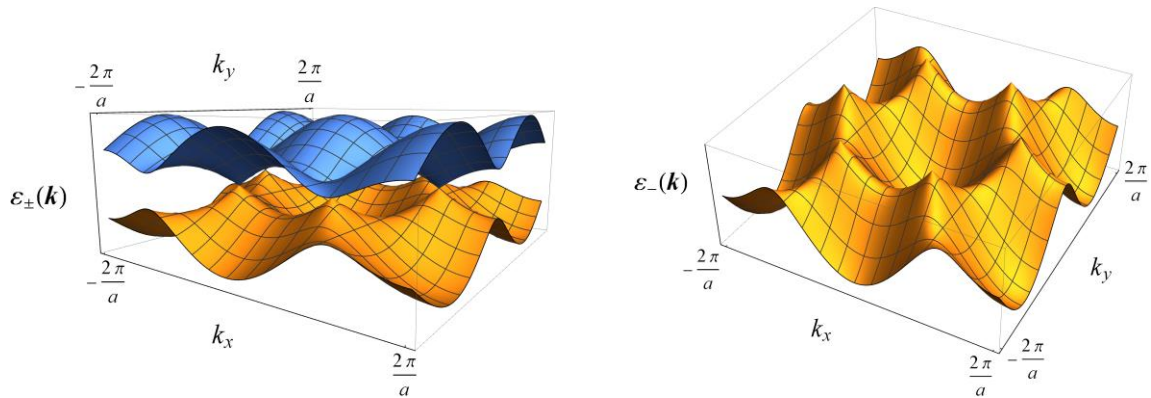


Figure 7-10 The energy surfaces of graphene obtained from the tight-binding model

To find the position of the K and K' points along the k_x axis, one should solve the equation $\varepsilon_{\pm}(\mathbf{k})|_{k_y=0} = 0$, which is

$$1 + 4\cos^2\left(\frac{1}{2}k_x a\right) + 4\cos\left(\frac{1}{2}k_x a\right) = 0. \quad (7.48)$$

The solution is $k_x = \pm 4\pi/3a$, hence

$$\mathbf{k}_K = \frac{4\pi}{3a}(1, 0) \text{ and } \mathbf{k}_{K'} = \frac{4\pi}{3a}(-1, 0). \quad (7.49)$$

Now, let us expand the Hamiltonian $\hat{\varepsilon}(\mathbf{k})$ in the vicinity of the K' point. For this purpose, we substitute $\mathbf{k} = \mathbf{k}_{K'} + \delta\mathbf{k}$ in Eq. (7.46) for $\Delta(\mathbf{k})$ and expand to second order in $\delta\mathbf{k}$. The result is

$$\Delta(\mathbf{k}) \simeq \frac{\sqrt{3}t}{2}(\delta k_x - i\delta k_y) + \frac{t}{8}(\delta k_x + i\delta k_y)^2 + \dots \quad (7.50)$$

thus

$$\hat{\varepsilon}(\mathbf{k}) = \begin{pmatrix} 0 & \Delta(\mathbf{k}) \\ \Delta^*(\mathbf{k}) & 0 \end{pmatrix} \simeq \frac{\sqrt{3}t}{2}\delta\mathbf{k} \cdot \boldsymbol{\tau}^{AB} + \frac{t}{8}\left[(\delta k_x^2 - \delta k_y^2)\tau_x^{AB} - 2\delta k_x\delta k_y\tau_y^{AB}\right]. \quad (7.51)$$

Notice that this local approximation has the same structure as that we obtain in Eq. (5.16) from pure symmetry arguments. Comparing the parameters obtained here with the ones found using the group theory approach, we see that $\hbar v = \sqrt{3}t/2$ and $h_w = -t/8$. Thus, the tight-binding approximation provides the relation between these phenomenological parameters.

Example: Degenerate bands in a square lattice

The tight-binding approximation has been derived under the assumption that energy bands are sufficiently far apart such that the effect of higher energy bands on the lower one is negligible. Namely, when the energy gap between nearby bands is larger than the hopping energy t . However, there are situations where on each lattice point, there are few degenerate orbitals. Examples of such materials are *iron pnictides*. These materials consist of weakly coupled two-dimensional layers. In each layer, the iron atoms form a square lattice. The states that contribute the charge carriers to the system are the d_{xz} and the d_{yz} degenerate orbitals of each iron atom. These orbitals are illustrated schematically in Fig. 7-11 by red and green colors.

The generalization of a tight-binding model for such a system is obtained by defining two wave functions associated with these orbitals, $c_{i_x, i_y}^{(x)}$ and $c_{i_x, i_y}^{(y)}$. Here the upper index refers to the type of the orbital (by indicating its direction in space), and the lower indices denote the position on the lattice.

There are two sorts of hopping matrix elements between nearest neighbors orbitals: $-t_1$ associated with σ -bond and $-t_2$ associated with π -bonds. These are represented, respectively, by the solid and the black dashed lines in Fig. 7-11. Thus, assuming hopping only to nearest neighbor sites, the tight-binding model is defined by the equations:

$$\begin{aligned} -t_1 \left(c_{i_x+1, i_y}^{(x)} + c_{i_x-1, i_y}^{(x)} \right) - t_2 \left(c_{i_x, i_y+1}^{(x)} + c_{i_x, i_y-1}^{(x)} \right) &= \varepsilon c_{i_x, i_y}^{(x)} \\ -t_1 \left(c_{i_x, i_y+1}^{(y)} + c_{i_x, i_y-1}^{(y)} \right) - t_2 \left(c_{i_x+1, i_y}^{(y)} + c_{i_x-1, i_y}^{(y)} \right) &= \varepsilon c_{i_x, i_y}^{(y)} \end{aligned} \quad (7.52)$$

Here, as in the previous example, we set the onsite energy to be zero. Notice that the symmetry of the orbitals implies no matrix element connecting x -type and y -type orbitals within this nearest neighbor approximation. Hence the energy levels obtained from the above equation include two bands where each one corresponds to a different type of orbital:

$$\begin{aligned} \varepsilon^{(x)}(\mathbf{k}) &= -2t_1 \cos(k_x a) - 2t_2 \cos(k_y a), \\ \varepsilon^{(y)}(\mathbf{k}) &= -2t_1 \cos(k_y a) - 2t_2 \cos(k_x a). \end{aligned} \quad (7.53)$$

Here a is the lattice constant.

These energy bands are degenerate along the $k_x = k_y$ line, as shown in Fig. 7-12. Here the energy levels (7.53) are plotted along the contour $\Gamma \rightarrow X \rightarrow W \rightarrow \Gamma$ in the Brillouin zone (the special points in the Brillouin zone of a square lattice are defined in Fig. 5-13), and the line $k_x = k_y$ is the one from Γ point to W -point. This degeneracy is lifted once hopping into the next nearest neighbors is taken into

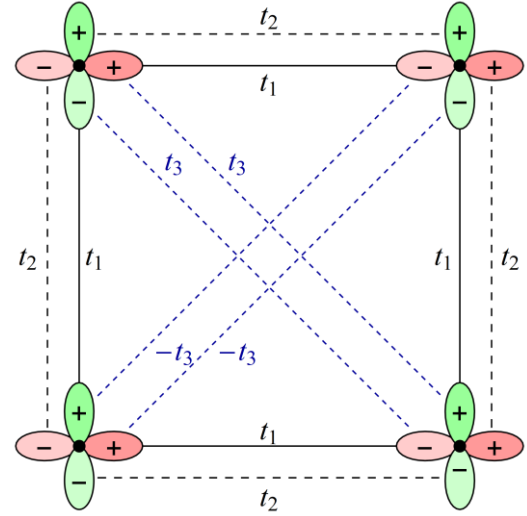


Figure 7-11 The orbitals d_{xz} and d_{yz} in a square lattice

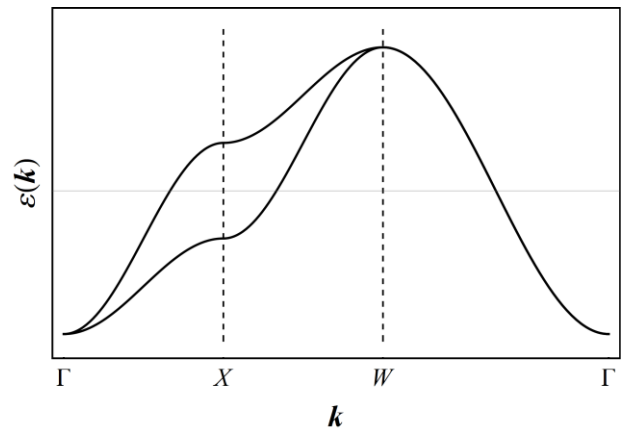


Figure 7-12 The energy bands of Eqs. (7.53)

account. The corresponding hopping matrix elements are denoted by $\pm t_3$ in Fig. 7-11. They generate coupling between the two types of orbitals, and consequently between Eqs. (7.53), see Exercise 2.

7.3 The $\mathbf{k} \cdot \mathbf{p}$ approximation, Kane's model, heavy holes and light holes

The $\mathbf{k} \cdot \mathbf{p}$ approximation is designed to treat situations where energy bands are close to each other, at some points in the Brillouin zone, by providing a local description near these points. Starting from the Schrödinger equation of a particle moving in the periodic potential $u(\mathbf{r})$,

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) \right] \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \varepsilon_j(\mathbf{k}) \psi_{\mathbf{k}}^{(j)}(\mathbf{r}), \quad (7.54)$$

we express the Bloch wave function of the j -th band in the form $\psi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \phi_{\mathbf{k}}^{(j)}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})$, where $\phi_{\mathbf{k}}^{(j)}(\mathbf{r})$ is an unknown periodic function. Then the left-hand side of the equation is expanded as follows:

$$\begin{aligned} \left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) \right] \psi_{\mathbf{k}}^{(j)}(\mathbf{r}) &= \exp(i\mathbf{k} \cdot \mathbf{r}) \left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) \right] \phi_{\mathbf{k}}^{(j)}(\mathbf{r}) \\ &+ \phi_{\mathbf{k}}^{(j)}(\mathbf{r}) \left[-\frac{\hbar^2}{2m} \nabla^2 \right] \exp(i\mathbf{k} \cdot \mathbf{r}) - \frac{\hbar^2}{m} \nabla \exp(i\mathbf{k} \cdot \mathbf{r}) \cdot \nabla \phi_{\mathbf{k}}^{(j)}(\mathbf{r}) \\ &= \exp(i\mathbf{k} \cdot \mathbf{r}) \left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) + \frac{\hbar^2 \mathbf{k}^2}{2m} \right] \phi_{\mathbf{k}}^{(j)}(\mathbf{r}) + \frac{\hbar}{m} \mathbf{k} \cdot (-i\hbar \nabla) \phi_{\mathbf{k}}^{(j)}(\mathbf{r}). \end{aligned} \quad (7.55)$$

From (7.54) and (7.55), we obtain an equation for $\phi_{\mathbf{k}}^{(j)}(\mathbf{r})$:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) + \frac{\hbar^2 \mathbf{k}^2}{2m} \right] \phi_{\mathbf{k}}^{(j)}(\mathbf{r}) + \frac{\hbar}{m} \mathbf{k} \cdot (-i\hbar \nabla) \phi_{\mathbf{k}}^{(j)}(\mathbf{r}) = \varepsilon_j(\mathbf{k}) \phi_{\mathbf{k}}^{(j)}(\mathbf{r}). \quad (7.56)$$

The second term on the left-hand side of this equation is the reason for the name of the approximation that we are about to develop. Let us assume that the functions $\phi_0^{(j)}(\mathbf{r})$, associated with $\mathbf{k} = 0$, and any j , are known from the solution of the equation

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) \right] \phi_0^{(j)}(\mathbf{r}) = \varepsilon_j(0) \phi_0^{(j)}(\mathbf{r}). \quad (7.57)$$

These functions form a basis for all functions within one unit cell with periodic boundary conditions. Here we choose them to satisfy the normalization condition:

$$\int_{uc} d^d r \phi_0^{(j)*}(\mathbf{r}) \phi_0^{(j)}(\mathbf{r}) = \delta_{jj'}, \quad (7.58)$$

where the integral is over the volume of one unit cell of the lattice. Now we shall use this basis in order to expand any other function $\phi_k^{(j)}(\mathbf{r})$ corresponding to a non-zero value of $\mathbf{k}$,

$$\phi_k^{(j)}(\mathbf{r}) = \sum_{j'} c_{j'} \phi_0^{(j')}(\mathbf{r}), \quad (7.59)$$

where $c_{j'}$ are the expansion coefficients that we seek to find. Substituting Eq. (7.59) in (7.56), multiplying the equation by $\phi_0^{(j)*}(\mathbf{r})$ and integrating over $\mathbf{r}$ in one unit cell we obtain (using Eq. (7.58)):

$$\left[\varepsilon_j(\mathbf{k}) - \varepsilon_j(0) - \frac{\hbar^2 \mathbf{k}^2}{2m} \right] c_j = -\frac{\hbar}{m} \sum_{j'} \mathbf{k} \cdot \mathbf{p}_{jj'} c_{j'}, \quad (7.60)$$

where

$$\mathbf{p}_{jj'} = \int_{uc} d^d r \phi_0^{(j)*}(\mathbf{r}) (-i\hbar \nabla) \phi_0^{(j')}(\mathbf{r}). \quad (7.61)$$

Equation (7.60) is an infinite matrix equation that is still exact. Thus, its solution is equivalent to an exact solution of the problem.

The approximation scheme that one can build using Eq. (7.60) is based on the property that several energy levels might come close together near special points of the Brillouin zone. In this situation, one can neglect all other energy levels of much higher (or lower) energy. Thus the $\mathbf{k} \cdot \mathbf{p}$ approximation is obtained by truncating the infinite matrix, $\mathbf{p}_{jj'}$, to a small finite matrix. The larger the dimension of this matrix is, the better is the approximation.

Another approximation, usually employed in this framework, is to replace $\phi_0^{(j)}(\mathbf{r})$ by the local electronic orbitals. This approximation is valid when the orbitals are well localized within each unit cell. Finally, we comment that although we have considered the expansion near the Γ point, a similar approximation can be constructed near any other special point of the Brillouin zone.

Example: Kane's model

To demonstrate the $\mathbf{k} \cdot \mathbf{p}$ approximation, consider a lattice where each unit cell contains four (relevant) orbitals with nearby energies: One orbital is associated with the conduction band,

$\varepsilon_0(0) = \varepsilon_g/2$, and three orbitals with degenerate energy $\varepsilon_x(0) = \varepsilon_y(0) = \varepsilon_z(0) = -\varepsilon_g/2$ (i.e., orbitals of holes) that create the valance bands. It is instructive to think about the first orbital as the $|s\rangle$ orbital of some atom while considering the other three as the degenerate $|p_x\rangle$, $|p_y\rangle$, and $|p_z\rangle$ orbitals. These four orbitals span the subspace of the $\mathbf{k} \cdot \mathbf{p}$ approximation. Selection rules imply that all diagonal matrix elements of $\mathbf{p}_{ij}$ vanish and, similarly, $\mathbf{p}_{xy} = \mathbf{p}_{xz} = \mathbf{p}_{yz} = 0$, where $\mathbf{p}_{ij} = \langle p_i | \mathbf{p} | p_j \rangle$, and $i, j = x, y, z$. The only nonzero matrix elements are $\mathbf{p}_{0x} = p\hat{x}$, $\mathbf{p}_{0y} = p\hat{y}$, $\mathbf{p}_{0z} = p\hat{z}$, and their hermitian conjugates. Here $\mathbf{p}_{0i} = \langle s | \mathbf{p} | p_i \rangle$, while $\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit vectors in the direction of the axes. Symmetry implies that these nonzero matrix elements have the same absolute value. From the above considerations, we get that Eq. (7.60) can be approximated by:

$$\begin{pmatrix} \frac{\varepsilon_g}{2} + \frac{\hbar^2 \mathbf{k}^2}{2m} & \hbar v k_x & \hbar v k_y & \hbar v k_z \\ \hbar v^* k_x & -\frac{\varepsilon_g}{2} - \frac{\hbar^2 \mathbf{k}^2}{2m} & 0 & 0 \\ \hbar v^* k_y & 0 & -\frac{\varepsilon_g}{2} - \frac{\hbar^2 \mathbf{k}^2}{2m} & 0 \\ \hbar v^* k_z & 0 & 0 & -\frac{\varepsilon_g}{2} - \frac{\hbar^2 \mathbf{k}^2}{2m} \end{pmatrix} \begin{pmatrix} c_0 \\ c_x \\ c_y \\ c_z \end{pmatrix} = \varepsilon(\mathbf{k}) \begin{pmatrix} c_0 \\ c_x \\ c_y \\ c_z \end{pmatrix}, \quad (7.62)$$

where $v = p/m$, and for the three hole-orbitals, we have changed the sign of the mass. This equation is the *Kane model*. In this form, it does not take into account spin-orbit interaction, but this generalization can be obtained by doubling the matrix size.

To solve Kane's model, it is convenient to define the vector $\mathbf{c} = (c_x, c_y, c_z)$. With the help of this vector, the above equation can be written in the form

$$\begin{pmatrix} \Delta & \hbar v \mathbf{k} \cdot \\ \hbar v^* \mathbf{k} \cdot & -\Delta \end{pmatrix} \begin{pmatrix} c_0 \\ \mathbf{c} \end{pmatrix} = \varepsilon(\mathbf{k}) \begin{pmatrix} c_0 \\ \mathbf{c} \end{pmatrix}, \quad \text{with } \Delta = \frac{\varepsilon_g}{2} + \frac{\hbar^2 \mathbf{k}^2}{2m}. \quad (7.63)$$

Now, one can identify two cases: The first is when $\mathbf{c}$ is parallel to $\mathbf{k}$ while the other is when these vectors are perpendicular to each other,

$$\begin{pmatrix} c_0 \\ c\hat{\mathbf{k}} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ \mathbf{c}_\perp \end{pmatrix}, \quad (7.64)$$

respectively. Here $\hat{k}$ is a unit vector in the direction of $\mathbf{k}$ and $\mathbf{c}_\perp = \mathbf{c} - \hat{k} \cdot \mathbf{c}$ is perpendicular to $\mathbf{k}$. Notice that these two vectors are orthogonal as required by solutions of the Schrödinger equation. For the first solution, Eq. (7.63) reduces to a 2×2 matrix equation,

$$\begin{pmatrix} \Delta & \hbar v k \\ \hbar v^* k & -\Delta \end{pmatrix} \begin{pmatrix} c_0 \\ c \end{pmatrix} = \varepsilon_\parallel(k) \begin{pmatrix} c_0 \\ c \end{pmatrix}, \quad (7.65)$$

whose straightforward diagonalization yields

$$\varepsilon_{\parallel,\pm}(k) = \pm \sqrt{\Delta^2 + \hbar^2 |v|^2 k^2}. \quad (7.66)$$

Expansion of these energy levels, up to second order in k , in the vicinity of $k = 0$ gives:

$$\varepsilon_{\parallel,\pm}(k) \approx \pm \frac{\varepsilon_g}{2} \pm \left(\frac{\hbar^2}{2m} + \frac{\hbar^2 |v|^2}{\varepsilon_g} \right) k^2. \quad (7.67)$$

Consider now the case where $\mathbf{c}$ is perpendicular to $\mathbf{k}$. Now, Eq. (7.63) reduces to $-\Delta \mathbf{c}_\perp = \varepsilon_\perp(k) \mathbf{c}_\perp$ which has two degenerate solutions (associated with the two possible directions of the vector $\mathbf{c}_\perp$) describing the holes spectrum:

$$\varepsilon_\perp(k) = -\frac{\varepsilon_g}{2} - \frac{\hbar^2 k^2}{2m}. \quad (7.68)$$

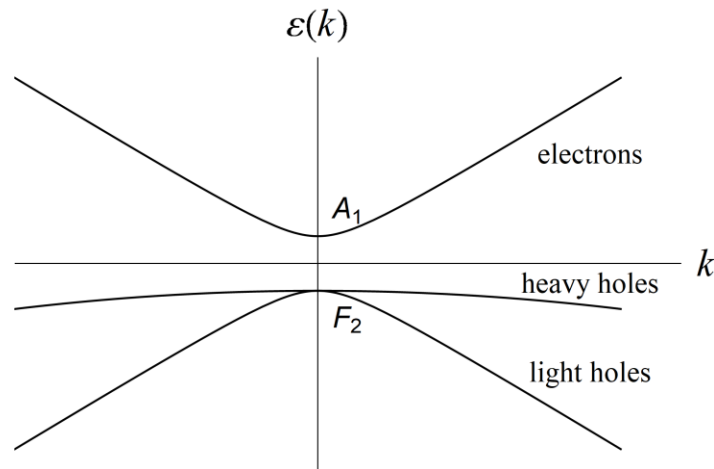


Figure 7-13 The energy levels of Kane's model

The spectrum that we obtained in Eqs. (7.66) and (7.68) is presented in Fig. 7-13. To understand its meaning, notice that the effective mass of the particle obtained from the first couple of solutions in Eq. (7.67) depends on the energy gap ε_g . When this gap is small, the effective mass is approximately

$$m_{\text{eff},\pm} \simeq \pm \frac{\varepsilon_g}{2|V|^2}. \quad (7.69)$$

Namely, it can be very small depending on ε_g . In this case, the energy band with negative mass describes particles known as “light holes”. The effective mass of the other solutions (7.68) is $m_{\text{eff}} = -m$. These solutions are associated with particles called “heavy holes”.

As a concrete example, let us present the experimental data of two semiconductors: InSb (indium antimonide) and GaAs (gallium arsenide). In GaAs, the energy gap is $\varepsilon_g \simeq 1.42\text{eV}$, and the measured effective masses are $m_e \simeq 0.045m_0$ for electrons, $m_{lh} \simeq 0.06m_0$ for light holes, and $m_{hh} \simeq 0.4m_0$ for the heavy holes, where m_0 is the free electron mass. In InSb the energy gap is much smaller, $\varepsilon_g \simeq 0.17\text{eV}$, and the effective masses are: $m_e \simeq m_{lh} \simeq 0.014m_0$, and $m_{hh} \simeq 0.4m_0$.

Let us look at the problem, once again, from a group theory perspective. The unit cell of both semiconductors mentioned above is displayed in Fig. 7-14. Its symmetry is the symmetry of a regular tetrahedron (see also Figs. 2-40 and 4-20). Therefore, its point group is the tetrahedral group T_d , whose character table is listed on the next page.

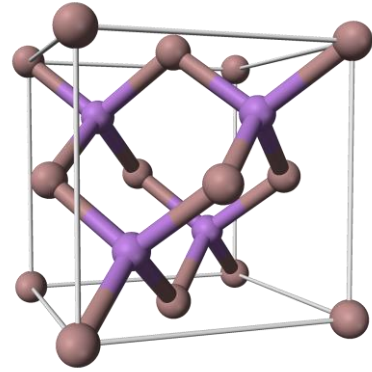


Figure 7-14 The unit cell of InSb and GaAs

The four energy levels of the Kane model, at $\mathbf{k} = 0$, are associated with the A_1 irreducible representation and the F_2 three-dimensional irreducible representation, as denoted in Fig. 7-13. Now consider the matrix element $\langle \psi_i | p_\alpha | \psi_j \rangle$, where $\psi_{i,j}$ are selected from the four wave functions associated with the aforementioned energy levels, while p_α is a component of the momentum vector. This matrix element is a scalar that must be invariant under all symmetry operations of the group. Therefore, it should belong to the identity representation, A_1 .

T_d	E	$8c_3$	$3c_2$	$6S_4$	$6\sigma_d$			
A_1	1	1	1	1	1	-	$x^2 + y^2 + z^2$	xyz
A_2	1	1	1	-1	-1	-	-	-
E	2	-1	2	0	0	-	$[2z^2 - x^2 - y^2, \sqrt{3}(x^2 - y^2)]$	-
F_1	3	0	-1	1	-1	(R_x, R_y, R_z)	-	$[x(z^2 - y^2), y(z^2 - x^2), z(x^2 - y^2)]$
F_2	3	0	-1	-1	1	(x, y, z)	(yz, xz, xy)	$(x^3, y^3, z^3), (x^2 + y^2 + z^2)(x, y, z)$

From the character table of T_d , we see that the momentum (being a vector) belongs to the F_2 irreducible representation. Therefore, the matrix element of the momentum between two wavefunctions that also belong to F_2 is nonzero if the direct product $F_2 \otimes F_2 \otimes F_2$ contains the identity representation. Similarly, the matrix element between wave functions, where one belongs to F_2 and the other to A_1 , is nonzero if $F_2 \otimes F_2 \otimes A_1$ contains the identity representation. Finally, the momentum matrix element between wavefunctions where both belong to the identity representation is non-zero only if $A_1 \otimes F_2 \otimes A_1$ includes the singlet representation (which is, clearly, not the case). These conditions set the *selection rules* for the matrix element p_{ij} . Using the following products of irreducible representations,

$$\begin{aligned}
 E \otimes F_1 &= F_1 + F_2 & E \otimes F_2 &= F_1 + F_2 \\
 F_1 \otimes F_1 &= A_1 + E + F_1 + F_2 & F_2 \otimes F_2 &= A_1 + E + F_1 + F_2 & F_1 \otimes F_2 &= A_2 + E + F_1 + F_2
 \end{aligned} \tag{7.70}$$

we see that the matrix element of the momentum between wave functions where one belongs to the identity representation A_1 while the other to the F_2 representation are non-zero. Those where both wave functions belong to A_1 vanish. On the other hand, the direct product $F_2 \otimes F_2 \otimes F_2$ contains the singlet and, apparently, implies that the matrix element $\langle \psi | \mathbf{p} | \psi' \rangle$ where both ψ and ψ' belong to F_2 is non-zero. However, time reversal symmetry of the problem implies that $\langle \psi | \mathbf{p} | \psi' \rangle = -\langle \psi | \mathbf{p} | \psi' \rangle$ because the wave functions belong the same representation as $\mathbf{p}$; hence under time reversal symmetry they change sign. Another way to get the result is by noticing that the only way of obtaining a scalar from three vectors is by $\langle \psi | \mathbf{p} \times \mathbf{p} | \psi' \rangle$ which is not the quantity we need. These group theory considerations prove the structure of Kane's Hamiltonian presented in Eq. (7.62).

7.4 Exercises

1. Prove Eqs. (7.18) (7.19) and (7.20).
2. Solve the tight-binding model of a square lattice with two degenerate orbitals. Take into account hopping to next nearest neighbors as illustrated by the dashed diagonal lines in Fig. 7-11. Plot the energy levels in the same way they are plotted in Fig. 7-12. Compare your results with those of Exercise 1 in chapter 5.

3. Solve the tight-binding model for the Kagome lattice shown in Fig. 7-15, assuming hopping only to nearest neighbors sites. The Bravais lattice of this system is Hexagonal, and each unit cell contains three sites. In other words, the Kagome lattice is made of three hexagonal sublattices. These are indicated, in Fig. 7-15, by different colors.

Denoting the primitive basis vectors of the Bravais lattice by

$$\mathbf{a}_1 = a(1,0) \quad \text{and} \quad \mathbf{a}_2 = \frac{a}{2}(1, \sqrt{3}) \quad (7.71)$$

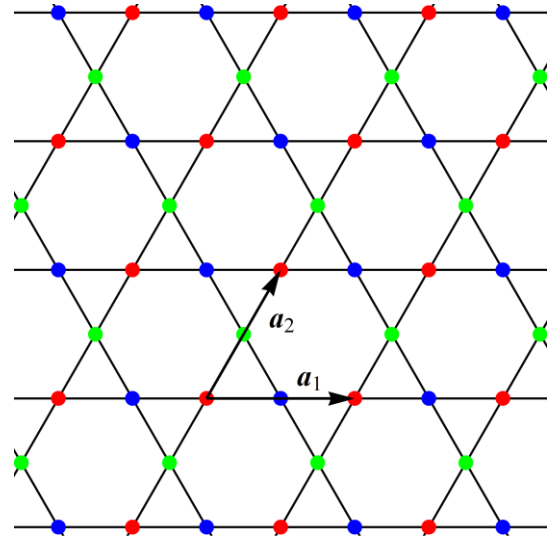


Figure 7-15 Kagome lattice

where a is the lattice constant (see Fig 7-15), the positions of the red, green, and blue points of the Kagome lattice are, respectively, given by

$$\mathbf{s}_1 = (0,0), \quad \mathbf{s}_2 = \frac{\mathbf{a}_2}{2} \quad \text{and} \quad \mathbf{s}_3 = \frac{\mathbf{a}_1}{2}. \quad (7.72)$$

Calculate the energy levels of this model and plot them along the line $\Gamma \rightarrow K \rightarrow M \rightarrow \Gamma$ in the Brillouin zone.

4. The purpose of this exercise is to derive the Kane-Mele term from the tight-binding model of graphene. The generalization of the tight-binding model for a system with spin-orbit interaction requires doubling the number of components in the wave functions to account for the two possible spin states. Hence the size of the matrix that describes the Hamiltonian is also doubled. For graphene, assuming hopping is only to nearest neighbors sites, the Hamiltonian that describes spin-orbit interaction is of the form:

$$H_{\text{so}}(\mathbf{k}) = \begin{pmatrix} 0 & \boldsymbol{\tau}^S \cdot \mathbf{d}(\mathbf{k}) \\ \boldsymbol{\tau}^S \cdot \mathbf{d}^*(\mathbf{k}) & 0 \end{pmatrix}, \quad (7.73)$$

where $\boldsymbol{\tau}^S$ are Pauli matrices acting on the spin space, while $\mathbf{d}(\mathbf{k})$ are function determined by spin-orbit interaction. The blocks of this matrix act in the space of sublattices of the graphene.

(a) Use inversion symmetry, i.e. $\mathbf{k} \leftrightarrow -\mathbf{k}$, and the symmetry to exchange of sublattices $A \leftrightarrow B$ to show that $\mathbf{d}(\mathbf{k}) = \mathbf{d}^*(-\mathbf{k})$.

(b) Use time-reversal symmetry to show that $\mathbf{d}(\mathbf{k}) = 0$.

From the above results it follows that to include spin-orbit interaction, one should go beyond nearest neighbors hopping and include the next-nearest neighbors. In graphene, it implies hopping between sites belonging to the same sublattice. Hence, assume that the spin-orbit interaction associated with a transition from two sites of sublattice A is of the form

$$H_{\text{so}}^{A \rightarrow A}(\mathbf{k}) = -it_2 \sum_{a'(nnn)} v_{aa'}(\tau_z^S) \exp[i\mathbf{k} \cdot (\mathbf{a}' - \mathbf{a})]. \quad (7.74)$$

Here t_2 is a constant that characterizes the coupling, and the factor $v_{aa'}(\tau_z^S) = \pm 1$ is determined according to the direction of the spin and the path of the electrons between the points. In Fig. 7-16, we show the sign of this factor for one of the spin states (say spin up). The rule is the following: $v_{aa'}(\uparrow) = +1$ if while passing from A sublattice point to its neighboring point on the same sublattice, a nearby point of sublattice B is on the right side of the path, while $v_{aa'}(\uparrow) = -1$ if the point is on the left side of the path. For a spin pointing down, these signs should be reversed. This choice respects the symmetries of spin-orbit interaction (3.23).

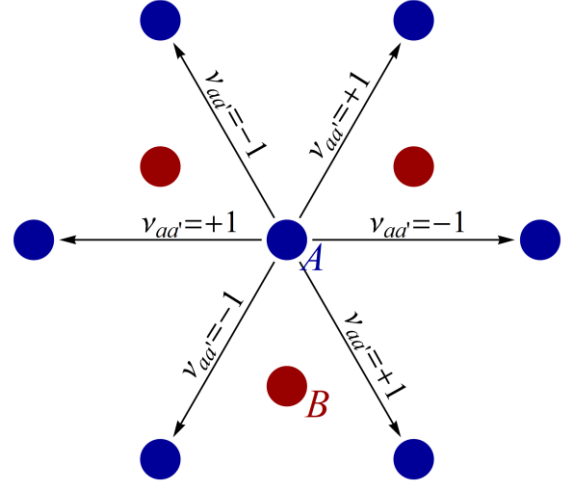


Figure 7-16 Extension of the tight binding model to include next nearest neighbor hopping in order to account for spin-orbit interaction

(c) Calculate the term (7.74), expand it to the leading (zeroth) order in the distance from the K and K' points, and show that it is identical to the Kane-Mele term given by Eq. (6.24). Find the relation between the parameters λ_{so} and t_2 .

8 Topological metals

This course is centered around the role of symmetry in condensed matter physics. Among other issues, its role in identifying degeneracy points in the Brillouin zone. However, another aspect of degeneracy points also applies to cases where the degeneracy is accidental - the aspect of topology. In a nutshell, topology is concerned with characterizing geometric objects, such as the surfaces illustrated in Fig. 8-1, by numbers. In the examples of Fig. 8-1, it is the number of holes in the surface. The left panel of the figure shows a surface that can be continuously deformed into a sphere (with no tearing or gluing), and therefore we say that it has the topology of a sphere. The object on the right panel of the figure has the topology of a torus, which is characterized by a single hole (a mug with one handle has the same topology). The number of holes in a closed surface is a topological (integer) number called the *genus* of the surface.

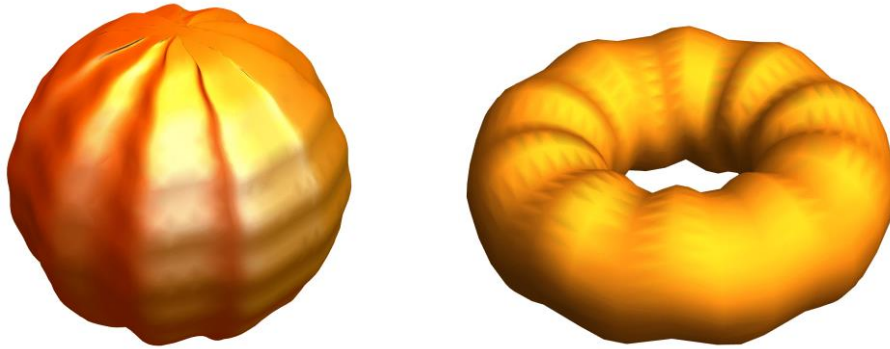


Figure 8-1 Surface having the topology of a sphere (left panel) and a torus (right panel)

As we shall see in this chapter, in some cases, degeneracy points of electronic spectra can also be associated with topological numbers. The importance of this characterization rests on the fact that, in general, these degeneracy points are not affected by small perturbations. Similar to smooth deformations of surfaces that do not change the genus of the surface, degeneracy points associated with topological numbers preserve their identity under perturbation like impurity scattering or application of some external forces. It is simply because integer numbers cannot be changed continuously. In these situations, we say that the degeneracy points are “protected by topology”.

To introduce the main ideas of this field, we begin this chapter with a qualitative discussion of mercury telluride (HgTe), showing how to obtain accidental (but robust) degeneracy points in the spectrum, called Weyl points. Next, we present some general ideas of topology, such as parallel transport and curvature, to motivate the introduction of Berry’s curvature in quantum mechanical systems. The latter allows us to calculate the topological numbers associated with band touching points. Finally, we discuss the exotic behavior of the surface states in materials with the Weyl points spectrum.

8.1 Energy bands in mercury telluride – a qualitative discussion.

Mercury telluride has the structure of a zinc blende crystal. Thus, the symmetry group of this crystal is that of the tetrahedral point group T_d . This group is also the little group at the Γ point where $\mathbf{k} = 0$. The point group T_d has two irreducible representations of dimension 3, one of dimension 2, and two additional one-dimensional representations. Thus ignoring the spin degree of freedom, the spectrum near the Γ point is expected to contain the typical band structure of light and heavy holes with three-fold degeneracy emerging from the three p -type orbitals of the Te atoms and the s -type orbital of the Hg atoms, as demonstrated in the left panel of Fig. 8-2 (see also Fig. 7-13). In this figure, we denote degenerate energy bands by thick lines.

Now let us take into account the spin degree of freedom. In the absence of spin-orbit interaction, all energy bands become doubly degenerate. However, in the presence of spin-orbit interaction, one expects some of the degeneracy to be lifted. To identify the dimensions of the irreducible representations of the double group T'_d , one can see from the rules on page 116 that it has eight conjugacy classes: $\{E\}$, $\{Q\}$, $\{8C_3\}$, $\{8QC_3\}$, $\{3C_2, 3QC_2\}$, $\{6S_4\}$, $\{6QS_4\}$, and $\{6\sigma, 6Q\sigma\}$. Hence it has eight irreducible representations. Five of these must be those of the original group (without spin), containing 24 symmetry elements. Therefore, the dimensions of the additional three representations satisfy the condition $\ell_6^2 + \ell_7^2 + \ell_8^2 = 24$. The only way to satisfy this equation is by choosing the dimensions of 4, 2, and 2. Thus the highest possible normal degeneracy is fourfold.

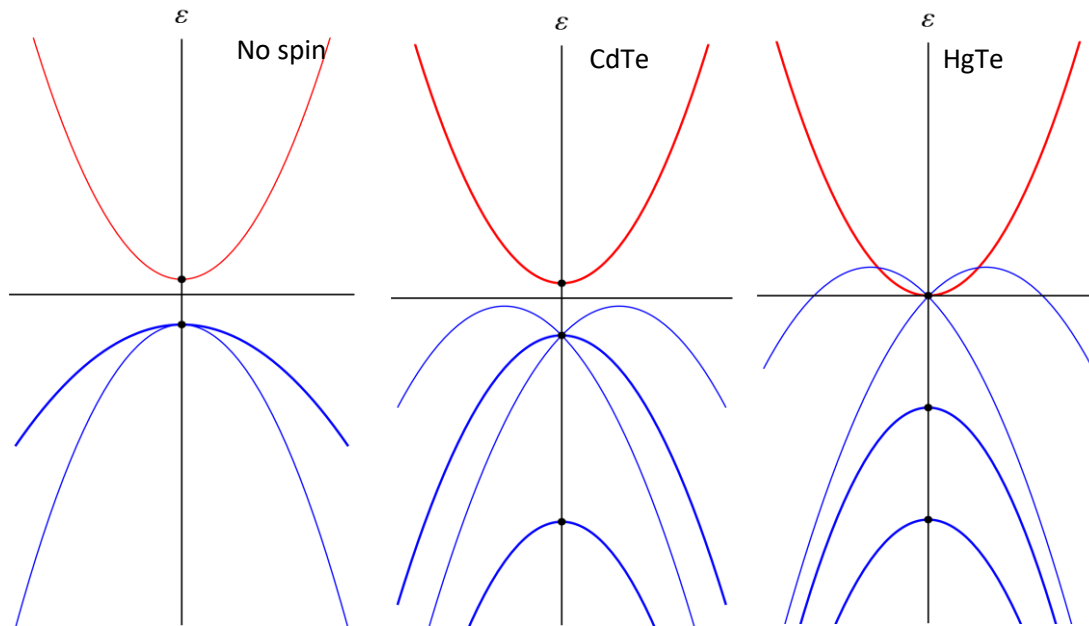


Figure 8-2 A schematic illustration of the energy levels in a zinc blende crystal near the Γ point. Twofold degenerate levels are drawn by thick lines, whereas nondegenerate levels are drawn by thin lines. Left panel is the band structure in the absence of spin. Middle panel shows the normal ordering of levels in CdTe with spin-orbit interaction, and the right panel shows *band inversion* in HgTe that appears when spin-orbit coupling is strong.

From the above analysis, one expects that the typical band structure in the presence of spin-orbit interaction will be similar to that depicted in the middle panel of Fig. 8-2. This is indeed the situation in cadmium telluride which has the same crystal structure of HgTe. However, in mercury telluride, relativistic effects are much stronger (because the atomic number of mercury is 80 while that of cadmium is only 48), and these push the s -type band down such that it crosses the two p -type bands, as demonstrated in the right panel of Fig. 8-2. This phenomenon is called *band inversion*.

Band inversion generates two accidental degeneracy points along the $[1,1,1]$ direction. They are accidental because they are not located at any symmetry point of the Brillouin zone. To understand why, recall that Zinc blende crystals are made of two interpenetrating fcc sublattices (see Fig. 7-14), therefore, the Bravais lattice is fcc, and the reciprocal lattice is bcc. The corresponding Brillouin zone with its special points is presented in Fig. 8-3. The global band structure of HgTe, obtained from *ab initio* calculations, and its magnification near the Γ point, is shown in Fig. 8-4. Notice the difference between the energy scales of the two panels.

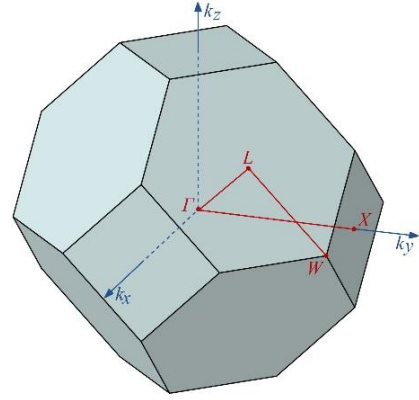


Figure 8-3 The Brillouin zone of HgTe

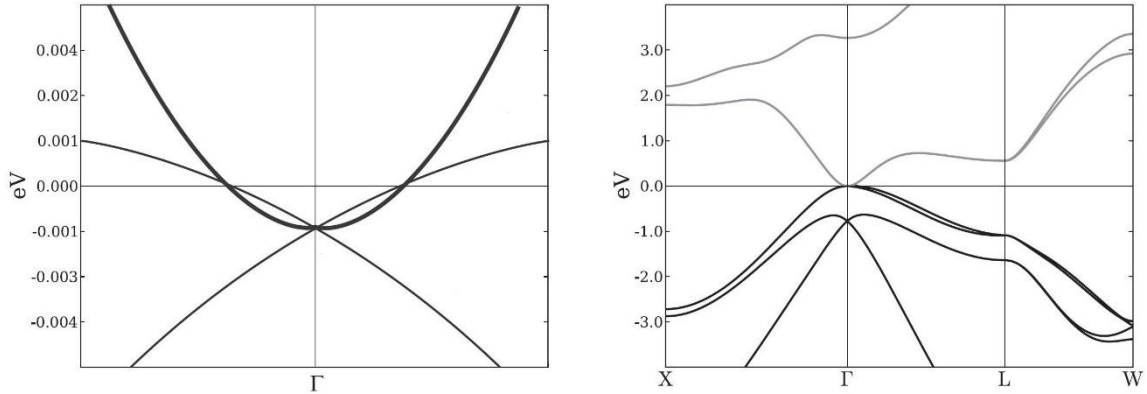


Figure 8-4 The band structure of HgTe. The left panel shows the local behavior near the Γ point along the $(1,1,1)$ direction (i.e., the line passing through the Γ and the L points in Fig. 8-3). The right panel shows the global band structure along the red path presented in Fig. 8-3. Taken from Zaheer et al. Phys. Rev B **87** 045202 (2013).

The fourfold degeneracy at the Γ point results from the high symmetry of the crystal. One may lift this degeneracy by reducing the symmetry of the lattice, say by applying external stress. A stress applied along the $[1,1,1]$ direction can be either compressive or tensile. In both cases, it reduces the tetrahedral symmetry to C_{3v} , and lifts the degeneracy. However, the degeneracy is lifted in two different manners, as demonstrated in Fig. 8-5. The compressive stress opens a gap in the spectrum, while tensile stress leaves lift the

degeneracy into two degenerate points below and above the Fermi level, as shown in the magnified view of the region near the Γ point shown in the lower panel of the figure.

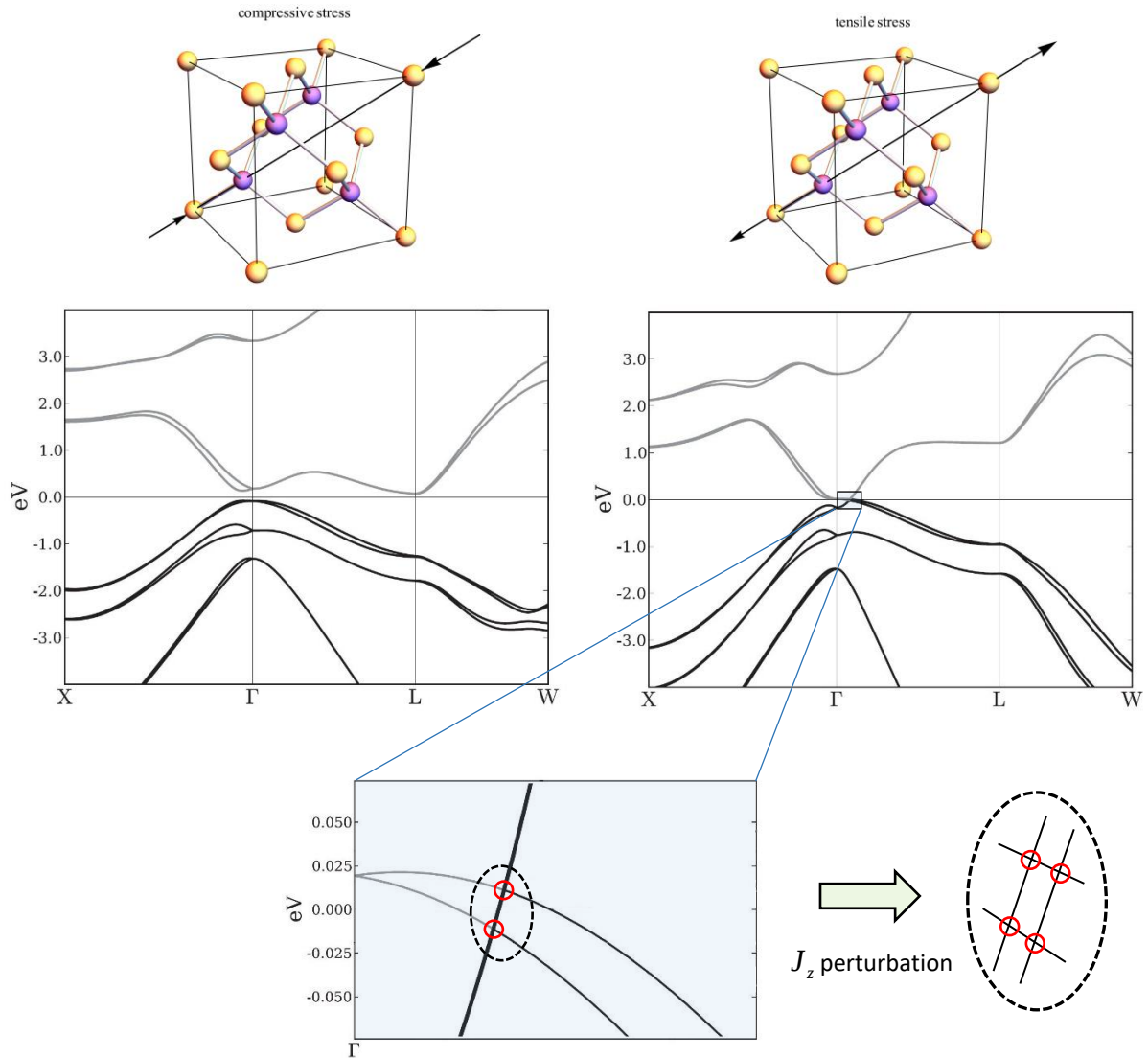


Figure 8-5 The band structure of HgTe under compressive (left) and tensile (right) stress along the $[1,1,1]$ direction of the crystal. The lower right panel shows the splitting into four degeneracy points due to addition J_z perturbation. Adapted from Zaheer et al. Phys. Rev B **87**045202 (2013).

Each of the points marked by the red circles in the lower panel of Fig. 8-5 is threefold degenerate due to the mirror symmetry of the deformed crystal about the plane that contains the $[1,1,1]$ axis. However, this symmetry can be broken by adding a perturbation proportional to J_z - the component of the angular momentum of the p -type orbitals in the z direction.

The four degeneracy points that we obtained here are *Weyl points* or *Weyl nodes*, and a system with a spectrum that comprises Weyl points is called *Weyl semimetal*. It should be emphasized that these are accidental degeneracy points obtained by “fine-tuning” of

parameters and not due to symmetry. However, once formed, they are robust to perturbations; namely, they will not disappear due to additional system factors that cannot be controlled, such as impurity scattering or application of weak external forces.

8.2 Weyl points

The distinctive feature of Weyl nodes is that their linear spectrum is in the vicinity of the band touching point. A 2×2 matrix describes the local spectrum near such a point (in the $k \cdot p$ approximation) because only two energy levels are involved. Hence, the most general form of the local Hamiltonian is

$$H(\delta \mathbf{k}) = \hbar \sum_{ij} v_{ij} \delta k_i \tau_j \quad \text{with} \quad \det v_{ij} \neq 0, \quad (8.1)$$

where $\delta \mathbf{k}$ is the wavenumber vector measured from the Weyl point, τ_j are Pauli matrices, v_{ij} are some system-dependent parameters (having the dimension of velocity), and without loss of generality, we choose the energy at the degeneracy point to be zero.

The Hamiltonian $H(\delta \mathbf{k})$ is robust to perturbations, i.e., there is nothing we can do to get rid of the band touching point. The reason is that we have used all 3 Pauli matrices; therefore, any perturbation (which is not a trivial constant shift of the energy levels) is of the form $\boldsymbol{\mu} \cdot \boldsymbol{\tau}$, where $\boldsymbol{\mu}$ is some general vector function of $\delta \mathbf{k}$. However, this perturbation can only shift the position of the Weyl point to a different location in the Brillouin zone or change the slopes of the dispersion by altering the parameters v_{ij} . Denoting by $v_F^{(i)}$ the eigenvalues of the matrix v_{ij} and redefining the directions of the coordinate system according to the corresponding eigenvectors, we obtain that the energy spectrum is

$$\varepsilon = \pm \sqrt{\sum_i \left(v_F^{(i)} \delta k_i \right)^2}. \quad (8.2)$$

This spectrum is linear, similar to that obtained at the K -point of graphene (in the absence of spin-orbit interaction). An illustration of this spectrum along the line, $\delta k_z = 0$, is shown in Fig. 8-6.

In order to have a spectrum with Weyl points, one must have two non-degenerate energy bands at any value of $\mathbf{k}$ other than that of the Weyl point. This condition, however, cannot be satisfied in a system with both time-reversal

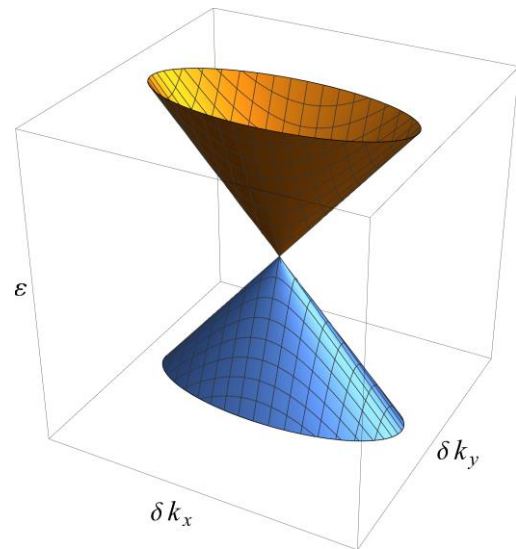


Figure 8-6 The spectrum near Weyl point

and inversion symmetry. The reason is that the time-reversal operator, Θ , takes $\mathbf{k} \rightarrow -\mathbf{k}$ and flips the direction of the spin, while the inversion operator, i , takes $\mathbf{k} \rightarrow -\mathbf{k}$ without flipping the spin direction. Thus, applying the two operations, $i\Theta$, on a wave function yields another state with the same wave number vector but with a flipped spin. However, if the system is symmetric to both operations, this new state must have the same energy as the original one. In other words, the energy at each value of $\mathbf{k}$ is doubly degenerate. From these considerations, it follows that either inversion symmetry or time-reversal symmetry (or both) must be broken in a system with Weyl points spectrum.

Consider now the isotropic limit of Hamiltonian (8.1), where $v_{ij} = \pm v_F \delta_{ij}$, and assume that Weyl point is at the origin:

$$H_{\pm} = \pm \hbar v_F \mathbf{k} \cdot \boldsymbol{\tau} \quad (8.3)$$

One obtains the same spectrum, $\varepsilon = \pm \hbar v_F |\mathbf{k}|$, for either choice of sign in this Hamiltonian. However, the wavefunctions corresponding to the same energy are different. Treating $\boldsymbol{\tau}$ similar to a spin degree of freedom (which we call pseudospin), the eigenstates of the system can be divided into two categories: One, when the pseudospin is parallel to $\mathbf{k}$ (right-handed chirality), and the second is when it is antiparallel (left-handed chirality). A state with positive energy will have $\mathbf{k}$ parallel to $\boldsymbol{\tau}$ for H_+ while antiparallel for H_- . Accordingly, one can associate a number (chiral charge) ± 1 to Weyl points described by $H_{\pm}$. As we shall see below, this number is a topological number similar to the genus of a surface.

Comment: When $\boldsymbol{\tau}$ is taken to be the real spin of a fermion and v_F replaced by the speed of light, Hamiltonian (8.3) becomes the *Weyl Hamiltonian* for massless relativistic fermions - called *Weyl fermions*.

8.3 Curvature, parallel transport, and topological numbers

My goal now is to weave the relation between Weyl points and topology. To this end, it will be instructive to start by presenting some of the basic ideas of topology with the help of two-dimensional surfaces. These ideas will be generalized to quantum mechanical systems in the next section

One of the intrinsic properties of a smooth surface is its *Gaussian curvature*. To define it, consider a point, $\mathbf{r}$, on the surface. The normal to the surface at this point, $\mathbf{n}$, is a vector perpendicular to the tangent plane at the same

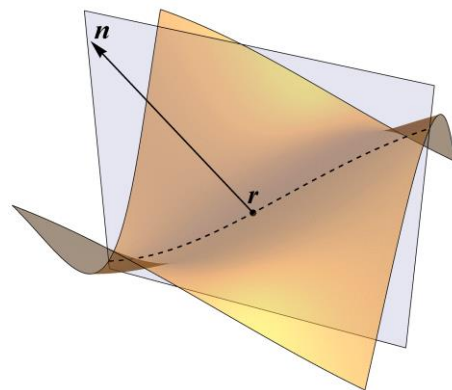


Figure 8-7 The normal vector to a surface, a normal plane and the corresponding normal section a surface (dashed line)

point. Planes that contain the normal vector are called *normal planes*, and their intersections with the surface are curves called *normal sections*, see Fig. 8-7. For each normal section, one can define the osculating circle at the point, $\mathbf{r}$. It is the circle residing on the normal plane and touches the point $\mathbf{r}$ additional pair of points on the normal section (one from each side of $\mathbf{r}$) that are infinitesimally close to $\mathbf{r}$, see Fig. 8-8.

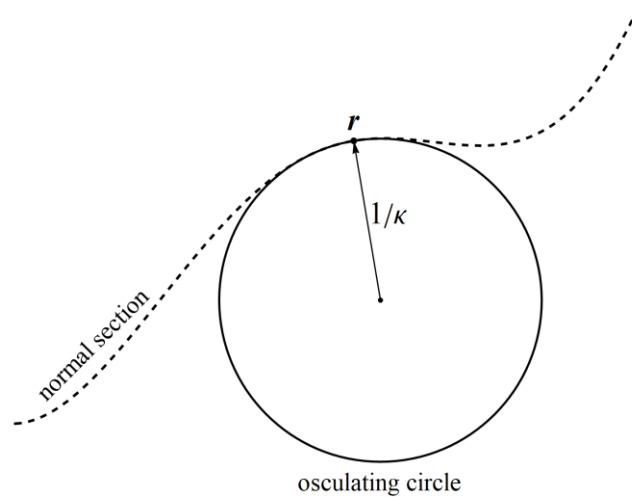


Figure 8-8 the osculating circle at point $\mathbf{r}$ of a normal section

The normal curvature, κ , associated with a normal plane is the inverse of the radius of the osculating circle. In general, this curvature depends on the direction of the normal plane. The maximal and minimal curvatures, κ_1 and κ_2 , are called the *principal curvatures*. The *Gaussian curvature* of the surface at the point $\mathbf{r}$ is defined as the product $K = \kappa_1 \kappa_2$. The sign of the Gaussian curvature is positive if both osculating circles, associated with the principal curvatures, reside on the same side of the surface, while it is negative if they are located on both sides of the surface, see Fig. 8-9.

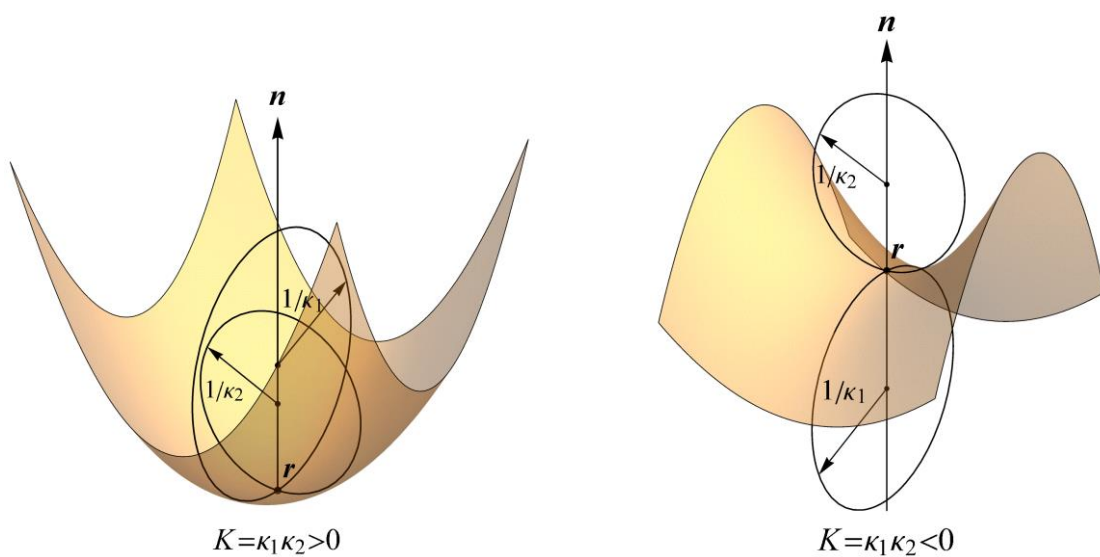


Figure 8-9 Examples for points on a surface with positive (left) and negative (right) Gaussian curvatures

The above definition of the Gaussian curvature utilizes the accommodation of the surface in a three-dimensional space. But imagine the existence of intelligent two-dimensional creatures living on a two-dimensional universe in the form of a surface - would they be able to know and measure the Gaussian curvature of their world even though they cannot move into the third dimension? The answer is yes! and it relies on an alternative definition of the Gaussian curvature based on a procedure called “*parallel transport*”.

Parallel transport is a method for transporting geometrical objects (say vectors) along curves on manifolds (say surfaces). Let $C(t)$ denote a curve on a manifold parameterized by t , and let $V(t)$ be a vector field defined along the curve. In a flat space, we say that $V(t)$ is kept parallel on $C(t)$ if $dV(t)/dt = 0$. In other words, when the direction of $V(t)$ is independent of its position along the curve, as demonstrated in the left panel of Fig. 8-9. However, if the $C(t)$ resides on a non-flat surface, the situation becomes more complicated. We say that $V(t)$ is parallel transported along the curve if the orthogonal projection of $dV(t)/dt = 0$ on the tangent plane to the surface at $C(t)$ is zero, see right panel of Fig. 8-10.

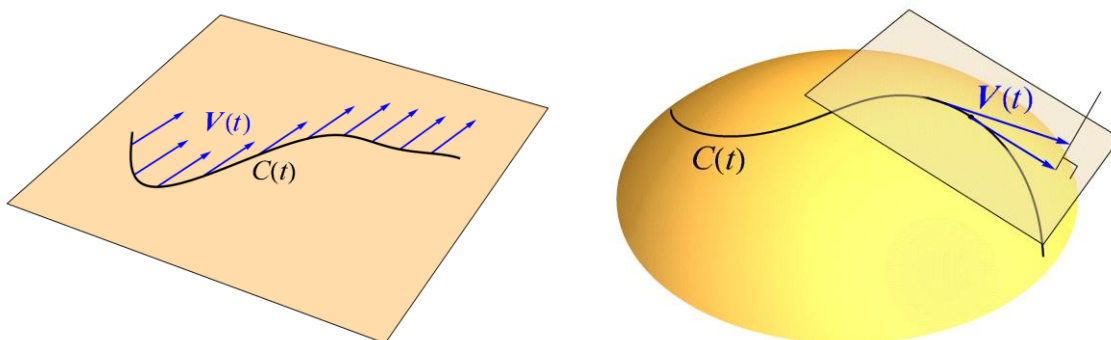


Figure 8-10 Parallel transport of vectors along a curve in a flat (left) and curved (right) surface

This definition ensures that the two-dimensional creatures walking along the curve $C(t)$ will see no change in the direction of the vector $V(t)$. Notice that to characterize parallel transport one needs only the curve and the tangent planes to the surface. Therefore, parallel transport of a vector along a curve, $C(t)$, is the same for any two surfaces tangent along $C(t)$. We shall use this property in the following example.

Example: Parallel transport along a latitude circle on a sphere.

Parallel transport along some general curves can be tricky to visualize. As an example, let us consider the parallel transport along a latitude circle on a sphere, as illustrated in Fig.

8-11. Here one can use the fact that the same parallel transport is obtained for a circle on a conical surface that is tangent to the sphere, as shown in the figure. However, the curvature of a conical surface is zero everywhere (except at the cone apex where it is not defined); therefore, we may cut open the cone along the dashed line shown in the left panel of the figure and flatten it onto a plane as demonstrated in the right panel. On a plane, parallel transport becomes easy because one has to keep the vector in the same direction as shown in the right panel of the figure. Gluing the cone back along the dashed line gives the parallel transport of the vector on the sphere. Notice that, in general, the parallel transport of a vector along a closed curve on a surface (which is not flat) results in a finite angle δ between the initial and final directions of the vector. Simple geometrical calculation shows that for a latitude circle, with angle θ , the angle δ defined in the figure is $\delta = 2\pi(1 - \cos \theta)$ ¹.

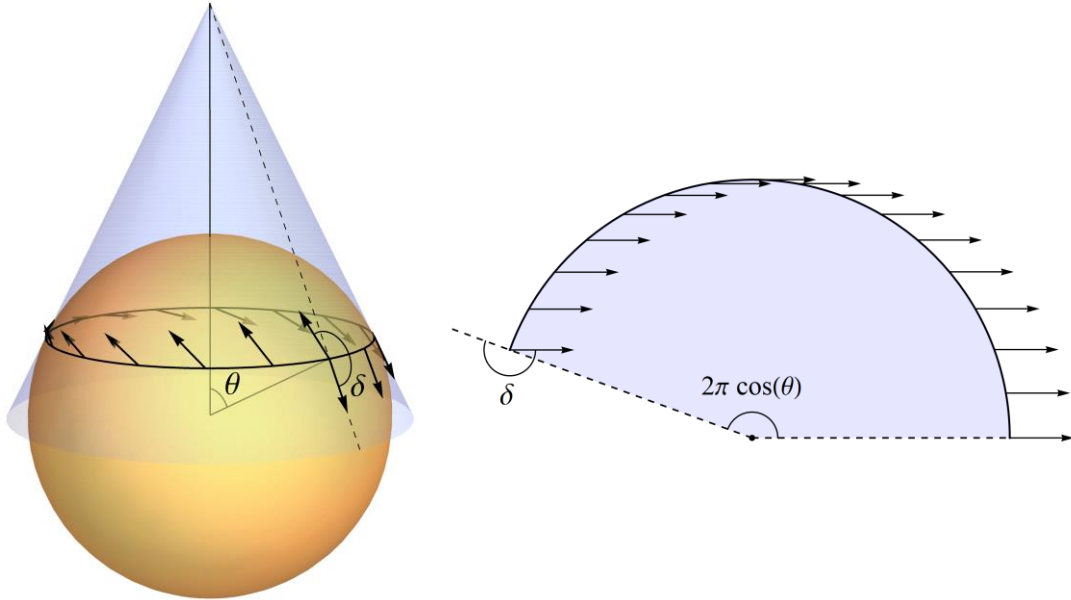


Figure 8-11 Parallel transport on a latitude circle on a sphere

Parallel transport becomes much easier on geodesics. A geodesic curve is defined as a curve along which parallel transport keeps a fixed angle between the vector and the tangent to the curve. It can also be shown that a geodesic going through nearby points is the shortest possible trajectory between these points (when the points are far apart, a geodesic is a stationary path, namely a path whose length is, essentially, not affected by small deformations).

¹ The angle $-2\pi \cos \theta$ is the precession angle of Foucault pendulum after 24 hours. Also, the contribution of Thomas precession to spin-orbit interaction comes from a similar reason. It is due to curvature effect which results in rotation of the spin by an angle $-2\pi \cosh(v/c)$ (where v is the particle velocity and c is the speed of light) upon completing a closed trajectory.

Example: Parallel transport along geodesics on a sphere.

The geodesics on a sphere are great circles, i.e. intersections of the sphere with planes passing through its center. One can construct a closed-loop from few geodesics, say by starting from the north pole and descending along a longitude to the equator; then follow the equator for some distance, and come back to the north pole along a different longitude, as shown in Fig. 8-12. Starting from a vector parallel to the latitude we see, that parallel transport along this closed path results in a finite angle between the initial and final vectors.

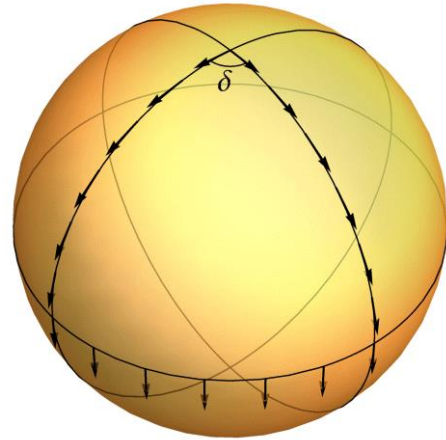


Figure 8-12 Parallel transport along geodesics

Clearly, the curvature of the surface affects the value of the angle δ obtained by parallel transport because it is always zero in a flat plane. Moreover, it turns out that the Gaussian curvature can be defined in terms of this angle as follows: To calculate the curvature at some point, $\mathbf{r}$, on the surface, we construct a closed-loop (on the surface) around that point, call it C_r , and calculate the mismatch angle, δ_{C_r} , obtained by parallel transport along that loop. If the area enclosed by the loop is A_{C_r} then the Gaussian curvature is obtained from the limit where the loop length shrinks to zero by

$$K = \lim_{\|C_r\| \rightarrow 0} \frac{\delta_{C_r}}{A_{C_r}}. \quad (8.4)$$

In the example of the parallel transport along the latitude, we obtained that $\delta = 2\pi(1 - \cos\theta)$. On the other hand, the area enclosed by the loop is $A = R^2 2\pi(1 - \cos\theta)$, where R is the sphere radius, hence the ratio $\delta/A = 1/R^2$ (which is independent of θ because we consider a simple case) is the curvature of the sphere.

Finally, the relation between the Gaussian curvature, K , and the genus, g , of a closed surface, Σ , is given by the *Gauss-Bonnet theorem* (which we present without proof):

$$\frac{1}{4\pi} \oint_{\Sigma} ds K = 1 - g. \quad (8.5)$$

Here the integral is over the whole surface and ds is an infinitesimal area element on this surface. The Gauss-Bonnet theorem shows that the topological number g (the number of holes in the surface) can be obtained from an integral of the curvature over the surface. It is straightforward to check that the above formula applies to a sphere.

8.4 Berry's curvature

Consider the Hamiltonian, H_x , which depends on a set of parameters, $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$. These parameters can be, for example, the three components of an external magnetic field, the strength and the direction of tensile stress acting on a system, some microscopic parameters that determine the structure of the potential, etc. For simplicity, we assume the system size to be finite, such that the energy levels are discrete and that the system does not possess symmetries so that all degeneracies are accidental. For each typical value of $\mathbf{x}$, diagonalization of the Hamiltonian, H_x , yields a different set of non-degenerate energy levels and wave functions:

$$H_x \left| \psi_x^{(n)} \right\rangle = \varepsilon_x^{(n)} \left| \psi_x^{(n)} \right\rangle. \quad (8.6)$$

Now suppose that we can change these parameters, $\mathbf{x}(t)$, very slowly in time (as slow as we wish), then adiabatic theorem tells us that assuming the system does not pass through a degeneracy point, it will remain in the same n -th eigenstate, $\left| \psi_{\mathbf{x}(t)}^{(n)}(t) \right\rangle$, at any time t , if it has been prepared in the n -th eigenstate when $t=0$, $\left| \psi_{\mathbf{x}(0)}^{(n)}(0) \right\rangle$. This property also holds when the energy of the n -th state changes significantly over time.

Now consider the adiabatic evolution of the system along a closed contour in the parameter space, $C(t)$, such that at times $t=0$ and $t=T$ the control parameters are equal, $\mathbf{x}(T) = \mathbf{x}(0)$. In this case, the wave function of the system must return to its initial state up to a phase factor:

$$\left| \psi_{\mathbf{x}(T)}^{(n)}(T) \right\rangle = \exp(i\phi) \left| \psi_{\mathbf{x}(0)}^{(n)}(0) \right\rangle. \quad (8.7)$$

Naively, one would guess that if the change of the parameters is sufficiently slow, then for each infinitesimal time interval the energy of the particle, $\varepsilon_{\mathbf{x}(t)}^{(n)}$, can be assumed to be constant; therefore the phase, ϕ , is the total dynamical phase accumulated during the evolution of the system, i.e.,

$$\phi_{\text{dynamical}} = -\frac{1}{\hbar} \int_0^T \varepsilon_{\mathbf{x}(t)}^{(n)} dt. \quad (8.8)$$

(In the case where the parameters are kept fixed in time, $\mathbf{x}(t) = \mathbf{x}$, this formula reduces to the familiar result, $\phi_{\text{dynamical}} = -\varepsilon_x^{(n)} T / \hbar$). However, Berry showed that there is an additional contribution to the phase which is geometric,

$$\phi = \phi_{\text{dynamical}} + \gamma_n(C), \quad (8.9)$$

where $\gamma_n(C)$ depends on the contour C in the parameter space (but independent of its time parametrization). This phase is called the *Berry phase* (Michal Berry calls it the *geometrical phase*).

Before we turn to calculate $\gamma_n(C)$, let us draw some parallels between quantum mechanical systems and two-dimensional geometrical surfaces: The parameter space in the quantum problem is analogous to the two-dimensional surface; Close contours on this parameter space correspond to closed loops on the surface; Vectors on the geometrical surface are analogous to eigenfunctions of the Hamiltonian, and parallel transport is the adiabatic evolution of the system. Finally, the mismatch phase that we have obtained by parallel transport along a closed contour, δ , is analogous to Berry's phase, $\gamma_n(C)$, as we shall see below.

This analogy suggests that one can also define the curvature of a quantum system in the parameter space similar to the definition of Gaussian curvature (8.4), and that the system may be characterized by topological numbers obtained by an integral of the curvature, analogously to the Gauss-Bonnet theorem (8.5).

To calculate Berry's phase, we solve the time-dependent Schrödinger equation,

$$i\hbar \frac{\partial}{\partial t} |\psi_{\mathbf{x}(t)}^{(n)}(t)\rangle = H_{\mathbf{x}(t)} |\psi_{\mathbf{x}(t)}^{(n)}(t)\rangle, \quad (8.10)$$

with a wave function of the form

$$|\psi_{\mathbf{x}(t)}^{(n)}(t)\rangle = \exp\left(-i \frac{1}{\hbar} \int_0^t \mathcal{E}_{\mathbf{x}(t')}^{(n)} dt' + i\gamma^{(n)}(t)\right) |\psi_{\mathbf{x}(t)}^{(n)}\rangle. \quad (8.11)$$

Here, $\gamma^{(n)}(t)$ is some unknown function of the time, while $|\psi_{\mathbf{x}(t)}^{(n)}\rangle$ is a solution of the time-independent Schrödinger equation (8.6) with $\mathbf{x} = \mathbf{x}(t)$. The left-hand side of Eq. (8.10) gives:

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} |\psi_{\mathbf{x}(t)}^{(n)}(t)\rangle &= i\hbar \frac{\partial}{\partial t} \exp\left(-i \frac{1}{\hbar} \int_0^t \mathcal{E}_{\mathbf{x}(t')}^{(n)} dt' + i\gamma^{(n)}(t)\right) |\psi_{\mathbf{x}(t)}^{(n)}\rangle \\ &= i\hbar \left[-\frac{i}{\hbar} \mathcal{E}_{\mathbf{x}(t)}^{(n)} + i \frac{d\gamma^{(n)}(t)}{dt} \right] |\psi_{\mathbf{x}(t)}^{(n)}(t)\rangle \\ &\quad + i\hbar \exp\left(-i \frac{1}{\hbar} \int_0^t \mathcal{E}_{\mathbf{x}(t')}^{(n)} dt' + i\gamma^{(n)}(t)\right) \frac{\partial \mathbf{x}(t)}{\partial t} \cdot \left| \frac{\partial}{\partial \mathbf{x}} \psi_{\mathbf{x}}^{(n)} \right\rangle \Big|_{\mathbf{x}=\mathbf{x}(t)} \end{aligned} \quad (8.12)$$

while from (8.6), the right-hand side is:

$$H_{\mathbf{x}(t)} |\psi_{\mathbf{x}(t)}^{(n)}(t)\rangle = \mathcal{E}_{\mathbf{x}(t)}^{(n)} |\psi_{\mathbf{x}(t)}^{(n)}(t)\rangle. \quad (8.13)$$

Hence equating the last two expressions we obtain:

$$-\hbar \frac{d\gamma^{(n)}(t)}{dt} \left| \psi_{\mathbf{x}(t)}^{(n)} \right\rangle + i\hbar \frac{\partial \mathbf{x}(t)}{\partial t} \cdot \left| \frac{\partial}{\partial \mathbf{x}} \psi_{\mathbf{x}}^{(n)} \right\rangle \bigg|_{\mathbf{x}=\mathbf{x}(t)} = 0. \quad (8.14)$$

Multiplying the above equation from the left by $\left\langle \psi_{\mathbf{x}(t)}^{(n)} \right|$ we obtain:

$$\frac{d\gamma^{(n)}(t)}{dt} = i \frac{\partial \mathbf{x}(t)}{\partial t} \cdot \left\langle \psi_{\mathbf{x}}^{(n)} \right| \frac{\partial}{\partial \mathbf{x}} \left| \psi_{\mathbf{x}}^{(n)} \right\rangle \bigg|_{\mathbf{x}=\mathbf{x}(t)}. \quad (8.15)$$

The Berry phase is obtained by integrating this formula along the closed contour in the parameter space,

$$\gamma_n(C) = \int_0^T dt \frac{d\gamma^{(n)}(t)}{dt} = i \int_0^T dt \frac{\partial \mathbf{x}(t)}{\partial t} \cdot \left\langle \psi_{\mathbf{x}}^{(n)} \right| \frac{\partial}{\partial \mathbf{x}} \left| \psi_{\mathbf{x}}^{(n)} \right\rangle \bigg|_{\mathbf{x}=\mathbf{x}(t)}, \quad (8.16)$$

and changing the variables of integration we obtain

$$\gamma_n(C) = i \oint_C d\mathbf{x} \cdot \left\langle \psi_{\mathbf{x}}^{(n)} \right| \frac{\partial}{\partial \mathbf{x}} \left| \psi_{\mathbf{x}}^{(n)} \right\rangle. \quad (8.17)$$

This formula shows that Berry's phase is purely geometrical because it is independent of time. It only depends on the contour in the parameter space. One can also prove that it is purely real (see Ex. 2), thus

$$\gamma_n(C) = -\text{Im} \oint_C d\mathbf{x} \cdot \left\langle \psi_{\mathbf{x}}^{(n)} \right| \frac{\partial}{\partial \mathbf{x}} \left| \psi_{\mathbf{x}}^{(n)} \right\rangle. \quad (8.18)$$

The integrand in the above formula is called *Berry's connection*.

Finally, for simplicity, let us restrict the discussion to a three-dimensional parameter space where $\mathbf{x} = (x_1, x_2, x_3)$. In this case, we can use Stokes' theorem to replace the above contour integral with an area integral:

$$\gamma_n(C) = -\text{Im} \iint ds \cdot \frac{\partial}{\partial \mathbf{x}} \times \left\langle \psi_{\mathbf{x}}^{(n)} \right| \frac{\partial}{\partial \mathbf{x}} \left| \psi_{\mathbf{x}}^{(n)} \right\rangle, \quad (8.19)$$

where ds is an infinitesimal area element in the parameter space whose direction is determined by the right-hand rule with respect to the direction of the contour C . The integral is taken over an arbitrary surface whose boundary is the contour C , as shown in Fig. 8-13.

Using the identity $\nabla \times \nabla f(\mathbf{x}) = 0$ we can rewrite Eq. (8.19) in the form:

$$\gamma_n(C) = -\text{Im} \iint ds \cdot \mathbf{V}_n(\mathbf{x}), \quad (8.20)$$

where

$$\mathbf{V}_n(\mathbf{x}) = \text{Im} \frac{\partial}{\partial \mathbf{x}} \times \left\langle \psi_x^{(n)} \left| \frac{\partial}{\partial \mathbf{x}} \right| \psi_x^{(n)} \right\rangle = \text{Im} \left\langle \frac{\partial}{\partial \mathbf{x}} \psi_x^{(n)} \left| \times \right| \frac{\partial}{\partial \mathbf{x}} \psi_x^{(n)} \right\rangle \quad (8.21)$$

For sufficiently small loop in the parameter space, $\mathbf{V}_n(\mathbf{x})$ is approximately constant. Thus, the integral (8.20) may be replaced with the product of $\mathbf{V}_n(\mathbf{x})$ by the area of the surface enclosed by the loop. Thus, in analogy with definition (8.4) of the Gaussian curvature, $\mathbf{V}_n(\mathbf{x})$ should be interpreted as the curvature of the parameter space. This curvature is called *Berry's curvature* or the *adiabatic curvature*.

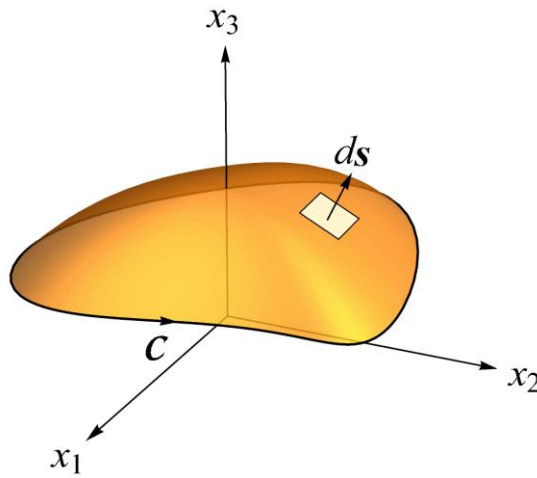


Figure 8-13 The surface integral for the Berry phase in a three dimensional parameter space

Example: The curvature near a band touching points (Weyl points)

In the vicinity of a band touching point, only two bands are essential, and the Hamiltonian of a system can be approximated by a 2×2 matrix of the form (8.3). This Hamiltonian contains three parameters which are the three components of Bloch's wavenumber vector. The number three is also the minimal number of parameters needed to obtain an accidental degeneracy point. In what follows, we calculate Berry's curvature of the Hamiltonian (8.3), where $\mathbf{x} = \mathbf{k} = (k_x, k_y, k_z)$ is the parameter space. Consider first, H_+ , and let us calculate Berry's curvature associated with the upper energy level $|\mathbf{d}\rangle$ for which

$$H_+ |\mathbf{d}\rangle = \hbar v_F k |\mathbf{d}\rangle. \quad (8.22)$$

In chapter 3 (see Eqs. (3.36), (3.37) and Ex. 2), we proved that this state is given by:

$$|\mathbf{d}\rangle = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right)\exp(i\phi) \end{pmatrix}, \quad (8.23)$$

where θ and ϕ are polar coordinates that parametrize the Bloch wavenumber:

$$\mathbf{k} = k(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta). \quad (8.24)$$

From formula (8.21), it follows that the l -th component of Berry's curvature is given by

$$[\mathbf{V}(\mathbf{k})]_l = \left[\text{Im} \frac{\partial}{\partial \mathbf{k}} \times \langle \mathbf{d} | \frac{\partial}{\partial \mathbf{k}} | \mathbf{d} \rangle \right]_l = \text{Im} \varepsilon_{ijl} \frac{\partial}{\partial k_i} \langle \mathbf{d} | \frac{\partial}{\partial k_j} | \mathbf{d} \rangle, \quad (8.25)$$

where ε_{ijl} is the antisymmetric tensor and repeated indices should be summed over. Using Eq. (8.23), we have

$$\left| \frac{\partial}{\partial k_j} \mathbf{d} \right\rangle = \frac{\partial}{\partial k_j} \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right)e^{i\phi} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -\sin\left(\frac{\theta}{2}\right) \\ \cos\left(\frac{\theta}{2}\right)e^{i\phi} \end{pmatrix} \frac{\partial \theta}{\partial k_j} + \begin{pmatrix} 0 \\ i \sin\left(\frac{\theta}{2}\right)e^{i\phi} \end{pmatrix} \frac{\partial \phi}{\partial k_j}, \quad (8.26)$$

therefore

$$\langle \mathbf{d} | \frac{\partial}{\partial k_j} | \mathbf{d} \rangle = i \sin^2\left(\frac{\theta}{2}\right) \frac{\partial \phi}{\partial k_j}. \quad (8.27)$$

Substituting this result in Eq. (8.25) we obtain

$$[\mathbf{V}(\mathbf{k})]_l = \sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)\varepsilon_{ijl} \frac{\partial \theta}{\partial k_i} \frac{\partial \phi}{\partial k_j}. \quad (8.28)$$

Thus

$$\mathbf{V}(\mathbf{k}) = \frac{\sin\theta}{2} \nabla_{\mathbf{k}} \theta \times \nabla_{\mathbf{k}} \phi, \quad (8.29)$$

where $\nabla_{\mathbf{k}}$ is the gradient operator in $\mathbf{k}$ space. Working with polar coordinates,

$$\nabla_{\mathbf{k}} = \hat{\mathbf{k}} \frac{\partial}{\partial k} + \frac{1}{k} \hat{\boldsymbol{\theta}} \frac{\partial}{\partial \theta} + \frac{1}{k \sin \theta} \hat{\boldsymbol{\phi}} \frac{\partial}{\partial \phi}, \quad (8.30)$$

we obtain

$$\mathbf{V}(\mathbf{k}) = \frac{1}{2k^2} \hat{\boldsymbol{\theta}} \times \hat{\boldsymbol{\phi}} = \frac{\hat{\mathbf{k}}}{2k^2}. \quad (8.31)$$

Repeating the calculation of H_- (see Ex. 3) yields the same result but with a minus sign. Thus, the Berry curvature of Weyl points described by the Hamiltonians (8.3) are

$$\mathbf{V}^{(\pm)}(\mathbf{k}) = \pm \frac{\hat{\mathbf{k}}}{2k^2}. \quad (8.32)$$

The vector fields associated with these curvatures are illustrated in Fig. 8-14. This figure shows that band touching points act as sources and sinks of Berry's curvature.

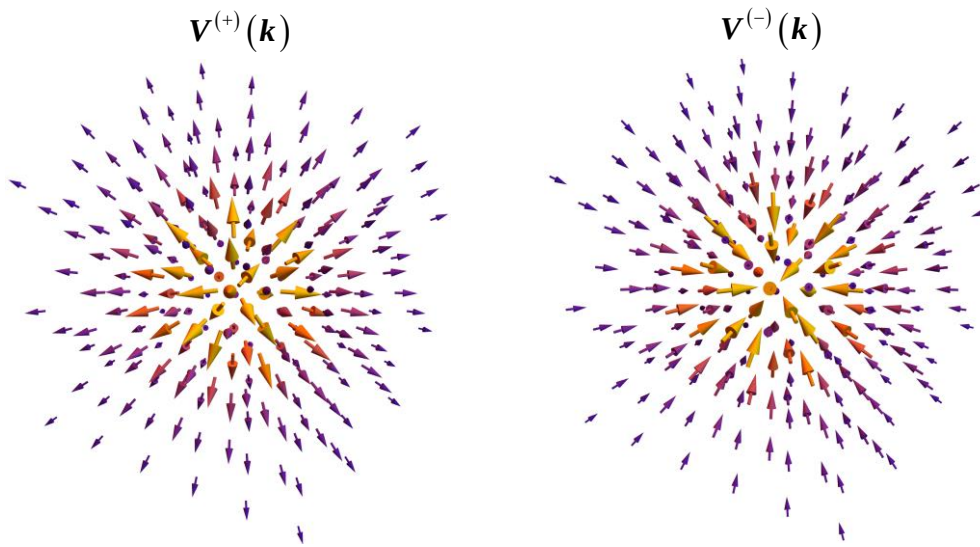


Figure 8-14 The vector fields of Berry curvature associated with Weyl points

8.5 Chern numbers and the chiral charge

Similar to the Gauss-Bonnet theorem (8.5) one can integrate $\mathbf{V}^{(\pm)}(\mathbf{k})$ over a closed surface surrounding the band touching points. Formula (8.32) has the form of an electric field generated by a charged particle, and from Gauss theorem, we obtain the analog of the Gauss-Bonnet theorem:

$$\mathcal{C} = \frac{1}{2\pi} \oint \mathbf{ds} \cdot \mathbf{V}^{(\pm)}(\mathbf{k}) = \pm 1 \quad (8.33)$$

The integer number, $\mathcal{C}$, obtained when integrating Berry's curvature over a closed surface in momentum space is called the Chern number. In the case of Weyl point, it is the chirality charge. This number is a topological number analogous to the genus of a two-dimensional closed surface. However, in general, the Chern number may be any integer, $\mathcal{C} = 0, \pm 1, \pm 2, \dots$.

An important constraint on the Chern numbers in a compact domain (such as the Brillouin zone, which is periodic) is that their total sum is zero. It is because the field lines of the

source of Berry curvature on a compact domain must end somewhere within the domain, and the only possibility is to have a sink. Thus, Weyl points must appear in pairs of opposite chirality charges. This property is known as the *Nielsen-Ninomiya theorem*. It implies that the only way of removing a Weyl point is by annihilating it with another point with opposite chirality.

The Nielsen-Ninomiya theorem dictates the minimal number of Weyl points in a system. As we saw in Sec. 8.2, one must break either time-reversal symmetry or inversion symmetry (or both) to obtain Weyl points. Time reversal symmetry reverses the momentum $\mathbf{k} \rightarrow -\mathbf{k}$ and the pseudospin $\tau \rightarrow -\tau$, and therefore does not change the chiral charge. Thus, a system with time-reversal symmetry must have at least four Weyl points to satisfy the Nielsen-Ninomiya theorem. On the other hand, inversion symmetry reverses the momentum, $\mathbf{k} \rightarrow -\mathbf{k}$, but does not change pseudospin $\tau \rightarrow \tau$. Hence the chiral charge is reversed by inversion. Thus, in a system with broken time-reversal symmetry, one can satisfy the Nielsen-Ninomiya theorem with only two Weyl points.

8.6 Dirac points

Dirac points are two Weyl points of opposite charge sitting at the same point in the Brillouin zone. The simplest Hamiltonian describing such a point is a 4×4 matrix of the form:

$$H = \begin{pmatrix} \hbar v_F \mathbf{k} \cdot \boldsymbol{\tau} & 0 \\ 0 & -\hbar v_F \mathbf{k} \cdot \boldsymbol{\tau} \end{pmatrix}. \quad (8.34)$$

However, adding generic perturbations, such as

$$\delta H = \begin{pmatrix} 0 & m \\ m & 0 \end{pmatrix}, \quad (8.35)$$

mixes the two Weyl points and opens a gap in the spectrum. An example of this scenario we saw when we added spin-orbit interaction to graphene. Thus, Dirac points are not genetic because they require fine-tuning of parameters to obtain a fourfold degeneracy of the energy levels. Nevertheless, breaking time-reversal or inversion symmetry may split a Dirac point into two Weyl points of opposite charges, which are topologically protected (A similar procedure was employed for HgTe, where degenerate points were separated by adding J_z perturbation that breaks time-reversal symmetry.)

8.7 Fermi arcs

Any physical system is finite, and one naturally asks what happens at the boundary of a system hosting the Weyl points spectrum. To keep the discussion simple, let us consider Weyl fermion with one chiral charge, described by the Hamiltonian:

$$H = -i\hbar v_F \boldsymbol{\tau} \cdot \nabla, \quad (8.36)$$

and assume that the system occupies half-space, $x > 0$. Namely, it has a boundary at $x = 0$. What are the corresponding boundary conditions for Weyl fermions? Consider a particle moving perpendicular to the $x = 0$ plane. Usually, one expects the particle to be reflected and reverse its direction, keeping the component of the angular momentum in the x direction intact. However, this will imply that the chirality of the Weyl fermion is reversed, as illustrated in Fig. 8-15.

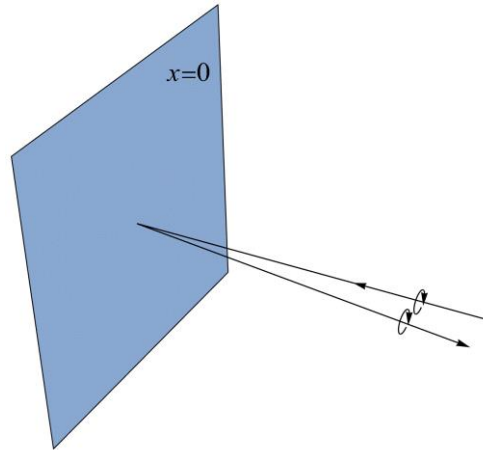


Figure 8-15 reflection that preserves angular momentum and reverses chirality

Thus, to preserve the chirality of the reflected Weyl fermion, the angular momentum cannot be preserved on the boundary. A boundary condition that can do the work is a condition that the wave function on the boundary is not an eigenfunction of τ_x . Instead, it is an eigenfunction of the spin in some perpendicular direction, such as:

$$\psi|_{x=0} = M \psi|_{x=0} \quad \text{with} \quad M = \tau_y \cos \varphi + \tau_z \sin \varphi \quad (8.37)$$

One can prove (see Ex. 5) that this boundary condition ensures that H is hermitian, i.e., $\langle \psi_1 | H \psi_2 \rangle = \langle H \psi_1 | \psi_2 \rangle$ for all wavefunctions that satisfy (8.37). In general, there will be an additional momentum-dependent contribution to the matrix M ; however, assuming the energy to be sufficiently close to the degeneracy point, the momentum is also small; therefore, this contribution can be neglected. The angle φ depends on the microscopic details of the crystal that we ignore. Here, without loss of generality, we rotate the coordinate system around the x -axis such that $\varphi = 0$ so that the boundary condition (8.37) reads:

$$\psi|_{x=0} = \tau_y \psi|_{x=0}. \quad (8.38)$$

Now let us look for a solution to the time-independent Schrödinger equation with zero energy,

$$-i\hbar v_F \left[\tau_x \frac{\partial}{\partial x} + \tau_y \frac{\partial}{\partial y} + \tau_z \frac{\partial}{\partial z} \right] \psi = 0, \quad (8.39)$$

assuming that $\psi = \tau_y \psi$ everywhere (namely, the particle's spin points in the y direction), and $\partial \psi / \partial y = 0$, i.e., the y component of the particle's momentum, vanishes. Using the following property of Pauli matrices, $\tau_z = -i\tau_x \tau_y$, the above equation reduces to

$$-i\hbar v_F \tau_x \left[\frac{\partial}{\partial x} - i \frac{\partial}{\partial z} \right] \psi = 0. \quad (8.40)$$

The solution to this equation is

$$\psi = \exp(-ik_z z + k_z x) \psi_0 \quad \text{with} \quad \psi_0 = \tau_y \psi_0. \quad (8.41)$$

More generally, the solution of the Schrödinger equation, $H\psi = \varepsilon\psi$, for non-zero (positive) energy, ε , is:

$$\psi = \exp(ik_y y) \exp(-ik_z z + k_z x) \psi_0, \quad \text{with} \quad \varepsilon = \hbar v_F k_y. \quad (8.42)$$

This solution is normalizable only when $k_z < 0$. Therefore, it describes edge states localized near the boundary of the sample. When $k_z \rightarrow 0$ the solution (8.42) reduces to a plane wave moving in the y direction- indistinguishable from the bulk solutions.

The energy ε is independent of k_z along the line $k_z < 0$. Hence, on the two-dimensional Brillouin zone associated with the surface states, the line $k_z < 0$, $k_y = \varepsilon / \hbar v_F$, should be part of the Fermi surface. This line is called "*Fermi arc*". The description that we built for this line holds only sufficiently close to zero energy where the Hamiltonian (8.36) provides a valid approximation to the system.

Finally, repeating the above calculation for a Weyl fermion with opposite chirality shows that normalizable solutions exist for $k_z > 0$.

Now consider a system with broken time-reversal symmetry having two Weyl points of opposite chirality located, say at $(0, 0, \pm k_0)$. From each one of these points, a Fermi arc emerges. However, since these arcs should end at points where surface states become bulk states, the two arcs should merge and form a single curve connecting the two Weyl nodes, as illustrated in Fig. 8-16. This behavior is very different from the traditional view of the Fermi surface, which should form closed loops in the Brillouin zone of a two-dimensional system. However, it is because the system is not indeed two-dimensional.

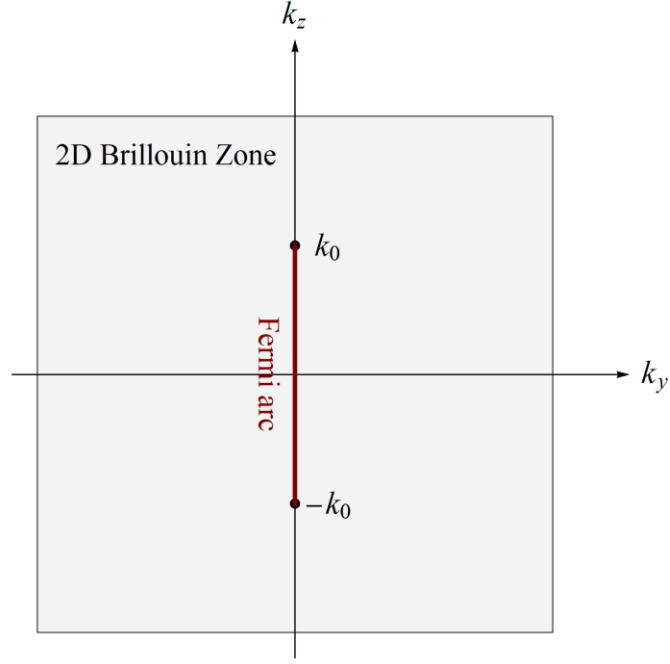


Figure 8-16 The Fermi arc in a system containing two Weyl points corresponding to Fermi energy $\varepsilon = 0$.

8.8 Exercises

1. Find the geodesic on a conical surface and use it to calculate δ .
2. Prove that Berry's phase is real.
3. Calculate the Berry curvature of $H_- = -\hbar v_F \mathbf{k} \cdot \boldsymbol{\tau}$ associated with the upper energy band.
4. Prove the following formula for Berry's curvature:

$$\mathbf{V}_n(\mathbf{x}) = \text{Im} \sum_{m \neq n} \frac{\left\langle \psi_x^{(n)} \left| \frac{\partial H_x}{\partial \mathbf{x}} \right| \psi_x^{(m)} \right\rangle \times \left\langle \psi_x^{(m)} \left| \frac{\partial H_x}{\partial \mathbf{x}} \right| \psi_x^{(n)} \right\rangle}{\left(\varepsilon_x^{(n)} - \varepsilon_x^{(m)} \right)^2}. \quad (8.43)$$

Advice: insert the Identity operator, $\sum_m \left| \psi_x^{(m)} \right\rangle \left\langle \psi_x^{(m)} \right|$, to rewrite Eq. (8.21) in the form

$$\mathbf{V}_n(\mathbf{x}) = \text{Im} \sum_{m \neq n} \left\langle \frac{\partial}{\partial \mathbf{x}} \psi_x^{(n)} \left| \psi_x^{(m)} \right\rangle \times \left\langle \psi_x^{(m)} \left| \frac{\partial}{\partial \mathbf{x}} \psi_x^{(n)} \right\rangle. \quad (8.44)$$

Now take the derivative of the Schrödinger equation (8.6) with respect to $\mathbf{x}$, multiply it from the left by $\langle \psi_x^{(m)} |$ and show that

$$\left\langle \psi_x^{(m)} \left| \frac{\partial}{\partial \mathbf{x}} \psi_x^{(n)} \right. \right\rangle = \frac{\left\langle \psi_x^{(m)} \left| \frac{\partial H_x}{\partial \mathbf{x}} \right| \psi_x^{(n)} \right\rangle}{\mathcal{E}_x^{(n)} - \mathcal{E}_x^{(m)}}. \quad (8.45)$$

Using the last two equations, prove Eq. (8.43).

5. Prove that the Weyl Hamiltonian is Hermitian in the Hilbert space spanned by wave functions that satisfy the boundary condition (8.37). Use the anti-commutation property of Pauli matrices, which implies that

$$\{M, \tau_x\} = M \tau_x + \tau_x M = 0. \quad (8.46)$$

9 Crystals in a constant electric field

In this chapter, we study the response of the electrons to a constant (time-independent) electric field in crystals. This issue has many aspects, and we shall present five of them: We begin with the physical response of dielectric materials to an electric field. Here we shall see how symmetries constrain the response of these materials to the external electric field. Next, we consider metallic systems in the presence of a time-independent electric field, focusing our attention on the dynamics of the electrons. To this end, we shall develop the effective mass approximation that holds when the energy bands are far apart. Then we will use this approximation to explain the phenomenon of Wannier-Stark oscillations. Next, we consider situations where the effective mass approximation fails when two energy bands come too close to each other. Finally, we discuss the Landau-Zener tunneling, which describes the dielectric breakdown in the aforementioned situation.

9.1 Physical response in dielectric crystals

Consider a dielectric (an insulator) crystal under the influence of an external electric field whose components are E_α . In such materials, the electric field polarizes the system by creating a finite polarization vector, $\mathbf{P}$. Here we shall assume that the external force and the response are smooth functions in space, i.e., the spatial changes in these functions are on length scales that are much larger than the lattice constant. Generally, one expects that for a sufficiently weak electric field, the response of the system is given by the asymptotic expansion:

$$P_\alpha = \varepsilon_0 \left[\chi_{\alpha\beta}^{(1)} E_\beta + \underbrace{\chi_{\alpha;\beta\gamma}^{(2)} E_\beta E_\gamma}_{\text{Pockless}} + \underbrace{\chi_{\alpha;\beta\gamma\delta}^{(3)} E_\beta E_\gamma E_\delta}_{\text{Kerr}} + \dots \right], \quad (9.1)$$

where repeated indices are assumed to be summed over, and ε_0 is the dielectric constant of the vacuum. The first term of this expansion represents the linear response of the system. Here $\chi_{\alpha\beta}^{(1)}$ is the electric susceptibility tensor. The second term is responsible for the *Pockles effect*. The latter is an electro-optical effect that changes or produces birefringence by changing the refractive indices proportional to the electric field. The third term is responsible for the *Kerr effect*. It also affects the crystal's optical properties, but the change in the refractive indices is quadratic in the electric field. More generally, the term $\chi^{(n)}$ is called the n -th order susceptibility tensor.

Notice that the asymptotic expansion (9.1) does not describe a situation where the response has an essential singularity, such as $\exp(-E_*/|E|)$, where E_* is constant. In the last section of this chapter, we shall deal with a case of this type associated with dielectric breakdown. Here we

neglect these terms and see how symmetry constrains the number of parameters in the expansion (9.1). The nominal number of parameters in $\chi^{(1)}$ is $3 \times 3 = 9$, but time-reversal symmetry reduces this number to 6; $\chi^{(2)}$ contains 18 parameters and $\chi^{(3)}$ has 30 parameters.

The guiding principle for applying symmetry considerations is that the terms on the left-hand side of Eq. (9.1) should belong to the same irreducible representation of the terms on the right-hand side of the equation. If this is not the case, then applying some of the symmetry operations to Eq. (9.1) would give different results on both sides of the equation. As an example, consider a system with a symmetry C_{2v} . The character table of this point group is displayed below.

C_{2v}	E	c_2	$\sigma_v(xz)$	$\sigma'_v(yz)$			
A_1	1	1	1	1	z	x^2, y^2, z^2	z^3, x^2z, y^2z
A_2	1	1	-1	-1	R_z	xy	xyz
B_1	1	-1	1	-1	x, R_y	xz	xz^2, x^3, xy^2
B_2	1	-1	-1	1	y, R_x	yz	yz^2, y^3, x^2y

The component of the polarization vector, P_α , and those of the electric field, E_β , are associated with the x , y , and z basis functions. For example, P_z belongs to the A_1 irreducible representation; hence all the terms on the right-hand side of Eq. (9.1) (associated with P_z) must also belong to A_1 irreducible representation. Thus, using the basis functions of the identity representation, we obtain:

$$A_1: \quad P_z = \chi_z^{(1)} E_z + \chi_{z;zz}^{(2)} E_z^2 + \chi_{z;xx}^{(2)} E_x^2 + \chi_{z;yy}^{(2)} E_y^2 + \chi_{z;zzz}^{(3)} E_z^3 + \chi_{z;zxz}^{(3)} E_z E_x^2 + \chi_{z;zyy}^{(3)} E_z E_y^2. \quad (9.2)$$

Similarly, the expansions of the two other components of the polarization vector are:

$$B_1: \quad P_x = \chi_x^{(1)} E_x + \chi_{x;xz}^{(2)} E_x E_z + \chi_{x;zzz}^{(3)} E_x E_z^2 + \chi_{x;xxx}^{(3)} E_x^3 + \chi_{x;xyy}^{(3)} E_x E_y^2, \quad (9.3)$$

$$B_2: \quad P_y = \chi_y^{(1)} E_y + \chi_{y;yz}^{(2)} E_y E_z + \chi_{y;yzz}^{(3)} E_y E_z^2 + \chi_{y;yyy}^{(3)} E_y^3 + \chi_{y;yxx}^{(3)} E_y E_x^2. \quad (9.4)$$

Why P_x does not contain the term $\chi_{x;xx}^{(2)} E_x^2$ in its expansion? Because if we apply c_2 rotation on the x component of vectors, $P_x \rightarrow -P_x$ and $E_x \rightarrow -E_x$, but $\chi_{x;xx}^{(2)} E_x^2$ does not change sign. For the same reason P_x does not contain the term $\chi_{x;xy}^{(2)} E_x E_y$ in its expansion. On the other hand, the

z component of a vector is not affected by the c_2 rotation, therefore, P_x contains the term $\chi_{x;xz}^{(2)} E_x E_z$ in its expansion.

For another example, consider a crystal having the symmetry of a regular tetrahedron. The point group in this case is T_d , and its character table is listed on page 144. Looking at the basis function of this character table, we see that the polarization vector belongs to the F_2 irreducible representation, hence

$$\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} = \chi^{(1)} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} + \chi^{(2)} \begin{pmatrix} E_y E_z \\ E_z E_x \\ E_x E_y \end{pmatrix} + \chi_1^{(3)} \begin{pmatrix} E_x^3 \\ E_y^3 \\ E_z^3 \end{pmatrix} + \chi_2^{(3)} \mathbf{E}^2 \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \quad (9.5)$$

Here, the high symmetry of the system severely constrains the number of parameters that determine the response.

9.2 The effective mass approximation

In this section, we consider metallic crystals. Our aim is to characterize the response of the electrons in such systems to the application of an external electric field. Namely, we are interested in the solution of the Schrödinger equation,

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) - e\phi(\mathbf{r}) \right] \psi = \varepsilon \psi. \quad (9.6)$$

The first two terms on the left-hand side of the equation describe an electron moving in a periodic potential, while the third term represents its interaction with the external electric field, $\mathbf{E}$, represented by the scalar potential $\phi(\mathbf{r}) = -\mathbf{E} \cdot \mathbf{r}$. Here $-e$ is the electron charge and m is its mass. The primary simplifying assumption that allows solving this problem is that $\phi(\mathbf{r})$ changes very slowly in space - over distances much larger than the lattice constant. This condition, however, does not preclude situations where the scalar potential is large; therefore, its effect might be non-perturbative.

As a reference point for our future discussion, we first consider the problem of a free-electron under the influence of a constant electric field,

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + e\mathbf{E} \cdot \mathbf{r} \right] \psi = \varepsilon \psi. \quad (9.7)$$

To obtain the solution of this equation, it is convenient to express the wave function as a superposition of plane waves,

$$\psi(\mathbf{r}) = \int \frac{d^d k}{(2\pi)^d} c(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (9.8)$$

where $c(\mathbf{k})$ are the Fourier expansion coefficients. Substituting this formula in the kinetic energy term gives:

$$-\frac{\hbar^2}{2m} \nabla^2 \int \frac{d^d k}{(2\pi)^d} c(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) = \int \frac{d^d k}{(2\pi)^d} \frac{\hbar^2 k^2}{2m} c(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (9.9)$$

while the potential energy term may be manipulated using integration by parts:

$$\begin{aligned} e\mathbf{E} \cdot \mathbf{r} \int \frac{d^d k}{(2\pi)^d} c(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) &= e \int \frac{d^d k}{(2\pi)^d} c(\mathbf{k}) \mathbf{E} \cdot \frac{\partial}{i\partial \mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}) \\ &= ie \int \frac{d^d k}{(2\pi)^d} \mathbf{E} \cdot \left(\frac{\partial}{\partial \mathbf{k}} c(\mathbf{k}) \right) \exp(i\mathbf{k} \cdot \mathbf{r}). \end{aligned} \quad (9.10)$$

Thus, the Schrödinger equation in $\mathbf{k}$ space is a first-order differential equation:

$$\left[\frac{\hbar^2 \mathbf{k}^2}{2m} + ie\mathbf{E} \cdot \frac{\partial}{\partial \mathbf{k}} \right] c(\mathbf{k}) = \varepsilon c(\mathbf{k}). \quad (9.11)$$

The separation of variables in this equation yields a plane-wave solution in perpendicular directions to the electric field. In what follows, we shall ignore this trivial part of the solution and focus our attention on the wavenumber component parallel to the electric field. Thus, the interesting part of Eq. (9.11) is along the electric field direction, where the equation becomes

$$\left[\frac{\hbar^2 k^2}{2m} + ieE \frac{\partial}{\partial k} \right] c(k) = \varepsilon c(k). \quad (9.12)$$

The solution of this equation is:

$$c(k) = c_0 \exp \left[\frac{i}{eE} \int dk \left(\frac{\hbar^2 k^2}{2m} - \varepsilon \right) \right] = c_0 \exp \left[\frac{i}{eE} \left(\frac{\hbar^2 k^3}{6m} - \varepsilon k \right) \right], \quad (9.13)$$

where c_0 is the normalization constant. Returning to real space using the inverse Fourier transform gives

$$\begin{aligned} \psi(x) &= \int_{-\infty}^{\infty} \frac{dk_x}{2\pi} c_0 \exp \left[\frac{i}{3} \left(\frac{\hbar^2}{2meE} \right) k^3 + ik \left(x - \frac{\varepsilon}{eE} \right) \right] \\ &= c_0 \left(\frac{2meE}{\hbar^2} \right)^{1/3} \text{Ai} \left[\left(\frac{2meE}{\hbar^2} \right)^{1/3} \left(x - \frac{\varepsilon}{eE} \right) \right], \end{aligned} \quad (9.14)$$

where $\text{Ai}(z)$ is the Airy function, and we choose the electric field to point in the direction of the x axis. The behavior of this wave function is depicted in Fig. 9-1. In the above solution, the electron energy may assume any value. However, different energy values only shift the particle's turning point by ε/eE along the x axis. Beyond this turning point, the wave function decays exponentially.

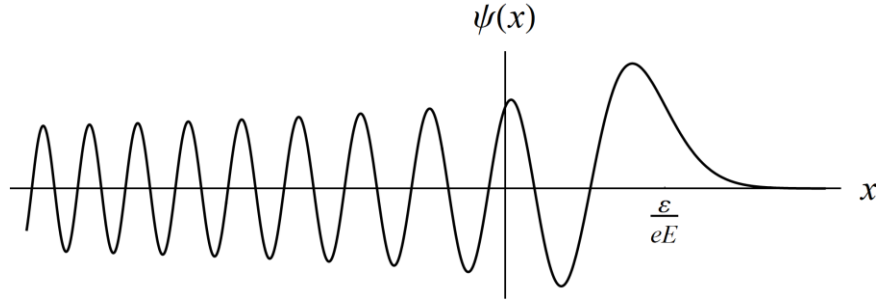


Figure 9-1 The Airy function describing an electron in a constant electric field

We turn now to discuss the case where the electron moves in a lattice. Here it is natural to expand the wave function of the electron in the basis of Bloch's wave functions:

$$\psi(\mathbf{r}) = \sum_j \int_{\text{BZ}} \frac{d^d k}{(2\pi)^d} c_j(\mathbf{k}) \phi_k^{(j)}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (9.15)$$

where $\phi_k^{(j)}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})$ is the Bloch function of the j -th band, $c_j(\mathbf{k})$ are expansion coefficients, and the integral is over the first Brillouin zone. In this basis, the part of the Schrödinger equation which does not include the electric field yields

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) \right] \psi(\mathbf{r}) = \sum_j \int_{\text{BZ}} \frac{d^d k}{(2\pi)^d} \varepsilon_j(\mathbf{k}) c_j^{(j)}(\mathbf{k}) \phi_k^{(j)}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (9.16)$$

where $\varepsilon_j(\mathbf{k})$ is the eigenenergy of the particle in the absence of the external electric field. Consider now the contribution from the electric field. Here integration by parts gives:

$$\begin{aligned} (e\mathbf{E} \cdot \mathbf{r}) \psi(\mathbf{r}) &= \sum_j \int_{\text{BZ}} \frac{d^d k}{(2\pi)^d} c_j(\mathbf{k}) \phi_k^{(j)}(\mathbf{r}) \left(e\mathbf{E} \cdot \frac{\partial}{i\partial \mathbf{k}} \right) \exp(i\mathbf{k} \cdot \mathbf{r}) \\ &= \sum_j \int_{\text{BZ}} \frac{d^d k}{(2\pi)^d} \left[\left(ie\mathbf{E} \cdot \frac{\partial}{\partial \mathbf{k}} \right) c_j(\mathbf{k}) \phi_k^{(j)}(\mathbf{r}) \right] \exp(i\mathbf{k} \cdot \mathbf{r}) \\ &= \sum_j \int_{\text{BZ}} \frac{d^d k}{(2\pi)^d} \left[ie\mathbf{E} \cdot \frac{\partial c_j(\mathbf{k})}{\partial \mathbf{k}} \phi_k^{(j)}(\mathbf{r}) + ie\mathbf{E} \cdot \frac{\partial \phi_k^{(j)}(\mathbf{r})}{\partial \mathbf{k}} c_j(\mathbf{k}) \right] \exp(i\mathbf{k} \cdot \mathbf{r}) \end{aligned} \quad (9.17)$$

The first term in this integral is analogous to the free particle problem, see Eq. (9.10), but there is an additional term proportional to $i\partial_{\mathbf{k}}\phi_{\mathbf{k}}^{(j)}(\mathbf{r})$. The latter is a periodic function of $\mathbf{r}$ with periodicity of the lattice because $\phi_{\mathbf{k}}^{(j)}(\mathbf{r})$ is the periodic component of Bloch's wave function. For any given value $\mathbf{k}$, the set of functions $\{\phi_{\mathbf{k}}^{(j')}(\mathbf{r})\}$ with $j' = 0, 1, \dots$ forms a complete basis for any function defined in the unit cell with periodic boundary conditions. In particular, one may expand the function $i\partial_{\mathbf{k}}\phi_{\mathbf{k}}^{(j)}(\mathbf{r})$ in this basis:

$$i\frac{\partial\phi_{\mathbf{k}}^{(j)}(\mathbf{r})}{\partial\mathbf{k}} = \sum_{j'} i \underbrace{\int_{\text{uc}} d^d r' \frac{\partial\phi_{\mathbf{k}}^{(j)}(\mathbf{r}')}{\partial\mathbf{k}} \phi_{\mathbf{k}}^{(j')*}(\mathbf{r}') \phi_{\mathbf{k}}^{(j')}(\mathbf{r})}_{\Omega_{jj'}(\mathbf{k})} = \sum_{j'} \Omega_{jj'}(\mathbf{k}) \phi_{\mathbf{k}}^{(j')}(\mathbf{r}), \quad (9.18)$$

where the integral is over the volume of a unit cell, and we choose the periodic components Bloch's wave functions to be normalized as follows:

$$\int_{\text{uc}} d^d r' \phi_{\mathbf{k}}^{(j')*}(\mathbf{r}') \phi_{\mathbf{k}}^{(j)}(\mathbf{r}') = \delta_{jj'}. \quad (9.19)$$

Thus, the difference between the equation of a free electron and that of an electron in a lattice is the vector matrix element:

$$\Omega_{jj'}(\mathbf{k}) = i \left\langle \phi_{\mathbf{k}}^{(j')} \left| \frac{\partial}{\partial\mathbf{k}} \right| \phi_{\mathbf{k}}^{(j)} \right\rangle. \quad (9.20)$$

The diagonal elements of this vector are the *Berry connections*. Substituting (9.16)-(9.20) in the Schrödinger equation (9.6), we obtain that, in the basis of Bloch's wave functions, it reduces to

$$\left\{ \left[\varepsilon_j(\mathbf{k}) + ie\mathbf{E} \cdot \frac{\partial}{\partial\mathbf{k}} \right] \delta_{jj'} + e\mathbf{E} \cdot \Omega_{jj'}(\mathbf{k}) \right\} c_{j'}(\mathbf{k}) = \varepsilon c_j(\mathbf{k}) \quad (9.21)$$

The first two terms on the left-hand side of the equation are diagonal in the band index, but the third is an infinite matrix. Nevertheless, in many cases, the contribution of this term is negligible. We can estimate it to be proportional to the lattice constant,

$$\left| \sum_{j'} \Omega_{jj'}(\mathbf{k}) \phi_{\mathbf{k}}^{(j')} \right| = \left| \frac{\partial\phi_{\mathbf{k}}^{(j)}}{\partial\mathbf{k}} \right| \sim \frac{|\phi_{\mathbf{k}}^{(j)}|}{b} \sim |\phi_{\mathbf{k}}^{(j)}| a, \quad (9.22)$$

while the term $i\partial_{\mathbf{k}}$ (which is, essentially, representation of the operator $\mathbf{r}$ in momentum space) can be of the order of the size of the sample. The *effective mass approximation* is the approximation where $e\mathbf{E} \cdot \Omega_{jj'}(\mathbf{k})$ is neglected altogether. Within this approximation, Eq. (9.21) reduces to

$$\left[\varepsilon_j(\mathbf{k}) + ie\mathbf{E} \cdot \frac{\partial}{\partial \mathbf{k}} \right] c_j(\mathbf{k}) = \varepsilon c_j(\mathbf{k}). \quad (9.23)$$

The meaning of this approximation is that we ignore possible coupling between different energy bands. Thus, we expect it to break when two energy bands intersect or become close to each other. We shall return to this issue later in section 9.4. Meanwhile, let us explore the implications of the above formula for electron dynamics.

9.3 Stark ladder and Wannier-Stark oscillations

Eq. (9.23) has the same structure as that of a free electron (9.12). The critical difference between the two cases is that, now, one has to impose periodic boundary conditions on the expansion coefficients, $c_j(\mathbf{k})$, at the edges of the Brillouin zone. To keep the discussion simple, in what follows, we assume that the structure of the Brillouin zone and the direction of the electric field allow separation of variables, such that the periodic boundary condition in the direction of the electric field is

$$c_j\left(-\frac{b}{2}\right) = c_j\left(\frac{b}{2}\right), \quad (9.24)$$

where b is the reciprocal lattice constant.

Integration of Eq. (9.23) gives

$$c_j(k) = c_0 \exp \left\{ \frac{i}{eE} \int_{-\frac{b}{2}}^k dk' [\varepsilon_j(k') - \varepsilon] \right\}, \quad (9.25)$$

and imposing the boundary condition (9.24) leads to

$$\exp \left\{ \frac{i}{eE} \int_{-\frac{b}{2}}^{\frac{b}{2}} dk' [\varepsilon_j(k') - \varepsilon] \right\} = \exp \left\{ \frac{i}{eE} \left[\int_{BZ} dk' \varepsilon_j(k') - \varepsilon b \right] \right\} = 1. \quad (9.26)$$

This condition is satisfied only for discrete values of the energy satisfying the equation:

$$\frac{1}{eE} \left[\varepsilon_n \frac{2\pi}{a} - \int_{BZ} dk' \varepsilon_j(k') \right] = 2\pi n \quad (9.27)$$

with integer n , i.e., for

$$\varepsilon_n = neEa + a \int_{\text{BZ}} \frac{dk'}{2\pi} \varepsilon_j(k'). \quad (9.28)$$

The second term on the right-hand side of the equation is a constant. Thus, we obtained a set of equidistant energy levels similar to those of a harmonic oscillator. This series of energy levels is called the *Stark ladder*. It implies that if we have prepared the electron in some typical wave packet, its motion should be periodic in time with frequency $\omega = eEa/\hbar$.

A physical picture of the periodic behavior of the electron follows from the semiclassical analysis of the Hamiltonian

$$H = \varepsilon(k) + eEx, \quad (9.29)$$

where $\varepsilon(k)$ is a periodic function of k . Hamilton's equations for this system,

$$\begin{aligned} \dot{x} &= v(k) = \frac{\partial \varepsilon(k)}{\hbar \partial k}, \\ \hbar \dot{k} &= -eE, \end{aligned} \quad (9.30)$$

yield a linear time dependence of the wavenumber, $k = -eEt/\hbar$. On the other hand, the velocity, being a derivative of $\varepsilon(k)$, is periodic in k ,

$$v(k+b) = v\left(k + \frac{2\pi}{a}\right) = v(k). \quad (9.31)$$

Hence $v(-eEt/\hbar)$ is a periodic function of time with a period τ given by

$$\frac{eE\tau}{\hbar} = \frac{2\pi}{a} \Rightarrow \tau = \frac{2\pi\hbar}{eEa}. \quad (9.32)$$

These oscillations of the electron in the lattice are called *Wannier-Stark oscillations*. Their intuitive explanation is as follows: Imagine an electron initially at the bottom of the band. In this region, the effective mass is positive, and the electric field accelerates the electron. This acceleration raises the electron's energy until it becomes close to the top of the band, where the effective mass becomes negative. Consequently, the electric field decelerates the electron, and its energy reduces until the effective mass becomes positive again and the process repeats itself. The excess momentum gained by the electron in each period of this motion is transferred to the lattice via the Bragg reflection mechanism (when the electron reaches the edge of the Brillouin zone, its wavelength is twice the lattice constant).

9.4 Beyond the effective mass approximation – the $k \cdot p$ approximation

In section 9.2, we mentioned that the effective mass approximation breaks down when energy bands become close to each other. This section aims to show how one can improve the description in these situations. Before that, let us illustrate the breakdown of the effective mass approximation with a simple example.

Consider a one-dimensional system with two energy bands, as illustrated in Fig. 9-2. We assume that the gap between the bands, 2Δ , is much smaller than the band's width, $\varepsilon_{\text{band}}$,

$$\frac{\Delta}{\varepsilon_{\text{band}}} \ll 1, \quad (9.33)$$

and that all other bands are sufficiently far so that their effect can be neglected.

Near the point where the gap is minimal, the Hamiltonian of the system can be approximated, locally, by a 2×2 matrix:

$$H = \begin{pmatrix} \varepsilon_b + \hbar v \delta k & \Delta \\ \Delta & \varepsilon_b - \hbar v \delta k \end{pmatrix}, \quad (9.34)$$

where δk is measured relative to the wavenumber at which the gap is minimal, and ε_b is the middle energy between the bands at this point. Diagonalization of this Hamiltonian gives the energy levels $\varepsilon_{\pm} = \varepsilon_b \pm \sqrt{(\hbar v \delta k)^2 + \Delta^2}$ and the wave functions:

$$\phi_k^{(+)} = N_+ \begin{pmatrix} \hbar v \delta k + \sqrt{(\hbar v \delta k)^2 + \Delta^2} \\ \Delta \end{pmatrix} \text{ and } \phi_k^{(-)} = N_- \begin{pmatrix} \hbar v \delta k - \sqrt{(\hbar v \delta k)^2 + \Delta^2} \\ \Delta \end{pmatrix}, \quad (9.35)$$

where $N_{\pm}$ are the normalization constants. Substituting these wave functions in Eq. (9.20) gives the vector-matrix elements $\Omega_{jj'}(\mathbf{k})$. The non-diagonal matrix element (responsible for the transition between the energy levels) is

$$\Omega_{+-}(\delta k) = i \left\langle \phi_k^{(-)} \left| \frac{\partial}{\partial k} \right| \phi_k^{(+)} \right\rangle = -i \frac{\hbar v \sqrt{\Delta}}{\sqrt{2} \left[(\hbar v \delta k)^2 + \Delta^2 \right]^{3/4}}, \quad (9.36)$$

and at the point where the gap reaches a minimum, i.e. $\delta k = 0$, it reduces to

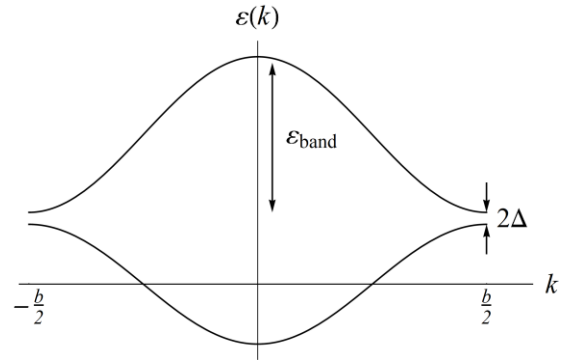


Figure 9-2 An illustration of a band structure where the effective mass approximation breaks down

$$\Omega_{+-}(0) = -i \frac{\hbar v}{\sqrt{2}\Delta}. \quad (9.37)$$

Now, the band's width can be estimated to be $\varepsilon_{\text{band}} \sim \hbar v b$, where $b = 2\pi/a$ is the reciprocal lattice constant, therefore

$$|\Omega_{+-}(0)| \sim \frac{\varepsilon_{\text{band}}}{\Delta} a \gg a. \quad (9.38)$$

This result shows that, at points where the gap between energy bands is small, the contribution from $\Omega_{jj'}(\mathbf{k})$ cannot be ignored. On the other hand, this problem usually appears only near a few isolated points in the Brillouin zone. Thus, one may construct an approximate description of the system focused only on these “dangerous” points. This is the $k \cdot p$ approximation that we turn to present now.

Let $\mathbf{k}_0$ denote the wave number vector at the point where the gap between bands is minimal, and let us expand the wave function as

$$\psi(\mathbf{r}) = \sum_j c_j(\mathbf{r}) \psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}), \quad (9.39)$$

where $\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) = \phi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) \exp(i\mathbf{k}_0 \cdot \mathbf{r})$ is the Bloch wave function associated with the j -th band and the wavenumber $\mathbf{k}_0$. The expansion coefficients $c_j(\mathbf{r})$ are functions that change slowly in space. Notice that here, unlike the $k \cdot p$ approximation discussed in chapter 7, the wavenumber vector is fixed. The $\mathbf{r}$ dependence of the expansion coefficients, $c_j(\mathbf{r})$, gives rise to deviation from the spatial behavior of the wave function $\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r})$.

To derive the equation for $c_j(\mathbf{r})$ we substitute (9.39) in the Schrödinger equation (9.7). Consider, first, the terms which do not include the electric field potential.

$$\begin{aligned} \left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) \right] \psi(\mathbf{r}) &= \sum_j c_j(\mathbf{r}) \underbrace{\left[-\frac{\hbar^2}{2m} \nabla^2 + u(\mathbf{r}) \right] \psi_{\mathbf{k}_0}^{(j)}(\mathbf{r})}_{\varepsilon_j(\mathbf{k}_0) \psi_{\mathbf{k}_0}^{(j)}(\mathbf{r})} \\ &+ \sum_j \psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) \left[-\frac{\hbar^2}{2m} \nabla^2 \right] c_j(\mathbf{r}) + \frac{\hbar}{m} (-i\nabla) \psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) \cdot (-i\hbar \nabla) c_j(\mathbf{r}) \end{aligned} \quad (9.40)$$

To calculate the last term on the right-hand side of this equation, we rewrite its first factor in the form

$$\begin{aligned}
-i\frac{\hbar}{m}\nabla\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) &= -i\frac{\hbar}{m}\nabla\left[\phi_{\mathbf{k}_0}^{(j)}(\mathbf{r})\exp(i\mathbf{k}_0\cdot\mathbf{r})\right] \\
&= \frac{\hbar\mathbf{k}_0}{m}\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) - i\frac{\hbar}{m}\exp(i\mathbf{k}_0\cdot\mathbf{r})\nabla\phi_{\mathbf{k}_0}^{(j)}(\mathbf{r}).
\end{aligned} \tag{9.41}$$

Now, as in section 9.2, we expand the periodic function $\nabla\phi_{\mathbf{k}_0}^{(j)}(\mathbf{r})$ in the basis of the periodic components of Bloch's wave functions, $\phi_{\mathbf{k}_0}^{(j')}(\mathbf{r})$. Using (9.19), we obtain

$$-i\frac{\hbar}{m}\nabla\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) = \frac{\hbar\mathbf{k}_0}{m}\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}) + \exp(i\mathbf{k}_0\cdot\mathbf{r})\sum_{j'}\int_{uc}d^d\mathbf{r}'\phi_{\mathbf{k}_0}^{(j')*}(\mathbf{r}')\left(-i\frac{\hbar}{m}\right)\nabla'\phi_{\mathbf{k}_0}^{(j)}(\mathbf{r}')\phi_{\mathbf{k}_0}^{(j')}(\mathbf{r}). \tag{9.42}$$

Defining the matrix

$$\mathbf{V}_{jj'} = \mathbf{v}_0\delta_{jj'} + \int_{uc}d^d\mathbf{r}'\phi_{\mathbf{k}_0}^{(j')*}(\mathbf{r}')\left(-i\frac{\hbar}{m}\right)\nabla'\phi_{\mathbf{k}_0}^{(j)}(\mathbf{r}'), \text{ with } \mathbf{v}_0 = \frac{\hbar\mathbf{k}_0}{m}, \tag{9.43}$$

the Schrödinger equation (9.10) becomes

$$\left[\varepsilon_j(\mathbf{k}_0) - \frac{\hbar^2\nabla^2}{2m} - e\varphi(\mathbf{r})\right]c_j(\mathbf{r}) + \sum_{j'}\mathbf{V}_{jj'}\cdot(-i\hbar\nabla)c_{j'}(\mathbf{r}) = \varepsilon c_j(\mathbf{r}). \tag{9.44}$$

The advantage of this equation is that it does not contain the dependence on the matrix $\Omega_{jj'}(\mathbf{k})$, which diverges when the gap between the energy bands closes. Its disadvantage is that it contains a different infinite matrix, $\mathbf{V}_{jj'}$, which complicates the solution. This matrix, however, can be handled in a similar way as in the $k\cdot p$ approximation introduced in chapter 7. Namely, it can be truncated to account only for the relevant energy levels. The more distant energy levels can be taken into account by perturbation theory (if necessary).

Example: Graphene in an electric field

We consider graphene in a constant electric field and focus our attention on the region of the K -point, i.e., we choose $\mathbf{k}_0 = \mathbf{k}_K$. In this case, one has to reinterpret the index j (in Eq. (9.44)) as also associated with the sublattice index of the pseudospin wave function. Adopting this view, we notice that the local description of the graphene near the K -point (5.16) has, already, the structure of Eq. (9.44) when $\varphi(\mathbf{r}) = 0$. Therefore, in the $k\cdot p$ approximation, the local equation of graphene, in the presence of an electric field, is

$$\left[\hbar v(-i\boldsymbol{\tau}^{AB}\cdot\nabla) - e\varphi(\mathbf{r})\right]\begin{pmatrix} c_A \\ c_B \end{pmatrix} = \varepsilon\begin{pmatrix} c_A \\ c_B \end{pmatrix}. \tag{9.45}$$

Setting E to point along the negative x direction, and separating variables, yields the following equation for the x component:

$$\begin{pmatrix} -eEx & -i\hbar v \frac{\partial}{\partial x} \\ -i\hbar v \frac{\partial}{\partial x} & -eEx \end{pmatrix} \begin{pmatrix} c_A \\ c_B \end{pmatrix} = \varepsilon \begin{pmatrix} c_A \\ c_B \end{pmatrix}. \quad (9.46)$$

The ε dependence of this equation can be eliminated by shifting the position coordinate: $x \rightarrow x - \varepsilon/(eE)$. Next, we define new combinations of the variables, c_A and c_B by the rotation

$$\begin{pmatrix} b_- \\ b_+ \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_A \\ c_B \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} c_A \\ c_B \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} b_- \\ b_+ \end{pmatrix} \quad (9.47)$$

These diagonalize the equation because:

$$\frac{1}{2} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -eEx & -i\hbar v \frac{\partial}{\partial x} \\ -i\hbar v \frac{\partial}{\partial x} & -eEx \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} b_- \\ b_+ \end{pmatrix} = \begin{pmatrix} i\hbar v \frac{\partial}{\partial x} - eEx & 0 \\ 0 & -i\hbar v \frac{\partial}{\partial x} - eEx \end{pmatrix} \begin{pmatrix} b_- \\ b_+ \end{pmatrix} = 0 \quad (9.48)$$

Thus, we obtained a pair of independent equations, $(\pm v\hat{p} - eE\hat{x})b_{\pm} = 0$ (where we have used operator notation). These equations describe a right moving electron (+ sign) and a left moving electron (− sign) without scattering at a constant velocity. This phenomenon is sometimes called *Klein tunneling*. The reason for this terminology will become apparent in the next section.

For future reference, we write down the wave functions obtained from the solution of Eq. (9.48)

$$b_{\pm} = b_{\pm}(0) \exp\left(\pm i \frac{eEx^2}{2\hbar v}\right), \quad (9.49)$$

where $b_{\pm}(0)$ are arbitrary constants that satisfy the normalization condition $|b_+(0)|^2 + |b_-(0)|^2 = 1$.

9.5 Dielectric breakdown (Landau-Zener tunneling)

Consider a one-dimensional system with almost degenerate energy bands in the presence of a constant electric field. In the framework of the $k \cdot p$ approximation, we focus our attention only on two energy levels in the vicinity of the point where the gap between the levels is minimal. In the absence of an electric field, Eq. (9.44) reduces to

$$\begin{pmatrix} -\Delta & -i\hbar v \frac{\partial}{\partial x} \\ -i\hbar v \frac{\partial}{\partial x} & \Delta \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \varepsilon \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad (9.50)$$

where the quadratic $\hbar^2 \nabla^2 / (2m)$ term is neglected, $V_{12} = V_{21} = v$ is assumed to be real, and we choose $\varepsilon_1(k_0) = -\Delta$ and $\varepsilon_2(k_0) = +\Delta$ (the middle point between the energy levels is set to be zero without loss of generality). Diagonalization of the above Hamiltonian gives the energy levels of the system,

$$\varepsilon_{\pm}(k) = \pm \sqrt{\Delta^2 + \hbar^2 v^2 k^2}, \quad (9.51)$$

where k is measured relative to k_0 . These energy levels are presented in Fig. 9-3.

Adding to Eq. (9.50) the potential energy resulting from an electric field, E , pointing in the negative x direction, we obtain the equation that we seek to solve in this section:

$$\begin{pmatrix} -\Delta - eEx & -i\hbar v \frac{\partial}{\partial x} \\ -i\hbar v \frac{\partial}{\partial x} & \Delta - eEx \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \varepsilon \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \quad (9.52)$$

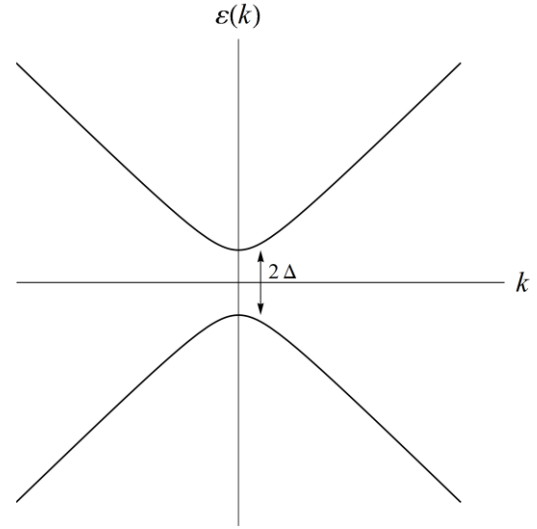


Figure 9-3 The energy levels obtained from the solution of Eq. (9.50)

Before deriving the solution of this equation, consider the dynamics it dictates in the framework of the effective mass approximation where the transition between different energy bands is forbidden. Suppose we have prepared the electron in a wave packet localized near some negative value of the wavenumber $k = -k_i$ in the lower energy band. The electric field exerts a force on the electron, and from Hamilton's equations, we see that the wave number increases linearly in time, $k = -k_i + eEt / \hbar$. The particle's velocity is positive until it reaches the point $k = 0$, because $\partial_k \varepsilon_-(k) > 0$, see Fig. 9-3. Beyond that point, assuming the particle remains in the lower energy band, the velocity reverses the sign, and its absolute value increases. In real space, it implies that the particle is reflected. This dynamical behavior is the familiar Wannier-Stark oscillations discussed in section 9.3.

Consider now the limit where the effective mass approximation is invalid because the gap between the energy levels is too small. Now there is another possibility: that the particle tunnels to the higher energy band. In this case, the particle velocity remains positive; namely, it continues with its motion without reflection (similar to the solution we found for graphene in an electric field). This phenomenon is known as *Landau-Zener breakdown* (or tunneling). It describes situations where the applied electric field is sufficiently strong to induce a current in an insulator by transferring charge carriers from the valance band to the conduction band. The physics of the Landau-Zener breakdown is analogous to tunneling through a potential barrier, as illustrated in Fig. 9-4. It is characterized by two parameters: the reflection and the transmission coefficients, r and t , respectively.

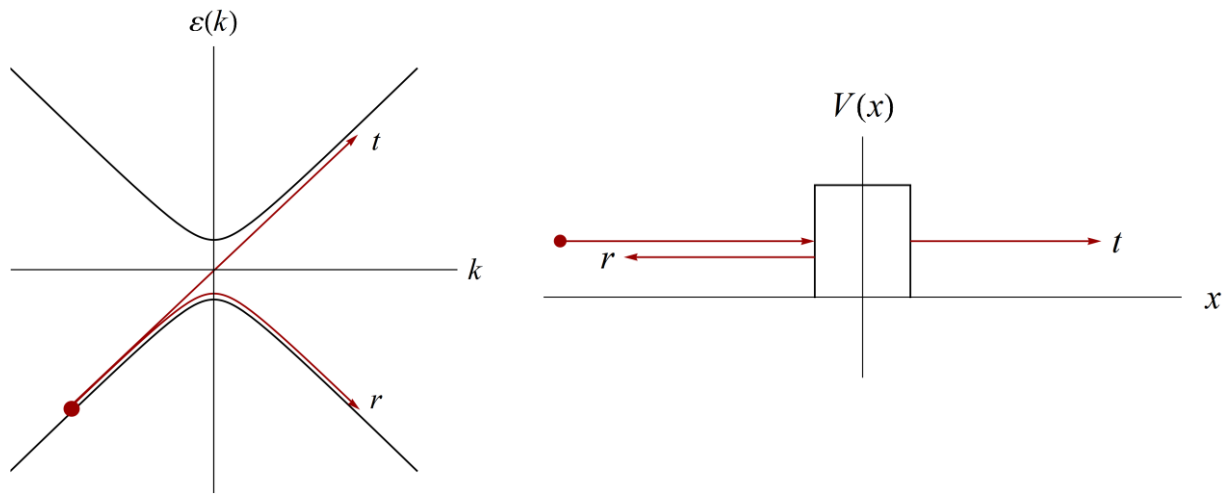


Figure 9-4 The analogy between Landau-Zener breakdown and tunneling under the barrier

Viewing the Landau-Zener breakdown phenomenon as tunneling under the barrier allows us to obtain the qualitative solution to the problem. Recall that the transmission coefficient for tunneling under the barrier, within the WKB approximation, is $t \sim \exp(-pw/\hbar)$, where p the (imaginary) momentum under the barrier, and w is the width of the barrier. In the analogous dielectric breakdown problem, the imaginary momentum is of order Δ/v , and the only length scale in the problem is $w = \Delta/|eE|$, hence

$$t \sim \exp\left(-\eta \frac{\Delta^2}{\hbar v |eE|}\right), \quad (9.53)$$

where η is a constant of order one that cannot be determined from these qualitative considerations. Nevertheless, this result highlights the singular dependence on the electric field in the limit $E \rightarrow 0$. In particular, a slight change in E results in a dramatic change in the

transmission coefficient. Notice that in the limit $\Delta \rightarrow 0$, i.e., when the gap between the bands closes, $t \rightarrow 1$, as we have obtained in the graphene example (the Klein tunneling).

The rest of this section is devoted to the proof of Eq. (9.53) and to the calculation of the numerical constant η (which gives $\eta = \pi/2$). We begin by transforming Eq. (9.52) into a more convenient form by repeating the steps presented in the graphene example on pages 178-179. Namely, first, we shift the position coordinate, $x \rightarrow x - \varepsilon/(eE)$, to eliminate the ε dependence, and then use rotated variables as defined by Eq. (9.47). The resulting equation is

$$\begin{pmatrix} i\hbar v \frac{\partial}{\partial x} - eEx & \Delta \\ \Delta & -i\hbar v \frac{\partial}{\partial x} - eEx \end{pmatrix} \begin{pmatrix} b_- \\ b_+ \end{pmatrix} = 0. \quad (9.54)$$

Next, we eliminate the trivial dependence of b_- on the electric field (see Eq. (9.49), in the limit $\Delta = 0$, by defining the new variables

$$\begin{pmatrix} b_- \\ b_+ \end{pmatrix} = \exp\left(-i \frac{eEx^2}{2\hbar v}\right) \begin{pmatrix} b'_- \\ b'_+ \end{pmatrix}. \quad (9.55)$$

With these variables, Eq. (9.54) becomes

$$\begin{pmatrix} i\hbar v \frac{\partial}{\partial x} & \Delta \\ \Delta & -i\hbar v \frac{\partial}{\partial x} - 2eEx \end{pmatrix} \begin{pmatrix} b'_- \\ b'_+ \end{pmatrix} = 0. \quad (9.56)$$

The Fourier transform of this equation is

$$\begin{pmatrix} -\hbar vk & \Delta \\ \Delta & \hbar vk - 2ieE \frac{\partial}{\partial k} \end{pmatrix} \begin{pmatrix} \tilde{b}'_- \\ \tilde{b}'_+ \end{pmatrix} = 0, \quad (9.57)$$

where $\tilde{b}'_{\pm}$ are the Fourier transforms of $b'_{\pm}$. From the first line of this matrix equation, we obtain $\tilde{b}'_- = \Delta \tilde{b}'_+ / (\hbar vk)$. Substituting this relation in the equation obtained from the second line of (9.57), gives an equation for $\tilde{b}'_+$,

$$\left(\frac{\Delta^2}{\hbar vk} + \hbar vk - 2ieE \frac{\partial}{\partial k} \right) \tilde{b}'_+ = 0. \quad (9.58)$$

Its solution is:

$$\ln \tilde{b}'_+ = -\frac{i}{2eE} \int_{-\frac{2\Delta}{\hbar v}}^k dk' \left(\frac{\Delta^2}{\hbar v k'} + \hbar v k' \right) = -\frac{i}{2eE} \left(\frac{\Delta^2}{\hbar v} \ln \left(-\frac{\hbar v k}{2\Delta} \right) + \frac{1}{2} \hbar v k^2 + \gamma \right), \quad (9.59)$$

where γ is an arbitrary phase that can be ignored. The choice of the lower bound in the above integral is merely for convenience. Changing it will only affect the phase γ . Taking the inverse Fourier transform of $\tilde{b}'_+$ and substituting the result in (9.55) we obtain an exact integral representation of the solution for the coefficient b_+ :

$$b_+ = b_0 \exp \left(-i \frac{eEx^2}{2\hbar v} \right) \int_{-\infty}^{\infty} \frac{dk}{2\pi} \exp \left[-i \frac{\Delta^2}{2eE\hbar v} \ln \left(-\frac{\hbar v k}{2\Delta} \right) - i \frac{\hbar v}{4eE} k^2 + ikx \right], \quad (9.60)$$

where b_0 is an arbitrary constant that depends on the way we prepared the system.

To simplify the above formula, we shall evaluate b_+ in the asymptotic limit of large distance, $|x| \gg \Delta / |eE|$. In this limit, the integral over k can be evaluated by the steepest descent method. In this technique, the integration path is deformed from the real k axis into the complex plane such that it passes through the saddle point of the phase and follows the steepest descent trajectory. The first step is to choose the branch cut of the logarithm in Eq. (9.60). It is convenient to choose is to overlap the negative imaginary axis of k as shown in Fig. 9-5.

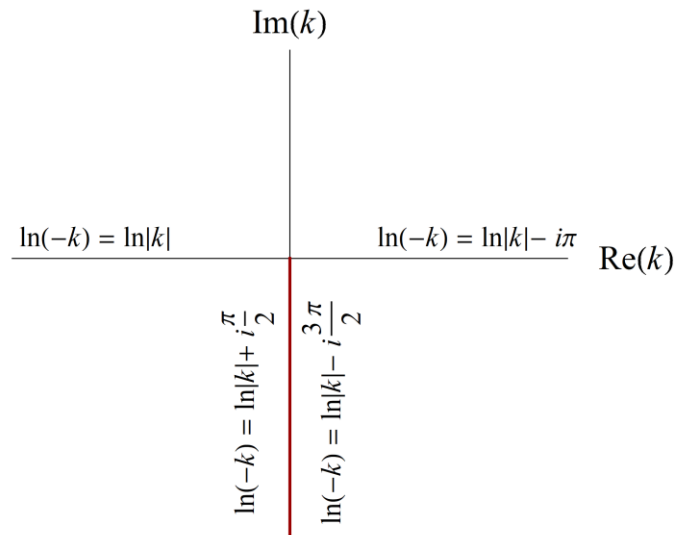


Figure 9-5 the choice of the logarithm branch cut in the Landau-Zener problem

Next, we identify the saddle point of the integral (9.60) by finding the stationary points of the phase (in the exponent of Eq. (9.60)),

$$\frac{d\phi(k)}{dk} = 0, \quad (9.61)$$

where

$$\phi(k) = -i \left[\frac{\Delta^2}{2eE\hbar v} \ln \left(-\frac{\hbar v k}{2\Delta} \right) + \frac{\hbar v}{4eE} k^2 - kx \right]. \quad (9.62)$$

From here, we obtain a quadratic equation whose solutions, in the limit $x \rightarrow \infty$, are

$$k_{\pm} = \frac{eEx}{\hbar v} \pm \frac{\sqrt{(eEx)^2 - \Delta^2}}{\hbar v} \xrightarrow{x \rightarrow \infty} \begin{cases} k_- = \frac{\Delta^2}{2eE\hbar v x} \\ k_+ = \frac{2eEx}{\hbar v} \end{cases} \quad (9.63)$$

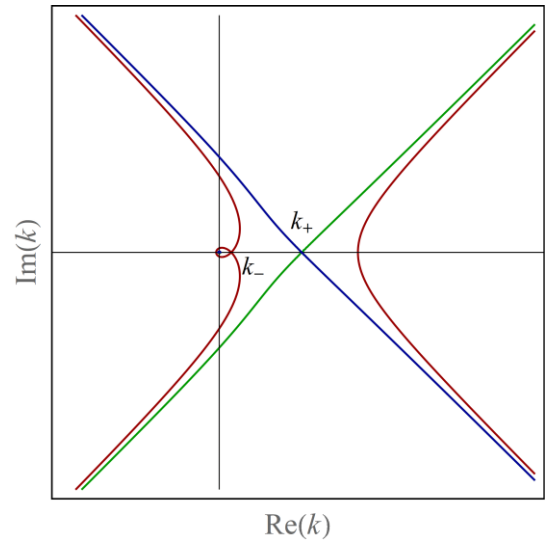
To identify the steepest descent paths that pass through the stationary points, we write the real and imaginary parts of $\phi(k + ik')$, where k and k' are real,

$$\phi(k + ik') = \alpha(k, k') + i\beta(k, k'). \quad (9.64)$$

Here both, $\alpha(k, k')$ and $\beta(k, k')$, are real functions. In particular, the imaginary part is given by

$$\beta(k, k') = -\frac{\hbar v(k^2 - k'^2)}{4eE} + kx - \frac{\Delta^2}{2\hbar v eE} \ln \left(\frac{\hbar v}{2\Delta} \sqrt{k^2 + k'^2} \right). \quad (9.65)$$

The steepest descent path going through a saddle point is the path along which the imaginary part of $\phi(k + ik')$ is constant (and hence equal to its value at the saddle point), i.e. $\beta(k, k') = \beta(k_{\pm}, 0)$. This equation, however, defines two trajectories: One is the steepest descent path along which the integrand decays at the highest rate with the distance from the saddle point. The second path is the steepest ascent path one we should avoid. Fig. 9-6 presents the steepest descent and steepest ascent paths obtained from the above condition. The red trajectory passing through the k_- saddle point is an artifact that results from the multivaluedness property of the logarithm function. It does not represent a steepest descent path. The blue and the green trajectories are the steepest descent and



איור 9-6 מסלולי המור והמעלה התלולים ביותר עבור $x \rightarrow +\infty$

steepest ascent paths going through k_+ , respectively. Thus, one has to deform the integration path from the real axis to the blue trajectory in the complex plane. One can easily see that for $x \rightarrow \infty$, $k_+ \rightarrow \infty$, and deformation of the integration path does not cross any pole or branch cut.

The advantage of integrating along the steepest descent path is that the contribution to the integrand is localized near the saddle point. Thus, to evaluate the integral, it is sufficient to expand $\phi(k)$ to second order in $\delta k = k - k_+$:

$$\phi(k_+ + \delta k) \approx i \frac{eEx^2}{\hbar v} - i \frac{\Delta^2}{2eE\hbar v} \left[\ln\left(\frac{eEx}{\Delta}\right) - i\pi \right] - i \frac{\hbar v}{4eE} \delta k^2. \quad (9.66)$$

Here, we have used the definition of the branch cut of the logarithm function shown in Fig. 9.5. With this approximation, the integral (9.60) becomes a Gaussian integral, and its evaluation yields:

$$b_+(x) = b_0 \sqrt{\frac{eE}{\pi\hbar v}} \exp\left(-\frac{\pi\Delta^2}{2eE\hbar v}\right) \exp\left[i \frac{eEx^2}{2\hbar v} - i \frac{\Delta^2}{2eE\hbar v} \ln\left(\frac{eEx}{\Delta}\right) - i \frac{\pi}{4}\right], \quad x \gg \frac{\Delta}{eE}. \quad (9.67)$$

We turn to evaluate $b_+(x)$ for negative values of x , when $x \ll -\Delta/(eE)$. With the analogy to the problem of tunneling under the barrier (see Fig. 9-4), we expect that in this case, there will be two contributions: one associated with the incoming wave, while the other describes the reflecting wave. Indeed, when $x \ll -\Delta/(eE)$ the saddle point, $k_* = -2eEx/\hbar v$, is located to the left of the branch cut, and to deform the contour along the steepest descent trajectory, it must go around the branch cut as illustrated in Fig. 9-7. This integration path gives two contributions: one from the saddle point and the other from the branch cut:

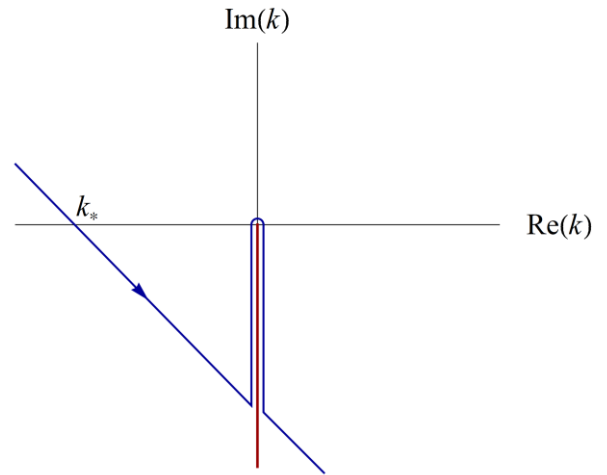


Figure 9-7 The integration path for the case $x \ll -\Delta/(eE)$

$$b_+(x) = b_+^{\text{sp}}(x) + b_+^{\text{cut}}(x) \quad (9.68)$$

The calculation of the contribution from the saddle point is similar to the above calculation. The only difference is that now the argument of the logarithm is positive and therefore does not contain an imaginary contribution. Thus

$$b_+^{\text{SP}}(x) = b_0 \sqrt{\frac{eE}{\pi \hbar v}} \exp \left[i \frac{eEx^2}{2\hbar v} - i \frac{\Delta^2}{2eE\hbar v} \ln \left(\frac{eE|x|}{\Delta} \right) - i \frac{\pi}{4} \right], \quad x \ll -\frac{\Delta}{eE}. \quad (9.69)$$

The second contribution, from the path that goes around the branch cut, can be calculated in the following way:

$$\begin{aligned} b_+^{\text{cut}} &= b_0 \exp \left(-i \frac{eEx^2}{2\hbar v} \right) \int_{\text{cut}} \frac{dk}{2\pi} \exp \left[-i \frac{\Delta^2}{2eE\hbar v} \ln \left(-\frac{\hbar vk}{2\Delta} \right) - i \frac{\hbar v}{4eE} k^2 + ikx \right] \\ &= b_0 \exp \left(-i \frac{eEx^2}{2\hbar v} \right) \int_{-i\infty+0_-}^{0_-} \frac{dk}{2\pi} \exp \left[-i \frac{\Delta^2}{2eE\hbar v} \ln \left(-\frac{\hbar vk}{2\Delta} \right) - i \frac{\hbar v}{4eE} k^2 + ikx \right] \\ &\quad + b_0 \exp \left(-i \frac{eEx^2}{2\hbar v} \right) \int_{0_+}^{-i\infty+0_+} \frac{dk}{2\pi} \exp \left[-i \frac{\Delta^2}{2eE\hbar v} \ln \left(-\frac{\hbar vk}{2\Delta} \right) - i \frac{\hbar v}{4eE} k^2 + ikx \right] \end{aligned} \quad (9.70)$$

Changing variables to $k = -iz$, and taking into account the jump in the function across branch cut (see Fig.9-5), we obtain

$$\begin{aligned} b_+^{\text{cut}} &= -ib_0 \exp \left(-i \frac{eEx^2}{2\hbar v} \right) \underbrace{\left[\exp \left(-\frac{3\pi\Delta^2}{4eE\hbar v} \right) - \exp \left(\frac{\pi\Delta^2}{4eE\hbar v} \right) \right]}_{-\exp \left(-\frac{\pi\Delta^2}{4eE\hbar v} \right) \sinh \left(\frac{\pi\Delta^2}{2eE\hbar v} \right)} \\ &\quad \times \int_0^\infty \frac{dz}{2\pi} \exp \left[-i \frac{\Delta^2}{2eE\hbar v} \ln \left(\frac{\hbar vz}{2\Delta} \right) + i \frac{\hbar v}{4eE} z^2 - z|x| \right]. \end{aligned} \quad (9.71)$$

From here, it follows that in the limit $x \rightarrow -\infty$, the only contribution to the integral comes from a small region (of order $1/|x|$) near $z = 0$. Thus, the quadratic term in z can be neglected.

Changing variables to $y = z|x|$, the integral may be approximated by

$$\begin{aligned} &\frac{1}{|x|} \int_0^\infty \frac{dy}{2\pi} \exp \left[-y - i \frac{\Delta^2}{2eE\hbar v} \ln(y) \right] \exp \left[-i \frac{\Delta^2}{2eE\hbar v} \ln \left(\frac{\hbar v}{2\Delta|x|} \right) \right] \\ &= \frac{1}{2\pi|x|} \exp \left[-i \frac{\Delta^2}{2eE\hbar v} \ln \left(\frac{\hbar v}{2\Delta|x|} \right) \right] \Gamma \left(1 - i \frac{\Delta^2}{2eE\hbar v} \right). \end{aligned} \quad (9.72)$$

Next, we use the following identity of the gamma function,

$$\Gamma(1-i\alpha) = \left[\frac{\Gamma(1-i\alpha)}{\Gamma(1+i\alpha)} \right]^{\frac{1}{2}} \left[\underbrace{\Gamma(1-i\alpha)\Gamma(1+i\alpha)}_{\frac{\pi\alpha}{\sinh(\pi\alpha)}} \right]^{\frac{1}{2}}, \quad (9.73)$$

collect the various terms

$$\begin{aligned}
 b_+^{\text{cut}}(x) &= ib_0 \exp\left(-i \frac{eEx^2}{2\hbar v}\right) \exp\left(-\frac{\pi\Delta^2}{4eE\hbar v}\right) \sinh\left(\frac{\pi\Delta^2}{2eE\hbar v}\right) \\
 &\times \frac{1}{2\pi|x|} \exp\left[-i \frac{\Delta^2}{2eE\hbar v} \ln\left(\frac{\hbar v}{2\Delta|x|}\right)\right] \left[\frac{\Gamma\left(1-i \frac{\Delta^2}{2eE\hbar v}\right)}{\Gamma\left(1+i \frac{\Delta^2}{2eE\hbar v}\right)} \right]^{\frac{1}{2}} \left[\frac{\pi \frac{\Delta^2}{2eE\hbar v}}{\sinh\left(\frac{\pi\Delta^2}{2eE\hbar v}\right)} \right]^{\frac{1}{2}}, \quad (9.74)
 \end{aligned}$$

and rearrange them

$$\begin{aligned}
 b_+^{\text{cut}}(x) &= b_0 \frac{\Delta}{4eE|x|} \sqrt{\frac{eE}{\pi\hbar v}} \exp\left[-i \frac{eEx^2}{2\hbar v} + i \frac{\Delta^2}{2eE\hbar v} \ln\left(\frac{2eE|x|}{\hbar v}\right) + i \frac{\pi}{4}\right] \\
 &\times \left[1 - \exp\left(-\frac{\pi\Delta^2}{2eE\hbar v}\right)\right]^{\frac{1}{2}} \exp\left[-i \frac{\Delta^2}{2eE\hbar v} \ln\left(\frac{2\Delta^2}{\hbar v e E}\right) + i \frac{\pi}{4}\right] \left[\frac{\Gamma\left(1-i \frac{\Delta^2}{2eE\hbar v}\right)}{\Gamma\left(1+i \frac{\Delta^2}{2eE\hbar v}\right)} \right]^{\frac{1}{2}}. \quad (9.75)
 \end{aligned}$$

This formula describes the contribution from the reflected wave (moving in the direction of the negative x axis) because the first factor in this expression (which contains the space dependence) is the complex conjugate of $b_+^{\text{sp}}(x)$ that describes the incoming wave.

From the ratio of the various components that we have calculated, one can identify the coefficients of transmission and reflection, respectively, given by

$$\begin{aligned}
 t &= \exp\left(-\frac{\pi\Delta^2}{2eE\hbar v}\right), \\
 r &= \left[1 - \exp\left(-\frac{\pi\Delta^2}{2eE\hbar v}\right)\right]^{\frac{1}{2}} \exp\left[-i \frac{\Delta^2}{2eE\hbar v} \ln\left(\frac{2\Delta^2}{\hbar v e E}\right) + i \frac{\pi}{4}\right] \left[\frac{\Gamma\left(1-i \frac{\Delta^2}{2eE\hbar v}\right)}{\Gamma\left(1+i \frac{\Delta^2}{2eE\hbar v}\right)} \right]^{\frac{1}{2}}. \quad (9.76)
 \end{aligned}$$

In particular, comparing the above formula for the transmission coefficient with (9.53), we see that $\eta = \pi/2$.

9.6 Exercises

1. The one dimensional tight-binding model of an electron in a constant electric field, E , is:

$$H = \sum_j (\bar{\varepsilon} + eEaj) \hat{c}_j^\dagger \hat{c}_j - t \sum_j (\hat{c}_j^\dagger \hat{c}_{j+1} + \hat{c}_j^\dagger \hat{c}_{j-1}), \quad (9.77)$$

where a is the lattice constant. Diagonalize this Hamiltonian and show that the energy levels are

$$\varepsilon_m = \bar{\varepsilon} + eaEm, \quad (9.78)$$

where m is an integer, while the corresponding wave functions are

$$\psi_m(j) = (-1)^j J_{m-j} \left(\frac{2t}{eEa} \right), \quad (9.79)$$

where $J_n(x)$ is the Bessel function of n -th order. Use the following integral:

$$J_n(x) = \int_{-\pi}^{\pi} \frac{d\kappa}{2\pi} \exp \left\{ -i \left[n(\kappa + \pi) + x \sin \kappa \right] \right\}. \quad (9.80)$$

Analyze and interpret the behavior of the wave function (9.79) in the limits of strong and weak electric fields using the following properties of the Bessel function:

$$J_n(0) = \delta_{n,0}, \quad (9.81)$$

and

$$\lim_{\substack{x, n \rightarrow \infty \\ n \sim x}} J_{-n}(x) = (-1)^n \left(\frac{2}{x} \right)^{\frac{1}{3}} \text{Ai} \left[\left(\frac{2}{x} \right)^{\frac{1}{3}} (n - x) \right] \quad (9.82)$$

2. Find the relation between the matrix Ω_{ij} , defined in Eq. (9.20), and the matrix V_{ij} defined in Eq.(9.43). Notice that the first describes matrix elements of the position operator $\mathbf{r}$, while the second is the matrix element of the velocity operator, $\mathbf{v}$.

10 Crystals in a constant magnetic field

This chapter discusses electrons moving in a periodic lattice and subjected to a constant and uniform magnetic field. Generally, the introduction of electric, $\mathbf{E}$, and magnetic, $\mathbf{B}$, fields in quantum mechanical systems is obtained using the scalar potential, φ , and the vector potential, $\mathbf{A}$, that satisfy the relations:

$$\begin{aligned}\mathbf{B} &= \nabla \times \mathbf{A}, \\ \mathbf{E} &= -\nabla \varphi - \frac{\partial \mathbf{A}}{\partial t}.\end{aligned}\tag{10.1}$$

The Hamiltonian of a particle with charge q moving in a periodic potential, $u(\mathbf{r})$, in the presence of an electromagnetic field is:

$$H = \frac{[\mathbf{p} - q\mathbf{A}(\mathbf{r})]^2}{2m} + q\varphi(\mathbf{r}) + u(\mathbf{r}).\tag{10.2}$$

This Hamiltonian is not periodic in space, even for the simplest case of a uniform magnetic field. For instance, for a constant B pointing in the z direction, a possible vector potential choice is $\mathbf{A} = Bx\hat{\mathbf{y}}$, where $\hat{\mathbf{y}}$ is a unit vector in y direction. The linear dependence of the vector potential in the x coordinate makes the Hamiltonian (10.2) non-periodic; hence one cannot employ Bloch's theorem to solve the problem.

Nevertheless, if the electromagnetic field is weak, one can assume that the vector and scalar potentials are locally constants and keep using Bloch's decomposition of the wave function as a leading-order approximation. This approach is valid for a large class of physical problems.

When the magnetic field is strong but uniform, it turns out that for specific values of its strength, the periodicity of the system can be restored, albeit by a larger unit cell which is an integer multiple of the original cell. These are the situations where the magnetic flux threading one unit cell is a rational fraction of the quantum flux unit, $\phi_0 = 2\pi\hbar/e$. The most prominent consequence of the multiplication of the unit cell is the disintegration of bands into minibands and the creation of self-similar fractal patterns in the spectrum.

This chapter is divided into two main parts. In the first, we discuss the limit of a weak magnetic field. Here we begin by presenting the gauge invariance property of the system and its relation to charge conservation. Next, we use gauge invariance to introduce the electromagnetic field into Bloch's Hamiltonian in the framework of the effective mass approximation. Then we present the dynamics of lattice electrons in the presence of a magnetic field and quantize their energy levels. Finally, we discuss the magnetic breakdown phenomenon.

In the second part of this chapter, we discuss the limit of a strong magnetic field. Here we, begin by discussing Moiré patterns to show how the increased size of the unit cell generates minibands. Moiré patterns are obtained, e.g., from superimposed layers of two lattices with slightly different lattice constants. Next, we define the magnetic translation operators that describe electrons in crystals subjected to a strong magnetic field and conclude by presenting the celebrated Hofstadter's butterfly describing the spectrum of electrons subjected to a uniform magnetic field in a two-dimensional lattice.

10.1 Gauge invariance and charge conservation

For a given electric $\mathbf{E}$ and magnetic $\mathbf{B}$ fields, the scalar and vector potentials are not unique. There is freedom in choosing them. From Eq. (10.1), it follows that changing potentials as:

$$\begin{aligned} \mathbf{A} &\rightarrow \mathbf{A} + \nabla f, \\ \varphi &\rightarrow \varphi - \frac{\partial f}{\partial t}, \end{aligned} \quad (10.3)$$

where f is some arbitrary function of space and time, yield precisely the same fields $\mathbf{E}$ and $\mathbf{B}$. This property is called *gauge invariance*. It tells us that the scalar and vector potentials themselves do not have a physical meaning. Only the electric and magnetic fields, which do not depend on gauge, are physical quantities.

On the other hand, the gauge invariance (10.3) is a continuous symmetry of the system, and from Noether's theorem, one expects this symmetry to manifest itself in a conserved quantity. In other words, similar to the symmetry for translation in time, which implies conservation of energy, or symmetry to rotations which leads to conservation of the angular momentum, it is expected that gauge invariance is associated with a conservation law. As we shall see below, this conserved quantity is the particle charge.

Reminder: Conservation laws in classical mechanics

In classical mechanics, the equations of motion, i.e., the Euler Lagrange equations, are derived by variation of the action:

$$S_{\text{cl}} = \int dt L(\mathbf{r}, \dot{\mathbf{r}}), \quad (10.4)$$

where $L(\mathbf{r}, \dot{\mathbf{r}})$ is the Lagrangian of the system. Namely

$$\frac{\delta S_{\text{cl}}}{\delta \mathbf{r}(t)} = 0 \quad \Rightarrow \quad \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{r}}} = \frac{\partial L}{\partial \mathbf{r}} \quad (10.5)$$

Now, let us define the following infinitesimal transformation of coordinates and time:

$$\mathbf{r}' = \mathbf{r} + \varepsilon \mathbf{R}(\mathbf{r}, \dot{\mathbf{r}}, t) \quad (10.6)$$

$$t' = t + \varepsilon T(\mathbf{r}, \dot{\mathbf{r}}, t) \quad (10.7)$$

where $\mathbf{R}(\mathbf{r}, \dot{\mathbf{r}}, t)$ and $T(\mathbf{r}, \dot{\mathbf{r}}, t)$ are some general functions, and ε is an infinitesimal dimensionless constant. This transformation is a symmetry of the system if it does not change the action at any time interval i.e., when

$$dt' L(\mathbf{r}', \dot{\mathbf{r}}') = dt L(\mathbf{r}, \dot{\mathbf{r}}) + O(\varepsilon^2). \quad (10.8)$$

Then the stationarity of the action with respect to the infinitesimal transformation (10.7) yields the conservation law:

$$\frac{dS_{cl}}{d\varepsilon} = 0 \quad \Rightarrow \quad \frac{d}{dt} \left[\frac{\partial L}{\partial \dot{\mathbf{r}}} \cdot (\mathbf{R} - T\dot{\mathbf{r}}) + TL \right] = 0 \quad (10.9)$$

Namely, the quantity in the square brackets is the conserved quantity associated with the symmetry (10.6-7). For instance, symmetry to translation in time, where $\mathbf{R} = 0$ and $T = \text{constant}$, shows that the Hamiltonian, i.e., the energy of the system, is conserved:

$$-T \frac{d}{dt} \left[\frac{\partial L}{\partial \dot{\mathbf{r}}} \cdot \dot{\mathbf{r}} - L \right] = 0. \quad (10.10)$$

Similarly, symmetry to translation in space, $\mathbf{R} = \mathbf{R}_0 = \text{constant}$ and $T = 0$ implies conservation of momentum:

$$\mathbf{R}_0 \cdot \frac{d}{dt} \left[\frac{\partial L}{\partial \dot{\mathbf{r}}} \right] = 0. \quad (10.11)$$

The proof of (10.9) can be found in the literature (see, for example, in section 6.12.1 of the book 'Waves & Optics', of the Open University).

To identify conservation laws in quantum mechanical systems, one can employ a procedure similar to classical systems. Let us define the action:

$$S_{qu} = \int dt d^d r \psi^* (i\hbar \partial_t - H) \psi, \quad (10.12)$$

where H is the Hamiltonian (10.2). Variation of this action with respect to ψ^* yields the Schrödinger equation of the problem:

$$\frac{\delta S}{\delta \psi^*} = 0 \quad \Rightarrow \quad i\hbar \frac{\partial \psi}{\partial t} = \left[\frac{1}{2m} (-i\hbar \nabla - q\mathbf{A})^2 + u(\mathbf{r}) + q\phi(\mathbf{r}) \right] \psi. \quad (10.13)$$

Changing the gauge by (10.3) transforms this equation to

$$i\hbar \frac{\partial \psi}{\partial t} = \left[\frac{1}{2m} (-i\hbar \nabla - q\mathbf{A} - q\nabla f)^2 + u(\mathbf{r}) + q\phi(\mathbf{r}) - q \frac{\partial f}{\partial t} \right] \psi. \quad (10.14)$$

It looks very different from the original one; however, this change does not affect the value of any physical quantity. In particular, one can check that if we substitute

$$\psi = \psi' \exp(iqf/\hbar) \quad (10.15)$$

in the above equation, it reduces to the original one (10.13) for ψ' .

The conservation law associated with the symmetry (10.3) is obtained from the stationarity of the action with respect to the gauge transformation, i.e.

$$\frac{\delta S_{\text{qu}}}{\delta f} = 0 \quad \Rightarrow \quad \frac{\partial}{\partial t} \frac{\delta S_{\text{qu}}}{\delta \phi} + \nabla \cdot \frac{\delta S_{\text{qu}}}{\delta \mathbf{A}} = 0 \quad (10.16)$$

This equation has the structure of a continuity equation (which is a conservation law). To identify its ingredients, notice that

$$\rho = \frac{\delta S_{\text{qu}}}{\delta \phi} = q\psi^* \psi \quad (10.17)$$

is the charge density, while

$$\mathbf{j} = \frac{\delta S_{\text{qu}}}{\delta \mathbf{A}} = \frac{-q}{2m} \left[\psi^* (-i\hbar \nabla - q\mathbf{A}) \psi + \psi (i\hbar \nabla - q\mathbf{A}) \psi^* \right], \quad (10.18)$$

is the electric current density. The equation obtained by substitution of (10.17) and (10.18) in (10.16),

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0, \quad (10.19)$$

implies that the charge cannot disappear; it can only move from one point to another. In particular, integrating the above equation over the whole space, using the divergence theorem, and assuming no charge or current at infinity, shows that

$$\frac{dQ}{dt} = 0, \quad \text{where} \quad Q = \int d^3r \rho \quad (10.20)$$

is the total charge of the system.

10.2 The effective mass approximation

To construct the effective mass approximation for a particle moving in periodic potential subjected to a weak magnetic field, let us first recall the case $\mathbf{A} = 0$. Namely, consider a particle of charge q moving in periodic lattice and subjected only to the force generated by the scalar electric potential, $\varphi(\mathbf{r})$ (that we consider to be a general function in space but time-independent). The wave function of this system can be expanded in the basis of Bloch's wave functions of the bare Hamiltonian, $H_0 = p^2/2m + u(\mathbf{r})$:

$$\psi(\mathbf{r}, t) = \sum_j \int_{\text{BZ}} \frac{d^d k}{(2\pi)^d} c_j(\mathbf{k}, t) \psi_k^{(j)}(\mathbf{r}), \quad (10.21)$$

where the integral is carried over the first Brillouin zone, $c_j(\mathbf{k}, t)$ are expansion coefficients, and $\psi_k^{(j)}(\mathbf{r}) = \phi_k^{(j)}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r})$ is Bloch's wave function of the j -th band and wave number $\mathbf{k}$. The latter is obtained from the solution of the Schrödinger equation, $H_0 \psi = \varepsilon \psi$. Substituting expansion (10.21) in the Schrödinger equation,

$$i\hbar \frac{\partial \psi}{\partial t} = [H_0 + q\varphi(\mathbf{r})] \psi, \quad (10.22)$$

we obtain an equation for the expansion coefficients $c_j(\mathbf{k}, t)$. Defining the vector, $\mathbf{c}(\mathbf{k}, t) = (c_0(\mathbf{k}, t), c_1(\mathbf{k}, t), c_2(\mathbf{k}, t), \dots)$, this equation takes the form,

$$i\hbar \frac{\partial \mathbf{c}(\mathbf{k}, t)}{\partial t} = \left[\hat{\varepsilon}(\mathbf{k}) + q\varphi \left(i \frac{\partial}{\partial \mathbf{k}} + \hat{\Omega}(\mathbf{k}) \right) \right] \mathbf{c}(\mathbf{k}, t), \quad (10.23)$$

where $\hat{\varepsilon}(\mathbf{k})$ is a diagonal matrix whose elements are the eigenvalues of H_0 , i.e. $[\hat{\varepsilon}(\mathbf{k})]_{ij} = \varepsilon_j(\mathbf{k}) \delta_{ij}$, while $\hat{\Omega}(\mathbf{k})$ is a matrix whose elements are

$$\Omega_{jj'}(\mathbf{k}) = i \left\langle \phi_k^{(j')} \left| \frac{\partial}{\partial \mathbf{k}} \right| \phi_k^{(j)} \right\rangle. \quad (10.24)$$

The proof of Eq. (10.23) is given as an exercise.

How to incorporate the magnetic field into this description? The natural way is by imposing the condition of gauge invariance (i.e., charge conservation). Namely, the change of gauge (10.3) manifests itself only in the phase of the wave function (10.15). However, the change of the wave function by a phase factor,

$$\mathbf{c}(\mathbf{k}) \rightarrow \exp\left[i\frac{q}{\hbar}f\left(i\frac{\partial}{\partial\mathbf{k}} + \hat{\Omega}\right)\right]\mathbf{c}(\mathbf{k}), \quad (10.25)$$

is due to the following transformation of the Hamiltonian,

$$\hat{\varepsilon}(\mathbf{k}) \rightarrow \exp\left[i\frac{q}{\hbar}f\left(i\frac{\partial}{\partial\mathbf{k}} + \hat{\Omega}\right)\right]\hat{\varepsilon}(\mathbf{k})\exp\left[-i\frac{q}{\hbar}f\left(i\frac{\partial}{\partial\mathbf{k}} + \hat{\Omega}\right)\right], \quad (10.26)$$

which ensures that $\hat{\varepsilon}(\mathbf{k})\mathbf{c}(\mathbf{k}) = \varepsilon\mathbf{c}(\mathbf{k})$. Thus, by identifying the dependence of the transformed Hamiltonian on the function f , one can deduce the dependence on the vector potential.

In the effective mass approximation, one neglects the matrix $\hat{\Omega}$. It is a justified approximation if $f(\mathbf{r})$ changes very slowly on the scale of the lattice constant and the energy bands are well separated from each other such that transitions between them are weak. Setting $\hat{\Omega} = 0$, the transformed Hamiltonian is given by

$$\hat{\varepsilon}(\mathbf{k}) \rightarrow \exp(i\hat{\gamma})\hat{\varepsilon}(\mathbf{k})\exp(-i\hat{\gamma}). \quad (10.27)$$

where

$$\exp(i\hat{\gamma}) = \exp\left[i\frac{q}{\hbar}f\left(i\frac{\partial}{\partial\mathbf{k}}\right)\right] \quad (10.28)$$

Assuming the energy bands to be analytic functions in the Brillouin zone (recall we assume there are no band touching points), $\hat{\varepsilon}(\mathbf{k})$ can be expanded as a Taylor series in $\mathbf{k}$. Then any term of this expansion can be represented in the following manner:

$$\begin{aligned} & \exp(i\hat{\gamma})k_\nu k_\mu \cdots k_\chi \exp(-i\hat{\gamma}) \\ &= \underbrace{\exp(i\hat{\gamma})k_\nu \exp(-i\hat{\gamma})}_{\text{}} \underbrace{\exp(i\hat{\gamma})k_\mu \exp(-i\hat{\gamma})}_{\text{}} \cdots \underbrace{\exp(i\hat{\gamma})k_\chi \exp(-i\hat{\gamma})}_{\text{}}. \end{aligned} \quad (10.29)$$

Thus, to obtain the transformed Hamiltonian it is sufficient to calculate the transformation of the wave number vector, $\mathbf{k} \rightarrow \exp[i\hat{\gamma}]\mathbf{k}\exp[-i\hat{\gamma}]$, which gives

$$\mathbf{k} \rightarrow \exp[i\hat{\gamma}]\mathbf{k}\exp[-i\hat{\gamma}] = \mathbf{k} - \frac{q}{\hbar}\nabla f\left(i\frac{\partial}{\partial\mathbf{k}}\right). \quad (10.30)$$

Before proving this result, let us discuss its meaning. The form of the gauge transformation (10.3) implies that the introduction of a vector potential into the system is obtained by Peierls substitution:

$$\mathbf{k} \rightarrow \mathbf{k} - \frac{q}{\hbar}\mathbf{A}\left(i\frac{\partial}{\partial\mathbf{k}}\right). \quad (10.31)$$

Thus, to include the magnetic field in the effective mass approximation, the Hamiltonian should be transformed according to:

$$\hat{\varepsilon}(\mathbf{k}) \rightarrow \hat{\varepsilon} \left[\mathbf{k} - \frac{q}{\hbar} \mathbf{A} \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right]. \quad (10.32)$$

Proof of Eq. (10.30)

To obtain Eq. (10.30) we shall use the following commutation formula that applies to an arbitrary pair of operators A and B :

$$e^{-B} [A; e^B] = \int_0^1 ds e^{-sB} [A; B] e^{sB}. \quad (10.33)$$

The proof of this formula follows from the fact that

$$e^{-tB} [A; e^{tB}] = \int_0^t ds e^{-sB} [A; B] e^{sB} \quad (10.34)$$

is trivially satisfied for $t = 0$, while its derivative with respect to time yields $e^{-tB} [A; B] e^{tB}$ in both sides of the equation.

Using Eq. (10.33) we have,

$$e^{i\hat{\gamma}} \mathbf{k} e^{-i\hat{\gamma}} = \mathbf{k} + e^{i\hat{\gamma}} [\mathbf{k}; e^{-i\hat{\gamma}}] = \mathbf{k} + \int_0^1 ds e^{is\hat{\gamma}} [\mathbf{k}; -i\hat{\gamma}] e^{-is\hat{\gamma}}. \quad (10.35)$$

To calculate the commutation of $[\mathbf{k}; -i\hat{\gamma}]$, let us consider, for simplicity, the one-dimensional case, and expand $\hat{\gamma} = i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial k} \right)$ in Taylor series, i.e. $f(x) = \sum_n f_n x^n$. Then

$$\begin{aligned} [k; -i\hat{\gamma}] &= \left[k; -i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial k} \right) \right] = -ik \frac{q}{\hbar} f \left(i \frac{\partial}{\partial k} \right) + i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial k} \right) k \\ &= -ik \frac{q}{\hbar} f \left(i \frac{\partial}{\partial k} \right) + i \frac{q}{\hbar} \sum_n f_n \left(i \frac{\partial}{\partial k} \right)^n k \\ &= -ik \frac{q}{\hbar} f \left(i \frac{\partial}{\partial k} \right) + ik \frac{q}{\hbar} \sum_n f_n \left(i \frac{\partial}{\partial k} \right)^n + i \frac{q}{\hbar} \sum_n i f_n n \left(i \frac{\partial}{\partial k} \right)^{n-1} \\ &= -\frac{q}{\hbar} f' \left(i \frac{\partial}{\partial k} \right), \end{aligned} \quad (10.36)$$

and a straightforward generalization to the tree-dimensional case gives

$$[\mathbf{k}; -i\hat{\gamma}] = -\frac{q}{\hbar} \nabla f \left(i \frac{\partial}{\partial \mathbf{k}} \right). \quad (10.37)$$

Substituting this formula in Eq. (10.35) and noticing that all three terms in the integrand are functions of $i\partial/\partial\mathbf{k}$ and therefore commute among themselves yields Eq. (10.30).

10.3 Perturbative corrections to the effective mass approximation

The implementation of gauge invariance to identify the correct form of the Hamiltonian, as discussed above, can be extended in order to calculate the perturbative corrections to the effective mass approximation. In this section, we derive the leading order correction to the Hamiltonian in the matrix $\hat{\Omega}$ defined in Eq. (10.24), and present a basic application of this correction in physical systems – a direct optical transition in crystals.

Consider the transformed Hamiltonian (10.26) and let us calculate it by keeping terms that are first order in $\hat{\Omega}$, thus

$$\begin{aligned} & \exp \left[i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial \mathbf{k}} + \hat{\Omega} \right) \right] \hat{\varepsilon}(\mathbf{k}) \exp \left[-i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial \mathbf{k}} + \hat{\Omega} \right) \right] \\ & \simeq \exp \left[i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right] \left(1 + i \frac{q}{\hbar} \nabla f \hat{\Omega} \right) \hat{\varepsilon}(\mathbf{k}) \left(1 - i \frac{q}{\hbar} \nabla f \hat{\Omega} \right) \exp \left[-i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right] \\ & \simeq \exp \left[i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right] \left\{ \hat{\varepsilon}(\mathbf{k}) + i \frac{q}{\hbar} [\nabla f \hat{\Omega}; \hat{\varepsilon}(\mathbf{k})] \right\} \exp \left[-i \frac{q}{\hbar} f \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right] \\ & = \hat{\varepsilon} \left[\mathbf{k} - \frac{q}{\hbar} \nabla f \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right] + i \frac{q}{\hbar} \left[\nabla f \hat{\Omega}; \hat{\varepsilon} \left[\mathbf{k} - \nabla f \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right] \right]. \end{aligned} \quad (10.38)$$

Also, expanding of the scalar potential to leading order in $\hat{\Omega}$ yields

$$\varphi \left(i \frac{\partial}{\partial \mathbf{k}} + \hat{\Omega} \right) = \varphi \left(i \frac{\partial}{\partial \mathbf{k}} \right) + \hat{\Omega} \nabla \varphi. \quad (10.39)$$

From the last two equations and the gauge invariance condition (10.3), it follows that the Hamiltonian of the system transforms according to:

$$\begin{aligned} \hat{\varepsilon}(\mathbf{k}) & \rightarrow \hat{\varepsilon} \left(\mathbf{k} - \frac{q}{\hbar} \mathbf{A} \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right) + q \varphi \left(i \frac{\partial}{\partial \mathbf{k}} \right) \\ & + \frac{iq}{\hbar} \left[\mathbf{A} \cdot \hat{\Omega}; \hat{\varepsilon} \left(\mathbf{k} - \frac{q}{\hbar} \mathbf{A} \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right) \right] + q \hat{\Omega} \nabla \varphi \end{aligned} \quad (10.40)$$

Here the first line is the Hamiltonian in the effective mass approximation, while the second line represents the first-order correction in $\hat{\Omega}$. This is the leading order correction in a/λ , where a is the lattice constant and λ is the typical scale of the spatial variations in the electromagnetic field. The improved form of the Hamiltonian (10.40) still holds only when energy bands are far apart. However, for cases where the system is subjected to an electromagnetic field that induces transitions between energy bands, the leading order correction that we have calculated plays a dominant role, as we shall see in the following example.

Example: Direct optical transitions in crystals

The general form of the Schrödinger equation associated with the Hamiltonian (10.40) is a set of an infinite number of coupled equations:

$$\begin{aligned} i\hbar \frac{\partial \mathbf{c}(\mathbf{k}, t)}{\partial t} = & \left\{ \hat{\varepsilon} \left(\mathbf{k} - \frac{q}{\hbar} \mathbf{A} \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right) + q\varphi \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right\} \mathbf{c}(\mathbf{k}, t) \\ & + \left\{ \frac{iq}{\hbar} \left[\mathbf{A} \cdot \hat{\Omega} ; \hat{\varepsilon} \left(\mathbf{k} - \frac{q}{\hbar} \mathbf{A} \left(i \frac{\partial}{\partial \mathbf{k}} \right) \right) \right] + q\hat{\Omega} \nabla \varphi \right\} \mathbf{c}(\mathbf{k}, t) \end{aligned} \quad (10.41)$$

However, when two energy bands become close to each other while all others remain sufficiently far, this system can be truncated to only two coupled equations.

For example, consider the problem of a crystal subjected to electromagnetic radiation that generates transitions between two energy bands, $\varepsilon_1(\mathbf{k})$ and $\varepsilon_2(\mathbf{k})$. We assume that the smallest gap between these bands is reached at $\mathbf{k} = \mathbf{k}_0$, as illustrated in Fig. 10-1. The electromagnetic wave's frequency, ω , is set to induce transitions between the two levels, i.e., $\varepsilon_2(\mathbf{k}_0) - \varepsilon_1(\mathbf{k}_0) = \hbar\omega$. It will also be assumed that the wavelength of the electromagnetic wave, $\lambda = 2\pi c / \omega$, is much larger than the lattice constant, a , in order to justify the leading order expansion of the Hamiltonian (10.40). Under these assumptions, the matrix $\hat{\Omega}$ reduces to a two by two matrix:

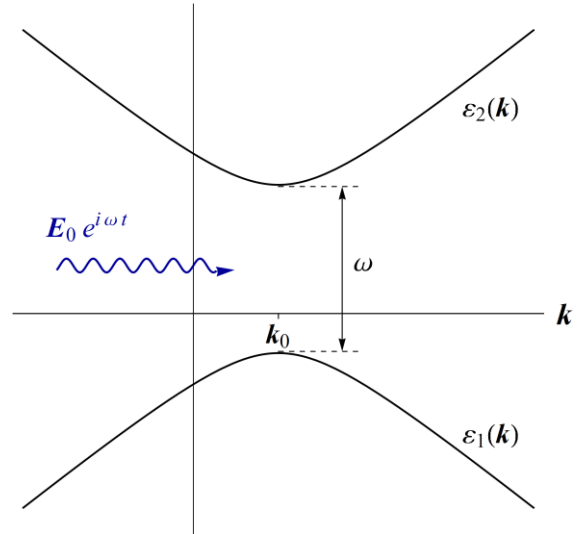


Figure 10-1 Direct transition in a crystal

$$\hat{\Omega} = \begin{pmatrix} 0 & \Omega_{12} \\ \Omega_{12}^* & 0 \end{pmatrix} \quad (10.42)$$

The diagonal matrix elements (which are the Berry connection) are zero because the energy levels are far apart; hence Berry's curvature is negligible.

Substituting (10.42) in the leading order correction of the Hamiltonian (second line of Eq. (10.40)) we obtain:

$$\begin{aligned} \frac{iq}{\hbar} [\mathbf{A} \cdot \hat{\Omega}; \hat{\varepsilon}] + q\hat{\Omega} \nabla \varphi &= \frac{iq}{\hbar} \left[\begin{pmatrix} 0 & \mathbf{A} \cdot \mathbf{\Omega}_{12} \\ \mathbf{A} \cdot \mathbf{\Omega}_{12}^* & 0 \end{pmatrix}; \begin{pmatrix} \varepsilon_1 & 0 \\ 0 & \varepsilon_2 \end{pmatrix} \right] + q \nabla \varphi \cdot \begin{pmatrix} 0 & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{12}^* & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & q \left[\frac{i}{\hbar} (\varepsilon_2 - \varepsilon_1) \mathbf{A} + \nabla \varphi \right] \cdot \mathbf{\Omega}_{12} \\ q \left[\frac{-i}{\hbar} (\varepsilon_2 - \varepsilon_1) \mathbf{A} + \nabla \varphi \right] \cdot \mathbf{\Omega}_{12}^* & 0 \end{pmatrix} \end{aligned} \quad (10.43)$$

Now using the assumption that $\varepsilon_2(\mathbf{k}_0) - \varepsilon_1(\mathbf{k}_0) = \hbar\omega$, we see that

$$q \left[\frac{i}{\hbar} (\varepsilon_2 - \varepsilon_1) \mathbf{A} + \nabla \varphi \right] \cdot \mathbf{\Omega}_{12} = q [i\omega \mathbf{A} + \nabla \varphi] \cdot \mathbf{\Omega}_{12} = q \left[\frac{\partial \mathbf{A}}{\partial t} + \nabla \varphi \right] \cdot \mathbf{\Omega}_{12} = -q\mathbf{E} \cdot \mathbf{\Omega}_{12}, \quad (10.44)$$

where to obtain the last equality, we have used Eq. (10.1), and the replacement of $i\omega \mathbf{A}$ by $\partial \mathbf{A} / \partial t$ follows from the assumption of coherent radiation at frequency ω implying that $\mathbf{A} = \mathbf{A}_0 \exp(i\omega t)$. Thus, the Schrödinger Eq. (10.41) reduces to

$$i\hbar \frac{\partial}{\partial t} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\mathbf{k}) & -q\mathbf{E} \cdot \mathbf{\Omega}_{12} \\ -q\mathbf{E} \cdot \mathbf{\Omega}_{12}^* & \varepsilon_2(\mathbf{k}) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \quad (10.45)$$

This equation shows that $-q\mathbf{\Omega}_{12}$ is the electric dipole moment of the system. Using Fermi's golden rule, we obtain that the rate of transition from the lower band to the upper one is:

$$\frac{1}{\tau_{12}} = \frac{2\pi}{\hbar} |q\mathbf{E}_0 \cdot \mathbf{\Omega}_{12}|^2 \rho_{12}, \quad (10.46)$$

where ρ_{12} is the density of states at the excited state of the system. This result shows that $\mathbf{\Omega}_{12}$ determines whether the direct optical transition is allowed ($\mathbf{\Omega}_{12} \neq 0$) or not ($\mathbf{\Omega}_{12} = 0$).

As a final comment, observe that the term "direct transition" refers to situations where the change in the electron momentum is negligible. It follows from our assumption $\lambda \gg a$ which implies that the momentum carried by the photon (and absorbed by the electron), $2\pi\hbar/\lambda$, is much smaller than the lattice momentum $2\pi\hbar/a$.

10.4 Charged particles in a crystal subjected to a magnetic field

In this section, we describe the dynamics of charged particles (electrons or holes) in a periodic lattice subjected to a static and uniform magnetic field, $\mathbf{B}$. Our starting point is the effective mass approximation,

$$H = \hat{\varepsilon}(\mathbf{p} - q\mathbf{A}), \quad (10.47)$$

where q is the particle charge, $\mathbf{p}$ is the canonical momentum, and $\hat{\varepsilon}(\mathbf{p})$ represents the Hamiltonian of the system in the absence of electromagnetic field.

The dynamical momentum of the particle is defined to be $\boldsymbol{\pi} = \mathbf{p} - q\mathbf{A}$ (In the case of a free particle with mass m , $\boldsymbol{\pi}/m$ is the particle's velocity). Let us calculate the commutation relations of its components:

$$\begin{aligned} [\pi_i, \pi_j] &= [p_i - qA_i, p_j - qA_j] = -[p_i, qA_j] - [qA_i, p_j] \\ &= i\hbar q \left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} \right) = i\hbar q \varepsilon_{ijk} B_k, \end{aligned} \quad (10.48)$$

where hereinafter summation over repeated indices is implied, and ε_{ijk} is the anti-symmetric tensor. It is convenient to present this result in the form:

$$[\pi_i, \pi_j] = i \frac{\hbar^2}{l_B^2} \frac{\varepsilon_{ijk} B_k}{B}, \quad (10.49)$$

where

$$l_B = \sqrt{\frac{\hbar}{qB}} \quad (10.50)$$

is the natural length scale of the problem called the *magnetic length*. The magnetic length is the radius of a circle threaded by half of the unit magnetic flux, $\pi l_B^2 B = \phi_0/2$, where $\phi_0 = 2\pi\hbar/e$. For a typical field of one Tesla, the magnetic length is 257Å. The effective mass approximation is valid when the magnetic length is much larger than the lattice constant.

We turn to derive the Heisenberg equations for the operators, $\boldsymbol{\pi}$ and $\mathbf{r}$. These are given by

$$\frac{\partial \pi_i}{\partial t} = \frac{-i}{\hbar} [\pi_i, H] = \frac{-i}{\hbar} \frac{\partial \hat{\varepsilon}}{\partial \pi_j} [\pi_i, \pi_j] = q \varepsilon_{ijk} \frac{\partial \hat{\varepsilon}}{\partial \pi_j} B_k = q \left(\frac{\partial \hat{\varepsilon}}{\partial \boldsymbol{\pi}} \times \mathbf{B} \right)_i, \quad (10.51)$$

and

$$\frac{\partial r_i}{\partial t} = \frac{-i}{\hbar} [r_i; H] = \frac{-i}{\hbar} \frac{\partial \hat{\mathcal{E}}}{\partial \pi_j} [r_i; \pi_j] = \frac{\partial \hat{\mathcal{E}}}{\partial \pi_i}. \quad (10.52)$$

The dynamics that result from the above equations ensure energy conservation because

$$\frac{\partial \hat{\mathcal{E}}}{\partial t} = \frac{\partial \hat{\mathcal{E}}}{\partial \pi_i} \frac{\partial \pi_i}{\partial t} = q \varepsilon_{ijk} \frac{\partial \hat{\mathcal{E}}}{\partial \pi_i} \frac{\partial \hat{\mathcal{E}}}{\partial \pi_j} B_k = q \mathbf{B} \cdot \left(\frac{\partial \hat{\mathcal{E}}}{\partial \boldsymbol{\pi}} \times \frac{\partial \hat{\mathcal{E}}}{\partial \boldsymbol{\pi}} \right) = 0. \quad (10.53)$$

Thus, the particle can move only on the energy surface.

From now on, we adopt a semiclassical approach, where functions replace operators. This approximation allows us to describe the particle trajectories in space. Multiplying Eq. (10.51) vectorially from the right by $\mathbf{B}$ and substituting Eq. (10.52), we obtain

$$\frac{\partial \boldsymbol{\pi}}{\partial t} \times \mathbf{B} = q \left(\frac{\partial \mathbf{r}}{\partial t} \times \mathbf{B} \right) \times \mathbf{B} = q \left(\mathbf{B} \cdot \frac{\partial \mathbf{r}}{\partial t} \right) \mathbf{B} - q B^2 \frac{\partial \mathbf{r}}{\partial t}, \quad (10.54)$$

where the second equality is obtained from the vector identity $(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = (\mathbf{A} \cdot \mathbf{C}) \mathbf{B} - (\mathbf{B} \cdot \mathbf{C}) \mathbf{A}$.

Thus, if we denote by $\mathbf{r}_\perp$ the component of the vector which is perpendicular to the magnetic field, then the above equation reduces to

$$\frac{\partial}{\partial t} \left[\frac{1}{q B^2} (\boldsymbol{\pi} \times \mathbf{B}) + \mathbf{r}_\perp \right] = 0. \quad (10.55)$$

Its integration over time yields

$$\mathbf{r}_\perp(t) = \mathbf{R}_\perp - \frac{1}{q B^2} (\boldsymbol{\pi} \times \mathbf{B}), \quad (10.56)$$

where $\mathbf{R}_\perp$ is a time-independent constant called the *guiding center*.

Multiplying Eq. (10.51) by B_i and summing over i shows that the component of the dynamical momentum, in the direction of the magnetic field, is time-independent. Thus, the particle's motion is restricted to the cross-section of the Fermi surface and the plane $\boldsymbol{\pi} \cdot \hat{\mathbf{B}} = \text{constant}$, where $\hat{\mathbf{B}}$ is a unit vector in the direction of the magnetic field, as demonstrated by the dashed line in Fig. 10-2.

In particular, if we choose a coordinate system where the magnetic field is pointing in the z direction, then $\pi_z = \bar{\pi} = \text{constant}$. On this plane, the particle is confined to move on the contour

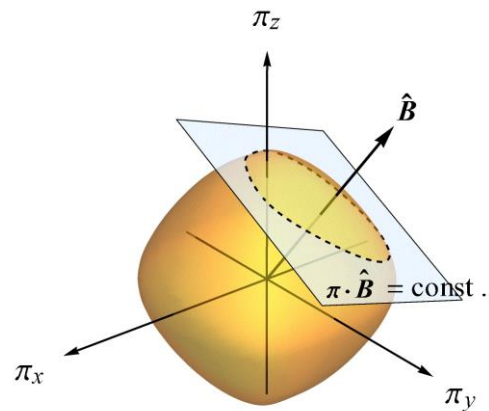


Figure 10-2 The cross section of the Fermi surface and the plane $\boldsymbol{\pi} \cdot \hat{\mathbf{B}} = \text{constant}$

$\varepsilon(\pi_x, \pi_y, \bar{\pi}) = \text{constant}$ in the Brillouin zone. Finally, Eq. (10.56) shows that the particle's trajectory in real space, projected on the plane perpendicular to the magnetic field, is the same as the trajectory in the (dynamical) momentum space up to rotation by 90° , and a change of the physical dimensions.

Examples

1. Consider a two-dimensional system of electrons ($q = -e$) moving in a lattice subjected to a constant magnetic field perpendicular to the system, $\mathbf{B} = B\hat{\mathbf{z}}$. The dynamical momentum, in this case, contains two components $\boldsymbol{\pi} = (\pi_x, \pi_y)$, and assuming that near the bottom of the band the spectrum is parabolic and isotropic, the Hamiltonian is

$$H = \frac{\boldsymbol{\pi}^2}{2m_{\text{eff}}}, \quad (10.57)$$

where m_{eff} is the effective mass. In two-dimensional systems, it is convenient to use complex coordinates, $r = x + iy$ and $\pi = \pi_x + i\pi_y$. In terms of these coordinates, the equations of motion (10.51) and (10.56) become

$$\frac{\partial \pi}{\partial t} = -i\omega_c \pi \quad \text{and} \quad \dot{r} = R - \frac{i\pi}{eB}, \quad (10.58)$$

where $\omega_c = eB/m_{\text{eff}}$ is the cyclotron frequency.

The solution of the above equations is straightforward:

$$\pi = \pi_0 \exp(-i\omega_c t) \quad \text{and} \quad r = R - \frac{i\pi_0}{eB} \exp(-i\omega_c t), \quad (10.59)$$

where π_0 is a vector whose magnitude is determined by the energy, $|\pi_0| = \sqrt{2m_{\text{eff}}\varepsilon}$, while the initial conditions fix its direction. This solution describes a circular motion, both in real and momentum space, as demonstrated in Fig. 10-3.

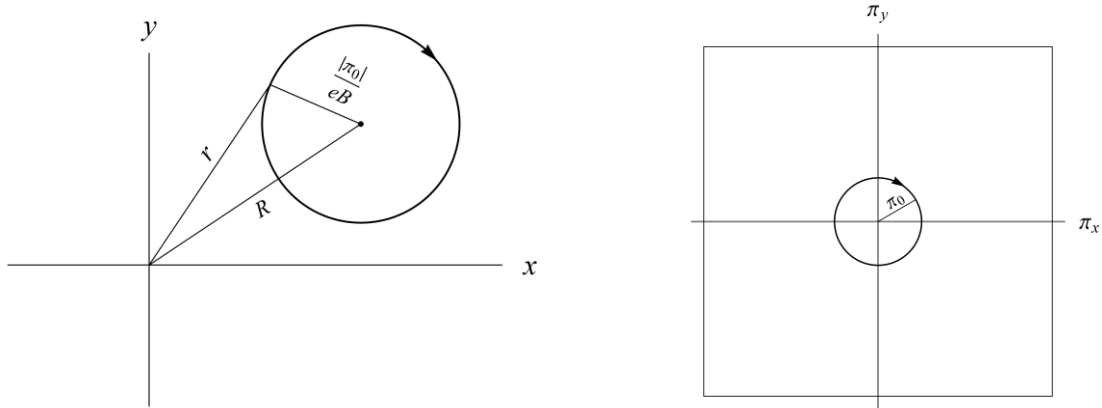


Figure 10-3 The circular motion of an electron in a two-dimensional system near the bottom of a parabolic band

- When the Fermi energy is lifted away from the bottom of the band, the Fermi surface may develop a more complicated structure. This structure is reflected in the electron motion because the electrons are constrained to move on this surface. In particular, if the Fermi surface (a line in two dimensions) is closed, the electron motion in real space follows a closed trajectory with the exact shape of the Fermi surface but rotated by 90° , as illustrated schematically in Fig. 10-4.

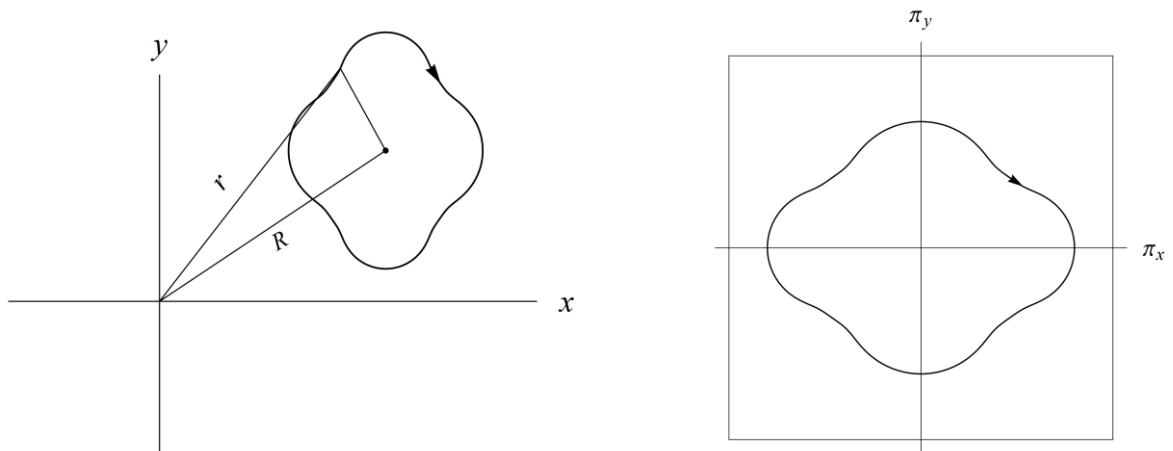


Figure 10-4 A schematic illustration of the electron trajectory in real space (left) and momentum space (right) for the case where the Fermi level is far from the bottom of the band (in two-dimensional systems)

- An intriguing behavior appears in systems with open Fermi surfaces, as illustrated in the right panel of Fig. 10-5. Assuming the particle is subjected to a uniform magnetic field and moves on the upper branch of the Fermi surface, its dynamical momentum in the y direction is positive at all times, but in the x direction it changes periodically in time. Therefore, the particle does not follow a closed contour in real space, as one might expect from the action

of Lorentz force. Instead, it forms a winding trajectory along the y axis, as shown in the left panel of Fig. 10-5.

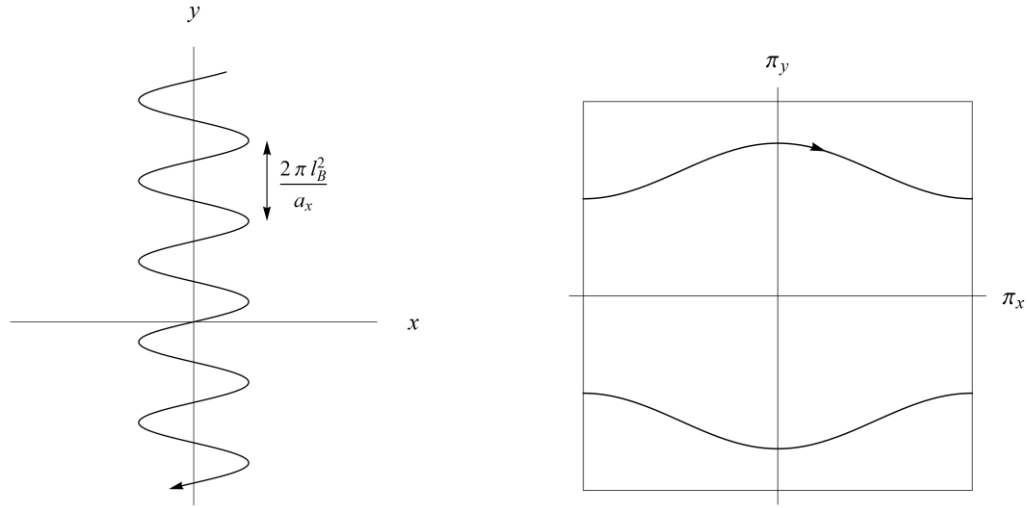


Figure 10-5 The trajectory of an electron subjected to uniform magnetic field in a two dimensional system with open Fermi surface.

To quantify the electron motion in real space, consider the following form of the upper branch of the Fermi surface:

$$\pi_y = \bar{g} + g\left(\frac{\pi_x}{b_x}\right), \quad (10.60)$$

where $\bar{g} > 0$ is a constant, $g(\eta)$ is some general periodic function (in the Brillouin zone) with zero mean, and b_x is the size of the Brillouin zone in the x direction. We also assume that $\max|g(\eta)| < \bar{g}$, to ensure that the upper branch of the Fermi surface remains positive for any value of π_x . For electrons ($q = -e$) moving in two-dimensional systems with a perpendicular magnetic field, the equations of motion (10.54) reduce to

$$\frac{\partial \pi_x}{\partial t} = -eB \frac{\partial y}{\partial t}, \quad \text{and} \quad \frac{\partial \pi_y}{\partial t} = eB \frac{\partial x}{\partial t}. \quad (10.61)$$

Integration of these equations give a parametric representation of the electron's trajectory, where the parameter of the representation is the x - component of the dynamic momentum:

$$\begin{cases} y = y_0 - \frac{1}{eB} \pi_x \\ x = x_0 + \frac{1}{eB} \left[\bar{g} + g \left(\frac{\pi_x}{b_x} \right) \right] \end{cases} \quad (10.62)$$

This representation of the electron trajectory shows that it is periodic in the x direction while linear in the y direction. From the above solution, one can deduce the distance that the electron advances along the y axis during a whole period of its motion in the x direction:

$$\Delta y = \frac{\hbar b_x}{eB} = \frac{2\pi\hbar}{eBa_x} = 2\pi \frac{l_B^2}{a_x}, \quad (10.63)$$

where $a_x = 2\pi / b_x$ is the size of the lattice cell in the x direction.

10.5 Bohr-Sommerfeld quantization and Landau levels

The classical solutions obtained above can be used to carry out a semiclassical quantization of the system, using the Bohr-Sommerfeld approach. Let us recall the basic idea of this approach. Consider a closed trajectory of an electron of length L , and assume that the electron's wavelength is λ , then constructive interference of the wave along the trajectory requires that L accommodates an integer number of wavelengths, $L/\lambda = n+1$, where $n = 0, 1, 2, \dots$. The momentum of the particle is $p/\hbar = 2\pi/\lambda$, therefore, one may write this condition in the form $pL/\hbar = 2\pi(n+1)$. Since the momentum is a function of the energy, ε , this condition yields a quantized set of energy levels.

Now consider a situation where the momentum changes along the trajectory. Assuming the wavelength to be sufficiently small, one can divide the trajectory into small segments and calculate the number of wavelengths in each one of them. It amounts to replacement of pL by an integral of p along the closed trajectory of the particle, $\mathcal{C}$. The latter is nothing but the action:

$$S(\varepsilon) = \oint_{\mathcal{C}} d\mathbf{r} \cdot \mathbf{p}. \quad (10.64)$$

This action, divided by $\hbar$, is, in principle, the total phase accumulated by the particle along a closed trajectory. However, one should also add contributions to this phase that come from turning points along its trajectory. Consider, for instance, Dirichlet boundary conditions on the edge of an infinite potential well. The reflected wave from this edge has an opposite sign to ensure that the total wave function vanishes on the boundary. It amounts to the accumulation

of a $-\pi$ phase. In the case of a soft boundary, as in parabolic potential well, the phase of the reflected wave is $-\pi/2$. This result is obtained by linearizing the potential near the turning point and analyzing the local solution that takes the form of an Airy function (see Eq. (9.7)).

Denoting the accumulated phase of the particle due to the turning points by $-\gamma$, the Bohr-Sommerfeld quantization condition takes the form:

$$\frac{1}{\hbar} S(\varepsilon_n) = 2\pi(n+1) - \gamma. \quad (10.65)$$

The phase γ is called the *Maslov phase*, and more generally, in high dimensions, it also contains contributions from focusing points, but for our purpose, it is enough to count the number of turning points and multiply them by the appropriate factor.

As an example for the application of the Bohr-Sommerfeld quantization, consider the harmonic oscillator:

$$H = \frac{p^2}{2m} + \frac{1}{2} m \omega^2 x^2. \quad (10.66)$$

The particle's momentum dependence on the energy is $p = \pm \sqrt{2m\varepsilon - m^2 \omega^2 x^2}$, where the sign depends on the direction of the motion. The two turning points are the points where the momentum vanishes. These are located at $x_{\pm} = \pm \sqrt{2\varepsilon/m\omega^2}$, and since the turning points are from a soft potential, $\gamma = \pi$.

The action along the closed trajectory of the particle is

$$S(\varepsilon) = \oint dx p = 2 \int_{x_-}^{x_+} dx \sqrt{2m\varepsilon - m^2 \omega^2 x^2} = \frac{2\pi\varepsilon}{\omega}. \quad (10.67)$$

Notice that this action is the area, $\mathcal{A}(\varepsilon)$, enclosed by the trajectory in phase space, see Fig. 10-6. Substituting (10.67) and $\gamma = \pi$ in the quantization condition (10.65) yields the energy levels of the harmonic oscillator:

$$\varepsilon_n = \hbar \omega \left(n + \frac{1}{2} \right). \quad (10.68)$$

Now let us return to the closed trajectories of charged particles subjected to magnetic field on the Fermi surface. The key observation for the semiclassical

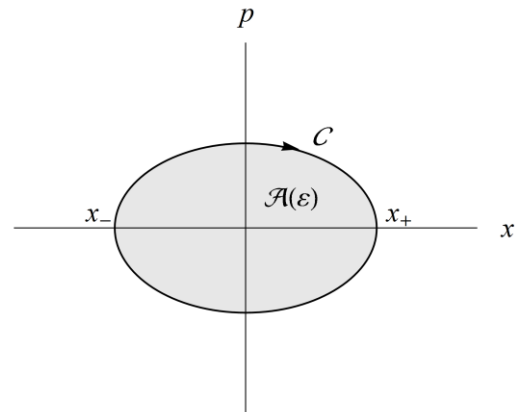


Figure 10-6 The trajectory of harmonic oscillator in phase space

quantization of these systems is that the components of the dynamical momentum in the plane perpendicular to the magnetic field behave as canonically conjugate variables. In particular, choosing the magnetic field to be in the z -direction, the x and the y components of the dynamical momentum satisfy the commutation relations (see Eq. (10.49)):

$$[\pi_i; \pi_j] = i \frac{\hbar^2}{l_B^2} \varepsilon_{ijz} \quad i, j = x, y, \quad (10.69)$$

where l_B is the magnetic length defined in Eq. (10.50). This equation implies that $\hbar$ in the quantization condition (10.65) should be replaced by $\hbar^2/l_B^2$; Hence the equitized energy levels are obtained from the formula:

$$\frac{l_B^2}{\hbar^2} \tilde{S}(\varepsilon_n) = 2\pi(n+1) - \gamma, \quad (10.70)$$

where

$$\tilde{S}(\varepsilon) = \iint d\pi_x d\pi_y = \mathcal{A}(\varepsilon) \quad (10.71)$$

is the area enclosed by the trajectory in the momentum space, see illustration in Fig. 10-7. Noticing that there are two turning points of the trajectory, the Bohr-Sommerfeld quantization condition yields the general formula:

$$\mathcal{A}(\varepsilon_n) = 2\pi \frac{\hbar^2}{l_B^2} \left(n + \frac{1}{2} \right). \quad (10.72)$$

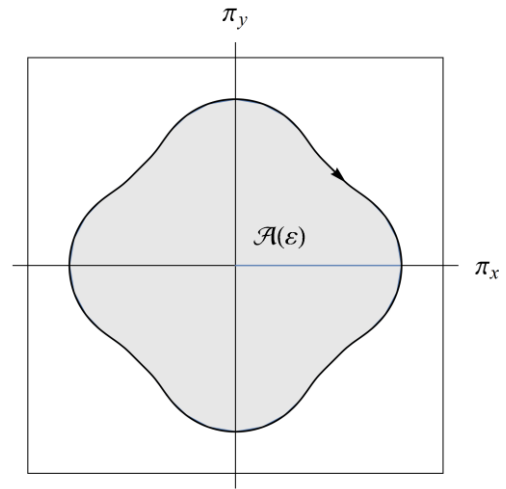


Figure 10-7 The area enclosed by trajectory in the momentum space

If the Fermi level is near the bottom of the band, the Hamiltonian of the system may be approximated by (10.57), and the particle's trajectory in momentum space is a circle of radius $\sqrt{2m_{\text{eff}}\varepsilon}$. The area enclosed by this trajectory is $\mathcal{A}(\varepsilon) = 2\pi m_{\text{eff}}\varepsilon$. Substituting it in formula (10.72) give the energy levels:

$$\varepsilon_n = \frac{\hbar^2}{m_{\text{eff}} l_B^2} \left(n + \frac{1}{2} \right) \quad (10.73)$$

Finally, substituting $l_B^2 = \hbar/|qB|$ and the formula for the cyclotron frequency, $\omega_c = eB/m_{\text{eff}}$, we obtain

$$\varepsilon_n = \hbar \omega_c \left(n + \frac{1}{2} \right) \quad (10.74)$$

These energy levels are called *Landau levels*.

10.6 Magnetic breakdown

Magnetic breakdown, similar to dielectric breakdown, is associated with situations where the particle trajectory deviates from the path dictated by the on-shell-energy condition, namely situations where the particle tunnels between trajectories separated by a potential barrier. There are several typical cases where a magnetic breakdown is likely to occur. One of them is when the Fermi surface passes near the edge of the Brillouin zone, as illustrated in Fig 10-8. In this case, the particle may tunnel between adjacent Brillouin zones and follow trajectories that are similar to those of the open Fermi surface. In Fig. 10-8, red discs denote the regions where such tunneling is expected.

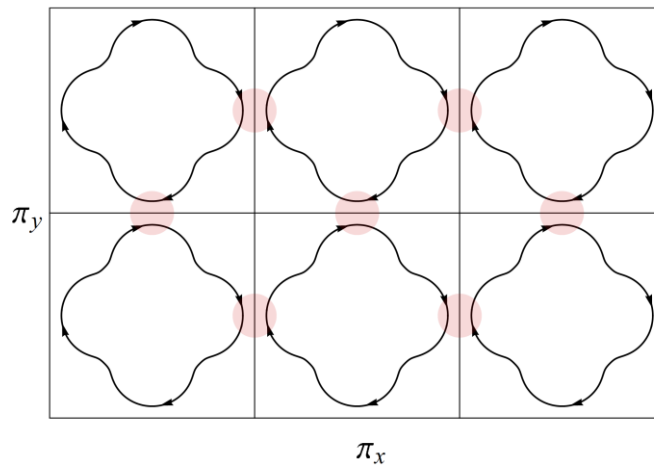


Figure 10-8 Magnetic breakdown at regions near the edge of the Brillouin zone where a transition between trajectories, by tunneling, becomes likely

Another situation where a magnetic breakdown might occur is in systems with open Fermi surfaces, at regions where two Fermi surfaces almost touch, as illustrated in Fig. 10-9.

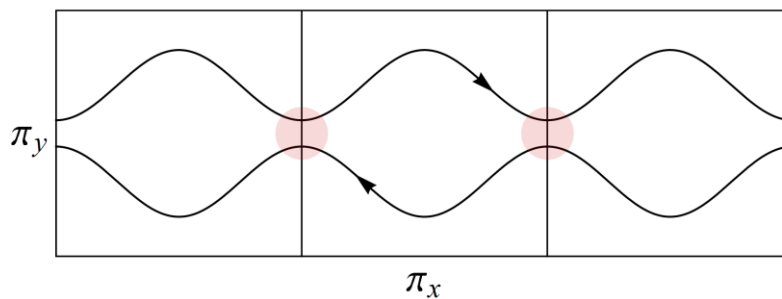


Figure 10-9 Hot spots of magnetic breakdown in systems with open Fermi surface

The calculation of the transition probability between trajectories is similar to that of dielectric breakdown and is given as an exercise. However, one can deduce the functional dependence of the transition coefficient, t , on the magnetic field without any calculation. The commutation relation of the dynamical momentum (10.69) indicates that the effective $\hbar$ in the problem is $\hbar^2/l_B^2$; hence, the transition coefficient should have the form $t \sim \exp(-l_B^2 \mathcal{A}/\hbar^2)$, where $\mathcal{A}$ is an area in the momentum space that characterizes the transition between two nearby trajectories. This area is a property of the Fermi surface; therefore, it is independent of the magnetic field. Thus, the primary dependence on the magnetic field comes from the magnetic length $l_B^2 = \hbar/|q\mathbf{B}|$, which implies that the transition coefficient takes the form

$$t \sim \exp(-B_0/B), \quad (10.75)$$

where B_0 is the typical value of the magnetic field above which a magnetic breakdown takes place. This value is system and energy dependent. Notice the singular dependence of the transition coefficient (10.75) on the magnetic field, which is similar to that of the electric field for the case of dielectric breakdown, see Eq. (9.53).

Example: Pippard's model

Pippard's model (1962) is a simple model for which the threshold magnetic field, B_0 , can be calculated analytically. It consists of a two-dimensional system with a potential that is periodic only in one direction, say, along the x axis. The periodic potential, $u(x)$, is assumed to be sufficiently weak to employ the nearly free electron approximation for calculating the band structure. In the empty lattice approximation, $u(x) = 0$, the Fermi surface is folded into the first Brillouin zone by duplicating the Fermi surface of a free electron in each Brillouin zone, as demonstrated in Fig. 10-10.

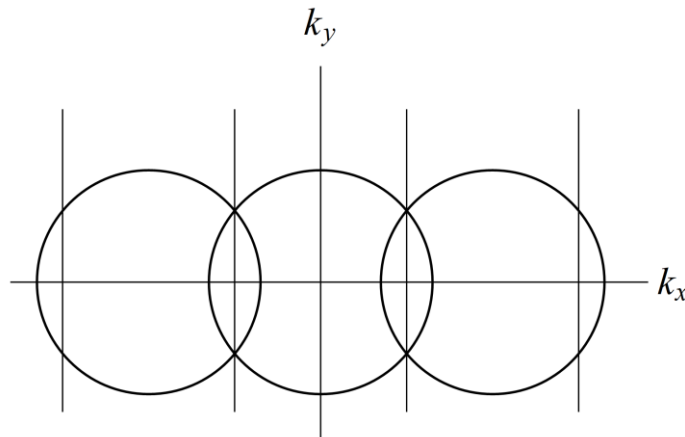


Figure 10-10 Fermi surface of Pippard's model in the empty lattice approximation

Each circle in this figure represents the Fermi surface of a free electron, and the vertical lines are the boundaries of the Brillouin zones, which in our case are vertical stripes. The main effect of the weak periodic potential, $u(x)$, is to lift the degeneracy at the points where Fermi surfaces intersect, as shown in Fig. 10-11. Thus, the Fermi surface contains both open and closed sectors. In the presence of a perpendicular magnetic field, depending on the initial conditions, electrons may follow closed or open trajectories in real space, similar to those shown in Figs. 10-4 and 10-5. The arrows in Fig. 10-11 show the electron trajectories in the momentum space.

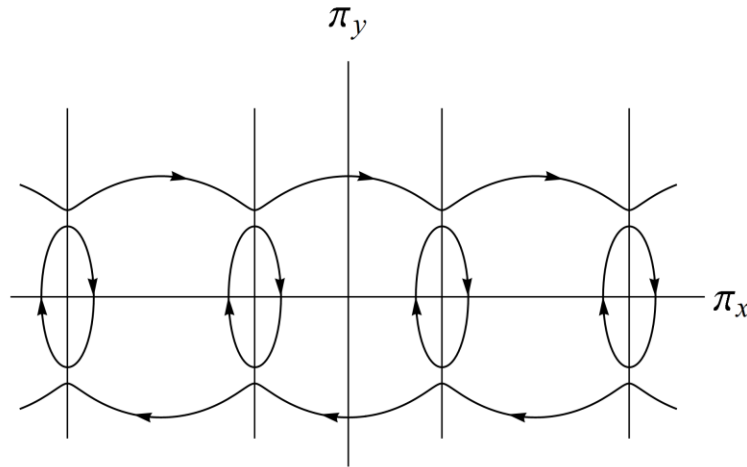


Figure 10-11 The Fermi surface in a two-dimensional electronic system with weak periodic potential, $u(x)$

Magnetic breakdown in this system is realized when an electron, prepared in the upper branch of the Fermi surface, tunnels into the closed elliptical sector of the Fermi surface, then into the lower branch of the Fermi surface, and back to the upper branch through the closed sector. If the magnetic field is sufficiently strong, the transition probability is high, and the particle completes a full circular motion, as shown by the dashed line in Fig. 10-12. In other words, the electron behaves as if it was a free particle in a magnetic field. It is what one should expect when the effect of the magnetic field overcomes that of the periodic potential. Thus, to estimate the typical magnetic field, B_0 , above which magnetic breakdown takes place, one should

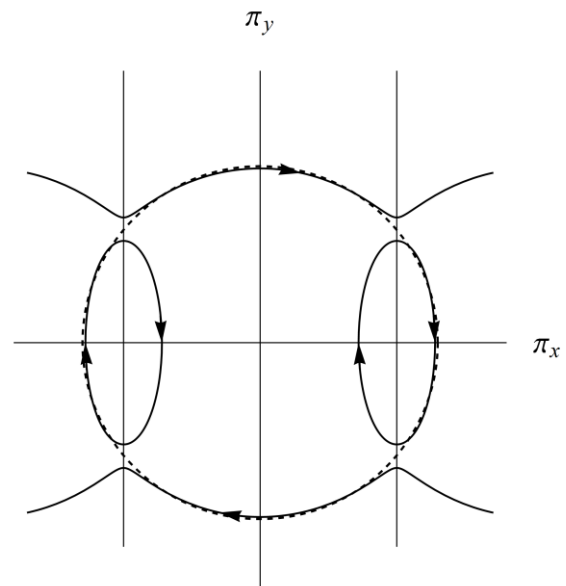


Figure 10-12 The trajectory (dashed line) of magnetic breakdown in Pippard's model

compare the energy correction to the Landau levels due to the periodic potential to the energy difference between adjacent Landau levels.

Since the periodic potential is assumed to be weak, the calculation of the energy correction to the Landau levels can be carried out in the framework of first-order perturbation theory which we turn to present now. To be concrete, we consider the Hamiltonian $H = H_0 + u(x)$ where

$$H_0 = \frac{(\mathbf{p} + e\mathbf{A})^2}{2m} = \frac{p_x^2 + (p_y + eBx)^2}{2m}, \quad (10.76)$$

is the Hamiltonian of a free electron moving in a two-dimensional plane with a perpendicular magnetic field (where we choose the gauge $\mathbf{A} = xB\hat{y}$), and

$$u(x) = u_0 \cos\left(\frac{2\pi}{a}x\right), \quad (10.77)$$

is the perturbation potential. Here u_0 is a constant that characterizes the strength of the potential, while a is the period of the potential.

Diagonalization of H_0 is obtained by separation of variables,

$$\psi = \exp(ik_y y) X(x), \quad (10.78)$$

which reduces the problem to the solution of the Harmonic oscillator,

$$-\frac{\hbar^2}{2m} \frac{\partial^2 X}{\partial x'^2} + \frac{1}{2} m \omega_c^2 x'^2 X = \varepsilon X, \quad (10.79)$$

where ε is the energy of the particle, while

$$\omega_c = \frac{eB}{m} \quad \text{and} \quad x' = x + \frac{\hbar k_y}{eB}. \quad (10.80)$$

Thus, the energies of H_0 are Landau levels given by Eq. (10.68), and the wave functions associated with the n -th Landau level are

$$\psi_{n,k_y}(x, y) = \frac{1}{\sqrt{L}} \exp(ik_y y) X_n(x'), \quad (10.81)$$

where L is the size of the system in the y direction, and

$$X_n(x) = \frac{1}{\sqrt{2^n n!}} \left(\frac{m\omega_c}{\pi\hbar}\right)^{1/4} \exp\left(-\frac{m\omega_c}{2\hbar} x^2\right) H_n\left(\sqrt{\frac{m\omega_c}{\hbar}} x\right), \quad (10.82)$$

Here $H_n(z)$ are the Hermite polynomials. Notice that these wave functions are normalized to unity. Thus, the first-order correction to the n -th energy level due to the potential (10.77) is:

$$\Delta\mathcal{E}_n = \left\langle \psi_{n,k_y} \left| u(x) \right| \psi_{n,k_y} \right\rangle = \int_{-\infty}^{\infty} dx X_n^2(x) u_0 \cos \left[\frac{2\pi}{a} \left(x - \frac{\hbar k_y}{eB} \right) \right]. \quad (10.83)$$

Comparing this correction to the distance between Landau levels, $\hbar\omega_c$, yields the value of B_0 .

Our goal, now, is to calculate $\Delta\mathcal{E}_n$ in the limit of high energies $n \gg 1$. For this purpose, one may use the semiclassical approximation for the wave function $X_n(x)$. Within this approximation, the amplitude of the wave function is the square root of the classical density of the particle in space:

$$\rho_{\text{cl}}(x) = \frac{\theta(x_*^2 - x^2)}{\pi\sqrt{x_*^2 - x^2}}, \quad (10.84)$$

where $x_* = \sqrt{2\varepsilon/m\omega_c^2}$ is the turning point of a particle with energy ε , while $\theta(z)$ is the Heaviside step function, which is one for $z > 0$ and zero otherwise¹. Notice that we choose to normalize $\rho_{\text{cl}}(x)$ such that its integral over space is unity.

The phase of the wave function is determined by the action (measured from some arbitrary point in space which here we choose to be the origin) divided by $\hbar$:

$$\phi(x) = \frac{1}{\hbar} \int_0^x dx' p(x'), \quad (10.85)$$

where the momentum is given by

$$p(x) = \sqrt{2m\varepsilon \left(1 - \frac{x^2}{x_*^2} \right)}. \quad (10.86)$$

Thus, the semiclassical approximation for the (normalized) wave functions is:

¹ The classical distribution is obtained by projecting the microcanonical distribution on the energy shell down to real space. For

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega_c^2 x^2,$$

The microcanonical distribution is $N\delta(\varepsilon - H)$, where N is the normalization constant, therefore

$$\rho_{\text{cl}}(x) = N \int_{-\infty}^{\infty} dp \delta \left(\varepsilon - \frac{p^2}{2m} - \frac{1}{2}m\omega_c^2 x^2 \right).$$

$$X(x) = \theta(x_*^2 - x^2) \sqrt{\frac{2}{\pi \sqrt{x_*^2 - x^2}}} \begin{Bmatrix} \cos[\phi(x)] \\ \sin[\phi(x)] \end{Bmatrix}, \quad (10.87)$$

where the cosine and the sine functions describe even and odd wave functions (as the system is symmetric to reflection through the origin).

Substituting formula (10.87) in Eq. (10.83), we obtain the first-order correction to energy:

$$\Delta \varepsilon = \int_{-x_*}^{x_*} dx \frac{u_0}{\pi \sqrt{x_*^2 - x^2}} \left[1 \pm \cos(2\phi(x)) \right] \cos\left(\frac{2\pi}{a} x\right), \quad (10.88)$$

where the + and – signs correspond to even and odd wave functions, respectively. Here, we have suppressed the shift of the cosine function by $\hbar k_y / eB$. This dependence is trivial and will be resorted in the final result.

The oscillatory component of the above integral changes very rapidly compared to the distance between the turning points, $\pm x_*$, therefore the integral can be calculated in the stationary phase approximation. Neglecting the contribution from the integral over $\cos(2\pi x/a)$ because it changes very rapidly in space, and using the identity $2 \cos \alpha \cos \beta = \cos(\alpha - \beta) + \cos(\alpha + \beta)$, the above integral can be approximated by

$$\Delta \varepsilon \simeq \int_{-x_*}^{x_*} dx \frac{u_0}{2\pi \sqrt{x_*^2 - x^2}} \left[\pm \cos\left(2\phi(x) - \frac{2\pi}{a} x\right) \pm \cos\left(2\phi(x) + \frac{2\pi}{a} x\right) \right]. \quad (10.89)$$

From here, we obtain the stationary phase conditions:

$$\frac{d}{dx} \left[2\phi(x) \pm \frac{2\pi}{a} x \right] = 0, \quad (10.90)$$

which give the stationary points, x_{sp} , as solutions of the equation

$$k(x_{sp}) \pm \frac{\pi}{a} = 0, \quad (10.91)$$

where $k(x) = p(x)/\hbar$ is the local wavenumber with the momentum defined in Eq. (10.86). Since $k(x)$ is positive within the range $|x| < x_*$, real solutions of the above equation are obtained only for the minus sign. The stationary points obtained in this case are:

$$x_{sp}^{(\pm)} = \pm x_* \sqrt{1 - \frac{\pi^2 \hbar^2}{2m \varepsilon a^2}}. \quad (10.92)$$

Notice that they lie within the integration range, i.e. $|x_{\text{sp}}| < x_*$.

It is convenient to define the angle θ by the relation:

$$\cos \theta = \frac{\pi \hbar}{a \sqrt{2m\varepsilon}}. \quad (10.93)$$

Fig. 10-13 shows the geometrical meaning of this angle: It is the angle in a right-angled triangle whose hypotenuse is the particle wavenumber while its leg is half of the lattice wavenumber. With this angle, the stationary points can be written in the form

$$x_{\text{sp}}^{(\pm)} = \pm x_* \sin \theta. \quad (10.94)$$

The solutions of the stationary equation (10.91) with the plus sign yield imaginary saddle points, and one can show that their contribution is exponentially small in the limit of large energy.

From the above analysis, it follows that the main contribution to the integral (10.89) comes from the first term in the square brackets; therefore, we may rewrite it in the form:

$$\Delta \varepsilon \simeq \pm \text{Re} \int_{-x_*}^{x_*} dx \frac{u_0}{2\pi \sqrt{x_*^2 - x^2}} \exp \left(2i\phi(x) - i \frac{2\pi}{a} x \right). \quad (10.95)$$

To evaluate this integral in the stationary phase approximation, we set $x = x_{\text{sp}}^{(\pm)}$ in the pre-exponential factor and approximate the argument of the exponential term by expanding it to second-order around the stationary points:

$$2\phi(x) - \frac{2\pi}{a} x \simeq \pm \beta \mp \frac{\sqrt{2m\varepsilon}}{\hbar x_*} \tan(\theta) \left(x - x_{\text{sp}}^{(\pm)} \right)^2, \quad (10.96)$$

where

$$\beta = 2\phi(x_{\text{sp}}^{(+)}) - \frac{2\pi}{a} \left(x_{\text{sp}}^{(+)} - \frac{\hbar k_y}{eB} \right). \quad (10.97)$$

Here we have restored the dependence on $\hbar k_y / eB$ (see comment below Eq. (10.88)).

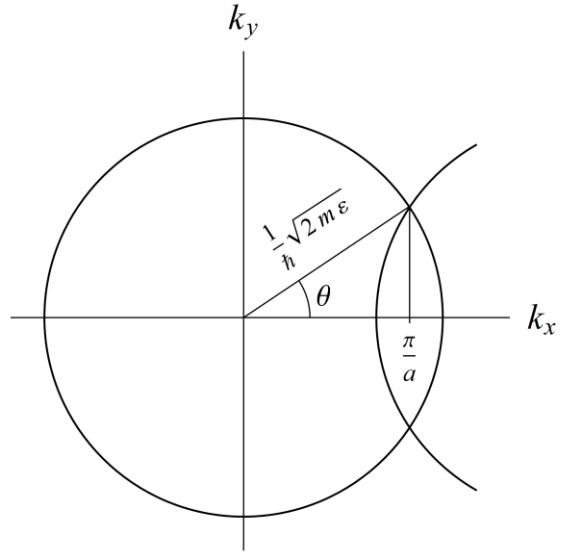


Figure 10-13 The geometrical interpretation of the stationary phase condition in Pippard's model

With these approximations, the integral (10.95) becomes a simple Gaussian integral. Collecting the contributions from both stationary points we obtain:

$$\Delta\varepsilon = \pm u_0 \sqrt{\frac{\hbar\omega_c}{\pi\varepsilon \sin(2\theta)}} \cos\left(\beta - \frac{\pi}{4}\right). \quad (10.98)$$

The energy correction, $\Delta\varepsilon$, depends on k_y through its dependence in β ; hence the condition for the validity of perturbation theory is $\max|\Delta\varepsilon| < \hbar\omega_c$ where the maximum value of $|\Delta\varepsilon|$ is obtained at values of k_y for which $|\cos(\beta - \pi/4)| = 1$. Thus, the threshold value of the magnetic field, above which the system experiences a magnetic breakdown, is obtained from the condition $\max|\Delta\varepsilon| = \hbar\omega_c$, i.e.

$$B_0 = \frac{mu_0^2}{\pi\hbar\varepsilon \sin(2\theta)}. \quad (10.99)$$

The B_0 dependence on the strength of the periodic potential and the particle energy is what one should expect: The threshold for magnetic breakdown is lowered as the periodic potential becomes weaker or as the particle energy increases. The quadratic dependence on u_0 is also expected because B_0 cannot depend on the sign of the potential. The more interesting ingredient in the above formula is the angle θ : Formula (10.99) shows that when $\theta \rightarrow 0$, $B_0 \rightarrow \infty$; namely, there is no magnetic breakdown. This strange behavior is because when $\theta = 0$ the electron wave number equals half of the lattice wave number (see Fig. 10-13); namely, the Bragg reflection condition is satisfied. Bragg reflection implies that the electron always gets the required lattice momentum needed to pass from one edge of the Brillouin zone to the opposite edge (i.e., passing to the next Brillouin zone); therefore, the magnetic breakdown is suppressed.

10.7 Strong magnetic field – Preliminary discussion (Moiré patterns)

Until now, we considered the weak magnetic field limit where the breakdown of the lattice translation symmetry is ignored. This limit is realized when the magnetic flux threading one unit cell of the lattice is much smaller than the quantum unit of magnetic flux $\phi_0 = 2\pi\hbar/e$. This regime applies to most physical systems, but it breaks down when the magnetic field is strong. Nevertheless, it turns out that if the magnetic flux threading one unit cell is a rational fraction of ϕ_0 , the lattice translation symmetry can be restored at the cost of increasing the size of the unit cell. The following section demonstrates how it is obtained by defining a new type of translation

operators called *magnetic translations*. This section will explore the implications of increasing lattice constant by an integer multiple on the band structure. We will do that with the help of a somewhat remote issue known as the *Moiré patterns*. As we shall see, increasing unit cell size by factor n leads to disintegrating of the energy band into n minibands.

Moiré patterns are large-scale patterns produced when superimposing two periodic objects with slightly different periods or at different angles of one with respect to the other. The patterns generated this way are similar to the phenomenon of beats in acoustics. Recall that superposition of two sinusoidal waves (say with the same amplitude) of nearby frequencies, ω_1 and ω_2 , such that $|\omega_1 - \omega_2| \ll \omega_1, \omega_2$ yields a modulated sinusoidal wave:

$$A \cos(\omega_1 t) + A \cos(\omega_2 t) = 2A \cos\left(\frac{\omega_1 + \omega_2}{2} t\right) \cos\left(\frac{\omega_1 - \omega_2}{2} t\right), \quad (10.100)$$

shown in Fig. 10-14. The period of the modulation, $4\pi/|\omega_1 - \omega_2|$, is much larger than the period of each one of the waves.

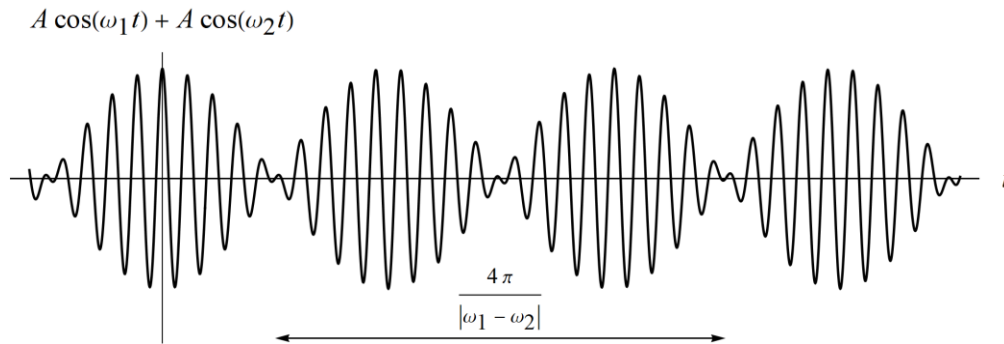


Figure 10-14 The beats phenomenon in acoustics

Similarly, a superlattice is obtained by superimposing two lattice layers with slightly different lattice constants. An example of such a system is when one layer is graphene while the other is boron nitride (BN). One of the reasons for preparing such heterostructures is to open a small (and controllable) gap at the K -points of graphene to make it a semiconductor. (In pure dichalcogenides, such as BN, the energy gap is too large, about 6 eV; hence, they are good insulators.) However, the lattice constants of graphene and BN are slightly different. In graphene, the bond length is 1.42 Å, while in BN, it is 1.444 Å - a difference of 1.7% between the lattice constants. The superlattice obtained from the superposition of two layers of these materials is periodic on a much larger scale, as illustrated in Fig. 10-15.

An alternative way of obtaining superlattices is by taking two layers of the same lattice but slightly rotating one of the layers with respect to the other.

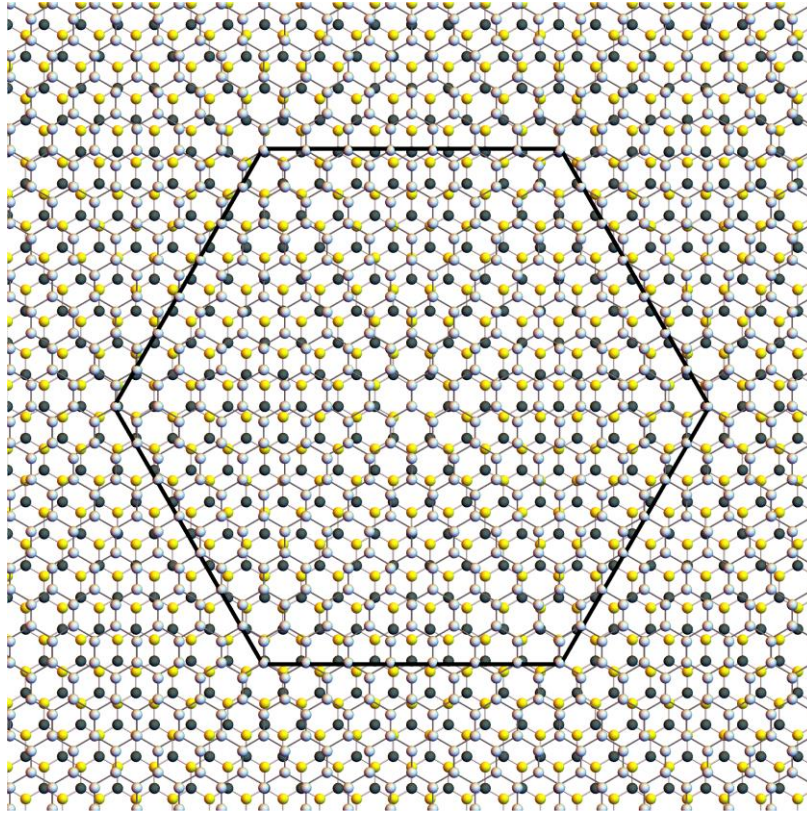


Figure 10-15 A superlattice obtained from two honeycomb lattices with slightly different lattice constants

Continuing with the example of graphene on boron nitride (BN), let $\mathbf{b}$ and $(1-\gamma)\mathbf{b}$ denote the vectors of the reciprocal lattices of graphene, and BN, respectively, where $\gamma \approx 0.017$. Also, let $u_{BN}(\mathbf{r})$ denote the potential that the BN layer generates at the point $\mathbf{r}$ of the graphene. Expanding this potential in Fourier series, we have

$$u_{BN}(\mathbf{r}) = \sum_{\mathbf{b}} u_{\mathbf{b}}^{(BN)} \exp[i(1-\gamma)\mathbf{b} \cdot \mathbf{r}], \quad (10.101)$$

where $u_{\mathbf{b}}^{(BN)}$ are the Fourier expansion coefficients. Evaluating this potential on the lattice points of the graphene layer, $\mathbf{r} = \mathbf{a}$, yields $u_{BN}(\mathbf{a}) = \sum_{\mathbf{b}} u_{\mathbf{b}}^{(BN)} \exp(-i\gamma 2\pi n)$ with an integer n . To obtain this result, we have used the relation between the vectors of the Bravais lattice and the reciprocal lattice, $\mathbf{a} \cdot \mathbf{b} = 2\pi n$. Thus the period of the potential created by the BN layer on the graphene, $u_{BN}(\mathbf{a})$, is a/γ . This period is much larger than the graphene lattice constant - of the order of $100\text{\AA}$.

Having a contribution to the potential energy with a periodicity that is γ^{-1} times larger than the original periodicity of the graphene implies that the Brillouin zone becomes smaller by a factor

of γ . Treating $u_{BN}(\mathbf{r})$ as a small perturbation, we can repeat the procedure of nearly free electrons. Namely, first, we fold the original spectrum of the graphene into the reduced Brillouin zone and then use perturbation theory to open gaps at the crossing points of the bands. The new bands obtained in this manner are called *minibands*. This procedure is schematically illustrated in Fig. 10-17 for a threefold increase of the lattice constant, $a \rightarrow 3a$, in a one-dimensional system.

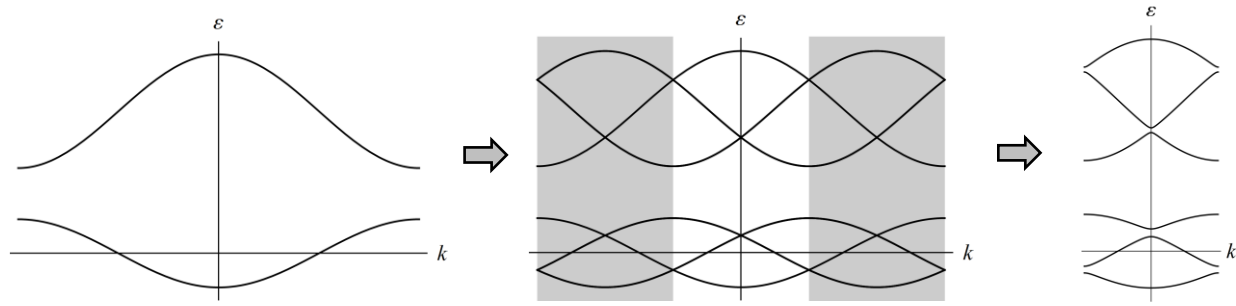


Figure 10-17 Disintegration of bands into minibands in a one-dimensional system due to a weak external potential that increases the lattice constant by a factor of 3. The left panel shows the original spectrum of the system. The middle panel shows the spectrum in the empty lattice approximation (here, the new Brillouin zone is highlighted by the bright middle stripe). The right panel is the spectrum obtained when the perturbation potential is included. This potential opens gaps in the intersection points of the energy levels shown in the middle panel.

10.8 Magnetic translations

Consider the problem of an electron moving in a two-dimensional periodic lattice subjected to a perpendicular magnetic field which is uniform in space and time-independent. The Hamiltonian of the system is

$$H = \frac{[\mathbf{p} + e\mathbf{A}(\mathbf{r})]^2}{2m} + u(\mathbf{r}), \quad (10.102)$$

and we choose to work with the symmetric gauge:

$$\mathbf{A} = \frac{1}{2} \mathbf{B} \times \mathbf{r}, \quad (10.103)$$

where $\mathbf{B}$ is constant.

It seems paradoxical that the Hamiltonian (10.102) is not invariant under the translation group of the lattice, while the physical quantities, i.e., the magnetic field and the potential, $u(\mathbf{r})$, are invariant. Namely for any lattice vector $\mathbf{a}$:

$$\mathbf{B}(\mathbf{r} + \mathbf{a}) = \mathbf{B}(\mathbf{r}) \quad \text{and} \quad u(\mathbf{r} + \mathbf{a}) = u(\mathbf{r}). \quad (10.104)$$

To reveal the cause of the problem, let us first apply the translation operator $T_{\mathbf{a}_1}$ on the Hamiltonian, where $\mathbf{a}_1$ is one of the primitive lattice vectors:

$$T_{\mathbf{a}_1} H T_{\mathbf{a}_1}^{-1} = H|_{\mathbf{r} \rightarrow \mathbf{r} + \mathbf{a}_1} = \frac{1}{2m} \left(-i\hbar \nabla + \frac{e}{2} \mathbf{B} \times \mathbf{r} + \frac{e}{2} \mathbf{B} \times \mathbf{a}_1 \right)^2 + u(\mathbf{r}). \quad (10.106)$$

The extra term that we got here is the source of the problem. However, it can be canceled out if we redefine the translation operators such that their operation includes multiplication by a phase factor:

$$\tilde{T}_{\mathbf{a}_1} = \exp \left(-i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_1) \cdot \mathbf{r} \right) T_{\mathbf{a}_1} \quad (10.107)$$

so that

$$\tilde{T}_{\mathbf{a}_1} H \tilde{T}_{\mathbf{a}_1}^{-1} = H. \quad (10.108)$$

This type of translation is called *magnetic translation* (Zack 1964, Brown, 1964). Notice that the order of the terms on the right-hand side of Eq. (10.107) is not important.

To check for consistency of the above definition of magnetic translations, let us apply them to wave functions:

$$\tilde{T}_{\mathbf{a}_1} \psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{a}_1) \exp \left(-i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_1) \cdot \mathbf{r} \right). \quad (10.109)$$

Operating on the above formula by another magnetic translation operator, $\tilde{T}_{\mathbf{a}_2}$, associated with the second primitive lattice vector, $\mathbf{a}_2$, gives

$$\tilde{T}_{\mathbf{a}_2} \tilde{T}_{\mathbf{a}_1} \psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{a}_1 + \mathbf{a}_2) \exp \left(-i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_1) \cdot (\mathbf{r} + \mathbf{a}_2) \right) \exp \left(-i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_2) \cdot \mathbf{r} \right). \quad (10.110)$$

Now, by reversing the order of the translations, we obtain

$$\tilde{T}_{\mathbf{a}_1} \tilde{T}_{\mathbf{a}_2} \psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{a}_1 + \mathbf{a}_2) \exp \left(-i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_2) \cdot (\mathbf{r} + \mathbf{a}_1) \right) \exp \left(-i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_1) \cdot \mathbf{r} \right). \quad (10.111)$$

Thus

$$\begin{aligned} \tilde{T}_{\mathbf{a}_1} \tilde{T}_{\mathbf{a}_2} &= \tilde{T}_{\mathbf{a}_2} \tilde{T}_{\mathbf{a}_1} \exp \left(-i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_1) \cdot \mathbf{a}_2 + i \frac{e}{2\hbar} (\mathbf{B} \times \mathbf{a}_2) \cdot \mathbf{a}_1 \right) \\ &= \tilde{T}_{\mathbf{a}_1} \tilde{T}_{\mathbf{a}_2} \exp \left[-i \frac{e}{\hbar} \mathbf{B} \cdot (\mathbf{a}_1 \times \mathbf{a}_2) \right] \end{aligned} \quad (10.112)$$

But $|\mathbf{a}_1 \times \mathbf{a}_2|$ is the area of one unit cell, hence $\phi = \mathbf{B} \cdot (\mathbf{a}_1 \times \mathbf{a}_2)$ is the magnetic flux threading a unit cell. Using the definition of the quantum magnetic flux, $\phi_0 = 2\pi\hbar/e$, we can rewrite the above result in the form:

$$\tilde{T}_{a_2} \tilde{T}_{a_1} = \tilde{T}_{a_1} \tilde{T}_{a_2} \exp\left[-i2\pi \frac{\phi}{\phi_0}\right], \quad (10.113)$$

or alternatively

$$\tilde{T}_{-a_2} \tilde{T}_{-a_1} \tilde{T}_{a_2} \tilde{T}_{a_1} = \exp\left[-i2\pi \frac{\phi}{\phi_0}\right]. \quad (10.114)$$

The phase factor in the last equation can be interpreted as due to the Aharonov-Bohm phase (or the Berry phase) that the particle accumulates when taken along the closed trajectory illustrated in Fig. 10-18. Thus, the naïve argument for periodicity of the system as follows by the periodicity of the magnetic field and the potential (10.104) is wrong because one must take into account also the Aharonov-Bohm phase, which is a pure quantum effect.

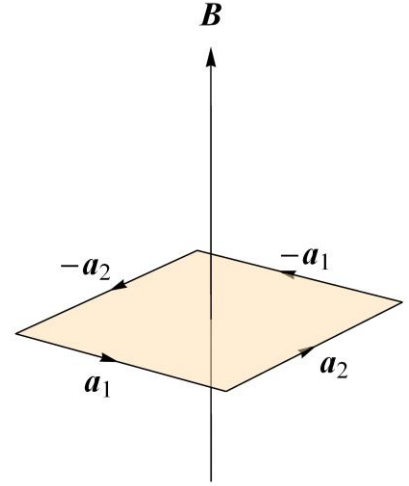


Figure 10-18 a closed trajectory along the edges of one unit cell

Notwithstanding this problem, there are situations where it does not exist. One possibility is when $\phi = p\phi_0$ where p is some integer. Namely, when the flux threading a unit cell is an integer multiple of the quantum unit flux, so that

$$\exp\left[-i2\pi \frac{\phi}{\phi_0}\right] = 1 \quad (10.115)$$

In this case, the magnetic translations commute and the translation groups with or without the magnetic field are identical.

Another possibility is when $\phi = \phi_0/q$ with $q = l^2$, where l is an integer. In this case, multiplying the unit cell by l in each direction of the primitive basis vectors, such that the new primitive lattice vectors are $l\mathbf{a}_1$ and $l\mathbf{a}_2$, produces a new lattice whose area is q times larger than the area of the original unit cell, $|l\mathbf{a}_1 \times l\mathbf{a}_2| = l^2 |\mathbf{a}_1 \times \mathbf{a}_2| = q |\mathbf{a}_1 \times \mathbf{a}_2|$. Therefore, the flux threading this new large cell is precisely ϕ_0 , and the translation group is restored, albeit for a larger lattice constant.

When $\phi = \phi_0/q$ where q is an integer but not a square of an integer, one can extend the unit cell by choosing the new primitive lattice vectors to be $q\mathbf{a}_1$ and $q\mathbf{a}_2$, so that the area of the unit cell is q^2 larger than that of the original lattice. The total magnetic flux threading the new lattice cell

is also an integer multiple of the quantum unit flux: $\mathbf{B} \cdot (q\mathbf{a}_1 \times q\mathbf{a}_2) = q^2 B |\mathbf{a}_1 \times \mathbf{a}_2| = q^2 \phi_0 / q = q\phi_0$, and translation symmetry applies for the new lattice.

The new Brillouin zone is smaller than that of the original system by factor of q^2 , and as in the case of Moiré patterns, each band will be split into minibands. However, the unit cell area is q times larger than needed to obtain a phase of 2π . It suggests that the energy levels are q times degenerate. To prove that, consider the Bloch wave function of the lattice with the extended unit cell, which satisfies the conditions:

$$\psi_k(\mathbf{r} + \mathbf{a}_2) = \exp(i\mathbf{k}\mathbf{a}_2)\psi_k(\mathbf{r}) \quad \text{and} \quad \psi_k(\mathbf{r} + q\mathbf{a}_1) = \exp(i\mathbf{k}q\mathbf{a}_1)\psi_k(\mathbf{r}) \quad (10.116)$$

This function is a valid Bloch wave function because the magnetic translation operators $\tilde{T}_{q\mathbf{a}_1}$ and $\tilde{T}_{\mathbf{a}_2}$ commute, hence the translation group they define, is identical to that obtained by the usual translation operators $T_{q\mathbf{a}_1}$ and $T_{\mathbf{a}_2}$. Now let us define $q-1$ additional functions obtained from $\psi_k(\mathbf{r})$ by magnetic translations of distances $j\mathbf{a}_1$ where $j=1, 2, \dots, q-1$:

$$\tilde{\psi}_k^{(j)}(\mathbf{r}) = \psi_k(\mathbf{r} + j\mathbf{a}_1) \exp\left(i \frac{e}{2} \mathbf{r} \cdot (\mathbf{B} \times j\mathbf{a}_1)\right). \quad (10.117)$$

These functions have the same energy as $\psi_k(\mathbf{r})$ because the Hamiltonian is invariant under magnetic translations (see Eq. (10.108)), but as we shall see, they correspond to a different quasi-momentum. Translation by $\mathbf{a}_2$ gives

$$\begin{aligned} \tilde{\psi}_k^{(j)}(\mathbf{r} + \mathbf{a}_2) &= \psi_k(\mathbf{r} + j\mathbf{a}_1) \exp\left(i \frac{e}{2} (\mathbf{r} + \mathbf{a}_2) \cdot (\mathbf{B} \times j\mathbf{a}_1)\right) \exp(i\mathbf{k}\mathbf{a}_2) \\ &= \tilde{\psi}_k^{(j)}(\mathbf{r}) \exp\left(i \left[\mathbf{k} + \frac{e}{2} (\mathbf{B} \times j\mathbf{a}_1)\right] \cdot \mathbf{a}_2\right), \end{aligned} \quad (10.118)$$

while translation by $q\mathbf{a}_1$ is

$$\begin{aligned} \tilde{\psi}_k^{(j)}(\mathbf{r} + q\mathbf{a}_1) &= \psi_k(\mathbf{r} + j\mathbf{a}_1) \exp\left(i \frac{e}{2} (\mathbf{r} + q\mathbf{a}_1) \cdot (\mathbf{B} \times j\mathbf{a}_1)\right) \exp(i\mathbf{k}q\mathbf{a}_1) \\ &= \tilde{\psi}_k^{(j)}(\mathbf{r}) \exp\left(i \left[\mathbf{k} + \frac{e}{2} (\mathbf{B} \times j\mathbf{a}_1)\right] \cdot q\mathbf{a}_1\right). \end{aligned} \quad (10.119)$$

From here it follows that $\tilde{\psi}_k^{(j)}(\mathbf{r})$ Bloch functions whose quasi-momentum, is $\hbar\mathbf{k} + \frac{\hbar e}{2} (\mathbf{B} \times j\mathbf{a}_1)$, hence

$$\varepsilon[\mathbf{k}] = \varepsilon\left[\mathbf{k} + \frac{e}{2} (\mathbf{B} \times j\mathbf{a}_1)\right]. \quad (10.120)$$

Thus, the energy levels are q -fold degenerate.

10.9 Hofstadter's butterfly

Imagine an electron in a two-dimensional lattice subjected to a uniform magnetic field (perpendicular to the sample). Let us observe the behavior of the electronic spectrum when increasing the magnetic field. From the above discussion it flows that this spectrum undergoes a series of changes: Each time the magnetic flux, threading a unit cell, reaches a value that equals $\phi = p\phi_0/q$ (where q and p are coprime integers) the band disintegrates into q minibands. However, since the rational numbers are dense, one expects that the dependence of the spectrum on the magnetic field is fractal. In 1976 Hofstadter solved, numerically, the tight-binding model in a square lattice subjected to a uniform magnetic field. He plotted the energy levels (divided by the hopping matrix element t) as a function of the magnetic flux passing through a unit cell divided by quantum unit flux; see Fig. 10-19. In the picture's leftmost and rightmost sides, corresponding to $\phi = 0$ and $\phi = \phi_0$, the spectrum is identical and contains a single band. Between these sides, one can see a self-similar behavior of the spectrum of the Landau levels. For instance, at $\phi = \phi_0/3$ the spectrum contains three minibands and the local behavior near this value is similar to the behavior near $\phi = 0$.

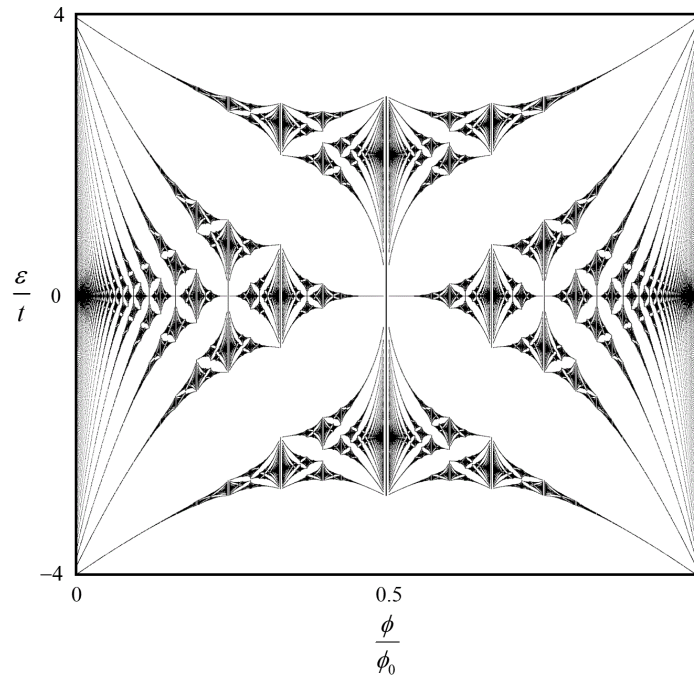


Figure 10-19 Hofstadter's butterfly

10.10 Exercises

1. Prove equation (10.23).
2. Show that in the framework of the tight-binding model, the hopping terms, $\tilde{t}$, to nearest neighbors in the presence of a weak magnetic field are:

$$\tilde{t} = t \exp \left(i \frac{q}{\hbar} \int_a^{a'} d\mathbf{r}' \cdot \mathbf{A}(\mathbf{r}') \right), \quad (10.121)$$

Where, t , is the hopping term without magnetic field, q is the particle charge, $\mathbf{a}$ and $\mathbf{a}'$ are neighboring lattice points and $\mathbf{A}(\mathbf{r}')$ is the vector potential.

Advice: Show that Wannier functions in the presence of the magnetic field, $\tilde{w}_a(\mathbf{r})$, are obtained from Wannier functions in the absence of magnetic field, $w_a(\mathbf{r})$, by

$$\tilde{w}_a(\mathbf{r}) = w_a(\mathbf{r}) \exp \left(i \frac{q}{\hbar} \int_a^{\mathbf{r}} d\mathbf{r}' \cdot \mathbf{A}(\mathbf{r}') \right), \quad (10.122)$$

and satisfy

$$\tilde{H} \tilde{w}_a(\mathbf{r}) = \exp \left(i \frac{q}{\hbar} \int_a^{\mathbf{r}} d\mathbf{r}' \cdot \mathbf{A}(\mathbf{r}') \right) H w_a(\mathbf{r}), \quad (10.123)$$

where $\tilde{H}$ and H are, respectively, the Hamiltonians of a particle moving in a periodic lattice in the presence or in the absence of a magnetic field.

3. Calculate Landau levels of graphene in a magnetic field (neglecting Zeeman effect)

Advice: Focus on the region of K -point and use Peierls substitution (10.31) to obtain the effective Hamiltonian within the effective mass approximation. Write down the time-independent Schrödinger equation as two coupled equations for the components of the dynamical momentum. Substitute one equation in the other and use the commutation relations (10.69).

4. Prove equations (10.98) and (10.99).

5. Calculate the transition coefficient in the problem of magnetic breakdown for the following model. Consider the point $\mathbf{k}_0$ in the $\mathbf{k}$ space, where two energy bands become close, as illustrated in Fig. 10-20. In the $\mathbf{k} \cdot \mathbf{p}$ approximation, the local Hamiltonian in the vicinity of this point is

$$H = \begin{pmatrix} \Delta + \hbar v_y k_y & \hbar v_x k_x \\ \hbar v_x k_x & -\Delta + \hbar v_y k_y \end{pmatrix}, \quad (10.124)$$

where $\mathbf{k}$ is measured from $\mathbf{k}_0$. Here 2Δ is the minimal energy gap between the levels (at $\mathbf{k} = 0$), v_x and v_y are parameters of the system, and without loss of generality, we set the Fermi energy to be zero. Now, assume the magnetic field is applied in the z axis direction and calculate the transition coefficient between the upper and the lower bands of the spectrum shown in the figure.

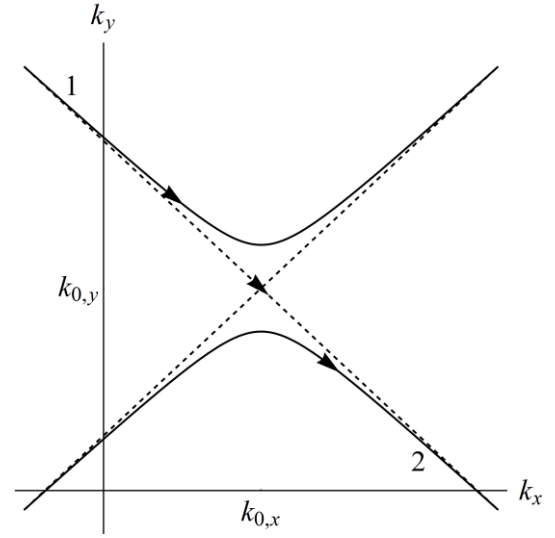


Figure 10-20 A model for magnetic breakdown

Advice: Chose the Landau gauge $A_y = Bx$ and $A_x = A_z = 0$, and reduce the problem to that of dielectric breakdown model discussed in the previous chapter.

6. Twisted bilayer graphene is obtained when superimposing two graphene layers with one of the layers twisted by a small angle α with respect to the other. Assuming a is the lattice constant of graphene, what is the lattice constant of the superlattice?

11 Elastic deformations, sound waves, and phonons

Until now, we have studied electronic properties of crystals, assuming them to have perfect periodic structures. However, there is no such thing as a perfect crystal (even for the mere reason that any crystal is finite in size). In reality, atoms are shifted from their equilibrium position, and the periodic structure of the crystal is destroyed. There are many reasons for that; For instance, defects and impurities in the crystal, thermal and quantum fluctuations, external forces acting on the crystal, and sound waves propagating within the crystal. The distortion created by these factors affects the behavior of the electrons in the system because Bloch's theorem does not apply anymore in its strict sense. In our quest to understand the physical properties of crystals, an important step is to clarify the nature of elastic deformations in crystals and sound waves that represent time-dependent deformations. We begin this chapter by developing the mathematical tools that describe elastic deformations in crystals. Next, we discuss the energy of such deformations and use it to derive the equations for sound propagation in a crystal. Then, we will introduce phonons obtained from quantizing these sound waves, and finally, discuss optical phonons that appear in crystals whose unit cells contain more than one atom.

11.1 The strain tensor

Lattice deformations are states of the lattice in which the atoms move from their equilibrium position, as illustrated in Fig. 11-1. In this figure, the black disks represent the positions of the atoms at equilibrium, while the gray disks are the locations of the atoms in the deformed lattice. The vector $\mathbf{u}(\mathbf{r})$ describes the shift of an atom from its equilibrium position at $\mathbf{r}$ into its new position at $\mathbf{r} + \mathbf{u}(\mathbf{r})$. It is called the *displacement vector*.

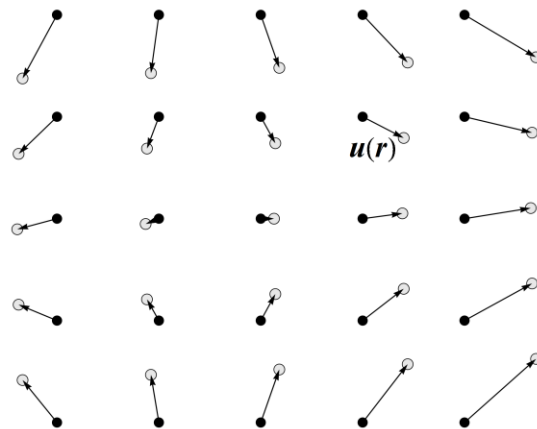


Figure 11-1 Deformation of the lattice structure as atoms shift from their equilibrium positions

Here we assume that the lattice deformations change very slowly in space, i.e.,

$$\left| \frac{\partial u_i(\mathbf{r})}{\partial r_j} \right| \ll 1. \quad (11.1)$$

This assumption allows us to ignore the discrete structure of the lattice and treat $\mathbf{r}$ as a continuous variable. Notice, however, that the above condition does not imply that the magnitude of the displacement vector, $|\mathbf{u}(\mathbf{r})|$, is smaller than the lattice constant. In fact, it can be much larger than the lattice constant because small changes in the distances between nearby atoms may accumulate to a large displacement vector.

The object that we need in order to understand the elastic properties of a crystal is its energy dependence on deformations. This energy comprises two main ingredients: the electrostatic repulsion of the ions and the electronic energy that compensates for the repulsion energy and a bit more so that all atoms are held together. The electron energy is traditionally calculated in the Born-Oppenheimer approximation (slow ions and fast electrons). Here the Schrödinger equation for the electrons is solved for a given static configuration of the ions, and the obtained eigenenergy serves as the potential energy for the motion of the ions. In particular, the global minimum of this potential energy determines the crystal structure.

Thus, to calculate the electronic contribution to the energy of deformations, one needs to consider the potential energy of the electrons in the deformed lattice:

$$V(\mathbf{r}) = \sum_j v_a[\mathbf{r} - \mathbf{R}_j - \mathbf{u}(\mathbf{R}_j)]. \quad (11.2)$$

Here the sum is over all ions of the lattice that, in equilibrium, are located at points $\mathbf{R}_j$, while $v_a(\mathbf{r})$ is the potential energy created by a single ion sitting at the origin. In general, crystals are made from several different types of ions that create different potential energies; however, this complication does not affect the results we will derive below.

In principle, one would like to expand the potential energy in the displacement vectors, as

$$v_a[\mathbf{r} - \mathbf{R} - \mathbf{u}(\mathbf{R})] \stackrel{?}{\approx} v_a(\mathbf{r} - \mathbf{R}) - \frac{\partial v_a(\mathbf{r} - \mathbf{R})}{\partial r_j} u_j(\mathbf{R}). \quad (11.3)$$

However, $v_a(\mathbf{r})$ typically changes over atomic scales, while $|\mathbf{u}(\mathbf{r})|$ may be much larger than the lattice constant. Thus, we cannot use the above expansion. To avoid this problem, let us define a new coordinate system that follows the ion's displacements. Namely, a new coordinate, $\tilde{\mathbf{r}}$, that satisfies the condition:

$$\tilde{\mathbf{r}} + \mathbf{u}(\tilde{\mathbf{r}}) = \mathbf{r}. \quad (11.4)$$

Taking into account that $v(\mathbf{r})$ is local and changes a scale of order of the lattice constant (i.e. approximately proportional to a δ -function) and the assumption expressed in Eq. (11.1), we see that in this coordinate system the potential energy is periodic in space with the original lattice periodicity:

$$V(\mathbf{r}) = \sum_j v[\mathbf{r} - \mathbf{R}_j - \mathbf{u}(\mathbf{R}_j)] = \sum_j v[\tilde{\mathbf{r}} + \mathbf{u}(\tilde{\mathbf{r}}) - \mathbf{R}_j - \mathbf{u}(\mathbf{R}_j)] = \sum_j v(\tilde{\mathbf{r}} - \mathbf{R}_j) \equiv V(\tilde{\mathbf{r}}). \quad (11.5)$$

However, the new coordinate system is not cartesian anymore. Namely, the distance scale between points depends on their position in space. Thus, in order to work with these coordinates, one should use the *metric tensor* g_{jk} that allows one to calculate distances in the new coordinate system. If ds represents an infinitesimal distance in the physical plane, then the new coordinate system satisfies the relation $ds^2 = g_{ij} d\tilde{r}_i d\tilde{r}_j$, where repeated indices are summed over. In our case:

$$\begin{aligned} ds^2 = d\mathbf{r} \cdot d\mathbf{r} &= \underbrace{\frac{dr_i}{d\tilde{r}_j} \frac{dr_i}{d\tilde{r}_k}}_{g_{jk}} d\tilde{r}_j d\tilde{r}_k = \left(\delta_{ij} + \frac{\partial u_i}{\partial \tilde{r}_j} \right) \left(\delta_{ik} + \frac{\partial u_i}{\partial \tilde{r}_k} \right) d\tilde{r}_j d\tilde{r}_k \\ &= (\delta_{jk} + 2u_{jk}) d\tilde{r}_j d\tilde{r}_k, \end{aligned} \quad (11.6)$$

where

$$u_{jk} = \frac{1}{2} \left[\frac{\partial u_k}{\partial \tilde{r}_j} + \frac{\partial u_j}{\partial \tilde{r}_k} + \frac{\partial u_i}{\partial \tilde{r}_j} \frac{\partial u_i}{\partial \tilde{r}_k} \right] \approx \frac{1}{2} \left[\frac{\partial u_k}{\partial \tilde{r}_j} + \frac{\partial u_j}{\partial \tilde{r}_k} \right] \quad (11.7)$$

is the *strain tensor*. Notice that the nonlinear term can be neglected using the assumption of slow changes in the displacement vector (11.1).

A non-trivial metric enforces changes in the form of the Laplacian operator. Using the chain rule for derivatives, one can show that the transformed Laplacian is:

$$\nabla^2 \rightarrow \frac{1}{\sqrt{g}} \frac{\partial}{\partial \tilde{r}_i} \sqrt{g} g^{ij} \frac{\partial}{\partial \tilde{r}_j}, \quad (11.8)$$

where $g^{ij} = g_{ij}^{-1}$ is the inverse of the metric matrix, while $g = \det g_{ij}$ is the determinant of the metric tensor. This operator is called the *Laplace-Beltrami operator*. We will not prove this formula but demonstrate it with the simple example of polar coordinates:

$$r = \sqrt{x^2 + y^2} \quad \text{and} \quad \varphi = \arctan\left(\frac{y}{x}\right). \quad (11.9)$$

The metric associated with these coordinates can be obtained by the following calculation:

$$dx^2 + dy^2 = dr^2 + r^2 d\varphi^2 = \underbrace{\begin{pmatrix} dr & d\varphi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \begin{pmatrix} dr \\ d\varphi \end{pmatrix}}_{g_{ij}} \quad (11.20)$$

Therefore $g = r^2$ and the inverse matrix of the metric tensor is

$$g^{ij} = g_{ij}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{r^2} \end{pmatrix}_{ij} \quad (11.21)$$

Substituting these results in the Laplace-Beltrami operator (11.8), we obtain

$$\frac{1}{\sqrt{g}} \frac{\partial}{\partial \tilde{r}_i} \sqrt{g} g^{ij} \frac{\partial}{\partial \tilde{r}_j} = \frac{1}{r} \left(\frac{\partial}{\partial r} r \frac{\partial}{\partial r} \right) + \frac{1}{r} \left(\frac{\partial}{\partial \varphi} r \frac{1}{r^2} \frac{\partial}{\partial \varphi} \right) = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \varphi^2}, \quad (11.22)$$

which is the familiar Laplacian in polar coordinates.

Let us return to the problem of finding the electronic contribution to the energy of a deformed lattice. Having the form of the Laplacian (11.8) and the potential energy (11.5) in the coordinate system that moves with the atoms, we can write the Schrödinger equation for an electron moving in the deformed lattice in the form

$$\left[-\frac{\hbar^2}{2m} \frac{1}{\sqrt{g}} \frac{\partial}{\partial \tilde{r}_i} \sqrt{g} g^{ij} \frac{\partial}{\partial \tilde{r}_j} + V(\tilde{r}) \right] \psi = \varepsilon \psi. \quad (11.23)$$

This equation is the starting point for a perturbative expansion of the electronic energy in the strain tensor (11.7), because the deviation of the metric $g_{jk} = \delta_{jk} + 2u_{jk}$ from the Cartesian metric $g_{jk}^{(0)} = \delta_{jk}$ is small (due to our assumption (11.1) that deformations are very smooth in space). Similar considerations also apply to the total energy of the system, which includes that of the ions.

11.2 The energy of elastic deformations

The energy of a deformed crystal is a functional of the strain tensor:

$$E_{\text{elastic}}[u_{ij}(\mathbf{r})] = \int d^d r \varepsilon_{\text{elastic}}[u_{ij}(\mathbf{r})], \quad (11.24)$$

where $\varepsilon_{\text{elastic}}[u_{ij}(\mathbf{r})]$ is the energy density for a local deformation of the lattice. It is clear that calculating this energy, from first principles, even in the framework of perturbation theory in the strain tensor, is a formidable task. However, assuming that deformations do not produce

dramatic changes in the band structures of the electrons, the above considerations suggest that the energy density can be expanded in powers of $u_{ij}(\mathbf{r})$

$$\mathcal{E}_{\text{elastic}}(u_{ij}) = \frac{1}{2} \Xi_{ij;kl} u_{ij} u_{kl} + \dots, \quad (11.25)$$

where, as usual, repeated indices should be summed over. Elasticity theory is obtained when this expansion terminates at the second order. Notice that the linear term does not appear in this expansion because we are interested only in small deviations from the minimal energy of the equilibrium configuration.

In the above formula, $\Xi_{ij;kl}$ is the *elastic modulus tensor*. In general, a tensor with four indices in a three-dimensional space contains $3^4 = 81$ elements. However, this number is reduced to 21 independent elements due to the following properties: First, it is symmetric to index changes of $i \leftrightarrow j$ and to $k \leftrightarrow l$ (because the strain tensor is a symmetric matrix). Second, it is symmetric to an interchange of pairs of indices $ij \leftrightarrow kl$. Spatial symmetries of the lattice are expected to reduce the number of independent parameters of the elastic modulus tensor even more, as we shall see now.

Let us show how to employ group theory considerations to identify the number of independent parameters of $\Xi_{ij;kl}$. The starting point is our understanding that the elastic energy (11.25) is invariant under all symmetry operations of the lattice. Therefore, the right-hand side of Eq. (11.25) must belong to the identity representation. Thus, one should identify the irreducible representations of the strain tensor u_{ij} , and from their direct products (because the energy is quadratic in the strain tensor) select the identity representation. The number of times that the identity representation appears in these products is the number of independent parameters of $\Xi_{ij;kl}$.

To identify the irreducible representation associated with the strain tensor

$$u_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial r_j} + \frac{\partial u_j}{\partial r_i} \right), \quad (11.26)$$

notice that its symmetry is the same as the basis function $r_i r_j$ because both u_i and $\partial/\partial r_j$ behave as vector's components: u_i is a component of the displacement vector, while $\partial/\partial r_j$, is a component of the gradient operator which behaves the same as a vector concerning rotations, reflection, and inversion.

Example 1: The elastic energy of a two-dimensional lattice with C_{3v} symmetry

Consider the in-plane deformations of a two-dimensional lattice with C_{3v} symmetry (such as boron nitride). The quadratic basis functions of the irreducible representations of this group (associated only with the in-plane coordinates) are shown in the table to the right. This table shows that $u_{xx} + u_{yy}$ belongs to the identity representation, while $(u_{xx} - u_{yy}, 2u_{xy})$

C_{3v}	
A_1	$x^2 + y^2$
A_2	
E	$(x^2 - y^2, 2xy)$

is a basis function of the two-dimensional irreducible representation. For the first case, it is clear that $A_1 \otimes A_1 = A_1$; hence the elastic energy must contain a term proportional to $(u_{xx} + u_{yy})^2$. For the second case, the product $E \otimes E = A_1 \oplus A_2 \oplus E$ also contains the identity representation. The singlet associated with the latter product is the square of the norm of the basis functions. (one can see that the square of this norm $(x^2 - y^2)^2 + (2xy)^2 = (x^2 + y^2)^2$ is the same as the square of the A_1 basis function). Thus, one expects the elastic energy to have an additional term that is proportional to $(u_{xx} - u_{yy})^2 + 4u_{xy}^2$; hence the elastic energy of this lattice is

$$\varepsilon_{\text{elastic}} = \frac{1}{2} \Xi_{A_1} (u_{xx} + u_{yy})^2 + \frac{1}{2} \Xi_E [(u_{xx} - u_{yy})^2 + 4u_{xy}^2], \quad (11.27)$$

where Ξ_{A_1} and Ξ_E are system-dependent parameters. These considerations show that $\Xi_{ij;kl}$ depends only on two parameters in the case of a two-dimensional lattice with C_{3v} symmetry.

Comment: To identify the basis functions associated with the identity representation of products of higher dimensional representations (such as $E \otimes E$ in the above example), one should work with the correct normalization of the basis functions. That is a normalization for which unitary operators describe the actions of the symmetry operations. For this choice, the sum of the square of the components of the basis functions is a singlet, because it is invariant under all symmetry operations represented by unitary transformations. The factor 2 that appears in the second term of the basis function $(x^2 - y^2, 2xy)$ ensures this property; however, many character tables do not use this convention. Ex. 7 of chapter 4 explains how to build basis functions that possess this property.

Opening the brackets in the second term of the elastic energy (11.27), and rearranging the terms one obtains:

$$\begin{aligned}
\varepsilon_{\text{elastic}} &= \frac{1}{2} \Xi_{A_1} (u_{xx} + u_{yy})^2 + \frac{1}{2} \Xi_E [u_{xx}^2 + u_{yy}^2 + 4u_{xy}^2 - 2u_{xx}u_{yy}] \\
&= \frac{1}{2} \Xi_{A_1} (u_{xx} + u_{yy})^2 + \frac{1}{2} \Xi_E [2u_{xx}^2 + 2u_{yy}^2 + 4u_{xy}^2 - (u_{xx} + u_{yy})^2].
\end{aligned} \tag{11.28}$$

This elastic energy can also be written in terms of traces of the strain tensor and its square:

$$\varepsilon_{\text{elastic}} = \frac{1}{2} (\Xi_{A_1} - \Xi_E) (\text{tr } \hat{u})^2 + \Xi_E \text{tr } \hat{u}^2, \tag{11.29}$$

where $\hat{u}$ stands for the strain tensor matrix. Notice that writing the energy density only in terms of $\text{tr } \hat{u}$ and $\text{tr } \hat{u}^2$ implies that the elastic energy of this system is rotationally symmetric.

The first term on the right-hand side of the above formula is the energy associated with compression or decompression of the lattice, i.e., the energy associated with volume changing and shape-preserving deformation. It is because $\text{tr } \hat{u}$ is the relative change in the system volume due to deformation: To see why, recall the Jacobian, $\sqrt{g}$, of the transformation from $\mathbf{r}$ to $\tilde{\mathbf{r}}$, is the ratio of volume elements in both spaces. Therefore, the relative change in the volume can be evaluated as follows:

$$\begin{aligned}
\frac{V_{\text{deformed}} - V_{\text{equilibrium}}}{V_{\text{equilibrium}}} &= \sqrt{g} - 1 = \sqrt{\det(I + 2\hat{u})} - 1 = \exp\left[\frac{1}{2} \ln \det(I + 2\hat{u})\right] - 1 \\
&= \exp\left[\frac{1}{2} \text{tr} \ln(I + 2\hat{u})\right] - 1 \approx \exp\left[\frac{1}{2} \text{tr } 2\hat{u}\right] - 1 \approx \text{tr } \hat{u},
\end{aligned} \tag{11.30}$$

where we have used the identity $\ln \det A = \text{tr} \ln A$ which holds for any diagonalizable matrix A .

The second contribution to the energy density (11.29) is due to shear deformations.

Example 2: The elastic energy of a crystal with tetrahedral symmetry

The quadratic basis functions of the irreducible representation of the tetrahedral group, T_d , are presented in the table to the right. The basis functions of the E and F_2 irreducible representations were calculated in Ex. 7 of chapter 4. Here, the direct product of each irreducible representation by itself contains the identity representation. However, one should also check whether a direct product of two different irreducible representations contains the identity

T_d	
A_1	$x^2 + y^2 + z^2$
A_2	
E	$[2z^2 - x^2 - y^2, \sqrt{3}(y^2 - x^2)]$
F_1	
F_2	(xy, xz, yz)

representation. It is not the case here because $F_2 \otimes E = F_1 \oplus F_2$. Using the same procedure presented in the previous example leads to

$$\begin{aligned} \varepsilon_{\text{elastic}} = & \frac{1}{2} \Xi_{A_1} (u_{xx} + u_{yy} + u_{zz})^2 + \frac{1}{2} \Xi_{F_2} (u_{xy}^2 + u_{xz}^2 + u_{yz}^2) \\ & + \frac{1}{2} \Xi_E \left[(2u_{zz} - u_{xx} - u_{yy})^2 + 3(u_{yy} - u_{xx})^2 \right], \end{aligned} \quad (11.31)$$

Thus, the elastic modulus tensor contains only three system-dependent parameters, Ξ_{A_1} , Ξ_{F_2} and Ξ_E . Rearranging the terms in this formula allows to write the elastic energy density in the form:

$$\varepsilon_{\text{elastic}} = \underbrace{\frac{1}{2} (\Xi_{A_1} - 2\Xi_E) (u_{ii})^2 + \frac{\Xi_{F_2}}{4} u_{ij} u_{ji}}_{\text{spherical symmetry}} + \frac{1}{2} \left(6\Xi_E - \frac{\Xi_{F_2}}{2} \right) \underbrace{(u_{xx}^2 + u_{yy}^2 + u_{zz}^2)}_{\text{cubic symmetry}}. \quad (11.32)$$

The first two terms in this formula are expressed in terms of $\text{tr } \hat{u}$ and $\text{tr } \hat{u}^2$, therefore, represent a rotationally symmetric contribution. The third term reflects the symmetry of a cube.

11.3 Sound waves in crystals

Sound waves in crystals are time-dependent smooth deformations. They come from the periodic conversion of potential energy into kinetic energy and vice versa. Thus, apart from the potential energy of the deformations one should include the kinetic energy of the atoms. The Hamiltonian that describes such a system is:

$$H = \int d^d r \left[\frac{\mathbf{p}^2(\mathbf{r})}{2\rho} + \frac{1}{2} \Xi_{ij;kl} u_{ij}(\mathbf{r}) u_{kl}(\mathbf{r}) \right] \quad (11.33)$$

where ρ is the mass density and $\mathbf{p}(\mathbf{r})$ is the momentum density that satisfies the commutation relation:

$$[p_n(\mathbf{r}); u_m(\mathbf{r}')] = -i\hbar \delta_{nm} \delta(\mathbf{r} - \mathbf{r}') \quad (11.34)$$

(recall that $\mathbf{u}(\mathbf{r})$ is the displacement vector at the point $\mathbf{r}$.)

We turn now to derive the equations of motion that follow from Hamiltonian (11.33). The time derivative of the displacement vector is obtained from its commutation with the Hamiltonian. Using Eq. (11.34) we obtain:

$$\begin{aligned}
\frac{du_m}{dt} &= \frac{i}{\hbar} [H; u_m(\mathbf{r})] = \int d^d r' \frac{p_n(\mathbf{r}')}{\rho} \frac{i}{\hbar} [p_n(\mathbf{r}'); u_m(\mathbf{r})] \\
&= \int d^d r' \frac{p_n(\mathbf{r}')}{\rho} \delta_{mn} \delta(\mathbf{r} - \mathbf{r}') = \frac{p_m(\mathbf{r})}{\rho}.
\end{aligned} \tag{11.35}$$

A similar calculation for the time derivative of the momentum density gives

$$\begin{aligned}
\frac{dp_m}{dt} &= \frac{i}{\hbar} [H; p_m(\mathbf{r})] = -\frac{i}{\hbar} \Xi_{ij;kl} \int d^d r' u_{ij}(\mathbf{r}') [p_m(\mathbf{r}); u_{kl}(\mathbf{r}')] \\
&= -\frac{i}{2\hbar} \Xi_{ij;kl} \int d^d r' u_{ij}(\mathbf{r}') \left\{ \frac{\partial}{\partial r'_k} [p_m(\mathbf{r}); u_l(\mathbf{r}')] + \frac{\partial}{\partial r'_l} [p_m(\mathbf{r}); u_k(\mathbf{r}')] \right\} \\
&= -\frac{1}{2} \Xi_{ij;kl} \int d^d r' u_{ij}(\mathbf{r}') \left\{ \delta_{ml} \frac{\partial}{\partial r'_k} \delta(\mathbf{r} - \mathbf{r}') + \delta_{mk} \frac{\partial}{\partial r'_l} \delta(\mathbf{r} - \mathbf{r}') \right\} \\
&= \frac{1}{2} \Xi_{ij;kl} \int d^d r' \left\{ \delta_{ml} \frac{\partial u_{ij}(\mathbf{r}')}{\partial r'_k} + \delta_{mk} \frac{\partial u_{ij}(\mathbf{r}')}{\partial r'_l} \right\} \delta(\mathbf{r} - \mathbf{r}') \\
&= \frac{\partial}{\partial r_k} \Xi_{ij;km} u_{ij}(\mathbf{r}).
\end{aligned} \tag{11.36}$$

Defining the *stress tensor* to be

$$\sigma_{km} = \Xi_{ij;km} u_{ij}(\mathbf{r}) = \frac{\delta H}{\delta u_{km}(\mathbf{r})} \tag{11.37}$$

we can write equation (11.36) in the form of a continuity equation:

$$\frac{dp_m}{dt} - \frac{\partial}{\partial r_k} \sigma_{km} = 0 \tag{11.38}$$

showing that the stress tensor is the momentum flux density.

Taking the time derivative of Eq. (11.35) and substituting Eq. (11.36), we obtain the wave equation:

$$\frac{d^2 u_m}{dt^2} = \frac{1}{\rho} \frac{\partial}{\partial r_n} \Xi_{ij;mn} u_{ij}(\mathbf{r}). \tag{11.39}$$

To solve this equation, we substitute a solution in the form of a traveling wave:

$$u_m = U e_m \exp(i\mathbf{q} \cdot \mathbf{r} - i\omega t), \tag{11.40}$$

where U is the wave amplitude, $\mathbf{e}$ is a unit vector that defines the direction of the displacement vector, $\mathbf{q}$ is the wavenumber, and ω is the frequency. With this choice, the strain tensor (11.7) takes the form

$$u_{ij} = U \frac{i}{2} (e_i q_j + e_j q_i) \exp(i\mathbf{q} \cdot \mathbf{r} - i\omega t). \quad (11.41)$$

Substituting (11.41) and (11.42) in the wave equation (11.40), we obtain the dispersion relation of the sound wave

$$\omega^2 e_m = \frac{1}{2\rho} \Xi_{ij;mn} (e_i q_j + e_j q_i) q_n. \quad (11.42)$$

To identify the sound velocity,

$$c(\hat{\mathbf{q}}) = \frac{\omega}{q}, \quad (11.43)$$

of a wave moving direction of the unit vector $\hat{\mathbf{q}} = \mathbf{q}/q$, we use the symmetry of $\Xi_{ij;mn}$ to the index change $i \leftrightarrow j$ to obtain

$$c^2(\hat{\mathbf{q}}) e_m = \frac{1}{\rho} \Xi_{ij;mn} \hat{q}_j \hat{q}_n e_i. \quad (11.44)$$

The above equation is an eigenvalue equation for the 3×3 matrix

$$M_{mi} = \frac{1}{\rho} \Xi_{ij;mn} \hat{q}_j \hat{q}_n. \quad (11.45)$$

Its three eigenvalues are the squares of the sound velocities, $c_\alpha^2(\hat{\mathbf{q}})$ ($\alpha=1,2,3$), for a given direction of the wave propagation, $\hat{\mathbf{q}}$. The eigenvectors of M_{mi} are the normal modes of the vibrations, $\mathbf{e}^{(\alpha)}(\hat{\mathbf{q}})$ (which also depend on the propagation direction of the wave). These modes are orthogonal

$$\mathbf{e}^{(\alpha)} \cdot \mathbf{e}^{(\beta)} = \sum_{j=1}^3 e_j^{(\alpha)} e_j^{(\beta)} = \delta_{\alpha\beta}, \quad (11.46)$$

and form a complete basis

$$\sum_{\alpha=1}^3 e_i^{(\alpha)} e_j^{(\alpha)} = \delta_{ij}. \quad (11.47)$$

The role of the unit vector $\mathbf{e}$ is similar to polarization in electromagnetic waves; therefore, it is called the *wave polarization vector*.

Example 1: The spherical approximation of sound waves in a lattice

Consider the case where the elastic modulus tensor is rotationally invariant. Such a tensor has the form

$$\Xi_{ij;kl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{kj}), \quad (11.48)$$

where λ and μ are constants called *Lamé parameters*. This tensor satisfies the required index symmetries, and the elastic energy obtained from this tensor (11.25) is expressed only in term of traces of the strain tensor and its square:

$$\begin{aligned} \varepsilon_{\text{elastic}} &= \frac{1}{2} \Xi_{ij;kl} u_{ij} u_{kl} = \frac{1}{2} [\lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{kj})] u_{ij} u_{kl} \\ &= \frac{\lambda}{2} u_{ii} u_{kk} + \frac{\mu}{2} (u_{kj} u_{kj} + u_{ij} u_{ji}) = \frac{\lambda}{2} (\text{tr} \hat{u})^2 + \mu \text{tr} \hat{u}^2 \end{aligned} \quad (11.49)$$

This form of the elastic modulus tensor provides an approximation for zinc blende crystals when the third term in Eq. (11.32) is negligible. Substituting it in Eq. (11.44), we obtain:

$$\begin{aligned} c^2(\hat{\mathbf{q}}) e_m &= \frac{1}{\rho} \Xi_{ij;mn} \hat{q}_j \hat{q}_n e_i = \frac{\lambda}{\rho} \delta_{ij} \delta_{mn} \hat{q}_j \hat{q}_n e_i + \frac{\mu}{\rho} (\delta_{im} \delta_{jn} + \delta_{in} \delta_{jm}) \hat{q}_j \hat{q}_n e_i \\ &= \frac{\lambda}{\rho} (\hat{\mathbf{q}} \cdot \mathbf{e}) \hat{q}_m + \frac{\mu}{\rho} (e_m + (\hat{\mathbf{q}} \cdot \mathbf{e}) \hat{q}_m), \end{aligned} \quad (11.50)$$

or in the vector form:

$$c^2(\hat{\mathbf{q}}) \mathbf{e} = \frac{\lambda + \mu}{\rho} (\hat{\mathbf{q}} \cdot \mathbf{e}) \hat{\mathbf{q}} + \frac{\mu}{\rho} \mathbf{e}. \quad (11.51)$$

Now we can identify two cases: Longitudinal waves where the wave propagates in the same direction of the displacement vector, $\hat{\mathbf{q}} \cdot \mathbf{e} = 1$, and two degenerate modes of transverse waves where the displacement vector is perpendicular to the propagation direction, $\hat{\mathbf{q}} \cdot \mathbf{e} = 0$. From the above equation, we obtain that for longitudinal waves,

$$c_{\parallel} = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad (11.52)$$

while for transverse waves

$$c_{\perp} = \sqrt{\frac{\mu}{\rho}}. \quad (11.53)$$

Thus, transverse waves propagate at a lower velocity than longitudinal waves. Typically, the ratio of the sound velocity of the transverse waves to that of the longitudinal wave is small:

$$c_{\perp}/c_{\parallel} \sim 10^{-2} \div 10^{-1}.$$

The dispersion of sound waves in this spherical approximation is illustrated in Fig. 11-2.

We remark that this figure only describes the depression of sound waves at small wave numbers, namely wave numbers that are much smaller than the lattice wave number. It is because our derivation is based on the assumption of slow and smooth deformations in space.

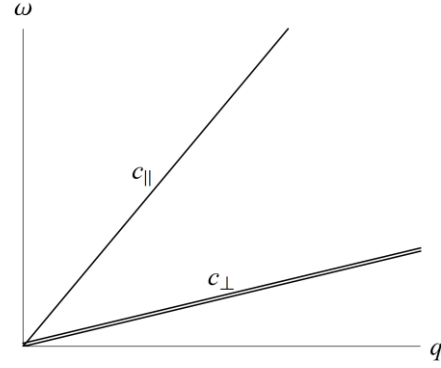


Figure 11-2 The dispersion relation of sound waves in the spherical approximation

Example 2: Sound waves in zinc-blende crystals

The elastic energy of a zinc-blende crystal was derived in the previous section and is given by Eq. (11.32). Redefining its constants, we write it in the form:

$$\mathcal{E}_{\text{elastic}} = \frac{\Xi_1}{2} (u_{ii})^2 + \frac{\Xi_2}{2} u_{ij} u_{ji} + \frac{\Xi_3}{2} (u_{xx}^2 + u_{yy}^2 + u_{zz}^2). \quad (11.54)$$

The elastic modulus tensor, in this case, is

$$\Xi_{ij;kl} = \Xi_1 \delta_{ij} \delta_{kl} + \frac{\Xi_2}{2} (\delta_{ik} \delta_{jl} + \delta_{jl} \delta_{ik}) + \Xi_3 \sum_{\nu=x,y,z} \delta_{i\nu} \delta_{j\nu} \delta_{k\nu} \delta_{l\nu}, \quad (11.55)$$

and it can be used in order to calculate the matrix M by formula (11.45). Alternatively, one can construct this matrix directly from the elastic energy using the formula:

$$M_{mn} = \frac{1}{\rho} \frac{\partial^2 \tilde{\mathcal{E}}_{\text{elastic}}}{\partial e_m \partial e_n} \quad (11.56)$$

where

$$\tilde{\mathcal{E}}_{\text{elastic}} = \frac{1}{2} \Xi_{ij;kl} \tilde{u}_{ij} \tilde{u}_{kl} \quad \text{and} \quad \tilde{u}_{ij} = \frac{1}{2} (e_i \hat{q}_j + e_j \hat{q}_i) \quad (11.57)$$

The proof of this formula is given as an exercise. Here we use it to calculate the velocity of sound of waves that move in the direction $[1,1,1]$ and for general direction on the $(1,0,0)$ plane.

In the first case $\hat{q}_x = \hat{q}_y = \hat{q}_z = 1/\sqrt{3}$, therefore $\tilde{u}_{ij} = (e_i + e_j)/(2\sqrt{3})$. Now, let us calculate the various contributions to the M_{mn} matrix coming from the elastic energy (11.54). The first term is:

$$\tilde{\epsilon}_{\text{elastic}}^{(1)} = \frac{\Xi_1}{2\rho} \tilde{u}_{ii} \tilde{u}_{jj} = \frac{\Xi_1}{6\rho} \sum_{ij} e_i e_j, \quad (11.58)$$

and from here, we obtain

$$M_{mn}^{(1)} = \frac{\partial^2 \tilde{\epsilon}_{\text{elastic}}^{(1)}}{\partial e_m \partial e_n} = \frac{\Xi_1}{6\rho} \sum_{ij} (\delta_{im} \delta_{jn} + \delta_{jm} \delta_{in}) = \frac{\Xi_1}{3\rho} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}_{mn}. \quad (11.59)$$

The second contribution to the elastic energy (associated with the middle term in Eq. (11.54)) is

$$\tilde{\epsilon}_{\text{elastic}}^{(2)} = \frac{\Xi_2}{2\rho} \tilde{u}_{ij} \tilde{u}_{ji} = \frac{\Xi_2}{24\rho} \sum_{ij} (e_i + e_j)(e_j + e_i) \quad (11.60)$$

which gives

$$\begin{aligned} M_{mn}^{(2)} &= \frac{\partial^2 \tilde{\epsilon}_{\text{elastic}}^{(2)}}{\partial e_m \partial e_n} = \frac{\Xi_2}{12\rho} \sum_{ij} (\delta_{im} + \delta_{jm})(\delta_{in} + \delta_{jm}) = \frac{\Xi_2}{6\rho} (\delta_{mn} + 1) \\ &= \frac{\Xi_2}{6\rho} \left[\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right]_{mn}. \end{aligned} \quad (11.61)$$

Finally, the last contribution to the elastic energy is,

$$\tilde{\epsilon}_{\text{elastic}}^{(3)} = \frac{\Xi_3}{2\rho} \left[(\tilde{u}_{xx})^2 + (\tilde{u}_{yy})^2 + (\tilde{u}_{zz})^2 \right] = \frac{\Xi_3}{6\rho} (e_x^2 + e_y^2 + e_z^2), \quad (11.62)$$

hence

$$M_{mn}^{(3)} = \frac{\partial^2 \tilde{\epsilon}_{\text{elastic}}^{(3)}}{\partial e_m \partial e_n} = \frac{\Xi_3}{3\rho} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{mn}. \quad (11.63)$$

Collecting all terms, we obtain that the matrix M_{mn} for a wave propagating in the $[1,1,1]$ direction is:

$$M = M^{(1)} + M^{(2)} + M^{(3)} = \frac{1}{3\rho} \left(\frac{\Xi_2}{2} + \Xi_3 \right) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{2\Xi_1 + \Xi_2}{6\rho} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (11.64)$$

The diagonalization of this matrix is simple because a matrix whose all elements are ones has one eigenvalue that equals three and two degenerate eigenvalues that equal zero. Taking into account this property, we obtain that the wave velocities are

$$c_{1,2} = \sqrt{\frac{\Xi_2}{6\rho} + \frac{\Xi_3}{3\rho}} \quad \text{and} \quad c_3 = \sqrt{\frac{\Xi_1}{\rho} + \frac{2\Xi_2}{3\rho} + \frac{\Xi_3}{3\rho}}. \quad (11.65)$$

One can check that c_3 is the velocity of the longitudinal wave that moves in the $[1,1,1]$ direction, while $c_{1,2}$ are the velocities of the transverse waves.

Consider now the case of a sound wave moving in some general direction on the $(1,0,0)$ plane. For this case $\hat{q}_x = 0$ hence $\tilde{u}_{ij} = (e_i \hat{q}_j + e_j \hat{q}_i)/2$ for $i, j = y, z$, while $\tilde{u}_{xj} = \tilde{u}_{jx} = e_x \hat{q}_j/2$ for $j = y, z$ and $u_{xx} = 0$. Following the same procedure presented above, we have:

$$\tilde{\varepsilon}_{\text{elastic}}^{(1)} = \frac{\Xi_1}{2\rho} \tilde{u}_{ii} \tilde{u}_{jj} = \frac{\Xi_1}{2\rho} (e_y \hat{q}_y + e_z \hat{q}_z)^2, \quad M^{(1)} = \frac{\Xi_1}{\rho} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \hat{q}_y^2 & \hat{q}_y \hat{q}_z \\ 0 & \hat{q}_z \hat{q}_y & \hat{q}_z^2 \end{pmatrix} \quad (11.66)$$

$$\tilde{\varepsilon}_{\text{elastic}}^{(2)} = \frac{\Xi_2}{8\rho} \sum_{ij=y,z} (e_i \hat{q}_j + e_j \hat{q}_i)^2 + \frac{\Xi_2}{4\rho} \sum_{j=y,z} (e_x \hat{q}_j)^2, \quad (11.67)$$

$$M^{(2)} = \frac{\Xi_2}{2\rho} \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \hat{q}_y^2 & \hat{q}_y \hat{q}_z \\ 0 & \hat{q}_z \hat{q}_y & \hat{q}_z^2 \end{pmatrix} \right], \quad (11.68)$$

and

$$\tilde{\varepsilon}_{\text{elastic}}^{(3)} = \frac{\Xi_3}{2\rho} [(e_y \hat{q}_y)^2 + (e_z \hat{q}_z)^2], \quad M^{(3)} = \frac{\Xi_3}{\rho} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \hat{q}_y^2 & 0 \\ 0 & 0 & \hat{q}_z^2 \end{pmatrix} \quad (11.69)$$

Thus

$$M = \begin{pmatrix} \frac{\Xi_2}{2\rho} & 0 & 0 \\ 0 & \frac{\Xi_2}{2\rho} + \left(\frac{\Xi_2}{2\rho} + \frac{\Xi_1}{\rho} + \frac{\Xi_3}{\rho} \right) \hat{q}_y^2 & \left(\frac{\Xi_1}{\rho} + \frac{\Xi_2}{2\rho} \right) \hat{q}_y \hat{q}_z \\ 0 & \left(\frac{\Xi_1}{\rho} + \frac{\Xi_2}{2\rho} \right) \hat{q}_y \hat{q}_z & \frac{\Xi_2}{2\rho} + \left(\frac{\Xi_2}{2\rho} + \frac{\Xi_1}{\rho} + \frac{\Xi_3}{\rho} \right) \hat{q}_z^2 \end{pmatrix}. \quad (11.70)$$

From here, it is clear that one wave velocity is:

$$c_1 = \sqrt{\frac{\Xi_2}{2\rho}} \quad (11.71)$$

while the two others are obtained from diagonalization of the 2×2 submatrix :

$$c_{2,3}^2 = \frac{3\Xi_2}{4\rho} + \frac{1}{2} \left(\frac{\Xi_1}{\rho} + \frac{\Xi_3}{\rho} \right) \mp \frac{1}{2} \sqrt{\left(\frac{\Xi_2}{2\rho} + \frac{\Xi_1}{\rho} + \frac{\Xi_3}{\rho} \right)^2 \cos^2(2\theta) + \left(\frac{\Xi_1}{\rho} + \frac{\Xi_2}{2\rho} \right)^2 \sin^2(2\theta)}, \quad (11.72)$$

where we choose $\hat{q}_y = \cos \theta$ and $\hat{q}_z = \sin \theta$ (so that $|\hat{\mathbf{q}}|=1$ because $\hat{q}_x=0$). Thus, we obtained three different wave velocities for this case, which is the typical situation as demonstrated in Fig. 11-3.

Finally, we comment that, in general, the polarization vector need not be parallel or perpendicular to the direction of wave propagation. In other words, the characterization of waves as longitudinal or transverse is, in general, only approximate.

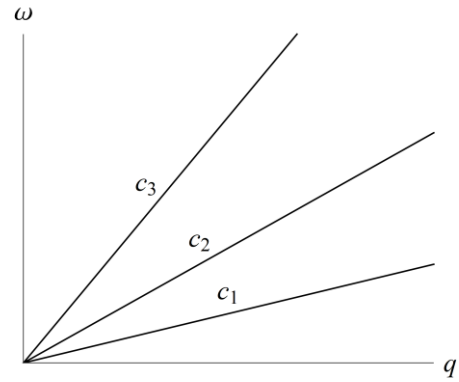


Figure 11-3 Typical dispersion of sound waves in crystals

11.4 Phonons

Phonons are obtained from the quantization of sound waves. Each normal mode of the lattice vibrations (assuming its amplitude to be sufficiently small) is essentially a simple harmonic oscillator. Quantization of the latter, as we know, can be easily obtained by expressing the position operator, $\hat{x}$, and the momentum operator, $\hat{p}$, in terms of the creation and annihilation operators, $\hat{a}^\dagger$ and $\hat{a}$, defined by

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega}} (\hat{a} + \hat{a}^\dagger), \quad \text{and} \quad \hat{p} = -i\sqrt{\frac{\hbar}{2}} m\omega (\hat{a} - \hat{a}^\dagger), \quad (11.73)$$

where m is the oscillator mass and ω is its frequency. With these variables, the Hamiltonian of the harmonic oscillator takes the form: $H = \hbar\omega(\hat{a}^\dagger\hat{a} + 1/2)$.

Similarly, to quantize the sound waves of a crystal, one has to expand the displacement vector and the momentum density in terms of the system's normal modes and then quantize each normal mode treating it as a simple harmonic oscillator. In analogy to Eq. (11.73), the displacement vector and the momentum density are given by:

$$u_j(\mathbf{r}) = \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q})\text{Vol}}} e_j^{(\alpha)}(\hat{\mathbf{q}}) [\hat{a}_\alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{r}) + \hat{a}_\alpha^\dagger(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r})], \quad (11.74)$$

$$p_j(\mathbf{r}) = -i \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar\rho\omega_\alpha(\mathbf{q})}{2\text{Vol}}} e_j^{(\alpha)}(\hat{\mathbf{q}}) [\hat{a}_\alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{r}) - \hat{a}_\alpha^\dagger(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r})],$$

where $\omega_\alpha(\mathbf{q}) = c_\alpha(\hat{\mathbf{q}})|\mathbf{q}|$. Here we assume the system to be finite in size (but large) and denote

its volume by Vol. This assumption implies that the wavenumber vector $\mathbf{q}$ can take only discrete values set by the boundary conditions. The operators $\hat{a}_\alpha^\dagger(\mathbf{q})$ and $\hat{a}_\alpha(\mathbf{q})$ are the creation and annihilation operators of a phonon with a wavenumber $\mathbf{q}$ and polarization α . Since these are bosons, their commutation relations are :

$$[\hat{a}_\alpha(\mathbf{q}); \hat{a}_\beta(\mathbf{q}')] = [\hat{a}_\alpha^\dagger(\mathbf{q}); \hat{a}_\beta^\dagger(\mathbf{q}')] = 0, \text{ and } [\hat{a}_\alpha(\mathbf{q}); \hat{a}_\beta^\dagger(\mathbf{q}')] = \delta_{\alpha\beta} \delta_{\mathbf{q},\mathbf{q}'} . \quad (11.75)$$

The above representation of the displacement vector and the momentum density ensures that these quantities satisfy the commutation relation (11.34) (see Ex. 4), and since each vibrational mode of the lattice is an independent harmonic oscillator, substituting Eqs. (11.74) in the Hamiltonian (11.33) yields:

$$H = \sum_{\alpha, \mathbf{q}} \hbar \omega_\alpha(\mathbf{q}) \left[\hat{a}_\alpha^\dagger(\mathbf{q}) \hat{a}_\alpha(\mathbf{q}) + \frac{1}{2} \right]. \quad (11.76)$$

The expectation value of the number operator $\hat{n}_\alpha(\mathbf{q}) = \hat{a}_\alpha^\dagger(\mathbf{q}) \hat{a}_\alpha(\mathbf{q})$ is the number of phonons that occupy the normal mode with wave $\mathbf{q}$ number and polarization α . Thus, if we denote by $|n_\alpha(\mathbf{q})\rangle$ the occupation state with $n_\alpha(\mathbf{q})$ phonons, then similar to the harmonic oscillator,

$$\langle n_\alpha + 1 | a_\alpha^\dagger | n_\alpha \rangle = \sqrt{n_\alpha + 1} \text{ and } \langle n_\alpha - 1 | a_\alpha | n_\alpha \rangle = \sqrt{n_\alpha} . \quad (11.77)$$

At thermal equilibrium, the average occupation of phonons is given by Planck's distribution:

$$\langle n_\alpha(\mathbf{q}) \rangle = \frac{1}{\exp\left(\frac{\hbar \omega_\alpha(\mathbf{q})}{k_B T}\right) - 1}, \quad (11.78)$$

where k_B is the Boltzmann constant, while T is the temperature. From here, it follows that the typical frequency of phonons at equilibrium is $\omega \simeq k_B T / \hbar$, and their typical wavenumber is $q \simeq k_B T / \hbar c$, where c is the speed of sound. Since the total vibrational energy of a system is obtained from an integral $d^d q$ over Planck's distribution multiplied by $\hbar \omega_\alpha(\mathbf{q}) \simeq k_B T$ (Here d is the system's dimensionality), it is of order $U \sim q^d T \propto T^{d+1} / c^d$. Hence the variation of this energy, due to a ΔT change in the temperature, is $\Delta U \sim (T/c)^d \Delta T$. Thus, the heat capacity of the phonons is: $C_V = \Delta U / \Delta T \propto T^d$.

There is nothing mysterious about the second quantization procedure that we used here to characterize sound waves in quantum systems. If we had treated the position and momentum of each particle in the lattice as quantum operators, all the way through, the resulting energy levels and eigenstate would be precisely those of the Hamiltonian (11.76) - see Ex. 5.

11.5 Optical vibrations in crystals (optical phonons)

In the previous sections, we discuss lattice vibrations associated with deformation that, essentially, do not affect the internal structure of each unit cell. For a simple crystal made from one type of atom situated on Bravais lattice, these vibrational modes, called *acoustic modes* or *acoustic phonons*, are all one can have. However, if each unit cell contains N_{uc} atoms, it has $N_{uc}d$ degrees of freedom, where d is the system dimension. Thus, there must be $(N_{uc} - 1)d$ additional vibrational modes associated with all possible internal deformations of a unit cell. These vibrational modes are called *optical vibrations* or *optical phonons* (when quantized). A schematic illustration of the phonon spectrum in three-dimensional systems where each unit cell contains two atoms is shown in Fig. 11-4.

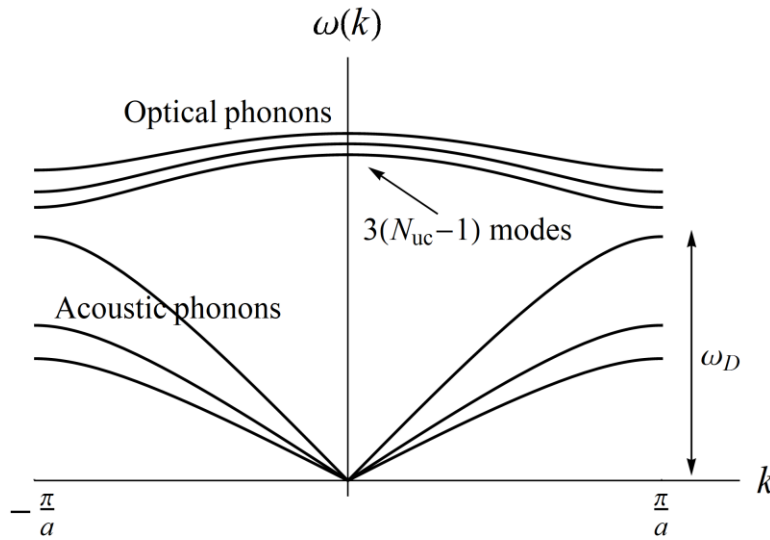


Figure 11-4 A schematic illustration of the phonon spectrum in a three-dimensional lattice

All phonon branches, whether acoustic or optical, satisfy the condition $\omega^2(\mathbf{k}) = \omega^2(-\mathbf{k})$ dictated by time-reversal symmetry. Translational symmetry implies that $\omega^2(\mathbf{k} + \mathbf{b}) = \omega^2(\mathbf{k})$ where $\mathbf{b}$ is an arbitrary vector of the reciprocal lattice. The highest frequency of the acoustic branches of the spectrum appears at the edge of the Brillouin zone ($k = \pi/a$ in the above figure). The highest frequency among them is called *Debye frequency* and denoted by ω_D . At temperatures much smaller than Debye's frequency, $T \ll \hbar\omega_D$, the heat capacitance is $C_V \sim T^d$, as we saw above. However, when $T \gg \hbar\omega_D$ the law of equipartition implies $C_V \sim k_B N_{uc} d / 2$.

Example: Waves in a one-dimensional system

Consider a one-dimensional system made of particles connected by springs having identical spring constants, K . Each unit cell of this lattice contains two particles with masses m_1 and m_2 as shown in Fig. 11-5. We shall consider only longitudinal vibrations of this system and denote by $u_j^{(1)}$ and $u_j^{(2)}$ the displacement vectors of the particles in the j -th unit cell, along the chain, as indicated in the figure below.

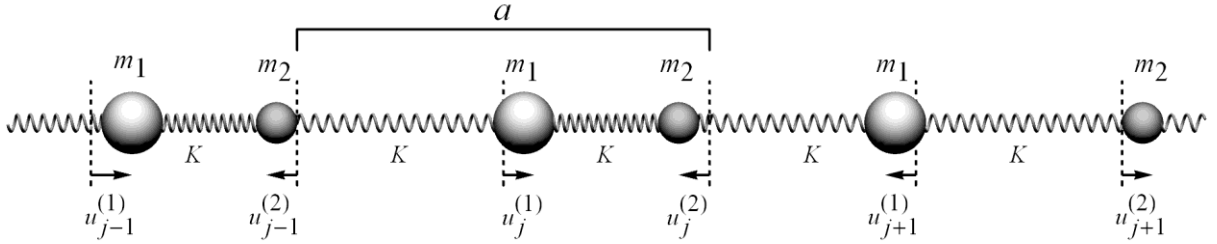


Figure 11-5 A one-dimensional chain made of two types of particles connected by identical springs

Newton's second law yields the equations of motion:

$$\begin{aligned} m_1 \frac{d^2 u_j^{(1)}}{dt^2} &= K [u_j^{(2)} - 2u_j^{(1)} + u_{j-1}^{(2)}], \\ m_2 \frac{d^2 u_j^{(2)}}{dt^2} &= K [u_{j+1}^{(1)} - 2u_j^{(2)} + u_j^{(1)}]. \end{aligned} \quad (11.79)$$

Defining $\omega_\nu^2 = K/m_\nu$ ($\nu = 1, 2$) and substituting solution of the type $u_j^{(\nu)} = A_\nu \exp(ikaj - i\omega t)$, where k is wave number and a is the lattice constant, reduces the above equations to

$$\begin{pmatrix} \omega^2 - 2\omega_1^2 & \omega_1^2 [1 + \exp(-ika)] \\ \omega_2^2 [1 + \exp(ika)] & \omega^2 - 2\omega_2^2 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (11.80)$$

A nontrivial solution of this equation is obtained only when the determinant of the above matrix vanishes. This condition yields the biquadratic equation:

$$\omega^4 - 2(\omega_1^2 + \omega_2^2) + 4\omega_1^2 \omega_2^2 \sin^2\left(\frac{ka}{2}\right) = 0 \quad (11.81)$$

whose solutions are:

$$\omega = \omega_a(k) = \sqrt{\omega_1^2 + \omega_2^2 - \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4\omega_1^2\omega_2^2 \sin^2\left(\frac{ka}{2}\right)}}, \quad (11.82)$$

and

$$\omega = \omega_o(k) = \sqrt{\omega_1^2 + \omega_2^2 + \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4\omega_1^2\omega_2^2 \sin^2\left(\frac{ka}{2}\right)}}. \quad (11.83)$$

The first solution (11.82) represents the acoustic branch of the vibrational spectrum. Expanding it to leading order in k yields a linear dispersion:

$$\omega_a(k) \approx \frac{\omega_1\omega_2 a}{\sqrt{2(\omega_1^2 + \omega_2^2)}} k. \quad (11.84)$$

The second solution (11.83) represents the optical branch of the spectrum. Its expansion, to leading order in k , yields:

$$\omega_o(k) \approx \sqrt{2(\omega_1^2 + \omega_2^2)} - \frac{\omega_1^2\omega_2^2 a^2}{4\sqrt{2}(\omega_1^2 + \omega_2^2)^{3/2}} k^2 \quad (11.85)$$

Thus, the frequency of the optical branch in the limit $k \rightarrow 0$ is a positive constant. It is easy to check that in the limit, $k \rightarrow 0$ the eigenvectors of the acoustic modes describe a rigid translation of the whole unit cell with $A_1 = A_2$, while for the optical mode, the particles move one against the other, keeping the center of mass of each unit cell fixed, $\omega_2^2 A_1 = -\omega_1^2 A_2$, see Fig. 11-6. Notice that identifying acoustic modes with a motion of the whole unit cell and optical modes with situations where particles move one against the other becomes meaningless as k it approaches the edge of the Brillouin zone.

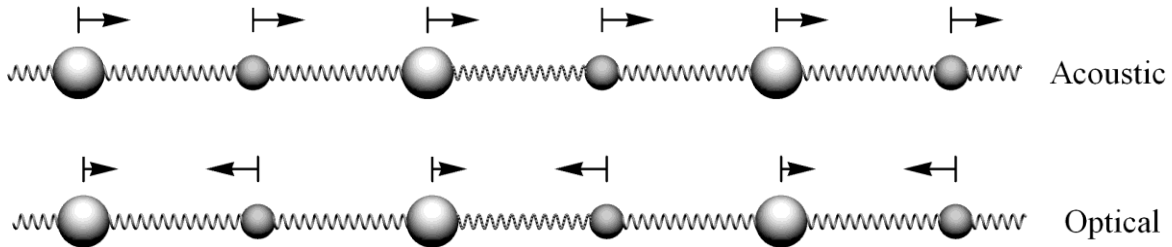


Figure 11-6 Acoustic and optical modes in a one-dimensional mass chain in the limit of vanishing wavenumber

11.6 Symmetry approach to optical modes in crystals

The observation that optical modes, in the limit of vanishing wavenumber are, essentially, the vibrational modes of an individual unit cell suggests harnessing symmetry considerations for revealing their properties - similar to our study of molecules in chapter 4. Identifying the symmetries of these optical modes is also crucial for the characterization of the lattice optical response to the electromagnetic field discussed in the next chapter. To identify the irreducible representations of the optical modes, one may use the following procedure:

- (a) Identify the symmetry group of the system at $k = 0$.
- (b) Construct the character table associated with the lattice displacement vectors representation. Here to simplify the analysis, it is convenient to represent the lattice displacement vector representation, which we denote by Γ_{lattice} , in the form:

$$\Gamma_{\text{lattice}} = \Gamma_{\text{atom-sites}} \otimes \Gamma_{\text{vector}}, \quad (11.86)$$

where Γ_{vector} is the irreducible representation associated with symmetry operations on vectors, while $\Gamma_{\text{atom-sites}}$ is the atom-site representation determined only by the positions of the atoms and not by their displacements (which are taken into account by Γ_{vector}). The character of the atom-site representation for a given symmetry operation is the number of atoms that remain at their positions or moved to points that differ by one of the lattice vectors.

- (c) Find the composition of the displacement vectors representation in terms of the irreducible representations of the group.
- (d) Identify the irreducible representations of the optical phonon by subtracting the ones associated with the acoustic modes from the direct sum of irreducible representations obtained in the previous step. Notice that here one should not subtract irreducible representation associated with rotations since, in a lattice, local rotations of unit cells are also part of the phonon spectrum.

Similar to the normal coordinates describing the small oscillations of a system near its equilibrium, the optical vibration modes of a lattice are associated with normal coordinates $\xi_i^{(\alpha)}$, where the index α denotes the irreducible representation of the vibrational mode, and $i = 1, \dots, \ell_\alpha$, where ℓ_α is the dimension of the representation. The Hamiltonian that describes these normal coordinates is:

$$H = \sum_{i=1}^{\ell_\alpha} \int d^d r \left[\frac{p_i^2}{2} + \frac{\omega_\alpha^2}{2} \xi_i^2 \right], \quad (11.87)$$

where $[p_i(\mathbf{r}); \xi_j(\mathbf{r}')] = -i\hbar \delta_{ij} \delta(\mathbf{r} - \mathbf{r}')$, and the equations of motion are those of simple harmonic oscillator whose frequency, ω_α , is ℓ_α -fold degenerate:

$$\begin{aligned} \dot{\xi}_i &= p_i, \\ \dot{p}_i &= -\omega_\alpha^2 \xi_i. \end{aligned} \quad (11.88)$$

Example: Optical phonons in graphene

As we already know, the symmetry group of graphene is C_{6v} , and each unit cell contains two atoms belonging to two sublattices, A and B. The characters for the atom-site representation depend on whether the symmetry operation replaces the two sublattices or not. If the symmetry operation replaces them, then the character is zero, while if it leaves atoms on the same sublattice, the character is two because there are two atoms in each unit cell. This consideration yields the following character table:

C_{6v}	E	$2C_6$	$2C_3$	C_2	$3\sigma_v$	$3\sigma_d$
$\Gamma_{\text{atom-sites}}$	2	0	2	0	2	0

Now, with the help of the character table of C_{6v} on page 94 we can identify the composition of this representation:

$$n_{A_1} = \frac{1}{12} (2 \cdot 1 + 2 \cdot 2 \cdot 1 + 3 \cdot 2 \cdot 1) = 1 \quad (11.89)$$

and

$$n_{B_1} = \frac{1}{12} (2 \cdot 1 + 2 \cdot 2 \cdot 1 + 3 \cdot 2 \cdot 1) = 1. \quad (11.90)$$

Thus $\Gamma_{\text{atom-sites}} = A_1 \oplus B_1$ and since the irreducible vector representation of is $\Gamma_{\text{vector}} = E_1$ (because the linear basis function of E_1 is a vector (x, y)) we obtain that the composition of the lattice displacement vector representation is:

$$\Gamma_{\text{lattice}} = (A_1 \oplus B_1) \otimes E_1 = E_1 \oplus E_2 \quad (11.91)$$

Acoustic phonons are associated with simple displacement of the unit cells; therefore, their irreducible representation is E_1 . Thus, the two optical vibrations modes of the graphene belong to E_2 , and clearly, their frequencies are degenerate. One can identify them from the condition that the center of mass in each cell does not move, and from their properties under the symmetry operations of the group (in particular that under c_2 rotation the basis functions of the E_2 representation return to themselves - see table on page 94). These optical modes are shown in Fig. 11-7.

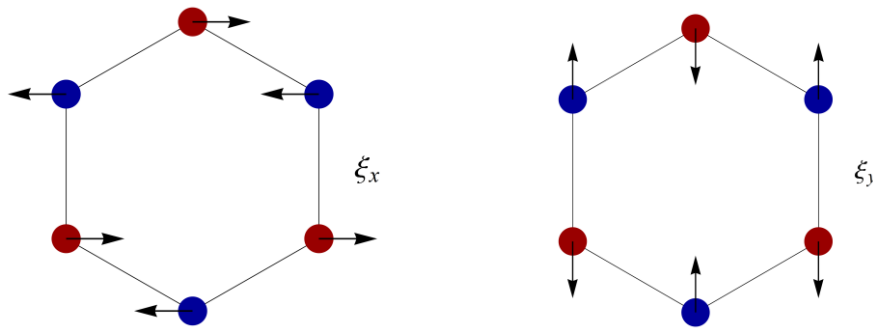


Figure 11-7 Optical modes of graphene

Example: Optical phonons in zinc blende crystals

The symmetry of zinc-blende crystals is the symmetry of a regular tetrahedron, see Fig. 11-8. It is made of two interpenetrating fcc sublattices (see Fig. 7-14). The symmetry group is T_d , and its character table can be found on page 144. One can check that all symmetry operations of the group leave the atoms on their sites modulo a change in the lattice vector; hence the character table of the atom site representation is:

T_d	E	$8c_3$	$3c_2$	$6S_4$	$6\sigma_d$
$\Gamma_{\text{atom-sites}}$	2	2	2	2	2

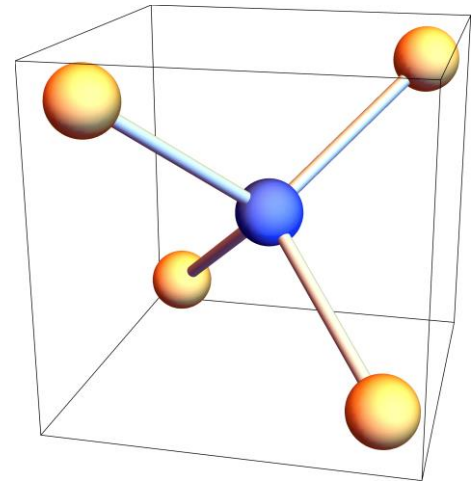


Figure 11-8 An element of a zinc blende crystal

Since the vector irreducible representation is $\Gamma_{\text{vector}} = F_2$, we obtain the composition of the lattice displacement representation: $\Gamma_{\text{lattice}} = 2A_1 \otimes F_2 = 2F_2$. From here, it is clear that the optical

phonons are basis functions of the E_2 which is threefold degenerate. These modes are shown in Fig. 11-9.

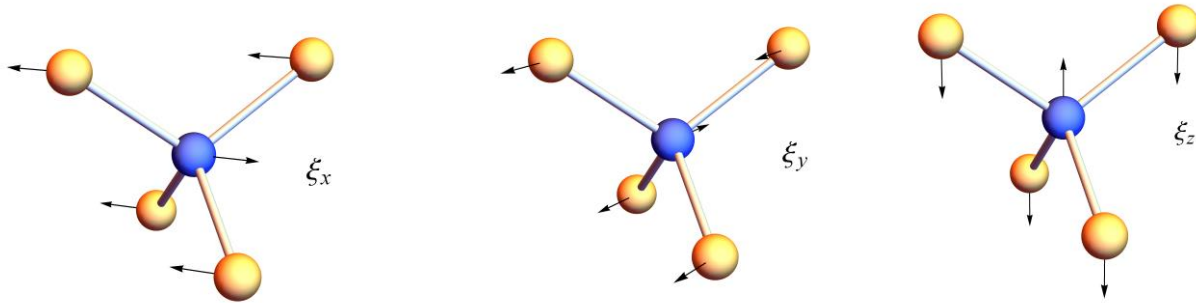


Figure 11-9 The optical modes of zinc-blende crystals at $\mathbf{k} = 0$ (Γ -point)

Comment: The threefold degeneracy of the optical phonon modes at $\mathbf{k} \rightarrow 0$ appears only in systems of infinite size. In reality, there is no degeneracy between the longitudinal optical modes (LO) and the transverse optical modes (TO) as one can see, for example, in the phonon spectrum of GaAs shown in Fig. 11-10. This situation where the system's behavior at $\mathbf{k} = 0$ is different from the behavior in the limit $\mathbf{k} \rightarrow 0$ is called “anomaly”.

This anomaly is due to the slight ionic nature of the crystal and the long-range Coulomb interactions. To explain the mechanism, consider waves moving in the $[1,1,1]$ direction, such that the gallium and the arsenic atoms reside on different $(1,1,1)$ planes. The plane of the gallium and the arsenic atoms are slightly charged: gallium planes are positive while the arsenic planes are negative. In transverse optical waves, these planes glide parallel to each other,

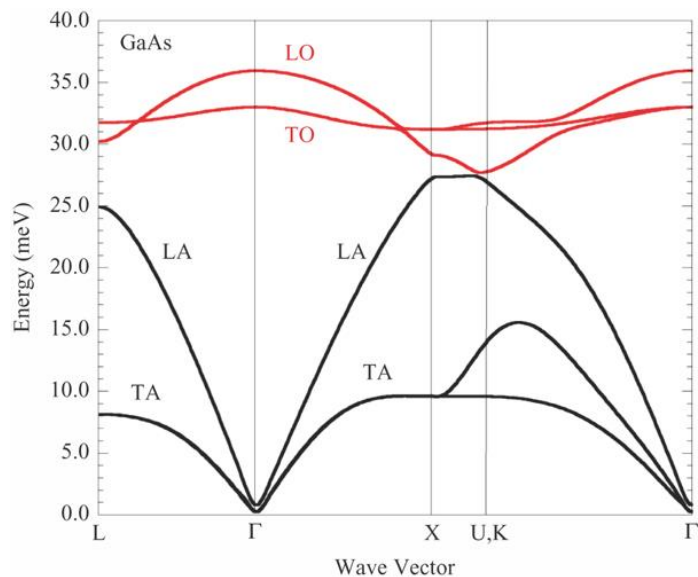


Figure 11-10 The phonon spectrum of GaAs

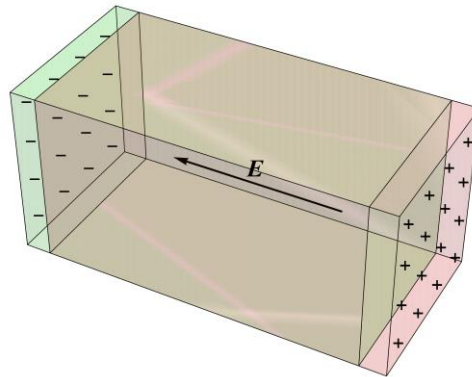


Figure 11-11 An illustration of the charge accumulated on the surface of a system due to optical vibration modes

and there is no net charge accumulated on the system's boundary. On the other hand, in longitudinal optical waves, the same planes move towards or away from each other such that a net charge is accumulated on the surface of the system. This charge accumulation is similar to that of plasma waves of electrons in metals, see Fig. 11-11. The surface charge creates an electric field that exerts an additional restoring force on the planes that increases the frequency of the longitudinal optical mode with respect to that of the transverse modes. In the next chapter, we will return to this feature and discuss the anomaly from a different perspective.

11.7 Exercises

1. Identify the components of the elastic modulus tensor for a three-dimensional hexagonal lattice.

Advice: The point group of hexagonal lattice in three dimensions is D_{6h} ; however it is sufficient to use the irreducible representations of the C_{6v} and then to verify that the resulting elastic energy is invariant to reflection through a plane perpendicular to the principle symmetry axis.

2. A trigonal crystal in three dimensions is described by D_{3d} point group. Find all components of the elastic modulus tensor of this crystal.

Hint: Recall that the subgroup D_3 of D_{3d} is isomorphic to C_{3v} .

3. Calculate the sound velocities in a three-dimensional hexagonal lattice for a wave moving in an arbitrary direction.

Advice: Show that the elastic energy has a full rotational symmetry around the principal axis of symmetry (in any angle) and use it to set to zero one of the components of the wavenumber vector in the direction perpendicular to this axis.

4. Prove that Eqs. (11.74) for the displacement vector and the momentum density satisfies the commutation relations (11.34).
5. Consider a one-dimensional chain of identical particles connected by identical springs and described by the Hamiltonian:

$$H = \sum_{j=1}^N \frac{p_j^2}{2m} + \frac{1}{2} K (x_j - x_{j-1})^2, \quad (11.92)$$

where m is the mass of each particle, K is the spring constant, x_j and p_j are, respectively, the displacement and momentum of the j -th particle, while N is the number of particles in the chain. For simplicity, assume it to be an odd number. Also, choose periodic boundary conditions, i.e., $x_0 = x_{N+1}$ and $p_0 = p_{N+1}$.

(a) Define the Fourier series representation of the variables by

$$x_j = \frac{1}{\sqrt{N}} \sum_{\nu} X_{\nu} \exp\left(i \frac{2\pi}{N} \nu j\right) \text{ and } p_j = \frac{1}{\sqrt{N}} \sum_{\nu} P_{\nu} \exp\left(-i \frac{2\pi}{N} \nu j\right) \quad (11.93)$$

where

$$\nu = -\frac{N-1}{2}, -\frac{N-3}{2}, \dots, 0, \dots, \frac{N-3}{2}, \frac{N-1}{2}. \quad (11.94)$$

Use the inverse Fourier transform of the above quantities and the commutation relations $[x_j, p_k] = i\hbar \delta_{jk}$ to show that X_{ν} and P_{ν} are canonically conjugated variables, i.e.

$$[X_{\nu}, P_{\mu}] = i\hbar \delta_{\nu\mu}. \quad (11.95)$$

(b) Calculate the Hamiltonian in the new variables and show that

$$H = \sum_{\nu} \left[\frac{P_{\nu} P_{-\nu}}{2m} + \frac{1}{2} m \Omega_{\nu}^2 X_{\nu} X_{-\nu} \right]. \quad (11.96)$$

$$\text{where } \Omega_{\nu}^2 = 4 \frac{K}{m} \sin^2\left(\frac{\pi \nu}{N}\right).$$

(c) Now express the canonical variables in terms of creation and annihilation operators:

$$X_{\nu} = \sqrt{\frac{\hbar}{2m\Omega_{\nu}}} (a_{\nu} + a_{-\nu}^{\dagger}) \quad \text{and} \quad P_{\nu} = -i \sqrt{\frac{\hbar m \Omega_{\nu}}{2}} (a_{-\nu} - a_{\nu}^{\dagger}). \quad (11.97)$$

Verify that with this definition, the commutation relation (11.95) is satisfied and show that the Hamiltonian of the system reduces to the second quantized form:

$$H = \sum_{\nu} \hbar \Omega_{\nu} \left(a_{\nu}^{\dagger} a_{\nu} + \frac{1}{2} \right). \quad (11.98)$$

6. Consider a two-dimensional lattice made of two interpenetrating square sublattices of two different atoms, as shown in Fig. 11-12. The symmetry group of this lattice is C_{4v} , and its character table is provided below. In this exercise, assume that deformations only occur in the lattice plane.

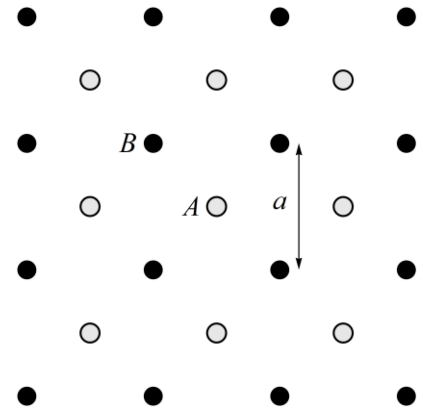


Figure 11-12 Interpenetrating square lattices

- Identify the elastic modulus tensor of this lattice..
- Assuming the mass density to be ρ , calculate the wave velocity of acoustic waves in an arbitrary direction.
- Find the character table of the atom-site representation and identify the composition of the Lattice displacement representation.
- Draw the optical phonon modes of this lattice.

C_{4v}	E	c_4, c_4^3	c_2	2σ	$2\sigma'$			
A_1	1	1	1	1	1	z	$x^2 + y^2, z^2$	$z^3, z(x^2 + y^2)$
A_2	1	1	1	-1	-1			
B_1	1	-1	1	1	-1		$x^2 - y^2$	$z(x^2 - y^2)$
B_2	1	-1	1	-1	1		xy	xyz
E	2	0	-2	0	0	(x, y)	$z(x, y)$	$z^2(x, y), xy(y, x), (x^3, y^3)$

12 Crystals in an electromagnetic field

The following chapter focuses on the interaction between lattice vibrations and electromagnetic waves. This subject is vast and deserves a course of its own because it includes many interesting optical phenomena such as absorption, dispersion, anomalous dispersion, double refraction, conical refraction, optical activity, dichroism, reflection, skin effect, and Faraday effect. Here, however, we consider only a few features associated with the interaction of electromagnetic waves with optical phonons. There are three main aspects associated with the interaction of electromagnetic waves and optical phonons: One is the absorption of electromagnetic waves by excitation of lattice vibrations – a phenomenon called *infrared activity*. The second is the scattering of electromagnetic waves, which involves either excitation or absorption of an optical phonon - an effect called *Raman scattering*. Finally, the third aspect concerns the propagation of electromagnetic waves in a crystal. Here we shall see that when the coupling between the optical lattice vibrations and the electromagnetic field is strong, a new type of excitation appears, which mixes both. This excitation is called a *polariton*.

12.1 Infrared activity

One of the essential experimental tools for investigating materials is the study of their absorption spectrum of electromagnetic waves. In atoms, this absorption is associated with the excitation of electrons into higher energy levels. These transitions are dictated by selection rules, namely a nonzero value of the transition dipole moment between the atom's initial and final states. Similarly, when an electromagnetic field impinges a crystal, it may excite one of its optical vibrational modes. This process that results in absorption (typically in the infrared regime of the electromagnetic spectrum) is called *infrared activity*. Our goal in this section is to formulate the selection rules for infrared absorption and identify which crystals are infrared active.

The transition matrix element required for the absorption of an electromagnetic wave is of the dipole operator

$$H_{\text{dipole}} = -\int d^3r \mathbf{E} \cdot \mathbf{P}, \quad (12.1)$$

where $\mathbf{E}$ is the oscillatory electric field acting on the crystal, and $\mathbf{P}$ is the induced electric dipole moment that results from the shifts of the atoms from their equilibrium positions. This dipole moment appears only in optical phonons, where atoms with a different number of valence electrons move in opposite directions to create a dipole moment. Thus, in general,

$$\mathbf{P} = \rho_p \boldsymbol{\xi} \quad (12.2)$$

where ξ is one of the normal coordinates of the optical vibrational modes and ρ_p is the charge density generated due to the atom shift. Crystals in which $\rho_p \neq 0$ are called polarizable. From the above equation, we see that in order to satisfy this condition, the irreducible representation of the optical phonons (ξ) should be the same as the irreducible representations of a vector ($\mathbf{P}$).

An alternative way of presenting this condition is by looking at the transition matrix element $\langle \psi_{ES} | H_{\text{dipole}} | \psi_{GS} \rangle$ between the ground state $|\psi_{GS}\rangle$ and the excited state $|\psi_{ES}\rangle$ of the system, represented by the diagram in Fig. 12-1. The selection rule for absorption is that this matrix element is nonzero. Now, the ground state of a system is generally the most symmetric state belonging to the identity irreducible representation A_1 . On the other hand, the operator H_{dipole} belongs to the vector irreducible representation since the dipole moment is a vector. Now, in order to have $\langle \psi_{ES} | H_{\text{dipole}} | \psi_{GS} \rangle \neq 0$, this matrix element must be invariant under all group operations. Hence, the excited state must also belong to the vector irreducible representation (because the direct product of the vector irreducible representation by itself contains the identity representation). However, the excited state $|\psi_{ES}\rangle$ has the same symmetry of ξ ; therefore, the selection rule is that ξ should belong to the vector irreducible representation.

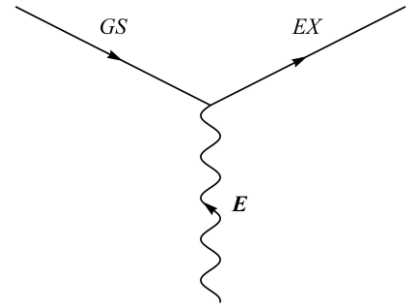


Figure 12-1 A diagram representing the absorption of electromagnetic wave in a crystal

Example: Zinc blende crystals

We have seen that the optical phonons of zinc-blende crystals belong to the F_2 irreducible representation (see example on page 244). This is also the irreducible vector representation of T_d ; therefore, zinc-blende crystals are infrared active.

Example: Diamond

A diamond is a crystal made only from only carbon atoms in the form of a zinc-blende crystal. Namely, it is made of two interpenetrating fcc lattices, as shown in Fig. 12-2. In this figure, the two colors denote the two sublattices, where one is shifted with respect to the other by the vector $\mathbf{l} = a(1,1,1)/4$, where a is the lattice constant. Being made of only carbon atoms, such that each one of

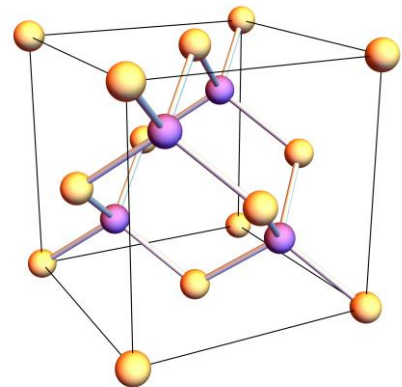


Figure 12-2 the structure of a diamond as two interpenetrating fcc sublattices

them is connected to four other atoms in a tetrahedral shape, one expects this material to be infrared *inactive*. Let us verify that using the group theory approach.

The space group of a diamond is nonsymmorphic. In addition to the 24 symmetry operations of the tetrahedral group, T_d it contains symmetry operations such as glide reflection and rotary translation. The conjugacy classes of the additional symmetry operations are: $\{i|I\}$, $8\{ic_3|I\}$, $3\{ic_2|I\}$, $6\{c'_2|I\}$, and $6\{c_4|I\}$. Altogether the group contains 48 elements (apart from translations), and the group is isomorphic to the octahedral group O_h . Below is the character table of the group.

Diamond	E	$8c_3$	$6\{c'_2 I\}$	$6\{c_4 I\}$	$3c'_2$	$\{i I\}$	$6S_4$	$8\{ic_3 I\}$	$3\{ic_2 I\}$	$6c_2$		
O_h	E	$8c_3$	$6c_2$	$6c_4$	$3c_2 = c_4^2$	i	$6S_4$	$8S_6$	$3\sigma_h$	$6\sigma_d$		
A_{1g}	1	1	1	1	1	1	1	1	1	1		$x^2 + y^2 + z^2$
A_{2g}	1	1	-1	-1	1	1	-1	1	1	-1		
E_g	2	-1	0	0	2	2	0	-1	2	0		$\begin{pmatrix} 2z^2 - x^2 - y^2 \\ \sqrt{3}(x^2 - y^2) \end{pmatrix}$
F_{1g}	3	0	-1	1	-1	3	1	0	-1	-1		
F_{2g}	3	0	1	-1	-1	3	-1	0	-1	1		(yz, xz, xy)
A_{1u}	1	1	1	1	1	-1	-1	-1	-1	-1		
A_{2u}	1	1	-1	-1	1	-1	1	-1	-1	1		
E_u	2	-1	0	0	2	-2	0	1	-2	0		
F_{1u}	3	0	-1	1	-1	-3	-1	0	1	1	(x, y, z)	
F_{2u}	3	0	1	-1	-1	-3	1	0	1	-1		

To identify the irreducible representation of the optical phonons, we need to construct the character table of the atom-site representation. Recall that, in this construction, a symmetry operation that transforms atoms between the same sublattice is considered the identity operation. Hence all symmetry operations of the T_d group leave the atoms on their sites, and taking one atom from each sublattice, the corresponding character is 2. On the other hand, all other operations that involve translation by I swap atoms between the two sublattices; therefore, the corresponding characters vanish. Thus, the characters of the atoms-site representation are:

Diamond	E	$8c_3$	$6c_2$	$6c_4$	$3c'_2$	$\{i I\}$	$\{6c_4 I\}$	$\{8ic_3 I\}$	$\{3ic_2 I\}$	$\{6c'_2 I\}$
$\Gamma_{\text{atom-sites}}$	2	2	2	2	2	0	0	0	0	0

From the character table of the O_h group, one can see that $\Gamma_{\text{atom-sites}} = A_{1g} \oplus A_{2u}$, hence the composition of the lattice representation is

$$\Gamma_{\text{lattice}} = (A_{1g} \oplus A_{2u}) \otimes F_{1u} = F_{1u} \oplus F_{2g}. \quad (12.3)$$

We see that the optical phonons belong to the F_{2g} irreducible representation, which is not the vector representation, F_{1u} . Thus, diamond is a nonpolarizable crystal, i.e., infrared inactive. Silicon crystal has precisely the same crystal structure as diamond; hence it is also a nonpolarizable crystal.

However, it is worth emphasizing that, contrary to the impression obtained from the example of diamond, monoatomic crystals may be infrared active. For example, graphite made of graphene layers stacked one on top of the other is infrared active, see Ex. 1.

12.2 Raman scattering

Raman scattering is an inelastic scattering of electromagnetic waves by molecules or crystals, where the scattered wave has a different frequency from that of the incoming wave. Usually, the frequency of the incoming and outgoing waves is within the visible spectrum. The small difference in their frequencies is due to the absorption or emission of an optical phonon. The process where the scattered wave has a lower frequency than that of the incoming wave,

$$|\text{photon}, \hbar\omega\rangle \longrightarrow |\text{photon}, \hbar\omega_1\rangle + |\text{phonon}, \mathbf{k} \approx 0\rangle, \quad (12.4)$$

is called "Stokes' process". The wavenumber of the emitted phonon is approximately zero because the momentum transferred to the phonon is $\hbar(\omega - \omega_1)/c$, where $\omega - \omega_1$ is very small while the speed of light c is very large. Another form of the Stokes process is by the creation of two phonons with opposite momenta:

$$|\text{photon}, \hbar\omega\rangle \longrightarrow |\text{photon}, \hbar\omega_1\rangle + |\text{phonon}, \mathbf{k}\rangle + |\text{phonon}, -\mathbf{k}\rangle \quad (12.5)$$

Here the total momentum transfer to the lattice is approximately zero as follows from the above argument, but the momentum of each phonon may be significant. However, the matrix element for two phonon emissions is much smaller than that of a single phonon emission.

An anti-Stokes process is an inelastic scattering of an electromagnetic wave where the scattered wave has a higher frequency than the incoming wave due to the absorption of a phonon. This process becomes significant when the temperature of the crystal is sufficiently high such that many optical phonons are excited and may deliver their energy to the scattered photon.

The Raman scattering can be viewed as a two-step process: In the first step, the system absorbs the photon and in the second, it emits a photon with a slightly different frequency.

Since the frequency of the electromagnetic wave is much higher than the typical phonon spectrum, the intermediate state after the absorption of the photon is a virtual state of high energy. Figure 12-3 illustrates the various processes of absorption and scattering of light in crystals. The left side of the figure shows the infrared absorption process. The second from the right pair of arrows represents the elastic Rayleigh scattering where the incoming and the outgoing wave have the same frequencies. The Stokes and the Anti-Stokes Raman scattering are illustrated from both sides of the Rayleigh scattering process.

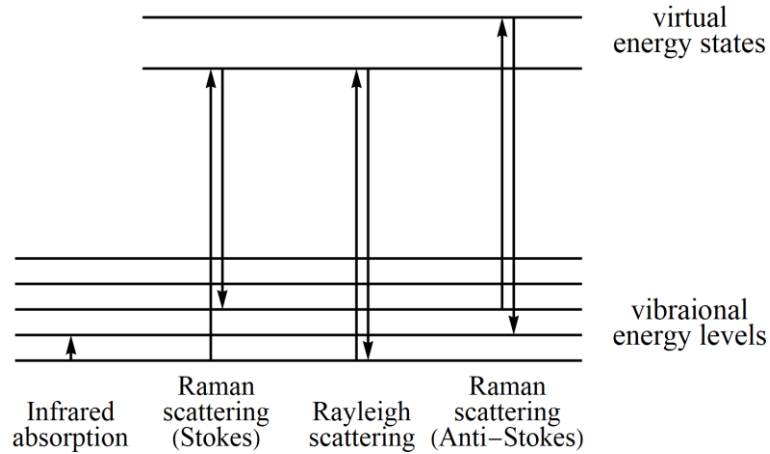


Figure 12-3 Illustration of the main processes of interaction of light and crystal vibrations

Being a two-step process, the Raman scattering is obtained by a second-order perturbation theory in the dipole Hamiltonian, H_{dipole} . A diagrammatic representation of this two-step process is shown in Fig. 12-4. Here IS denotes the initial state of the crystal, VS is its intermediate virtual state, and FS is the final state of the crystal. E and E' denote the incoming and outgoing photons, respectively.

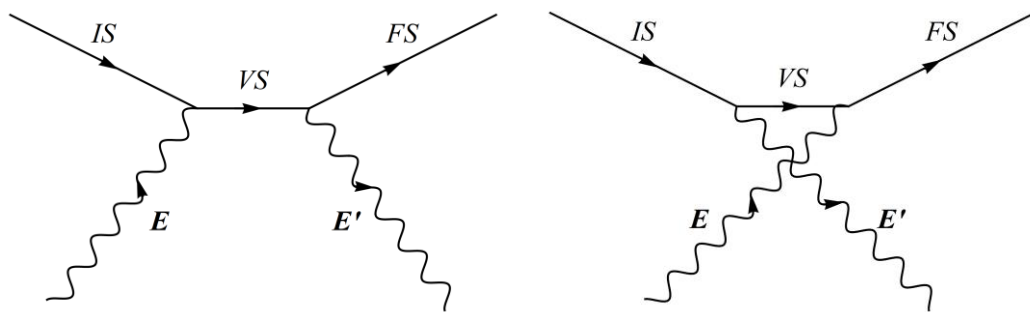


Figure 12-4 Diagrams of Raman scattering

Thus, the matrix element for Raman scattering is of the form:

$$\langle \psi_{FS} | \sum_{VS} \frac{H_{\text{dipole}} | VS \rangle \langle VS | H_{\text{dipole}}}{E_{IS} - E_{VS} + \hbar \omega} | \psi_{IS} \rangle + \langle \psi_{FS} | \sum_{VS} \frac{H_{\text{dipole}} | VS \rangle \langle VS | H_{\text{dipole}}}{E_{IS} - E_{VS} - \hbar \omega} | \psi_{IS} \rangle. \quad (12.6)$$

From here, one can deduce the selection rules for Raman scattering: Since H_{dipole} belongs to the irreducible vector representation, the above matrix element will be nonzero if the normal coordinate ξ belongs to irreducible representation with quadratic basis functions of the form $r_i r_j$. In principle, the product of two vector representations (each one of them stands for H_{dipole}) also contains the identity representation, and one may argue that a nonzero matrix element is associated with a final state that also belongs to A_1 . However, this process does not excite the lattice's optical vibrational modes required for Raman scattering. In other words, it is associated with Rayleigh scattering.

Examples

- (a) **Graphene:** We have seen that the in-plane optical phonons of graphene belong to the E_2 irreducible representation of C_{6v} . This irreducible representation has quadratic basis functions $(x^2 - y^2, 2xy)$ (see table on page 94); hence graphene is Raman active.
- (b) **Zinc blende crystals:** Here, we have seen that optical phonons belong to F_2 which is the irreducible vector representation. It contains the basis function (yz, xz, xy) (see character table on page 142); hence this family of materials is Raman active.
- (c) **Diamond:** The optical phonons of diamond belong to the F_{2g} irreducible representation, which has a basis function (yz, xz, xy) (see table on page 251). Hence diamond is Raman active.
- (d) **NaCl (salt):** Sodium chloride is an ionic crystal whose structure is shown in Fig. 12-5. Here each type of atom is located on an fcc sublattice, and the two sublattices are shifted from each other by half of the lattice constant. A unit cell of this crystal contains two atoms, and the corresponding point group is the octahedral group O_h . One can check that all symmetry operations of this group shift the atoms only within their sublattices. Hence the atom site representation is $\Gamma_{\text{atom-sites}} = 2A_{1g}$, and the lattice representation is $\Gamma_{\text{lattice}} = 2A_{1g} \otimes F_{1u} = 2F_{1u}$. This decomposition implies that the optical phonons belong to the F_{1u} irreducible representation, which does not have quadratic basis functions (see table on page 251). Hence, sodium chloride is Raman inactive.

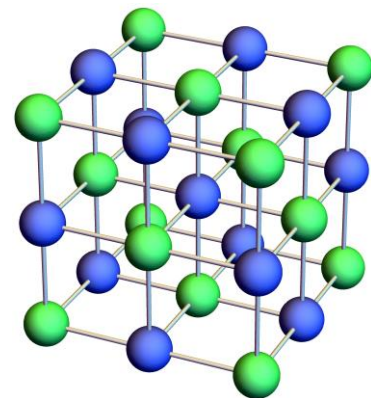


Figure 12-5 NaCl crystal

12.3 Electromagnetic waves in polarizable crystals and polaritons

This section discusses the propagation of electromagnetic waves in polarizable crystals. To begin with, let us calculate the polarization vector of the crystal due to the application of a time-dependent electric field. From Eqs. (11.87) , (12.1) and (12.2), we obtain that the Hamiltonian of a polarizable crystal subjected to electric field is:

$$H = \int d^d r \left\{ \sum_{i=1}^{\ell_\alpha} \left[\frac{p_i^2}{2} + \frac{\omega_\alpha^2}{2} \xi_i^2 \right] - \mathbf{E} \rho_p \boldsymbol{\xi} \right\}, \quad (12.7)$$

and the equations of motion are:

$$\dot{\boldsymbol{\xi}} = \mathbf{p} \quad \text{and} \quad \dot{\mathbf{p}} = -\omega_0^2 \boldsymbol{\xi} + \rho_p \mathbf{E}. \quad (12.8)$$

From now on, for simplicity, we assume that (at $k=0$) there are only three degenerate optical phonons with frequency $\omega_\alpha = \omega_0$ (hence the system is invariant to rotations). Taking the time derivative of the first equation of motion and substituting it in the second equation, we have:

$$\ddot{\boldsymbol{\xi}} = -\omega_0^2 \boldsymbol{\xi} + \rho_p \mathbf{E}. \quad (12.9)$$

The Fourier transform of this equation with respect to time allows one to solve for $\boldsymbol{\xi}$. substituting this solution in formula (12.2) for the polarization vector we obtain:

$$\mathbf{P}(\omega) = \frac{\rho_p^2 \mathbf{E}(\omega)}{\omega_0^2 - \omega^2}. \quad (12.10)$$

The electric displacement field is given by

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}_e + \mathbf{P}(\omega) = \varepsilon_0 \underbrace{\left[\varepsilon_{r,\infty} + \frac{\rho_p^2}{\varepsilon_0 (\omega_0^2 - \omega^2)} \right]}_{\varepsilon_r(\omega)} \mathbf{E}, \quad (12.11)$$

where $\mathbf{P}_e = \varepsilon_0 (\varepsilon_{r,\infty} - 1) \mathbf{E}$ is the polarization vector due to the electrons that occupy the atoms' outer shells in the crystal. We ignore the frequency dependence of $\mathbf{P}_e$ because it is only relevant at high frequencies ω - well above the phonon spectrum. The ratio of the electric displacement field to the electric field is the dielectric constant of the material. Dividing it by the permittivity of the vacuum, ε_0 , we obtain the relative dielectric constant:

$$\varepsilon_r(\omega) = \varepsilon_{r,\infty} + \frac{\rho_p^2}{\varepsilon_0 (\omega_0^2 - \omega^2)} = \varepsilon_{r,\infty} \frac{\omega^2 - \omega_L^2}{\omega^2 - \omega_0^2}, \quad (12.12)$$

where ω_L is the frequency at which $\varepsilon_r(\omega)$ vanishes, as demonstrated in Fig. 12-6. Notice that $\varepsilon_{r,\infty} = \lim_{\omega \rightarrow \infty} \varepsilon_r(\omega)$. We also denote $\varepsilon_{r,0} = \lim_{\omega \rightarrow 0} \varepsilon_r(\omega)$. In GaAs, for instance, $\varepsilon_{r,0} = 12.9$ while $\varepsilon_{r,\infty} = 10.9$.

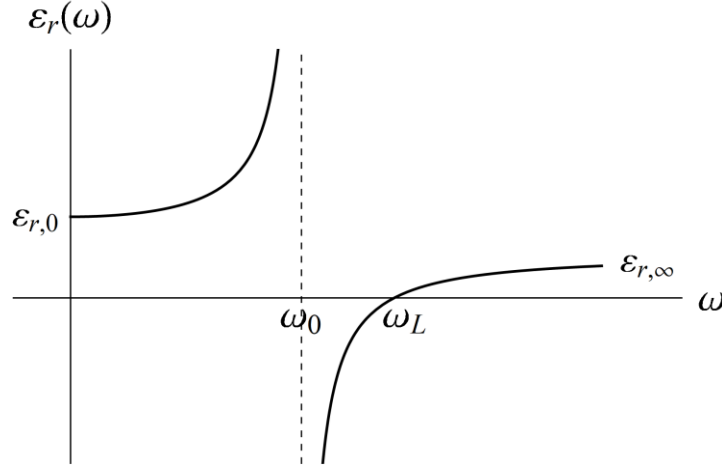


Figure 12-6 An illustration of the relative dielectric constant in polarizable crystal

To describe the propagation of electromagnetic waves in such a crystal, consider the Maxwell equations in a material that does not have free charges or free currents. Namely, when the effect of the electric field on the charge carriers can be accounted for by the polarization vector we have calculated above. Then, in Fourier space, the Maxwell equations take the form:

$$\begin{aligned} \mathbf{k} \cdot \mathbf{D} &= 0, & \mathbf{k} \cdot \mathbf{B} &= 0, \\ \omega \mathbf{B} &= \mathbf{k} \times \mathbf{E}, & \omega \mathbf{D} &= -\frac{1}{\mu_0} \mathbf{k} \times \mathbf{B}. \end{aligned} \quad (12.13)$$

Substituting Faraday's law in Ampere-Maxwell law and using (12.11) gives

$$\omega^2 \mathbf{D} = \omega^2 \varepsilon_0 \varepsilon_r(\omega) \mathbf{E} = -\frac{1}{\mu_0} \mathbf{k} \times \omega \mathbf{B} = -\frac{1}{\mu_0} \mathbf{k} \times (\mathbf{k} \times \mathbf{E}) = \frac{1}{\mu_0} [\mathbf{k}^2 \mathbf{E} - (\mathbf{k} \cdot \mathbf{E}) \mathbf{k}]. \quad (12.14)$$

Now looking for a solution of transverse waves, $\mathbf{k} \cdot \mathbf{E} = 0$, and taking into account that the speed of light in a vacuum is $c^2 = 1/(\varepsilon_0 \mu_0)$, we obtain that the dispersion relation of electromagnetic waves in a polarizable medium takes the form $\omega^2 \varepsilon_r(\omega) = c^2 k^2$, i.e.

$$\varepsilon_{r,\infty} \omega^2 \frac{\omega^2 - \omega_L^2}{\omega^2 - \omega_0^2} = c^2 k^2. \quad (12.15)$$

This relation is a biquadratic equation for ω with solutions:

$$\omega_{\pm}^2 = \frac{1}{2\varepsilon_{r,\infty}} \left[c^2 k^2 + \varepsilon_{r,\infty} \omega_L^2 \pm \sqrt{(c^2 k^2 + \varepsilon_{r,\infty} \omega_L^2)^2 - 4\varepsilon_{r,\infty} \omega_0^2 c^2 k^2} \right]. \quad (12.16)$$

The behavior of these solutions, in the limits $k \rightarrow 0$ and $k \rightarrow \infty$, is:

$$\begin{aligned} \omega_- &\xrightarrow{k \rightarrow 0} \frac{c}{\sqrt{\varepsilon_{r,0}}} k, & \omega_- &\xrightarrow{k \rightarrow \infty} \omega_0, \\ \omega_+ &\xrightarrow{k \rightarrow 0} \omega_L + \frac{\omega_L^2 - \omega_0^2}{2\varepsilon_{r,0}\omega_0^2\omega_L} c^2 k^2, & \omega_+ &\xrightarrow{k \rightarrow \infty} \frac{c}{\sqrt{\varepsilon_{r,\infty}}} k. \end{aligned} \quad (12.17)$$

These solutions describe transverse waves, however, there is an additional solution describing longitudinal waves for which $\mathbf{k} \times \mathbf{E} = 0$. From Faraday's law it follows for such a solution $\mathbf{B} = 0$; hence the Ampere-Maxwell law implies $\omega \mathbf{D} = 0$. A nontrivial solution of this equation exists for $\mathbf{D} = 0$ (but $\mathbf{E} \neq 0$), i.e. when the polarization vector precisely compensates the electric field (see Eq. (12.11)). This condition is equivalent to $\varepsilon_r(\omega) = 0$, thus

$$\omega = \omega_L. \quad (12.18)$$

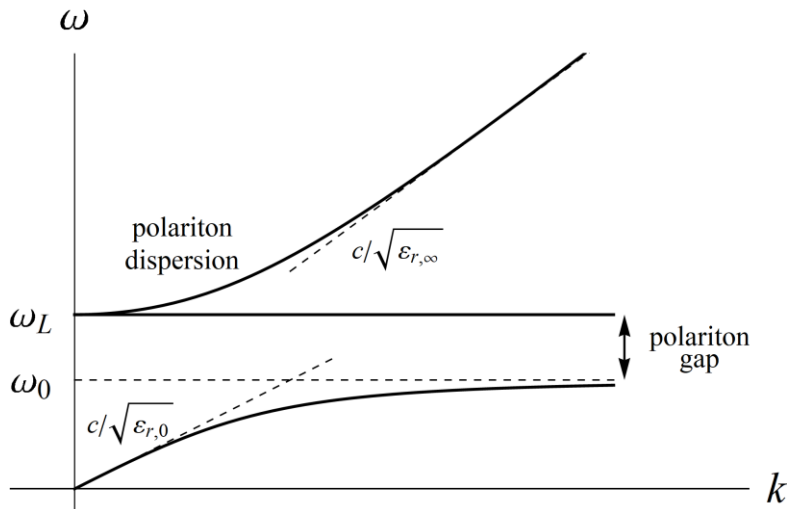


Figure 12-7 Dispersion relation curves in a polarizable crystal

The solutions that we have derived above are depicted in Fig. 12-7. Three different curves represent them. Two curves are associated with the solutions $\omega_{\pm}(k)$. Each one of these curves represents two degenerate solutions since they describe transverse waves characterized by two possible polarizations. The fifth solution (12.18) is the longitudinal wave. Here the magnetic field component vanishes. Thus, the oscillatory nature of this wave comes from two sources: The mechanical restoring force acting on the atoms and the force due to the electric field created by the motion of the atoms.

In the absence of coupling between the electromagnetic waves and lattice vibrations, these five solutions reduce to 3 phonon modes (two transverse and one longitudinal) and electromagnetic waves having two possible polarizations. However, the lesson from the above result is that in a polarizable material, one cannot decouple the lattice vibrations from the oscillations of the electromagnetic waves.

Nevertheless, in the limit $c \rightarrow \infty$ and $k \rightarrow 0$ such that $ck \gg \omega_L$ the solutions of the dispersion relations reduce to the following: Two degenerate solutions $\omega \equiv \omega_{\text{TO}} = \omega_0$ that describe the transverse optical vibrations of the lattice (essentially in the absence of electromagnetic field since the wavelength is very large), and an additional solution, $\omega \equiv \omega_{\text{LO}} = \omega_L$, describing the longitudinal optical vibrations of the lattice. These are the three solutions for the optical vibrations in zinc-blende materials discussed in the previous chapter. From the above result, one can show (the proof is given as an exercise) that the frequency ratio of the longitudinal to the transverse optical vibrations satisfies the Lyddane-Sachs-Teller (1941) relation:

$$\frac{\omega_{\text{LO}}^2}{\omega_{\text{TO}}^2} = \frac{\epsilon_{r,0}}{\epsilon_{r,\infty}}. \quad (12.19)$$

Two additional solutions describe the dispersion of electromagnetic waves $\omega = ck/\sqrt{\epsilon_{r,\infty}}$, with a refraction index $n = 1/\sqrt{\epsilon_{r,\infty}}$ that results from the electronic polarization of the crystal.

In the opposite limit, $ck \ll \omega_0$, the solution $\omega_- \simeq ck/\sqrt{\epsilon_{r,0}}$ describes the dispersion of a transverse electromagnetic wave moving in material with a refractive index determined by the crystal's electronic and ionic polarizations, $n = 1/\sqrt{\epsilon_{r,0}}$. On the other hand, $\omega_+(k)$ is the dispersion relation of a quasiparticle called *polariton* that behaves quadratically in the limit $k \rightarrow 0$. A polariton is an excitation of the system in the limit where the coupling between the electromagnetic field and the lattice vibrations is strong, because the frequency of the electromagnetic waves is in resonance with the frequency of the lattice vibrations. In this limit, the photon and the phonon cannot be considered to be separate particles - they are strongly entangled and form one quasiparticle.

12.4 Exercises

1. α - graphite is a crystal made carbon atoms obtained from staking graphene layers one on top of the other in a periodic form, ABABAB... such that layer B is obtained from the shift of layer A by $\mathbf{l} = (\mathbf{a}_1 + \mathbf{a}_2)/3 + \mathbf{c}/2$ where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the primitive basis vectors in the graphene plane, and $\mathbf{c}$ is the primitive basis vector in the perpendicular direction, see Fig. 12-8. Each unit cell contains four atoms which are represented by different colors in the figure.

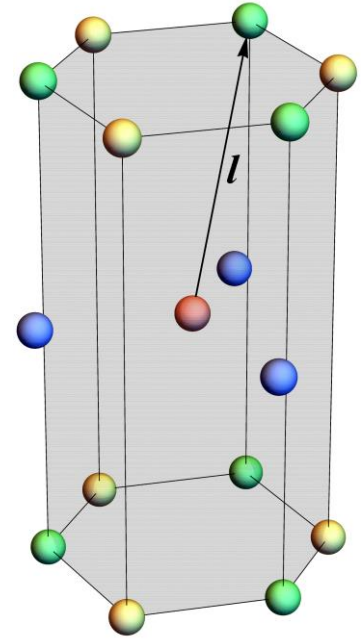


Figure 12-8 A unit cell of α graphite

The space group of this crystal is nonsymmorphic. In addition to the symmetry operations of D_{3h} it contains the following operations: $2\{c_6|\mathbf{l}\}$, $\{c_2|\mathbf{l}\}$, $3\{c_2''|\mathbf{l}\}$, $\{i|\mathbf{l}\}$, $2\{S_6|\mathbf{l}\}$, and $3\{\sigma_d|\mathbf{l}\}$. This group is isomorphic to D_{6h} group whose character table is given below

α graphite	E	$2\{c_2' \mathbf{l}\}$	$2c_3$	$\{c_2 \mathbf{l}\}$	$3c_2'$	$3\{c_2'' \mathbf{l}\}$	$\{i \mathbf{l}\}$	$2S_3$	$2\{S_6 \mathbf{l}\}$	σ_h	$3\{\sigma_d \mathbf{l}\}$	$3\sigma_v$		
D_{6h}	E	$2c_6$	$2c_3$	c_2	$3c_2'$	$3c_2''$	i	$2S_3$	$2S_6$	σ_h	$3\sigma_d$	$3\sigma_v$		
A_{1g}	1	1	1	1	1	1	1	1	1	1	1	1	R_z	$x^2 + y^2, z^2$
A_{2g}	1	1	1	1	-1	-1	1	1	1	1	-1	-1		
B_{1g}	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1		
B_{2g}	1	-1	1	-1	-1	1	1	-1	1	-1	-1	1		
E_{1g}	2	1	-1	-2	0	0	2	1	-1	-2	0	0	(R_x, R_y)	(xz, yz)
E_{2g}	2	-1	-1	2	0	0	2	-1	-1	2	0	0		
A_{1u}	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	z	$(x^2 - y^2, 2xy)$
A_{2u}	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1		
B_{1u}	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1		
B_{2u}	1	-1	1	-1	-1	1	-1	1	-1	1	1	-1		
E_{1u}	2	1	-1	-2	0	0	-2	-1	1	2	0	0	(x, y)	
E_{2u}	2	-1	-1	2	0	0	-2	1	1	-2	0	0		

Identify the characters of atom site representation and show that the composition of the lattice representation is given by

$$\Gamma_{\text{lattice}} = 2(A_{2u} \oplus B_{2g} \oplus E_{1u} \oplus E_{2g})$$

Identify the composition of the vibrational modes and show that α -graphite is infrared active.

2. Identify the optical vibrational modes of graphene, which also include those that are out-of-plane. Draw the out-of-plane optical phonon of the system. Is graphene infrared active?
3. Is graphite Raman active?
4. Prove the *rule of mutual exclusion*, which states that in a system with a center of symmetry (i.e., symmetric to inversion), vibrational modes that are infrared active are Raman inactive and vice-versa.
5. Prove the Lyddane-Sachs-Teller relation given by formula (12.9).

13 Piezoelectric and polar crystals

The previous chapter discussed crystals that become electrically polarized when subjected to an electric field or deformed by optical phonons. However, some crystals possess spontaneous electric polarization even at equilibrium. These materials are called *polar crystals*, and in this chapter, we briefly discuss them.

13.1 Piezoelectric crystals

According to their point group symmetries, all 230 crystal structures can be divided into 32 crystal classes. Of these, 21 classes are noncentrosymmetric, i.e., associated with point groups that do not possess inversion symmetry. Out of these 21 classes, 20 are associated with materials that may possess an electric dipole within the elementary unit cell of the crystal¹. These materials are called *piezoelectric crystals*. Their prominent property is the ability to change their electric polarization vector, $\mathbf{P}$, by mechanical stress. Ten crystals classes out of the piezoelectric class are called *polar-neutral*. In these materials $\mathbf{P} = 0$ in equilibrium but $\mathbf{P} \neq 0$ when the crystal is mechanically deformed. Namely,

$$P_i = \gamma_{i;kl} u_{kl}, \quad (13.1)$$

where u_{kl} is the strain vector. The tensor $\gamma_{i;kl}$ is called the *piezoelectric tensor*.

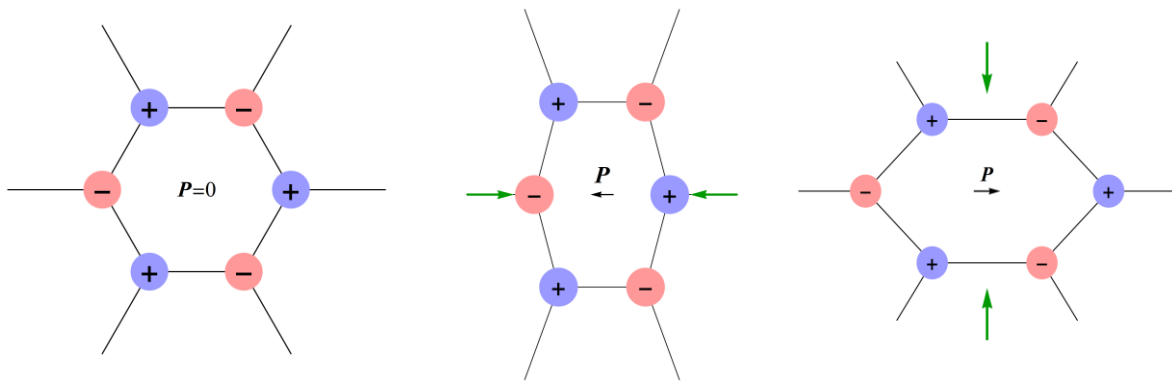


Figure 13-1 the mechanism of the piezoelectric effect in noncentrosymmetric crystals

The mechanism that generates polarization by stress is illustrated in Fig. 13-1. The left panel shows a unit cell of a crystal with no stress. Here, the “center of mass” of the negative and the positive charge is precisely at the center of the cell; hence the polarization vector vanishes. In the middle panel, compressive stress in the horizontal direction is applied to the crystal. Now the center of mass of the negative and positive charges shift in opposite

¹ The classes are: $C_1, C_2, C_{1h}, C_3, C_4, S_4, D_2, C_{2v}, C_{3h}, D_3, C_{3v}, C_6, D_4, C_{4v}, D_{2d}, D_{3h}, D_6, C_{6v}, T$, and T_d . The 21st class is associated with the O point group (see Ex. 1).

directions and create a polarization vector. In the right panel, the compressive stress is applied in the vertical direction. Here again, the lattice deformation shifts the centers of mass of the charges, resulting in polarization (albeit in the opposite direction from horizontal stress).

Identifying the non-zero elements of the piezoelectric tensor is obtained in a similar way as in the Pockless effect discussed in section 9.1. Namely, to ensure that Eq. (13.1) is satisfied under all symmetry operations, one has to look for irreducible representations having, both, linear and quadratic basis functions. Notice that it is impossible in centrosymmetric materials.

Example: Piezoelectricity in C_{3v} crystals

Consider the case of a lattice with C_{3v} symmetry (a two-dimensional version of such a crystal is shown in Fig. 13-1). Here the A_1 and the E irreducible representations have both linear and quadratic basis functions, therefore:

C_{3v}		
A_1	z	$x^2 + y^2, z^2$
A_2		
E	(x, y)	$(x^2 - y^2, 2xy), (xz, yz)$

$$P_z = \gamma_1^{(A_1)} (u_{xx} + u_{yy}) + \gamma_2^{(A_1)} u_{zz},$$

$$\begin{pmatrix} P_x \\ P_y \end{pmatrix} = \gamma_1^{(E)} \begin{pmatrix} u_{xx} - u_{yy} \\ 2u_{xy} \end{pmatrix} + \gamma_2^{(E)} \begin{pmatrix} u_{xz} \\ u_{yz} \end{pmatrix}. \quad (13.2)$$

The above equation shows that compression in the x or in the y direction results in polarization vector in the x direction (as illustrated in Fig. 13-1), while shear stress gives rise to a finite u_{xy} component of the strain tensor, which results in polarization in the y direction.

Example: Piezoelectricity in C_{6v} crystals

In crystals with C_{6v} group symmetry, the A_1 and the E_1 irreducible representations have both linear and quadratic basis functions, hence:

$$P_z = \gamma_1^{(A_1)} (u_{xx} + u_{yy}) + \gamma_2^{(A_1)} u_{zz}$$

$$\begin{pmatrix} P_x \\ P_y \end{pmatrix} = \gamma^{(E_1)} \begin{pmatrix} u_{xz} \\ u_{yz} \end{pmatrix} \quad (13.3)$$

C_{6v}		
A_1	z	$x^2 + y^2, z^2$
A_2		
B_1		
B_2		
E_1	(x, y)	(xz, yz)
E_2		$(x^2 - y^2, 2xy)$

Notice, however, that if the strain tensor does not have a component in the z direction, one cannot obtain polarization in the xy plane.

Example: Piezoelectricity in zinc-blende crystals

The symmetry of zinc-blende crystals is T_d , and from the character table on page 90, one can see that only F_2 irreducible representation has, both, linear and quadratic basis functions. Therefore, there is only one constant that characterizes the piezoelectric effect in these materials:

$$\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} = \gamma^{(F_2)} \begin{pmatrix} u_{yz} \\ u_{xz} \\ u_{xy} \end{pmatrix}. \quad (13.4)$$

The symmetry group T_d is the highest symmetry that enables piezoelectricity. Finally, we comment that diamond has a similar lattice structure; however, its symmetry group is O_h (see table on page 252), and the irreducible representation of vectors, F_{1u} , does not have quadratic basis functions. Therefore, diamond is not piezoelectric material.

13.2 Pyroelectric crystals

In the previous section, we mentioned the ten crystal classes of the 20 piezoelectric classes that are polar-neutral. The other 10 are crystals that exhibit spontaneous polarization, $\mathbf{P} \neq 0$, at equilibrium. These crystals are called *pyroelectric*. An example of such a crystal is zinc oxide, ZnO , whose crystal structure is shown in Fig. 13-2. It is made of two interpenetrating close-packed hexagonal lattices² of zinc and the other of oxygen. The oxygen layers are shifted from the middle point between the layers of zinc. Thus, each atom of one kind is surrounded by four atoms of the other kind, creating a tetrahedron. This structure is called *wurtzite crystal*. Since the valences of oxygen and zinc atoms are different, it has a nonzero polarization vector.

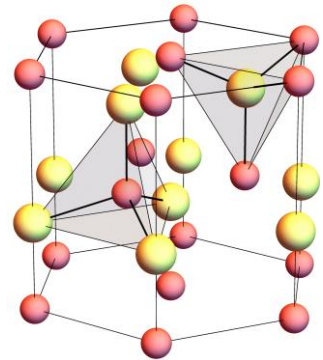


Figure 13-2 The crystal structure of ZnO

The polarization charge density is given by

$$\rho_{\text{pol}} = -\nabla \cdot \mathbf{P}, \quad (13.5)$$

and from Gauss law, $\epsilon_0 \nabla \cdot \mathbf{E} = \rho_{\text{pol}}$, we know that

$$\mathbf{E} = -\mathbf{P}/\epsilon_0, \quad (13.6)$$

² A closed packed hexagonal (cph) lattice is made from layers of two dimensional hexagonal lattice ordered in an alternating manner such that nearby layers are shifted from each other by half of the lattice constant (as shown by the red sphere in Fig. 13-2).

where ε_0 is the permittivity of the vacuum. An illustration of the microscopic behavior of the electric field in a pyroelectric crystal is presented in Fig. 13-3. This figure highlights the following main features:

- (a) The polarization charge is accumulated on the surface of the system. This property follows from Eq. (13.5) when $\mathbf{P}$ is assumed to be constant within the crystal and zero outside.
- (b) The electric field outside the crystal vanishes.
- (c) Within the sample, the electric field changes rapidly in space (on the scale of the lattice constant); however, its average is finite. This average is essentially the field generated by the effective accumulated charge on the surface of the crystal.

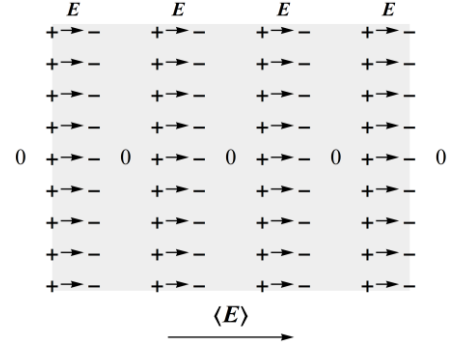


Figure 13-3 The electric field in pyroelectric crystal

An important point that deserves attention is that although the charge accumulates only on the surface, the energy associated with such a configuration is proportional to the system's volume. It is because the energy is given by the integral of the square of the electric field over the whole volume of the system. This property implies that in reality, “free” charges coming from impurities and the ambient atmosphere will adhere to the surface to compensate the electric field and lower the energy.

Thus, under normal circumstances, polar materials do not display a net electric dipole moment. Nevertheless, the polarization vector is temperature dependent; therefore, changes in the polarization vector, $\Delta \mathbf{P}$, of pyroelectric crystals can be detected by changing the temperature. In particular, if the temperature is changed by ΔT :

$$\Delta P_i = \kappa_i \Delta T; \quad i = x, y, z, \quad (13.7)$$

where κ_i , are the pyroelectric coefficients. This material property is the *pyroelectric effect*.

Since $\mathbf{P}$ is a vector that transforms according to the symmetry operations of the group, while κ_i is material property independent of symmetry operations, the only way of fulfilling the above relation is in crystals with a group symmetry in which the identity representation A_1 contains linear basis functions. The crystals (within the piezoelectric class) having this property are characterized by a group symmetry belonging to one of the *polar point groups*, $C_1, C_2, C_{1h}, C_{2v}, C_3, C_{3v}, C_4, C_{4v}, C_6$, and C_{6v} . This class is called the *pyroelectric class*.

13.3 Ferroelectric crystals

Pyroelectric crystals having the property that their polarization vector can be reversed by application of an external electric field are called *ferroelectric*. An additional property of these materials is that nonzero polarization appears only below some critical temperature, T_c , thus

$$\mathbf{P} = 0 \quad \text{for} \quad T > T_c, \quad (13.8)$$

while

$$\mathbf{P} \neq 0 \quad \text{for} \quad T < T_c. \quad (13.9)$$

A prototype family of ferroelectric materials is the family of perovskite oxides. These materials have a chemical composition ABO_3 . Above the critical temperature, and in ideal situations, they form a lattice of cubic symmetry, as shown on the left panel of Fig. 13-4. In this lattice, the A cations form a simple cubic lattice; the O anions are located on the corners of a regular octahedron, while the B cations are at the center of each cell. When the temperature reduces below some critical value, the unit cell undergoes a structural phase transition, as shown on the right panel of Fig. 13-4: The B cation shifts from the center of the cell, and lattice symmetry reduces to C_{4v} . This phase transition is accompanied by a finite polarization vector. Notice, however, that it is a *spontaneous symmetry breaking* since the energy for the case where the B cation moves up is the same as if it moves down or sideways (in any direction of one of the oxygen atoms).

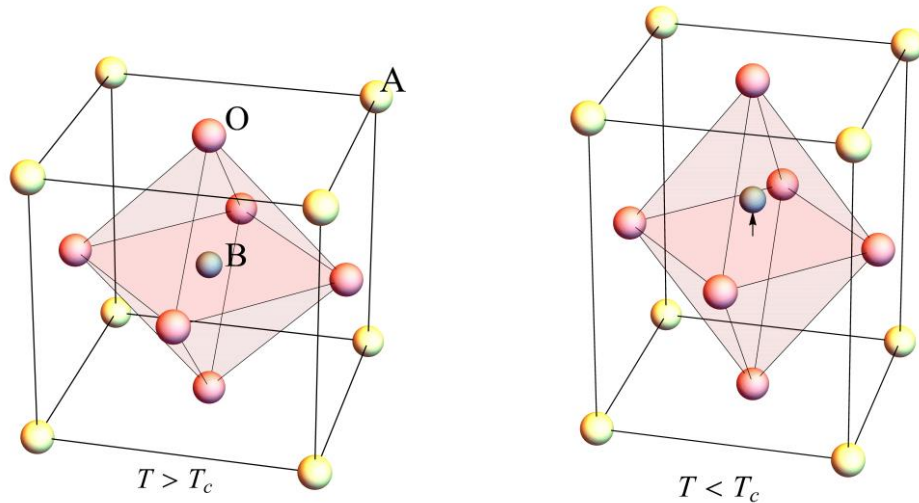


Figure 13-4 An illustration of the ferroelectric phase transition in perovskite oxides.

To understand the mechanism for the ferroelectric transition shown in the above figure, let us construct the system's energy using symmetry considerations. The energy contains three main terms:

$$\mathcal{E} = \mathcal{E}_P + \mathcal{E}_{\text{elastic}} + \mathcal{E}_{\text{coupling}} \quad (13.10)$$

The first is the energy due to electric polarization. To the lowest order in the polarization vector, it should be of the form:

$$\varepsilon_P = \sum_{ij} \chi_{ij}^{(2)} P_i P_j + \sum_{ijkl} \chi_{ijkl}^{(4)} P_i P_j P_k P_l + \dots \quad (13.11)$$

This expansion contains only even powers of the polarization vector because odd powers cannot produce a scalar. Demanding also that the above expansion is invariant under all symmetry operations of the O_h group, and using its quadratic basis functions, we obtain:

$$\varepsilon_P = \chi_1 (P_x^2 + P_y^2 + P_z^2) + \chi_2 (P_x^2 + P_y^2 + P_z^2)^2 + \chi_3 (P_x^2 P_y^2 + P_x^2 P_z^2 + P_y^2 P_z^2). \quad (13.12)$$

(Notice that the term associated with the square of the norm of the basis function of the E_g irreducible representation yield terms that can be absorbed into the quartic terms of the above expansion).

The second contribution to the total energy of the system is the elastic energy of deformation, which is given by:

O_h	
A_{1g}	$x^2 + y^2 + z^2$
E_g	$[2z^2 - x^2 - y^2, \sqrt{3}(y^2 - x^2)]$
F_{2g}	(xy, xz, yz)

$$\begin{aligned} \varepsilon_{\text{elastic}} = & \frac{1}{2} \Xi_{A_1} (u_{xx} + u_{yy} + u_{zz})^2 + \frac{1}{2} \Xi_{E_g} (u_{xy}^2 + u_{xz}^2 + u_{yz}^2) \\ & + \frac{1}{2} \Xi_{E_g} [(2u_{zz} - u_{xx} - u_{yy})^2 + 3(u_{yy} - u_{xx})^2]. \end{aligned} \quad (13.13)$$

This energy is the same as that of crystals with tetrahedral symmetry, T_d , (see Eq. (11.31)).

Finally, the coupling between the lattice deformation and the polarization vector is of the form:

$$\varepsilon_{\text{coupling}}(u_{ij}) = - \sum_{ijkl} \eta_{ij;kl} u_{ij} P_k P_l. \quad (13.14)$$

Requiring this energy to be a singlet we obtain:

$$\begin{aligned} \varepsilon_{\text{coupling}} = & -\eta_{A_1} (u_{xx} + u_{yy} + u_{zz}) (P_x^2 + P_y^2 + P_z^2) - \eta_{E_g} (u_{xy} P_x P_y + u_{xz} P_x P_z + u_{yz} P_y P_z) \\ & - \eta_{E_g} [(2u_{zz} - u_{xx} - u_{yy}) (2P_z^2 - P_x^2 - P_y^2) + 3(u_{yy} - u_{xx}) (P_y^2 - P_x^2)]. \end{aligned} \quad (13.15)$$

Now, let us assume that polarization may be generated along the z -axis, i.e., assume $P_x = P_y = 0$, and use the symmetry in the xy plane to set $u_{yy} = u_{xx}$. One can see that $P_x = P_y = 0$ implies that there is no coupling of the polarization vector to shear deformations; therefore, we also set $u_{xy} = u_{xz} = u_{yz} = 0$. With these assumptions, the energy reduces to:

$$\begin{aligned}
\varepsilon &= \chi_1 P_z^2 + \chi_2 P_z^4 + \frac{1}{2} \Xi_{A_1} (2u_{xx} + u_{zz})^2 + 2\Xi_{E_g} (u_{zz} - u_{xx})^2 \\
&\quad - \eta_{A_1} (2u_{xx} + u_{zz}) P_z^2 - 4\eta_{E_g} (u_{zz} - u_{xx}) P_z^2 \\
&= \chi_1 P_z^2 + \chi_2' P_z^4 + \frac{1}{2} \Xi_{A_1} \left(2u_{xx} + u_{zz} - \frac{\eta_{A_1} P_z^2}{\Xi_{A_1}} \right)^2 + 2\Xi_{E_g} \left(u_{zz} - u_{xx} - \frac{\eta_{E_g} P_z^2}{\Xi_{E_g}} \right)^2,
\end{aligned} \tag{13.16}$$

where to obtain the second equality, we have completed the squares and redefined the coefficient χ_2 to include the contribution of the quartic terms obtained by this procedure.

The condition for minimum energy follows from the equations:

$$\frac{\partial \varepsilon}{\partial P_z} = \frac{\partial \varepsilon}{\partial u_{xx}} = \frac{\partial \varepsilon}{\partial u_{zz}} = 0. \tag{13.17}$$

Taking first the derivatives with respect to the strain tensor, summing and subtracting the resulting equations, we obtain that minimization with respect to the strain vector implies that:

$$2u_{xx} + u_{zz} - \frac{\eta_{A_1} P_z^2}{\Xi_{A_1}} = 0, \quad u_{zz} - u_{xx} - \frac{\eta_{E_g} P_z^2}{\Xi_{E_g}} = 0, \tag{13.18}$$

with the solution

$$u_{xx} = u_{yy} = \frac{\chi_1 \Xi_{E_g} - \chi_2' \Xi_{A_1}}{3\Xi_{E_g} \Xi_{A_1}} P_z^2, \quad \text{and} \quad u_{zz} = \frac{\chi_1 \Xi_{E_g} + 2\chi_2' \Xi_{A_1}}{3\Xi_{E_g} \Xi_{A_1}} P_z^2. \tag{13.19}$$

Substituting Eqs. (13.18) in the energy function (13.16) one obtains

$$\varepsilon = \chi_1 P_z^2 + \chi_2' P_z^4. \tag{13.20}$$

Minimizing this energy with respect to P_z leads to the equation:

$$P_z (\chi_1 + 2\chi_2' P_z^2) = 0. \tag{13.21}$$

Now choosing $\chi_1 = 2\alpha(T - T_c)$, we see that for $T > T_c$, the only real solution is $P_z = 0$ which also implies that the strain tensor (13.19) vanishes. However, for $T < T_c$ there are two additional solutions:

$$P_z = \pm \sqrt{\frac{\alpha(T_c - T)}{\chi_2'}}, \quad T < T_c, \tag{13.22}$$

and one can quickly check that they are of lower energy than the solution $P_z = 0$. These solutions describe spontaneous symmetry breaking associated with a non-zero value of polarization and hence deformation of the lattice as follows from Eqs. (13.19), and illustrated in the right panel of Fig. 13-4.

The energy as a function of the electric polarization, at temperatures above and below the critical temperature, is illustrated in Fig. 13-5. Notice that the energy barrier between the two polarization states (13.22), depends on the temperature and becomes smaller as the temperature approaches the critical temperature. Thus, close enough to the critical temperature, the electric field required to flip the direction of the polarization can be very small. In contrast, the energy barrier required to flip the polarization in pyroelectric crystals (which are not ferroelectric) is very high. Here, the electric field required for reversing the polarization exceeds the threshold for dielectric breakdown.

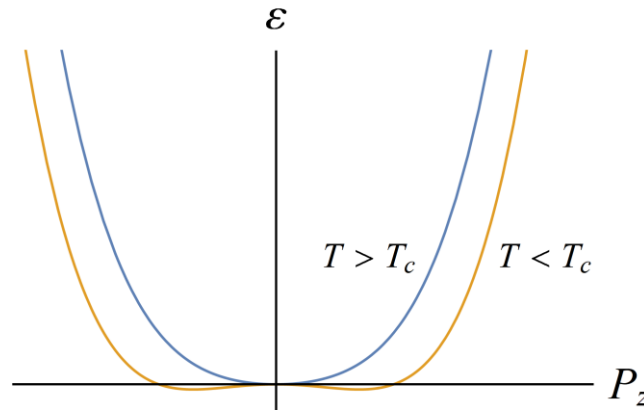


Figure 13-5 The energy as function of electric polarization in ferroelectric materials

Example: Landau-Lifshitz-Kittel domains in ferroelectric layers

Consider a layer of ferroelectric material with thickness h , width w , and length L , such that $h \ll w \ll L$. Let us assume that the crystal directions dictate the electric polarization vector to be perpendicular to the layer. Namely, the polarization charge is accumulated on the upper and lower surfaces of the layer.

Assuming constant polarization throughout the sample, the electrostatic energy required for such configuration is

$$\varepsilon_{\text{electric}}^{(0)} = \frac{\varepsilon_0}{2} \int d^3r \mathbf{E}^2 = \frac{\mathbf{P}^2}{2\varepsilon_0} Lwh, \quad (13.23)$$

where we have used Eq. (13.6).

However, it is possible to reduce this energy by creating domains with opposite polarization vectors, as illustrated in Fig. 13-6. In this configuration, the average charge accumulating on the surface is zero. In particular, if we assume that the width of each domain is a , then the electric field penetrates, essentially, only to a distance of order a into the sample. Therefore, the electrostatic energy of this configuration is obtained from Eq. (13.23) by replacing h with a :

$$\varepsilon_{\text{electric}}(a) \simeq \frac{\mathbf{P}^2}{2\varepsilon_0} Lwa. \quad (13.24)$$

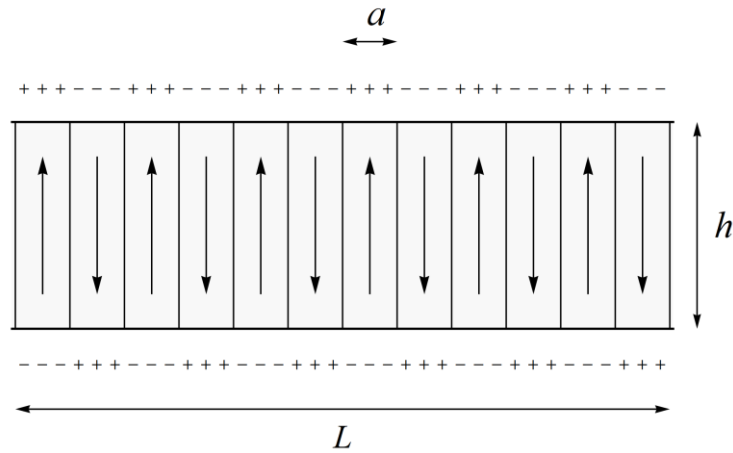


Figure 13-6 A cross-section of a ferroelectric layer in which the ferroelectric order disintegrates into domains

On the other hand, breaking the homogeneous configuration of the ferroelectric order into domains where the polarization vector points in opposite directions requires energy in order to create domain walls between two nearby domains. The energy of such a domain is proportional to its area. If we take into account that there are L/a such domain walls, the energy needed in order to create them is

$$\varepsilon_{\text{domain-wall}}(a) = \gamma h w \frac{L}{a}, \quad (13.25)$$

where γ is a constant with dimensions of surface tension energy. The last two equations show that the electrostatic energy prefers small domains, while the domain wall energy becomes smaller when domains are large. The optimal size of the domains is obtained by minimizing the total energy with respect to a :

$$\frac{d}{da} \left[\varepsilon_{\text{electric}}(a) + \varepsilon_{\text{domain-wall}}(a) \right] = \frac{d}{da} \left[\frac{\mathbf{P}^2}{2\varepsilon_0} L w a + \gamma h w \frac{L}{a} \right] = \frac{\mathbf{P}^2}{2\varepsilon_0} L w - \gamma h w \frac{L}{a^2} = 0, \quad (13.26)$$

which gives

$$a = \sqrt{\gamma \frac{2\varepsilon_0 h}{\mathbf{P}^2}}. \quad (13.27)$$

These domains are called Landau-Lifshitz-Kittel domains (Landau Lifshitz 1935, Kittel 1946).

The above result applies to the case of thin layers where anisotropy of the lattice dictates only two possible orientations of the polarization vector. In cases where the polarization vector can point in more directions and in three-dimensional samples, one may obtain different configurations of domains. An experimental picture of the domains in a three-dimensional system of the perovskite oxide BaTiO_3 is presented in Fig. 13-7.

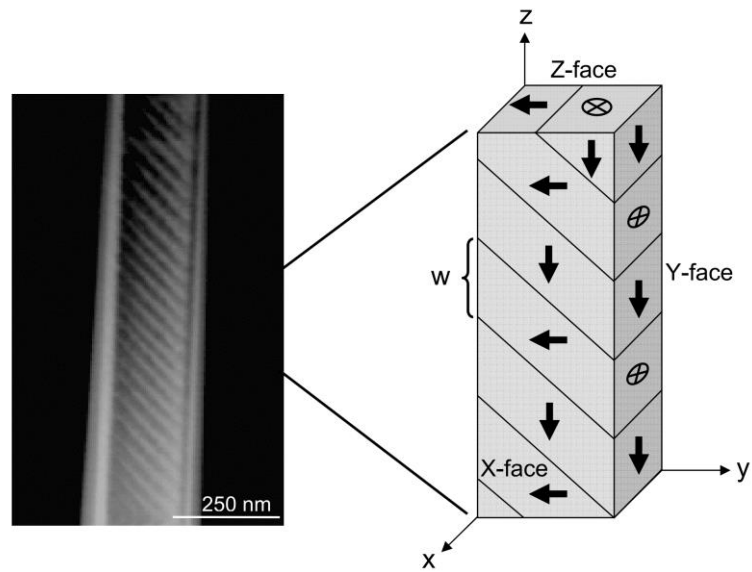


Figure 13-7 Domains in Ferroelectric rod of BaTiO_3
(Adapted from Catalan et. al, J. Phys.: Cond. Mat. **19**, 132201 (2007).)

13.4 Exercises

1. Explain why crystals described by O point group do not belong to the piezoelectric class even though this group lacks inversion symmetry.
2. Identify the piezoelectric tensor for crystals with D_4 symmetry
3. Write an expression for the total energy (due to polarization, elastic deformation, and the coupling between them) for lithium niobate³, LiNbO_3 having a rhombohedral unit cell with D_{3d} group symmetry at the paraelectric phase, see Fig. 13-8.

Minimize the energy and find the polarization value below the critical temperature. Consider two cases, one for which $P_z \neq 0$, while $P_x = P_y = 0$, and a second case where $P_z = 0$ while $P_x \neq 0$ and/or $P_y \neq 0$.

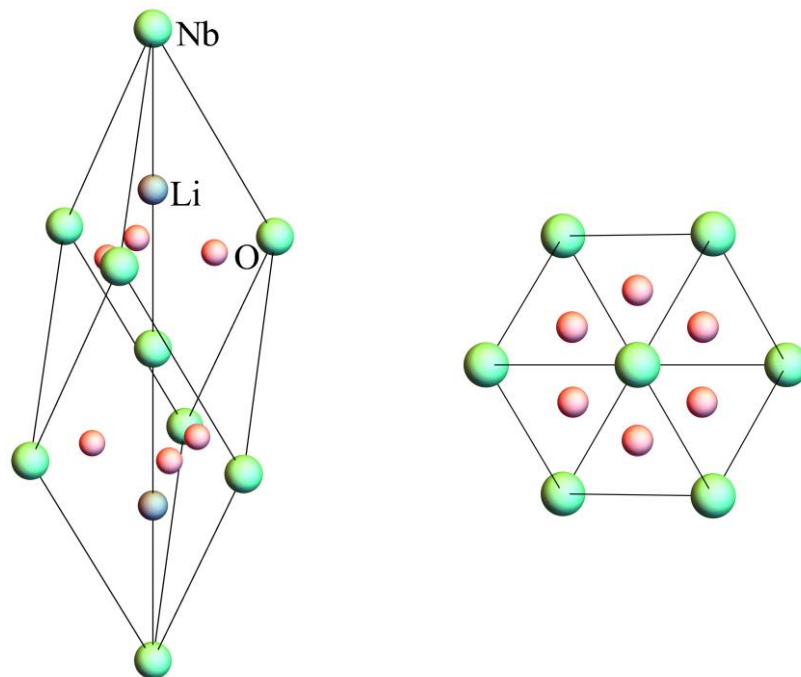


Figure 13-8 The unit cell of LiNbO_3 (at the paraelectric phase). Left panel side view, and right panel top view.

³ This unique material is used nowadays extensively in telecommunication, mobile phones, optical switches, optical waveguides, optical deflectors, surface acoustic devices, and many other applications.

14 Electrons in deformed crystals

Until now, we have discussed the electrons' behavior and the lattice deformations (whether static or dynamic) separately. However, to get a complete physical picture of a crystal, we need to consider the coupling between electrons and lattice deformation. Here we start presenting this issue by discussing the interplay between electrons and static deformations of the lattice. We begin the chapter with the effect of deformations induced by external stress on the electronic spectra of crystals and how they deform the Fermi surface. Next, we consider the effect of local deformations, created by dislocations, on the wavefunction of the electrons. Finally, we show that metallic crystals may spontaneously deform and become insulators due to the interaction between electrons and lattice deformations - a phenomenon called Peierls instability.

14.1 The effect of lattice deformation on the electron's energy spectrum

Generally, one expects lattice deformation to produce a perturbation proportional to the strain tensor. There are two types of contributions: One is a local contribution that results from a local change in the potential seen by the electrons due to the shift of atoms from their equilibrium position:

$$\delta H(\mathbf{r}) = \lambda_{ij} u_{ij}(\mathbf{r}), \quad (14.1)$$

where λ_{ij} are system-dependent constants. The second type is a nonlocal contribution that appears in piezoelectric materials. Here the polarization charge density induced by deformations,

$$\rho_{\text{pol}} = -\nabla \cdot \mathbf{P} = -\gamma_{i;kl} \frac{\partial}{\partial r_i} u_{kl}, \quad (14.2)$$

generates a long-range potential according to Coulomb's law:

$$\delta H(\mathbf{r}) = -\int d^3r' \frac{e \rho_{\text{pol}}(\mathbf{r}')}{4\pi\epsilon_0\epsilon_r |\mathbf{r} - \mathbf{r}'|} = \int d^3r' \frac{e}{4\pi\epsilon_0\epsilon_r |\mathbf{r} - \mathbf{r}'|} \gamma_{i;kl} \frac{\partial u_{kl}(\mathbf{r}')}{\partial r_i'}, \quad (14.3)$$

where ϵ_r is the relative dielectric constant of the material. Usually, this is the dominant contribution in piezoelectric materials; however, we focus only on the local contribution of lattice deformations.

Lattice deformations commonly reduce the system's symmetry; hence, they are expected to lift degeneracies in the electronic spectra. However, as we know, this should not always be the case because degenerate points may be protected by topology. In this section, we demonstrate both these scenarios by concrete examples.

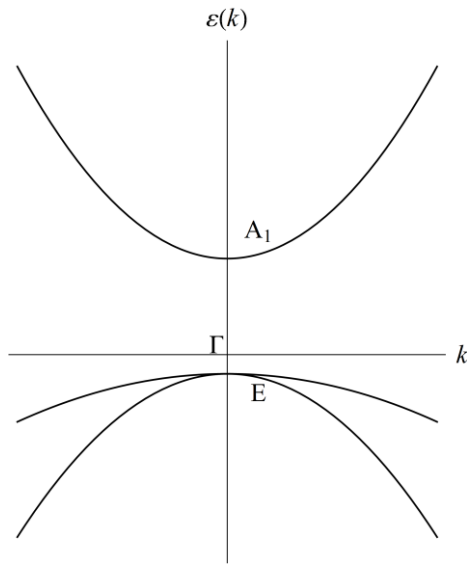


Figure 14-1 A typical behavior of energy bands near Γ point of a crystal with T_d symmetry

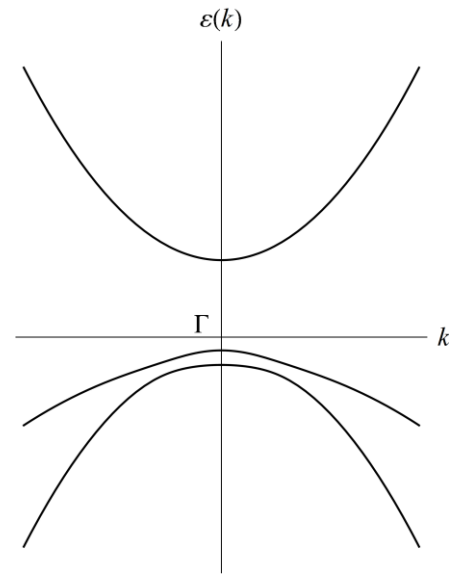


Figure 14-2 A gap opening in the spectrum of a crystal as a result of lattice deformation

Consider first the spectrum of electrons near the Γ point of a crystal with tetrahedral symmetry, T_d . The typical behavior of the energy bands in such materials is shown in Fig. 14-1. Here the upper band is nondegenerate and belongs to the A_1 irreducible representation, while the two other bands are degenerate and belong to the E irreducible representation.

To identify the structure of the Hamiltonian (14.1) of the system, one has to construct singlet terms that are linear in the strain tensor. The latter behaves as the quadratic basis functions of the irreducible representation. The quadratic basis functions of T_d are listed in the table on the right.

Consider first the energy level associated with the A_1 irreducible representations. From the table, it follows that the local Hamiltonian that takes into account the lattice deformation is of the form:

$$\varepsilon_{A_1}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m_{\text{eff}}} + \lambda_{\text{def}}^{(1)} u_{ii}, \quad (14.4)$$

T_d	
A_1	$x^2 + y^2 + z^2$
A_2	
E	$(2z^2 - x^2 - y^2, \sqrt{3}(x^2 - y^2))$
F_1	
F_2	(yz, xz, xy)

where $\lambda_{\text{def}}^{(1)}$, is a constant that characterizes the strength of the coupling between the electrons and the lattice deformation. (Recall that repeated indices are summed over). The term $\lambda_{\text{def}}^{(1)} u_{ii}$ is known as the *deformation potential*. This perturbation simply shifts the energy band.

Consider now the degenerate energy levels associated with the E representation. Assuming the Pauli matrices $\{\tau_x, \tau_y\}$ to be basis functions of this representation, the local Hamiltonian

takes the form:

$$\begin{aligned} \varepsilon_E(\mathbf{k}) = & -\frac{\hbar^2 k^2}{2m'_{\text{eff}}} + \beta \left[(2k_z^2 - k_x^2 - k_y^2) \tau_x + \sqrt{3} (k_x^2 - k_y^2) \tau_y \right] \\ & + \lambda_{\text{def}}^{(2)} \left[(2u_{zz} - u_{xx} - u_{yy}) \tau_x + \sqrt{3} (u_{xx} - u_{yy}) \tau_y \right], \end{aligned} \quad (14.5)$$

where m'_{eff} , β , and $\lambda_{\text{def}}^{(2)}$ are constants that characterize the system. The first two terms in this local Hamiltonian describe two quadratic energy levels that are degenerate at $\mathbf{k} = 0$. The third term is the perturbation due to deformation. Its form is similar to the second term, and therefore, it is a singlet. This term lifts the degeneracy at $\mathbf{k} = 0$, unless $u_{xx} = u_{yy} = u_{zz}$, as demonstrated in Fig. 14-2.

We now analyze the effect of lattice deformations on the electronic spectrum of graphene near the K and K' points of the Brillouin zone. Recall that the little group associated with these points is C_{3v} . The basis functions of this point group are listed in the table to the right. From this table, it follows that the local Hamiltonian near the K points, which takes into account planar lattice deformation, is:

C_{3v}		
A_1		$x^2 + y^2$
A_2	τ_z^{AB}	
E	$\begin{pmatrix} x+iy \\ x-iy \end{pmatrix}; \begin{pmatrix} \tau_x^{AB} + i\tau_y^{AB} \\ \tau_x^{AB} - i\tau_y^{AB} \end{pmatrix}$	$\begin{pmatrix} (x-iy)^2 \\ (x+iy)^2 \end{pmatrix}$

$$\begin{aligned} H = \hbar v \tau_z^{KK'} \otimes \tau^{AB} \cdot \delta \mathbf{k} + \lambda_{\text{def}}^{(1)} I^{KK'} \otimes I^{AB} (u_{xx} + u_{yy}) \\ + \lambda_{\text{def}}^{(2)} I^{KK'} \otimes \left[\tau_x^{AB} (u_{xx} - u_{yy}) + 2\tau_y^{AB} u_{xy} \right] \end{aligned} \quad (14.6)$$

The first term accounts for Dirac's spectrum of graphene. Here the wavenumber $\delta \mathbf{k}$ is a four-component vector describing the deviation from the degeneracy points, i.e.

$$\delta \mathbf{k} = \begin{pmatrix} \mathbf{k} - \mathbf{k}_K \\ \mathbf{k}' - \mathbf{k}_{K'} \end{pmatrix} \quad (14.7)$$

The second term, obtained from the quadratic basis function of the A_1 representation, is associated with compressive (or tensile) deformations of the lattice. This contribution shifts the whole energy spectrum and is not very interesting.

The third term is a singlet formed from the E irreducible representation, which has the same form as the triangular wrapping term (see Eq. (5.16)). It plays a role similar to vector potential because the Hamiltonian can be rewritten in the form

$$H = \hbar v \tau_z^{KK'} \otimes \tau^{AB} \cdot (\delta \mathbf{k} - \tau_z^{KK'} \otimes \mathbf{A}_{\text{def}}) + \lambda_{\text{def}}^{(1)} I^{KK'} \otimes I^{AB} (u_{xx} + u_{yy}), \quad (14.8)$$

where

$$\mathbf{A}_{\text{def}} = -\frac{\lambda_{\text{def}}^{(2)}}{v\hbar} \begin{pmatrix} u_{xx} - u_{yy} \\ 2u_{xy} \end{pmatrix}. \quad (14.9)$$

This contribution does not open a gap at the degeneracy points. It only shifts the Dirac points in the energy and the momentum space, as illustrated in Fig. 14-3. Notice that the K and K' points are shifted in opposite directions in the momentum space to preserve time-reversal symmetry.

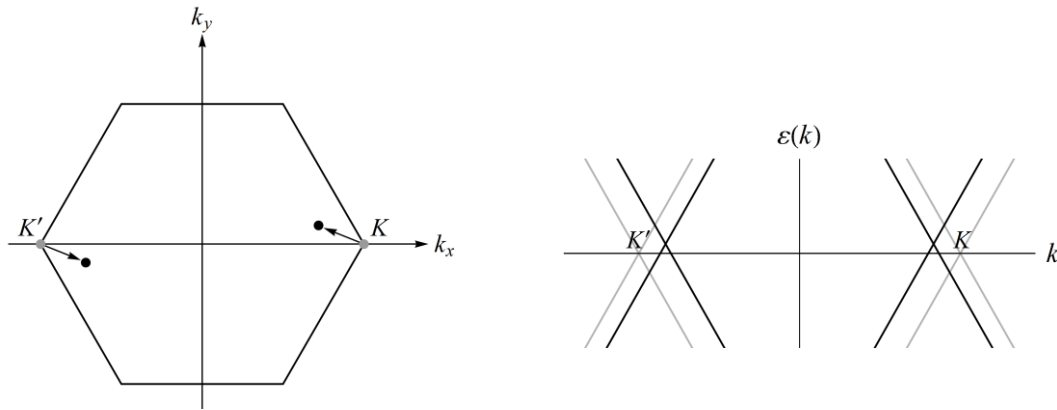


Figure 14-3 Changes in the electronic spectrum of graphene due to lattice deformations. The left panel shows the shift of Dirac points, while the right panel depicts a cross-section along k_x axis when $u_{xy} = 0$.

The Hamiltonian (14.8) implies that in the case where the lattice deformations are not homogenous in space, such that $\nabla \times \mathbf{A}_{\text{def}} \neq 0$, the system may be viewed as subject to magnetic field pointing perpendicular to the graphene layer, but with opposite signs for the particles occupying the K and the K' points (so that time-reversal symmetry is preserved).

14.2 Deformation of the Fermi surface of metals due to lattice deformations

In the previous section, we discussed the effect of lattice deformation on the degeneracy points of the spectrum. In metals, however, the Fermi surface is usually far from such points, and the effect of lattice deformation is mainly to deform the Fermi surface. To account for this effect, one cannot ignore a basic property of metal - its ability to screen electric charge effectively. As we shall see below, this feature constrains the form of deformation of the Fermi surface.

The simplest theoretical description of screening is by the Thomas-Fermi approximation. If we denote by $V(\mathbf{r})$ the potential seen by one electron due to all charges in the system (ions and other electrons) and assume that it changes slowly in space (such that a semiclassical approximation applies), then the charge density in space due to the electrons is given by

$$\rho_e(\mathbf{r}) = -e \int_{V(\mathbf{r})}^{\varepsilon_F} v d\varepsilon \simeq -ev [\varepsilon_F - V(\mathbf{r})], \quad (14.10)$$

where ν is the electron density of states which we assume to be essentially constant, for simplicity, and we have neglected temperature effects since it is usually much smaller than the Fermi energy (by a few orders of magnitude). A schematic illustration that explains the above formula is shown in Fig. 14-4.

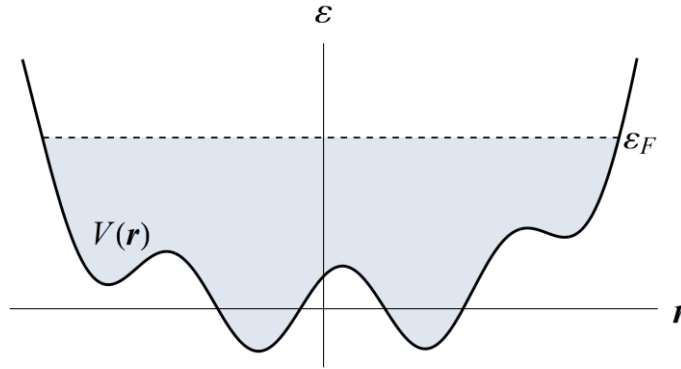


Figure 14-4 An illustration explaining formula (14.19): The density of electrons at a given point r is proportional to the number of levels between the bottom of the potential $V(r)$ and the Fermi energy ε_F .

The total electric charge in the system is

$$\rho_{\text{tot}} = \rho_e + \rho_{\text{ion}} + \rho_{\text{ext}}. \quad (14.11)$$

Here $\rho_{\text{ion}} = +e\nu\varepsilon_F$ is the positive charge due to the ions of the crystal (averaged over distances larger than the lattice constant; hence considered to be homogeneous), while ρ_{ext} is the external charge. Substituting (14.10) in (14.11), we obtain

$$\rho_{\text{tot}} = -\nu e^2 \varphi(r) + \rho_{\text{ext}} \quad (14.12)$$

where $\varphi(r) = -V(r)/e$ is the electric potential. Using Gauss law

$$\nabla \cdot \mathbf{E} = \frac{\rho_{\text{tot}}}{\varepsilon_0}, \quad (14.13)$$

where $\mathbf{E} = -\nabla \varphi(r)$, we obtain the self-consistent equation:

$$-\nabla^2 \varphi(r) = -\frac{\nu e^2}{\varepsilon_0} \varphi(r) + \frac{\rho_{\text{ext}}}{\varepsilon_0}. \quad (14.14)$$

In particular, the electric potential generated by a point charge, Q , located at the origin is obtained from the solution of

$$(-\nabla^2 + q_{\text{TF}}^2) \varphi(r) = \frac{Q}{\varepsilon_0} \delta(r), \quad (14.15)$$

where q_{TF} is the Thomas-Fermi wavenumber given by:

$$q_{TF}^2 = \frac{e^2 \nu}{\epsilon_0}. \quad (14.16)$$

The solution of Eq. (14.15) is¹ :

$$\varphi(r) = \frac{Q}{4\pi\epsilon_0 r} \exp(-q_{TF} r). \quad (14.17)$$

Thus, at distances much smaller than $r_s = 1/q_{TF}$, the potential is essentially the Coulomb potential of a point charge, while for $r \gg r_s$, the potential is essentially zero. This behavior manifests the screening of a charge by the conduction electrons and r_s is the corresponding *screening length*. In metals, r_s is of the order of the Fermi wavelength, and screening implies that these systems are quasi-neutral.

Crystal deformations do not introduce charge into the system; hence, quasi-neutrality imposes constraints on how the Fermi surface can change. In particular, the electronic charge density which is given by the integral

$$\rho = 2 \int \frac{d^d k}{(2\pi)^d} \quad (14.18)$$

over the volume enclosed by the Fermi surface in the first Brillouin zone (see illustration in Fig. 14-5), must remain unchanged. In the above formula, the factor of 2 is due to spin.

To formulate the quasi-neutrality constraint, let us assume that the Fermi surface is deformed as

$$\mathbf{k}_F(\hat{\mathbf{n}}) \rightarrow \mathbf{k}_F(\hat{\mathbf{n}}) + \delta \mathbf{k}_F(\hat{\mathbf{n}}), \quad (14.19)$$

where $\hat{\mathbf{n}}$ is a unit vector pointing in some arbitrary direction in $\mathbf{k}$ space (see figure), $\mathbf{k}_F(\hat{\mathbf{n}})$ is the Fermi wavenumber in that direction, and $\delta \mathbf{k}_F(\hat{\mathbf{n}})$ is the change in the Fermi surface

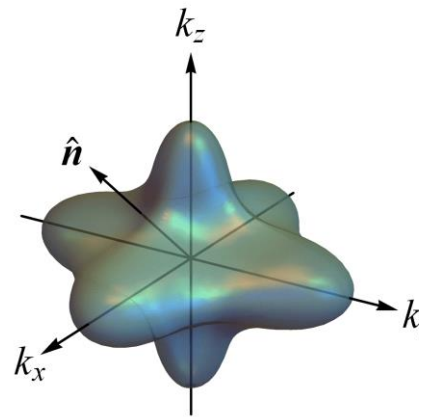


Figure 14-5 An illustration of Fermi surface in three dimensional system

¹ Taking into account the radial symmetry of the problem, and expressing the Laplacian in polar coordinate Eq. (14.15) reduces to

$$-\frac{1}{r} \frac{\partial^2}{\partial r^2} (r\varphi) + q_{TF}^2 \varphi = 0 \quad \text{for } r \neq 0.$$

Substituting $\varphi = \tilde{\varphi}/r$ one obtains a simple equation for $\tilde{\varphi}$, whose physical solution is

$$\varphi = \frac{b}{r} \exp(-q_{TF} r).$$

The constant b is now determined from the requirement that in the limit $r \rightarrow 0$, φ should reduce to the Coulomb potential of a point charge.

due to the lattice deformation. Quasi-neutrality requires that the change in the electron density vanishes, i.e.

$$\delta\rho = \frac{2}{(2\pi)^d} \oint d\hat{\mathbf{n}} \cdot \delta\mathbf{k}_F(\hat{\mathbf{n}}) k_F^2(\hat{\mathbf{n}}) = 0. \quad (14.20)$$

Here $d\hat{\mathbf{n}} = d\Omega\hat{\mathbf{n}}$ and $d\Omega$ is an infinitesimal element of the solid angle in $\mathbf{k}$ space. Now the most general distortion of the Fermi surface, which is linear in the strain tensor, is:

$$\delta k_{F,i} = \Lambda_{ij;kl} u_{kl} \hat{n}_j, \quad (14.21)$$

where $\Lambda_{ij;kl}$ is some general tensor. Substituting this formula in (14.19) yields the following condition on $\Lambda_{ij;kl}$:

$$\oint d\mathbf{n}_i \Lambda_{ij;kl} u_{kl} n_j k_F^2(\hat{\mathbf{n}}) = 0. \quad (14.22)$$

Example – The spherical approximation for $\Lambda_{ij;kl}$

The spherical approximation for $\Lambda_{ij;kl}$ is

$$\Lambda_{ij;kl} = \Lambda_1 \delta_{ij} \delta_{kl} + \frac{\Lambda_2}{2} (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}), \quad (14.23)$$

where at this point Λ_1 and Λ_2 are free parameters. However, substituting (14.23) in (14.22) yields:

$$\begin{aligned} & \oint d\mathbf{n}_i \left[\Lambda_1 \delta_{ij} \delta_{kl} + \frac{\Lambda_2}{2} (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \right] u_{kl} n_j k_F^2(\hat{\mathbf{n}}) \\ &= \oint [\Lambda_1 u_{kk} d\mathbf{n} \cdot \mathbf{n} + \Lambda_2 u_{il} n_l d\mathbf{n}_i] k_F^2(\hat{\mathbf{n}}) = 0 \end{aligned} \quad (14.24)$$

If we assume that $k_F^2(\hat{\mathbf{n}})$ is symmetric for reflections:

$$k_F^2(n_x, n_y, n_z) = k_F^2(-n_x, n_y, n_z) = k_F^2(n_x, -n_y, n_z) = k_F^2(n_x, n_y, -n_z), \quad (14.25)$$

then in the integral over $k_F^2(\hat{\mathbf{n}}) \Lambda_2 u_{il} n_l d\mathbf{n}_i$ only diagonal terms survive $k_F^2(\hat{\mathbf{n}}) \Lambda_2 (u_{ii} n_i d\mathbf{n}_i)$; thus quasi-neutrality implies

$$\Lambda_1 = -\frac{\Lambda_2}{3} \equiv \Lambda. \quad (14.26)$$

From here, it follows that the Fermi surface deformation is

$$\delta k_{F,i} = \Lambda \left[u_{il} n_l - \frac{1}{3} u_{kk} n_i \right]. \quad (14.27)$$

In particular, compressive deformation where $u_{ij} \propto \delta_{ij}$ implies that $\delta\mathbf{k}_F = 0$.

14.3 Dislocations

Perfect crystals are extremely rare. In reality, any crystal has defects that destroy its perfect periodic structure. A common defect is a *dislocation*, illustrated in Fig. 14-6 for a two-dimensional square lattice. Here an extra row of lattice points appears on the right side of the lattice, and far from the endpoint of this row (the dislocation center), the lattice seems to have a perfect structure². Thus, dislocation represents a nonlocal lattice deformation, which is noticeable only when encircling the dislocation. The mathematical characterization of dislocations is obtained using the *Burgers vector* defined by the following integral

$$b_k = \oint dr_i \frac{\partial u_k}{\partial r_i} \quad (14.28)$$

along a closed contour that encircles the dislocation, see Fig. 14-6. The value of this integral is independent of the contour's shape or its distance from the dislocation point as long as it encircles the dislocation center.

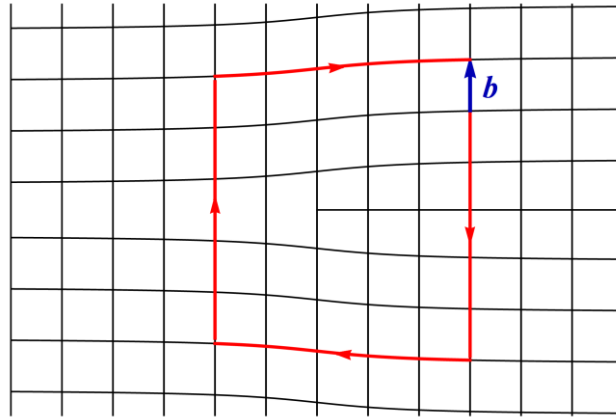


Figure 14-6 A dislocation in a two-dimensional square lattice and the definition of the Burgers vector

As we shall see in this section, dislocations may play a role similar to magnetic flux lines. We begin the section with a reminder of the Aharonov-Bohm effect due to magnetic flux lines, then discuss dislocations in graphene, and finally, in the framework of the $k \cdot p$ approximation.

Reminder: Magnetic flux lines and Aharonov-Bohm phase

A magnetic flux line in a two-dimensional system is obtained from a magnetic field perpendicular to the system and concentrated at a point. For instance, if the flux line is located at the origin of the xy plane, then

² Dislocations play a central role in explanation of the plastic properties of materials. In particular they explain large difference (of several orders of magnitude) between the experimental measurement of the force needed for plastic deformation, and the theoretical results which assume perfect lattice structure.

$$B_z(\mathbf{r}) = \phi \delta(\mathbf{r}), \quad (14.29)$$

so that at any point $\mathbf{B} = \nabla \times \mathbf{A} = 0$ but still

$$\oint \mathbf{A} \cdot d\mathbf{r} = \phi, \quad (14.30)$$

as flows from Green's theorem. Thus, although the magnetic field is zero, the vector potential is not, and the Hamiltonian of a particle with charge q is given by

$$H = \frac{(-i\hbar\nabla - q\mathbf{A})^2}{2m} + V(\mathbf{r}) \quad (14.31)$$

The wave function obtained from the solution of the Schrödinger equation with this Hamiltonian can be written in the form:

$$\psi(\mathbf{r}) = \exp\left(i \frac{q}{\hbar} \int \mathbf{r}' \cdot \mathbf{A}(\mathbf{r}')\right) \psi_0(\mathbf{r}). \quad (14.32)$$

Here $\psi_0(\mathbf{r})$ is the solution of Schrödinger's equation with $\mathbf{A} = 0$, and boundary conditions (around the flux line) that ensure a single-valued function, $\psi(\mathbf{r})$. The phase in the exponent of (14.32) is the *Aharonov-Bohm phase* (Aharonov & Bohm, 1959). Along a contour that encircles the flux line it equals $2\pi\phi/\phi_0$, where $\phi_0 = 2\pi\hbar/q$ is the unit quantum flux. Thus, when ϕ is not an integer multiple of ϕ_0 , the exponential factor in (14.32) must have a jump somewhere along the closed contour that encircles the flux. One can choose this jump along an arbitrary line that starts at the flux line and extends to infinity. The freedom of choosing this branch cut is due to the gauge invariance of the electromagnetic potential. In particular, for the problem considered above, one can choose it to be along the positive x axis, as demonstrated in Fig. 14-

7. Thus, if we denote by $\psi_0^{(\pm)}$ the wave function above and below the branch cut, as shown in Fig. 14-7, then:

$$\psi_0^{(-)} = \exp\left(i \frac{2\pi\phi}{\phi_0}\right) \psi_0^{(+)} \quad (14.33)$$

This condition and the requirement that $\psi(\mathbf{r})$ is a single-valued function are manifested in a shift of the energy levels and the existence of persistent currents.

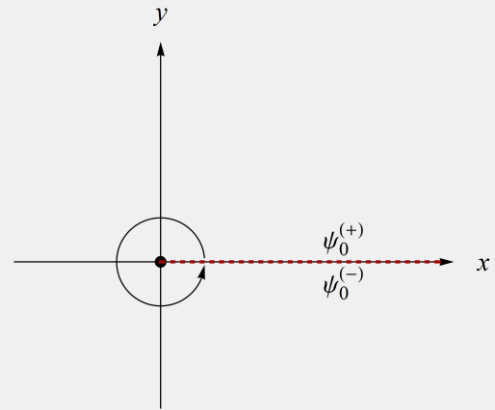


Figure 14-7 The branch cut of flux line

To illustrate these features, consider the simple example of a free particle of mass m moving in a ring of radius R threaded by a flux line, ϕ . The wave functions of the system are $\psi_0(x) = b \exp(ikx)$, where x is the coordinate along the ring, k is a free parameter, and b is the normalization constant. The condition (14.33) yields

$$b = b \exp(ik2\pi R) \exp\left(i \frac{2\pi\phi}{\phi_0}\right) \quad (14.34)$$

and, hence,

$$2\pi kR + \frac{2\pi\phi}{\phi_0} = 2\pi n, \quad (14.35)$$

where n is an integer. Thus,

$$k = \frac{n - \phi/\phi_0}{R}, \quad (14.36)$$

and the energy levels of the system of the particle are:

$$\varepsilon_n = \frac{\hbar^2}{2mR^2} \left(n - \frac{\phi}{\phi_0}\right)^2 \quad (14.37)$$

The electric current associated with the n -th state is

$$j_n = -e \frac{\hbar k_n}{m} = -e \frac{\hbar}{mR} \left(n - \frac{\phi}{\phi_0}\right) \quad (14.38)$$

Assuming $|\phi| < \phi_0/2$, and an even number of particles (with two possible spins) that occupy all the lowest energy states of the system, $n = -N, -N+1, \dots, 0, \dots, N-1, N$, the total current flowing in the system is

$$j_{\text{tot}} = 2 \sum_{n=-N}^N j_n = e \frac{(2N+1)\hbar}{mR} \frac{\phi}{\phi_0}, \quad (14.39)$$

This current is zero for $\phi = 0$ (due to time-reversal symmetry); however, it is finite for a non-zero value of the flux.

Notice that if we had an odd number of particles that occupy the lowest energy states of the system with $\phi = 0$, then the currents of all particles, except that at the highest energy level, will flow in opposite directions and cancel each other. In this case, the total current in the system is that of the last particle $j_{N+1} = e\hbar(N+1)/(mR)$. It shows that the persistent current (14.39), for $\phi \sim \phi_0/2$, is approximately the current carried by the most energetic particle.

We now show that dislocations behave as Aharonov-Bohm flux lines in some situations. It is convenient to start with the example of graphene, where we already know that deformations appear in the form of a vector potential, see Eq. (14.8), and then generalize the result to other systems.

In Fig. 14-8, we depict a dislocation in graphene lattice. It is realized by two disclinations: A cell with a polygon shape with seven sides attached to another cell with five sides. This distortion appears as a column of hexagonal unit cells inserted below the x -axis.

In the continuum limit, we can choose the negative y -axis to be the branch cut where the jump that gives the Burgers vector, occurs. Thus, over distances much larger than the size of a unit cell, the translation vector associated with the dislocation is given by:

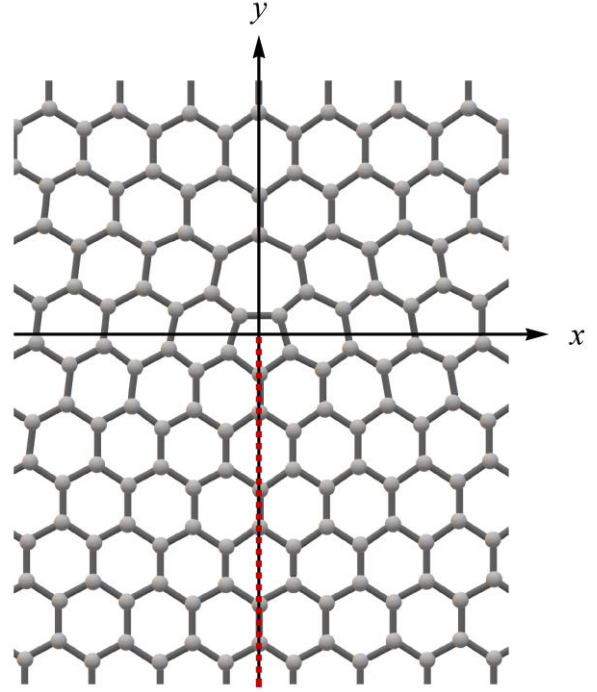


Figure 14-8 Dislocation in graphene

$$u_x = -\frac{a}{2\pi} \arctan\left(\frac{y}{x}\right) + \begin{cases} \frac{a}{4} & x > 0 \\ -\frac{a}{4} & x < 0 \end{cases}, \quad \text{and} \quad u_y = 0, \quad (14.40)$$

where a is the lattice constant. The function $u_x(x, y)$ is presented in Fig. 14-9. Substituting Eqs. (14.40) in formula (14.9) for the vector potential due to deformation we obtain:

$$\mathbf{A}_{\text{def}} = -\frac{\lambda_{\text{def}}^{(2)} a}{v\hbar} \frac{1}{x^2 + y^2} \begin{pmatrix} y \\ -x \end{pmatrix} = \frac{\lambda_{\text{def}}^{(2)} a}{v\hbar} \frac{\mathbf{z} \times \mathbf{r}}{r^2}. \quad (14.41)$$

Now one can quickly check that the magnetic field associated with this vector potential vanishes,

$$\mathbf{B}_{\text{def}} = \nabla \times \mathbf{A}_{\text{def}} = 0, \quad (14.42)$$

but the magnetic flux is finite:

$$\phi = \oint \mathbf{A}_{\text{def}} \cdot d\mathbf{r} = 2\pi \frac{\lambda_{\text{def}}^{(2)} a}{v\hbar}. \quad (14.43)$$

Thus, the effect of dislocation on the behavior of electrons is similar to that of magnetic flux

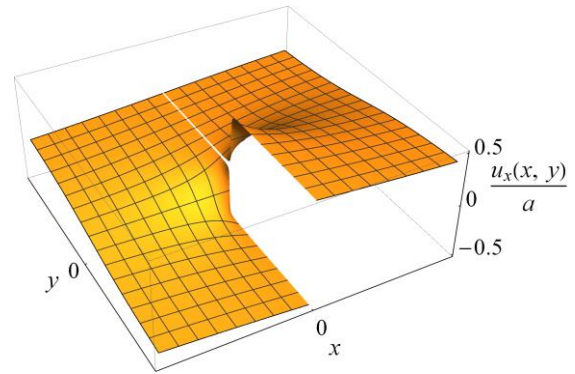


Figure 14-9 The continuum limit of the translation vector for the dislocation presented in Fig. 14-8.

line (with opposite signs for electrons in the K and the K' valleys, to endure time reversal symmetry).

Dislocations in the framework of the $k \cdot p$ approximation

We now show that the analogy between dislocations and magnetic flux lines goes beyond the particular example of graphene and also applies for cases where the $k \cdot p$ approximation holds near a point in the Brillouin zone which is not the Γ point (Iordanskii & Koshelev, 1985).

In the continuum limit, a dislocation can be described as a branch cut on which the translation vector, $\mathbf{u}(\mathbf{r})$, jumps as demonstrated in Fig. 14-9. Recall that the convenient coordinate system to work with is defined by $\tilde{\mathbf{r}} + \mathbf{u}(\tilde{\mathbf{r}}) = \mathbf{r}$ (see Eq. (11.4)). Thus if we denote by $\tilde{\mathbf{r}}_+$ and $\tilde{\mathbf{r}}_-$ the coordinates at the same point in space but from both sides of the branch cut, then $\tilde{\mathbf{r}}_+ - \tilde{\mathbf{r}}_- = \mathbf{b}$, where $\mathbf{b}$ is the Burgers vector. However, the wave function should be single-valued, $\psi(\tilde{\mathbf{r}}_+) = \psi(\tilde{\mathbf{r}}_-)$, therefore $\psi(\tilde{\mathbf{r}}_-) = \psi(\tilde{\mathbf{r}}_+ - \mathbf{b})$ as shown in the figure below.

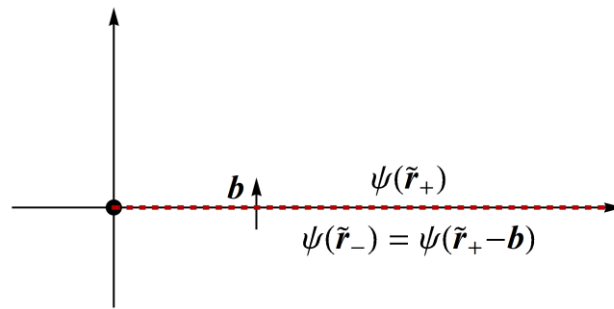


Figure 14-9 The branch cut associated with dislocation and the wave function on its both sides.

Now, in the framework of the $k \cdot p$ approximation, the wavefunction near a valley at $\mathbf{k}_0$ is approximated by a sum over bands:

$$\psi(\mathbf{r}) = \sum_j c_j(\mathbf{r}) \psi_{\mathbf{k}_0}^{(j)}(\mathbf{r}), \quad (14.44)$$

where $\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r})$ is the Bloch wave function of the j -th band at $\mathbf{k} = \mathbf{k}_0$, and $c_j(\mathbf{r})$ are functions that change slowly in space. Let us identify the change of the wave function when going along a contour that encircles the dislocation from the upper side of the branch cut to its lower side:

$$\psi_{\mathbf{k}_0}^{(j)}(\tilde{\mathbf{r}}_+) \rightarrow \psi_{\mathbf{k}_0}^{(j)}(\tilde{\mathbf{r}}_-) = \psi_{\mathbf{k}_0}^{(j)}(\tilde{\mathbf{r}}_+ - \mathbf{b}) = \psi_{\mathbf{k}_0}^{(j)}(\tilde{\mathbf{r}}_+) \exp(-i\mathbf{k}_0 \cdot \mathbf{b}), \quad (14.45)$$

where for the last equality we have used the fact that $\psi_{\mathbf{k}_0}^{(j)}(\mathbf{r})$ is a Bloch wave function and that the Burgers vector is a lattice vector (i.e., a linear combination of the primitive lattice vectors). From here, it follows that

$$c_j(\tilde{\mathbf{r}}_-) = c_j(\tilde{\mathbf{r}}_+) \exp(-i\mathbf{k}_0 \cdot \mathbf{b}) \quad (14.46)$$

Thus the wave function, $c_j(\tilde{\mathbf{r}})$, accumulates a phase $\phi = -\mathbf{k}_0 \cdot \mathbf{b}$ similar to the Aharonov-Bohm phase obtained when a particle encircles a flux line (see Eq. (14.33)). This phase is obtained by adding the vector potential $\mathbf{A}_{\text{def}} = -\nabla(\mathbf{k}_0 \cdot \mathbf{u})$ to the Hamiltonian by minimal substitution:

$$-i\hbar\nabla \rightarrow -i\hbar\nabla + \nabla(\mathbf{k}_0 \cdot \mathbf{u}). \quad (14.47)$$

Notice that the magnetic field associated with this vector potential vanishes, $\mathbf{B} = \nabla \times \mathbf{A}_{\text{def}} = 0$, but its integral around the dislocation is ϕ , hence dislocations can be regarded as magnetic flux lines.

14.4 Peierls instability

One of the dramatic manifestations of the interaction between electrons and lattice deformation is when this interaction drives a metal into an insulator. If the interaction is sufficiently strong, the system favors a lower energy state in which the lattice deforms such that the unit cell is doubled, and a gap opens at the Fermi level. This phenomenon, called *Peierls instability*, is generic for half-filled one-dimensional systems, which we now discuss.

Consider a one-dimensional lattice and let us assume that each unit cell, represented by a lattice point of the Bravais lattice, contributes one electron to the conduction band. The spin degree of freedom implies that the conduction band is half-filled, as illustrated in Fig. 14-10, and the system is, apparently, metallic.

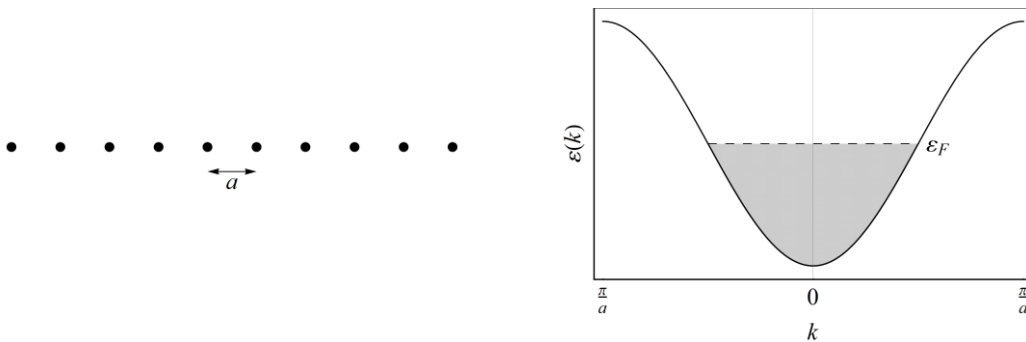


Figure 14-10 The spectrum of a one-dimensional lattice where each unit cell contributes one electron to the conduction band

However, consider a scenario where pairs of lattice units approach each other. This can be realized with alternating shifts of the original lattice points, by δ , to the left and right, as shown in Fig. 14-11. We shall assume that these shifts are much smaller than the original lattice constant a , i.e. $\delta \ll a$.

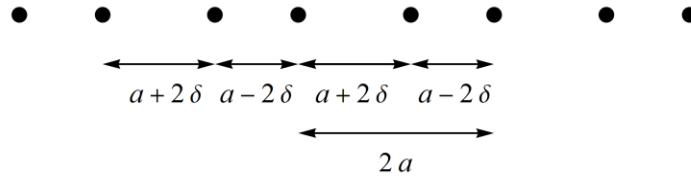


Figure 14-11 A deformation of one-dimensional lattice that doubles the unit cell

This deformation doubles the unit cell of the lattice; therefore, the Brillouin zone becomes smaller by a factor of 2. The electronic spectrum in the deformed lattice is obtained by folding the original spectrum into the new Brillouin zone and opening a gap at the degeneracy points, as shown in Fig. 14-12 (the lift of the degeneracy follows from the same arguments presented in section 7.1 for nearly free electrons).

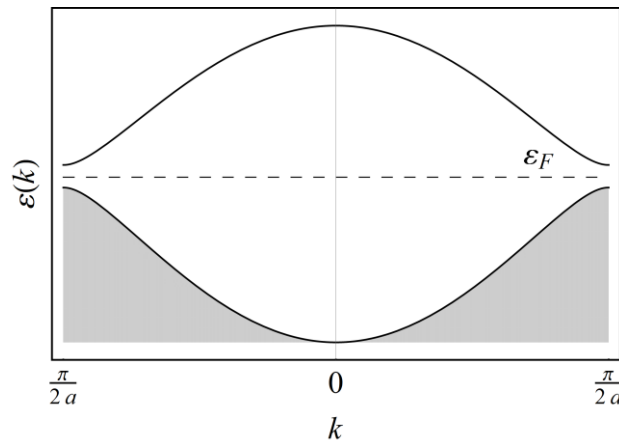


Figure 14-12 The electronic spectrum obtained from doubling the size of a unit cell as shown in Fig. 14-2

Now the electrons will fill the lower energy band; hence the gap that has been formed at the Fermi energy lowers the electronic energy, $\Delta\mathcal{E}_{\text{electron}} < 0$. Clearly, lattice deformations requires elastic energy, $\mathcal{E}_{\text{deform}} > 0$. However, if the total change in the system's energy is negative, $\Delta\mathcal{E} = \mathcal{E}_{\text{deform}} + \Delta\mathcal{E}_{\text{electron}} < 0$, the original lattice becomes unstable. Namely, it will deform, as illustrated above, and the system will become an insulator. This scenario is generic in one dimensional-systems (as we shall see below). In higher dimensions, Peierls instability depends on details of the system, such as the band structure, the elastic modulus tensor, and the strength of coupling between the electrons and the lattice deformation.

To be concrete, let us calculate the total energy of a one-dimensional lattice in the nearly free electron approximation. This energy is given by a sum of the electronic energy and the elastic deformation energy :

$$\mathcal{E}_{\text{tot}} = \mathcal{E}_{\text{deform}} + \mathcal{E}_{\text{electron}}. \quad (14.48)$$

Our goal is to calculate the dependence of these energies on the deformation parameter, δ , defined in Fig. 14-11.

The deformation energy can be deduced from general principles: It should be proportional to δ^2 (being a shift from an equilibrium position), and to the length of the system, L . Therefore

$$\mathcal{E}_{\text{deform}} = \frac{1}{2} \Xi L \left(\frac{\delta}{a} \right)^2, \quad (14.49)$$

where the constant Ξ (having the same physical dimensions as the elastic modulus tensor) accounts for the elastic restoring force. Here we normalized δ by the lattice constant of the undeformed system a .

The electronic energy is given by the integral

$$\mathcal{E}_{\text{electron}} = 2L \int_{-k_F}^{k_F} \frac{dk}{2\pi} \varepsilon(k) = 4L \int_0^{k_F} \frac{dk}{2\pi} \varepsilon(k), \quad (14.50)$$

where the factor 2 accounts for the spin degeneracy, $\varepsilon(k)$ is the lowest energy band, and

$$k_F = \frac{\pi}{2a} \quad (14.51)$$

is the Fermi wavenumber.

To calculate $\varepsilon(k)$, we employ a procedure similar to the nearly free approximation. Namely, we first fold the original spectrum into the reduced Brillouin zone and then treat the degeneracy point at the edge of the Brillouin zone by perturbation theory in $\Delta V(x)$ where the latter is the perturbation potential obtained from the shifts of the atoms from their positions by δ . Namely, if we represent the potential of the undeformed lattice, $V(x)$, as a sum over the potentials produced by the ions in each unit cell,

$$V(x) = \sum_n v(x - na), \quad (14.52)$$

then

$$\Delta V(x) = \sum_n v(x - na + (-1)^n \delta) - v(x - na) \simeq \sum_n \frac{\partial v(x - na)}{\partial x} (-1)^n \delta \quad (14.53)$$

where we assume $\delta/a \ll 1$.

Let us now focus on the two lowest bands within the range $0 \leq k \leq k_F = \pi/(2a)$. Furthermore, let $\psi_k^{(\pm)}(x)$ denote the wave functions corresponding to the lower and upper bands, as shown in Fig. 14-13.

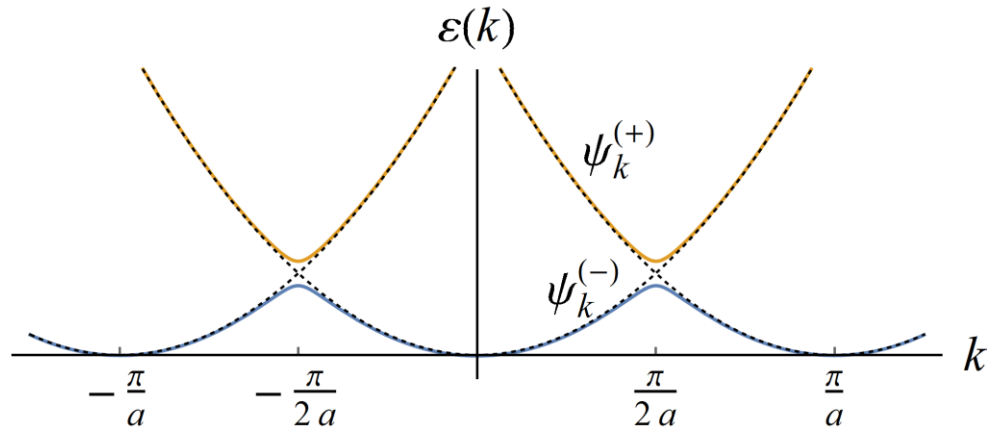


Figure 14-13 The energy spectrum of deformed one-dimensional lattice in the nearly free electron approximation

In the nearly free electron approximation, these wave functions are approximately given by

$$\psi_k^{(-)}(x) = \frac{1}{\sqrt{L}} \exp(ikx) \quad \text{and} \quad \psi_k^{(+)}(x) = \frac{1}{\sqrt{L}} \exp[i(k - 2k_F)x] \quad (14.54)$$

Notice that in the range $k_F > k > 0$, the higher energy band describes a left moving particle, while the lower energy branch is associated with a right moving particle. In the basis of these functions, the Hamiltonian takes the form:

$$H = \begin{pmatrix} \frac{\hbar^2 k^2}{2m} & \langle \psi_k^{(-)} | \Delta V | \psi_k^{(+)} \rangle \\ \langle \psi_k^{(+)} | \Delta V | \psi_k^{(-)} \rangle & \frac{\hbar^2 (k - 2k_F)^2}{2m} \end{pmatrix}, \quad (14.55)$$

where m is the electron mass. The diagonal matrix elements of the perturbation vanish because the space average of $\Delta V(x)$ is zero. The off-diagonal matrix elements can be expressed as Fourier coefficients of the Fourier series of the perturbation potential:

$$\begin{aligned} \langle \psi_k^{(-)} | \Delta V | \psi_k^{(+)} \rangle &= \lim_{L \rightarrow \infty} \frac{1}{L} \int_{-L/2}^{L/2} dx \Delta V(x) \exp(-2ik_F x) \\ &= \frac{1}{2a} \int_{-a}^a dx \Delta V(x) \exp(-2ik_F x) = \Delta V_{2k_F}^* \end{aligned} \quad (14.56)$$

where to obtain the second equality, we have used the $2a$ periodicity of the integrand. The Fourier coefficient, ΔV_{2k_F} , can be expressed as (see Ex. 2):

$$\Delta V_{2k_F} = -i2k_F v_{2k_F} \frac{\delta}{a}, \quad (14.57)$$

where

$$v_k = \int_{-\infty}^{\infty} dx v(x) \exp(ikx). \quad (14.58)$$

Substituting (14.55) and (14.57) in (14.55), and diagonalizing the Hamiltonian yields:

$$\varepsilon_{\pm}(k) = \frac{\hbar^2 (k - k_F)^2}{2m} + \frac{\hbar^2 k_F^2}{2m} \pm k_F \sqrt{\frac{\hbar^4 (k - k_F)^2}{m^2} + 4|v_{2k_F}|^2 \left(\frac{\delta}{a}\right)^2}. \quad (14.59)$$

Since the lower band is the one occupied by the electrons, the total energy density is:

$$\frac{\varepsilon_{\text{tot}}}{L} = \frac{1}{2} \Xi \left(\frac{\delta}{a}\right)^2 + 4 \int_0^{k_F} \frac{dk}{2\pi} \varepsilon_{-}(k) \quad (14.60)$$

To find the minimal value of this energy, we take its derivative with respect to the deformation parameter, δ , and demand it to be zero:

$$\begin{aligned} \frac{a}{L} \frac{d}{d\delta} \varepsilon_{\text{tot}} &= \Xi \frac{\delta}{a} - \frac{8}{\pi} \int_0^{k_F} dk \frac{mk_F |v_{2k_F}|^2 \frac{\delta}{a}}{\sqrt{\hbar^4 (k - k_F)^2 + 4m^2 |v_{2k_F}|^2 \left(\frac{\delta}{a}\right)^2}} \\ &= \Xi \frac{\delta}{a} - \frac{8k_F m |v_{2k_F}|^2 \delta}{\pi \hbar^2 a} \tanh^{-1} \left(\frac{k_F}{\sqrt{k_F^2 + \frac{4m^2 |v_{2k_F}|^2 \delta^2}{\hbar^4 a^2}}} \right) = 0 \end{aligned} \quad (14.61)$$

The solution of this equation (with minimal energy) is

$$\delta = \pm \frac{\hbar^2 k_F}{m |v_{2k_F}|} \frac{a}{2 \cosh \left(\frac{\pi \hbar^2 \Xi}{8mk_F |v_{2k_F}|^2} \right)} \simeq \pm \frac{\hbar^2 k_F a}{m |v_{2k_F}|} \exp \left(-\frac{\pi \hbar^2 \Xi}{8mk_F |v_{2k_F}|^2} \right). \quad (14.62)$$

It shows that no matter how stiff the elastic deformations are or how small the coupling of the electrons to the lattice, the system's total energy can be lowered by deforming the lattice, thereby opening a gap in the electronic spectrum makes the system insulating.

Comment: Mott instability

Doubling of the unit cell, which in the case of half-filling, transforms the system into an insulator, can also be achieved by spin ordering, as demonstrated in Fig. 14-14.

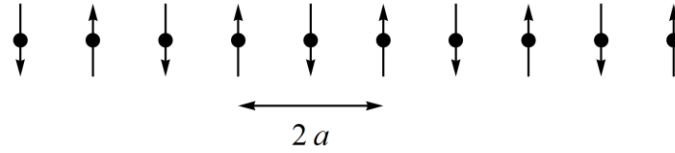


Figure 14-14 Doubling of the unit cell due to the antiferromagnetic ordering of spins

Such an antiferromagnetic spin ordering will open a gap in the conduction band and transform the system into an insulator - a phenomenon known as the *metal-insulator Mott transition*. Like the Peierls instability, the Mott instability always occurs in one-dimensional systems. In two and three dimensions, it depends on the strength of electron-electron interaction. In some cases, this transition can be induced by external forces such as compression stress.

14.5 Exercises

1. Consider the square lattice presented in the example on page 136 and Ex. 2 of the same chapter and assume it is subjected to shear stress as illustrated in Fig. 14-15 below.

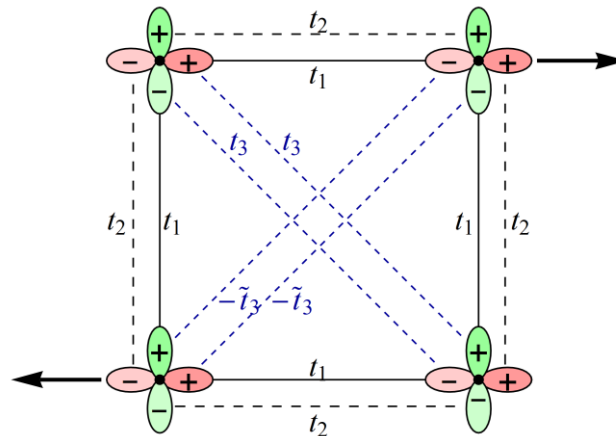


Figure 14-15 A square lattice with degenerate bands subjected to a shear stress

Let us model the effect of the shear by changing the hopping matrix elements shown in the figure such that

$$t_3 = \bar{t} (1 + \varepsilon) \quad \text{and} \quad \tilde{t}_3 = \bar{t} (1 - \varepsilon), \quad (14.63)$$

where $|\varepsilon| \ll 1$. Analyze the behavior of the energy levels near Γ as function of ε , by expanding the Hamiltonian to second order in the components of the wavenumber. In particular, check the possibility of obtaining accidental degeneracy points.

2. Prove Eq. (14.57).

Advice: Show that Fourier coefficients of the Fourier series for the periodic function $\Delta V(x)$ (with periodicity $2a$) is:

$$\begin{aligned} \Delta V_k &= \frac{1}{2a} \int_{-a}^a \Delta V(x) \exp(ikx) dx = \frac{\delta}{2a} \int_{-\infty}^{\infty} \left[\frac{\partial v(x)}{\partial x} - \frac{\partial v(x-a)}{\partial x} \right] \exp(ikx) dx, \\ &= -i \frac{k\delta}{2a} v_k [1 - \exp(ika)] \end{aligned} \quad (14.64)$$

and the set $k = 2k_F$.

3. In a tight-binding approach to Peierls instability, one considers the Hamiltonian

$$H = - \sum_j t_j c_j^\dagger c_{j+1} + h.c., \quad (14.65)$$

where hopping matrix elements between nearest neighbors site is

$$t_j = t \left[1 + \frac{\delta}{a} (-1)^j \right], \quad (14.66)$$

where a is the lattice constant and δ represents the deformation parameter.

(a) Calculate the energy bands of this tight-binding model.

(b) Use formula (14.49) for the elastic energy and the result of (a) to obtain the following equation for the minimum of the total energy of the system:

$$\frac{a^2}{L\delta} \frac{d}{d\delta} \varepsilon_{\text{tot}} = \Xi - \frac{4t}{\pi} \int_0^{k_F} dk \frac{\sin^2(ka)}{\sqrt{\cos^2(ka) + \left(\frac{\delta}{a}\right)^2 \sin^2(ka)}} = 0, \quad (14.67)$$

where $k_F = \pi/(2a)$.

(c) To evaluate the above integral in the limit $\delta \rightarrow 0$, notice that it diverges logarithmically due to divergence near $k = k_F = \pi/(2a)$. Expand the numerator and denominator of the integrand to leading order around $k = k_F$, evaluate the resulting integral and solve the equation for δ .

15 Electron-phonon interaction

In the previous chapter, we discussed the effect of a static deformation of the crystal on the electronic spectrum. Now we shall extend the discussion to the interaction of electrons with dynamical deformations of the crystal, i.e., phonons. There are few channels for the electron-phonon interaction that depend on the crystal's nature (e.g., whether it is piezoelectric, polar, or neither) and the type of phonons (e.g., acoustic or optical). We begin this chapter by discussing the case where the electron-phonon coupling is due to local deformations of the crystal and see how the kinetic equations that follow from this coupling take the system into an equilibrium state (hence breaking the time-reversal symmetry on a macroscopic level). Next, we consider piezoelectric coupling and interaction of electrons with optical phonons that give rise to an excitation called “polaron”.

15.1 Deformation interaction

The guiding principle for describing the interaction of electrons and phonons is the observation that electrons are light and move fast while phonons, associated with the dynamics of heavy ions, are slow. This property justifies the Born-Oppenheimer approximation in which the potential seen by the electrons, at any given time, is the one generated by the deformed lattice.

Deformation interaction refers to the coupling between electrons and acoustic phonons that is described by the interaction Hamiltonian:

$$H_{\text{e-ph}} = \lambda_{ij} u_{ij}(\mathbf{r}), \quad (15.1)$$

where $u_{ij}(\mathbf{r})$ is the strain tensor, while λ_{ij} are arbitrary coupling constants (at this stage), and, as usual, repeated indices should be summed over. Since $u_{ij}(\mathbf{r})$ is a symmetric tensor, λ_{ij} can also be chosen to be symmetric without loss of generality. The above Hamiltonian manifests the Born-Oppenheimer approximation because the electron-phonon interaction depends only on the deformation potential and does not involve coupling between the electrons and the momentum density of the lattice deformation.

Thus, the total Hamiltonian of the system contains three components:

$$H = H_{\text{e}} + H_{\text{ph}} + H_{\text{e-ph}}. \quad (15.2)$$

The first is the Hamiltonian of an electron in a periodic lattice. The second is the Hamiltonian of the phonons,

$$H_{\text{ph}} = \sum_{\alpha, q} \hbar \omega_{\alpha}(\mathbf{q}) \left[n_{\alpha}(\mathbf{q}) + \frac{1}{2} \right], \quad (15.3)$$

where $\omega_\alpha(\mathbf{q})$ is the phonon frequency on the α branch of the sound wave spectrum with wavenumber $\mathbf{q}$, while

$$n_\alpha(\mathbf{q}) = a_\alpha^\dagger(\mathbf{q}) a_\alpha(\mathbf{q}) \quad (15.4)$$

is the number operator expressed in terms of the creation and annihilation operators with the property:

$$a|n\rangle = \sqrt{n}|n-1\rangle, \quad a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle. \quad (15.5)$$

Finally, the third contribution to the Hamiltonian (15.2) is the coupling Hamiltonian (15.1).

In order to express the coupling Hamiltonian in terms of the creation and annihilation operators of phonons, we use Eq. (11.74) for the translation vector:

$$u_j(\mathbf{r}) = \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q})\text{Vol}}} e_j^{(\alpha)}(\hat{\mathbf{q}}) [\hat{a}_\alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{r}) + \hat{a}_\alpha^\dagger(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r})], \quad (15.6)$$

where ρ is the mass density of the crystal, Vol is the volume of the system, $e_j^{(\alpha)}(\hat{\mathbf{q}})$ is the j -th component of the polarization of a wave moving in the direction $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$ with frequency $\omega_\alpha(\mathbf{q})$. With this representation, the strain tensor takes the form:

$$u_{ij}(\mathbf{r}) = i \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q})\text{Vol}}} \frac{e_i^{(\alpha)} q_j + e_j^{(\alpha)} q_i}{2} [\hat{a}_\alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{r}) - \hat{a}_\alpha^\dagger(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r})]. \quad (15.7)$$

To shorten the notations, from now on, we suppress the dependence of the polarization vector on the propagation direction of the wave. Taking into account that the coupling tensor, λ_{ij} , is symmetric, we obtain that Eq. (15.1) can be written in the form:

$$H_{\text{e-ph}} = i \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q})\text{Vol}}} \lambda_{ij} e_i^{(\alpha)} q_j [\hat{a}_\alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{r}) - \hat{a}_\alpha^\dagger(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r})]. \quad (15.8)$$

Our goal now is to project this Hamiltonian on the basis of Bloch's wave functions, i.e., to calculate the transition matrix elements between two Bloch states:

$$H_{\text{e-ph}}^{k\mathbf{k}} = \langle \psi_{k'}^{(\nu')} | H_{\text{e-ph}} | \psi_k^{(\nu)} \rangle, \quad (15.9)$$

where $\psi_k^{(\nu)}(\mathbf{r})$ is Bloch wave function of the ν -th band and with wavenumber $\mathbf{k}$.

Apart from cases where energy bands become degenerate or very close to each other, energy conservation hinders transitions between different energy bands; hence, we shall consider only the matrix elements between Bloch wave functions of the same band and suppress the band index. Let us represent these functions in the form:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\text{Vol}}} \phi_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (15.10)$$

and expand the periodic component of this Bloch wave function in Fourier series

$$\phi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{b}} c_{\mathbf{b}}(\mathbf{k}) \exp(i\mathbf{b} \cdot \mathbf{r}), \quad (15.11)$$

where $c_{\mathbf{b}}(\mathbf{k})$ are the Fourier expansion coefficients, and the sum is over all vectors of the reciprocal lattice. The normalization of $\psi_{\mathbf{k}}(\mathbf{r})$ implies that

$$\begin{aligned} \int d^d r |\psi_{\mathbf{k}}(\mathbf{r})|^2 &= \frac{1}{\text{Vol}} \int d^d r \sum_{\mathbf{b}, \mathbf{b}'} c_{\mathbf{b}}(\mathbf{k}) c_{\mathbf{b}'}^*(\mathbf{k}) \exp[i(\mathbf{b} - \mathbf{b}') \cdot \mathbf{r}] \\ &= \sum_{\mathbf{b}, \mathbf{b}'} c_{\mathbf{b}}(\mathbf{k}) c_{\mathbf{b}'}^*(\mathbf{k}) \delta_{\mathbf{b}, \mathbf{b}'} = \sum_{\mathbf{b}} |c_{\mathbf{b}}(\mathbf{k})|^2 = 1. \end{aligned} \quad (15.12)$$

To calculate the matrix elements of the electron-phonon interaction (15.9), we need to calculate the following matrix element:

$$\begin{aligned} \langle \psi_{\mathbf{k}'} | \exp(\pm i\mathbf{q} \cdot \mathbf{r}) | \psi_{\mathbf{k}} \rangle &= \frac{1}{\text{Vol}} \sum_{\mathbf{b}, \mathbf{b}'} \int d^d r c_{\mathbf{b}}(\mathbf{k}) c_{\mathbf{b}'}^*(\mathbf{k}') \exp[i(\mathbf{k} - \mathbf{k}' \pm \mathbf{q} + \mathbf{b} - \mathbf{b}') \cdot \mathbf{r}] \\ &= \frac{1}{\text{Vol}} \sum_{\mathbf{b}, \mathbf{b}'} c_{\mathbf{b}}(\mathbf{k}) c_{\mathbf{b}'}^*(\mathbf{k}') (2\pi)^d \delta(\mathbf{k} - \mathbf{k}' \pm \mathbf{q} + \mathbf{b} - \mathbf{b}') \\ &= \sum_{\mathbf{b}, \mathbf{b}'} c_{\mathbf{b}}(\mathbf{k}) c_{\mathbf{b}'}^*(\mathbf{k}') \delta_{\mathbf{k} - \mathbf{k}' \pm \mathbf{q} + \mathbf{b} - \mathbf{b}'} \end{aligned} \quad (15.13)$$

In order to simplify this result, we shall assume that: (a) $\mathbf{b} = \mathbf{b}'$ (In the next chapter we will discuss situations where this is not the case), and (b) consider only acoustic phonons of long wavelength such that $\mathbf{k} - \mathbf{k}' = \mathbf{q} \approx 0$, hence $c_{\mathbf{b}}^*(\mathbf{k}') \approx c_{\mathbf{b}}^*(\mathbf{k})$. With these assumptions and the normalization condition (15.12), we obtain the momentum conservation condition:

$$\langle \psi_{\mathbf{k}'} | \exp(\pm i\mathbf{q} \cdot \mathbf{r}) | \psi_{\mathbf{k}} \rangle \approx \delta_{\mathbf{k} - \mathbf{k}' \pm \mathbf{q}}. \quad (15.14)$$

Substituting (15.8) in (15.9) and using (15.4) yields

$$H_{\text{e-ph}}^{k'k} = i \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_{\alpha}(\mathbf{q})\text{Vol}}} \lambda_{ij} e_i^{(\alpha)} q_j \left[\hat{a}_{\alpha}(\mathbf{q}) \delta_{\mathbf{k}' - \mathbf{k}, \mathbf{q}} - \hat{a}_{\alpha}^{\dagger}(\mathbf{q}) \delta_{\mathbf{k} - \mathbf{k}', \mathbf{q}} \right]. \quad (15.15)$$

The first term in this Hamiltonian describes an absorption of a phonon by an electron that changes its momentum from $\hbar\mathbf{k}$ to $\hbar(\mathbf{k} + \mathbf{q})$. The second term is associated with emission of a phonon when the electron changes its momentum, from $\hbar\mathbf{k}$ to $\hbar(\mathbf{k} - \mathbf{q})$. These contributions are represented by the diagrams shown in Fig. 15-1.

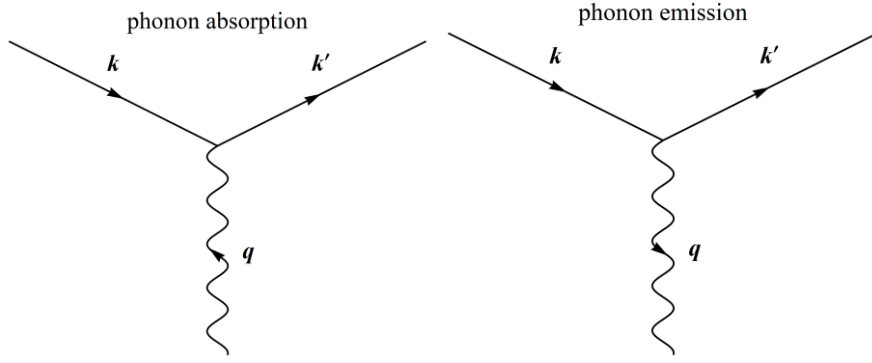


Figure 15-1 Diagrammatic representation of absorption and emission of a phonon when an electron scatters from lattice vibrations

The effects of electron-phonon scattering can be classified into two main groups:

- (a) Real scattering processes where an electron may change its energy by emitting and absorbing phonons - a mechanism that drives the system into thermal equilibrium. When the system is subjected to an electric field, these real scattering processes generate finite electrical resistance.
- (b) Virtual processes of phonon emission followed by absorption of the same phonon. These processes manifest themselves in the renormalization of parameters, such as the effective mass of the electron.

In what follows, we shall focus our attention on real scattering processes. The essential tool for describing these processes is Fermi's golden rule which gives the transition rate from one state to another. Hence, before turning to calculate the transition rate between electronic states (due to electron-phonon interaction), we remind Fermi's golden rule.

Reminder: Fermi's golden rule

Consider the Hamiltonian $H = H_0 + \Delta H$, where H_0 represents the unperturbed Hamiltonian that satisfies the time-independent Schrödinger equation

$$H_0 |n\rangle = \varepsilon_n |n\rangle, \quad (15.16)$$

while ΔH is some perturbation. The time-dependent Schrödinger equation of the system is

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle. \quad (15.17)$$

Its solution, to leading order perturbation theory in ΔH , is obtained by expanding $\psi(t)$ in the basis eigenfunctions of the unperturbed Hamiltonian:

$$|\psi(t)\rangle = \sum_n c_n(t) \exp\left(-i \frac{\varepsilon_n t}{\hbar}\right) |n\rangle. \quad (15.18)$$

Substituting (15.18) in (15.17), one obtains:

$$\sum_n i\hbar \frac{\partial c_n(t)}{\partial t} \exp\left(-i \frac{\varepsilon_n t}{\hbar}\right) |n\rangle = \sum_n c_n(t) \exp\left(-i \frac{\varepsilon_n t}{\hbar}\right) \Delta H |n\rangle. \quad (15.19)$$

Multiplying this equation by $\langle k|$ from the left leads to an equation for the expansion coefficients:

$$i\hbar \frac{\partial c_k(t)}{\partial t} = \sum_n c_n(t) \exp\left(-i \frac{(\varepsilon_n - \varepsilon_k)t}{\hbar}\right) \langle k | \Delta H | n \rangle. \quad (15.20)$$

This equation is still exact. To solve it, in the leading order perturbation theory, we assume that the system has been prepared (at the time $t = 0$) in an eigenstate of the unperturbed system $|i\rangle$ and replace $c_n(t)$, on the right-hand side of the equation, by the zeroth-order solution, $c_n(t) = \delta_{ni}$. Thus:

$$c_k(t) \approx \frac{1}{i\hbar} \int_0^t dt' \exp\left(-i \frac{(\varepsilon_i - \varepsilon_k)t'}{\hbar}\right) \langle k | \Delta H | i \rangle \quad (15.21)$$

Taking the time derivative of $|c_k(t)|^2$ we obtain that the rate of change of the probability of finding the system in the state $|k\rangle$ is

$$\frac{d}{dt} |c_k(t)|^2 = \frac{2}{\hbar} \frac{\sin\left(\frac{\Delta \varepsilon_{ik} t}{\hbar}\right)}{\Delta \varepsilon_{ik}} |\langle k | \Delta H | i \rangle|^2, \quad (15.22)$$

where $\Delta \varepsilon_{ik} = \varepsilon_i - \varepsilon_k$. Now let us focus our attention on the following order of limits: First, we take the system size to infinity so that the mean spacing between energy levels, Δ , (which is inversely proportional to the volume of the system) goes to zero, and then take time to infinity, $t \rightarrow \infty$, such that $\Delta \ll \hbar/t$. In this limit, the energy spectrum forms a continuum, and the sinc function can be replaced by a δ -function because

$$\int_{-\infty}^{\infty} d\Delta \varepsilon \frac{\sin\left(\frac{\Delta \varepsilon t}{\hbar}\right)}{\Delta \varepsilon} = \pi. \quad (15.23)$$

Thus, if we denote by dV_f an infinitesimal volume element of the final states of the system, associated with energy ε_f , then the rate of transition to the final states within this volume is

$$dW_{i \rightarrow f} = \frac{d}{dt} \sum_{k \in dV_f} |c_k(t)|^2 = \frac{2\pi}{\hbar} |\langle f | \Delta H | i \rangle|^2 \delta(\varepsilon_f - \varepsilon_i) dV_f. \quad (15.24)$$

This form of Fermi's golden rule will be used below. However, in the literature, one usually finds the formula obtained when setting $dV_f = \rho(\varepsilon_f) d\varepsilon_f$, where $\rho(\varepsilon_f)$ is the density of the final states of the system, and the integral over ε_f is carried out:

$$W_{i \rightarrow f} = \frac{2\pi}{\hbar} \int d\varepsilon_f \rho(\varepsilon_f) |\langle f | \Delta H | i \rangle|^2 \delta(\varepsilon_f - \varepsilon_i). \quad (15.25)$$

15.2 The transition rate between electronic states

We turn now to calculate the transition rate between electronic states due to electron-phonon interaction. As we saw above, these can occur due to absorption or emission of a phonon, see Fig. 15-1. Here we calculate the transition rate associated with absorption. The transition rate associated with emission can be obtained by a straightforward generalization of the same calculation.

Consider the process of absorption of a phonon with wavenumber $\mathbf{q}$ and polarization α :

$$| \mathbf{k}_i, n_\alpha(\mathbf{q}) \rangle \longrightarrow | \mathbf{k}_f, n_\alpha(\mathbf{q}) - 1 \rangle, \quad (15.26)$$

where $\mathbf{k}_i$ and $\mathbf{k}_f$ are the initial and the final wavenumber vectors of the electron, respectively. From Fermi's golden rule, we obtain that the rate of this transition is:

$$dW_{i \rightarrow f}^{\text{absorption}} = \frac{2\pi}{\hbar} |M_{if}|^2 \delta(\varepsilon_f - \varepsilon_i) dV_f, \quad (15.27)$$

where dV_f is an infinitesimal volume element in the final state of the system (which we shall specify below) while

$$M_{if} = \langle \mathbf{k}_f, n_\alpha(\mathbf{q}) - 1 | H_{\text{e-ph}} | \mathbf{k}_i, n_\alpha(\mathbf{q}) \rangle \quad (15.28)$$

is the transition matrix element between the initial and the final state of the system. Using Eqs. (15.15) and (15.5) we obtain that this matrix element is given by

$$M_{if} = i \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q})\text{Vol}}} \lambda_{ij} e_i^{(\alpha)} q_j \sqrt{n_\alpha(\mathbf{q})} \delta_{\mathbf{k}_f - \mathbf{k}_i, \mathbf{q}}. \quad (15.29)$$

The energy levels of the initial and the final states of the system are, respectively,

$$\varepsilon_i = \varepsilon(\mathbf{k}_i) + \hbar\omega_\alpha(\mathbf{q})n_q^{(\alpha)} \quad \text{and} \quad \varepsilon_f = \varepsilon(\mathbf{k}_f) + \hbar\omega_\alpha(\mathbf{q})(n_q^{(\alpha)} - 1), \quad (15.30)$$

thus

$$dW_{i \rightarrow f}^{\text{absorption}} = \frac{dV_f}{\text{Vol}} \frac{2\pi}{\hbar} \sum_{\alpha} \frac{\hbar |\lambda_{ij} e_i^{(\alpha)} q_j|^2}{2\rho\omega_{\alpha}(\mathbf{q})} n_{\alpha}(\mathbf{q}) \delta_{\mathbf{k}_f - \mathbf{k}_i, \mathbf{q}} \delta[\varepsilon(\mathbf{k}_f) - \varepsilon(\mathbf{k}_i) - \hbar\omega_{\alpha}(\mathbf{q})] \quad (15.31)$$

The infinitesimal volume element of the final state is associated with particular momentum states of the electron and the phonon; hence it is given by

$$dV_f = \frac{\text{Vol}}{(2\pi)^d} d^d k_f \frac{\text{Vol}}{(2\pi)^d} d^d q. \quad (15.32)$$

(Recall that $\text{Vol}/(2\pi)^d$ is the density of states in the wavenumber space of a d dimensional system.) However, in order to sum over the final states of the system with the help of an integral over $\mathbf{k}_f$ and $\mathbf{q}$, we should replace the Kronecker delta function in Eq. (15.31) by its continuum counterpart. This is obtained by the rule:

$$\delta_{\mathbf{k}_f - \mathbf{k}_i, \mathbf{q}} \rightarrow \frac{(2\pi)^d}{\text{Vol}} \delta(\mathbf{k}_f - \mathbf{k}_i - \mathbf{q}). \quad (15.33)$$

Thus

$$dW_{i \rightarrow f}^{\text{absorption}} = \frac{d^d k_f}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} \frac{(2\pi)^{d+1}}{\hbar} \sum_{\alpha} \frac{\hbar |\lambda_{ij} e_i^{(\alpha)} q_j|^2}{2\rho\omega_{\alpha}(\mathbf{q})} n_{\alpha}(\mathbf{q}) \times \delta(\mathbf{k}_f - \mathbf{k}_i - \mathbf{q}) \delta[\varepsilon(\mathbf{k}_f) - \varepsilon(\mathbf{k}_i) - \hbar\omega_{\alpha}(\mathbf{q})]. \quad (15.34)$$

A similar calculation for the transition rate for a process where the phonon is emitted yields:

$$dW_{i \rightarrow f}^{\text{emission}} = \frac{d^d k_f}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} \frac{(2\pi)^{d+1}}{\hbar} \sum_{\alpha} \frac{\hbar |\lambda_{ij} e_i^{(\alpha)} q_j|^2}{2\rho\omega_{\alpha}(\mathbf{q})} (n_{\alpha}(\mathbf{q}) + 1) \times \delta(\mathbf{k}_f - \mathbf{k}_i + \mathbf{q}) \delta[\varepsilon(\mathbf{k}_f) - \varepsilon(\mathbf{k}_i) + \hbar\omega_{\alpha}(\mathbf{q})] \quad (15.35)$$

In order to shorten the notations of the following discussion, it will be convenient to define the density of transition rate:

$$w_{\alpha}(\mathbf{k}_f, \mathbf{k}_i, \mathbf{q}) = \frac{|\lambda_{ij} e_i^{(\alpha)} q_j|^2}{2\rho\omega_{\alpha}(\mathbf{q})} (2\pi)^{d+1} \delta(\mathbf{k}_f - \mathbf{k}_i - \mathbf{q}) \delta[\varepsilon(\mathbf{k}_f) - \varepsilon(\mathbf{k}_i) - \hbar\omega_{\alpha}(\mathbf{q})] \quad (15.36)$$

so that

$$dW_{i \rightarrow f}^{\text{absorption}} = \sum_{\alpha} \frac{d^d k_f}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}_f, \mathbf{k}_i, \mathbf{q}) n_{\alpha}(\mathbf{q})$$

$$dW_{i \rightarrow f}^{\text{emission}} = \sum_{\alpha} \frac{d^d k_f}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}_i, \mathbf{k}_f, \mathbf{q}) [n_{\alpha}(\mathbf{q}) + 1] \quad (15.37)$$

Observe that in deriving the above formulas, we considered a single electron. Thus, we did not consider the possibility that another electron may already occupy the final state of the scattered electron. In the next section, we present the kinetic equations for the electron and phonon distributions, which account for this information.

15.3 The kinetic equations for distribution functions of the electrons and the phonons in a crystal

Let $f(\mathbf{k})$ be the *electron distribution function*, i.e., the probability to find an electron in a state defined by the wavenumber $\mathbf{k}$ (for simplicity, we assume a single energy band and ignore spin). In the absence of coupling of the electrons to the lattice vibrations, this distribution is independent of time (assuming no other reasons for electron scattering, such as defects in the crystal structure or impurities). However, the electron-phonon interaction induces transitions of the electron from one state to another, and, using the results of the previous section, one may write down the rate of change of $f(\mathbf{k})$ due to collision with the lattice vibrations:

$$\begin{aligned} \left(\frac{df(\mathbf{k})}{dt} \right)_{\text{coll}} = & - \sum_{\alpha} \int \frac{d^d k_1}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}_1, \mathbf{k}, \mathbf{q}) n_{\alpha}(\mathbf{q}) f(\mathbf{k}) [1 - f(\mathbf{k}_1)] \\ & + \sum_{\alpha} \int \frac{d^d k_1}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}, \mathbf{k}_1, \mathbf{q}) n_{\alpha}(\mathbf{q}) f(\mathbf{k}_1) [1 - f(\mathbf{k})] \left. \vphantom{\sum_{\alpha}} \right\} \text{absorption} \\ & - \sum_{\alpha} \int \frac{d^d k_1}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}, \mathbf{k}_1, \mathbf{q}) [n_{\alpha}(\mathbf{q}) + 1] f(\mathbf{k}) [1 - f(\mathbf{k}_1)] \\ & + \sum_{\alpha} \int \frac{d^d k_1}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}_1, \mathbf{k}, \mathbf{q}) [n_{\alpha}(\mathbf{q}) + 1] f(\mathbf{k}_1) [1 - f(\mathbf{k})] \left. \vphantom{\sum_{\alpha}} \right\} \text{emission} \end{aligned} \quad (15.38)$$

Let us explain this formula: The first two terms are associated with the absorption of a phonon. The first describes a process where an electron, at state $\mathbf{k}$, is scattered into a state $\mathbf{k}_1$ by absorbing a phonon of wavenumber $\mathbf{q}$. The factor $f(\mathbf{k})[1 - f(\mathbf{k}_1)]$ takes into account the requirement that the initial state of the system is occupied while the final state, into which the electron scatters, is unoccupied. The integral is over all possible final states of the scattered electron and the absorbed phonon. This contribution appears with a minus sign because it reduces the probability of finding the electron in the state $\mathbf{k}$. However, there is also a process that enhances the probability of finding an electron in the state $\mathbf{k}$. It is given by the second term of Eq. (15.38). Here an electron at state $\mathbf{k}_1$ is scattered into the state $\mathbf{k}$, while absorbing a phonon. It appears with a positive sign, and the factor $f(\mathbf{k}_1)[1 - f(\mathbf{k})]$ takes into account the probability of finding the states $\mathbf{k}_1$ and $\mathbf{k}$ occupied and unoccupied, respectively. The other two terms in Eq. (15.38) describe similar processes but with an

emission of a phonon rather than absorption. Finally, in the above equation, $n_\alpha(\mathbf{q})$ represents the average number of phonons with polarization α and wavenumber $\mathbf{q}$ - this quantity is the *phonon distribution function*.

The justification of Eq. (15.38) rests on several assumptions. First, it assumes no memory effects, namely, that an electron which leaves the state $\mathbf{k}$ can be considered as never coming back. Second, the initial and final states of the electron and the number of phonons in the system are assumed to be uncorrelated. This property implies that the joint distribution function is reduced to a product of the distribution functions. Third, Eq. (15.38) neglects other scattering processes which involve emission and/or absorption of more than one phonon at a time.

The kinetics equation of the system may be written in the form

$$\frac{df(\mathbf{k})}{dt} = \text{St}_e[f(\mathbf{k})], \quad (15.39)$$

where

$$\text{St}_e[f(\mathbf{k})] = \left(\frac{df(\mathbf{k})}{dt} \right)_{\text{coll}}. \quad (15.40)$$

This term is called the *collision integral* (and in German stoß). Rearranging the terms in the collision integral yields:

$$\begin{aligned} \text{St}_e[f(\mathbf{k})] = & \sum_{\alpha} \frac{d^d k_1}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}_1, \mathbf{k}, \mathbf{q}) \{ [n_{\alpha}(\mathbf{q}) + 1] f(\mathbf{k}_1) [1 - f(\mathbf{k})] - n_{\alpha}(\mathbf{q}) f(\mathbf{k}) [1 - f(\mathbf{k}_1)] \} \\ & + \sum_{\alpha} \frac{d^d k_1}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} w_{\alpha}(\mathbf{k}, \mathbf{k}_1, \mathbf{q}) \{ n_{\alpha}(\mathbf{q}) f(\mathbf{k}_1) [1 - f(\mathbf{k})] - [n_{\alpha}(\mathbf{q}) + 1] f(\mathbf{k}) [1 - f(\mathbf{k}_1)] \}. \end{aligned} \quad (15.41)$$

In a similar manner, one can obtain the kinetic equation for the phonons:

$$\frac{dn_{\alpha}(\mathbf{q})}{dt} = \text{St}_{\text{ph}}[n_{\alpha}(\mathbf{q})], \quad (15.42)$$

where the collision integral for the phonons is:

$$\begin{aligned} \text{St}_{\text{ph}}[n_{\alpha}(\mathbf{q})] = & \sum_{\alpha} \frac{d^d k_1}{(2\pi)^d} \frac{d^d k}{(2\pi)^d} w_{\alpha}(\mathbf{k}_1, \mathbf{k}, \mathbf{q}) \{ [n_{\alpha}(\mathbf{q}) + 1] f(\mathbf{k}_1) [1 - f(\mathbf{k})] - n_{\alpha}(\mathbf{q}) f(\mathbf{k}) [1 - f(\mathbf{k}_1)] \}. \end{aligned} \quad (15.43)$$

The collision integrals satisfy several properties associated with conservation laws:

$$1. \quad \int \frac{d^d k}{(2\pi)^d} \text{St}_e [f(\mathbf{k})] = 0. \quad (15.44)$$

This property follows from the conservation of the number of electrons in the system. Noticing that the integral of $f(\mathbf{k})$ over the wavenumbers in the Brillouin zone gives the electron density, we see that the integral of the left-hand side of Eq. (15.39) must vanish. Hence the integral of the right-hand side of the equation also vanishes, leading to (15.44).

Notice, however, that phonons do not share this property because excluding some special situations (such as an equilibrium state), the phonon number is not conserved:

$$\sum_{\alpha} \int \frac{d^d q}{(2\pi)^d} \text{St}_{\text{ph}} [n_{\alpha}(\mathbf{q})] \neq 0 \quad (15.45)$$

2. Conservation of the total energy of the system implies:

$$\int \frac{d^d k}{(2\pi)^d} \varepsilon(\mathbf{k}) \text{St}_e [f(\mathbf{k})] + \sum_{\alpha} \int \frac{d^d q}{(2\pi)^d} \hbar \omega_{\alpha}(\mathbf{q}) \text{St}_{\text{ph}} [n_{\alpha}(\mathbf{q})] = 0. \quad (15.46)$$

The following steps yield the above equation: First, we multiply Eq. (15.39) by the electron energy $\varepsilon(\mathbf{k})$ and integrate over $\mathbf{k}$. This integral gives the time derivative of the total electronic energy of the system. Next, we multiply Eq. (15.42) by $\hbar \omega_{\alpha}(\mathbf{q})$, sum over the phonon polarization and integrate over $\mathbf{q}$. The result is the time derivative of the total phonon energy of the system (which does not take into account the zero-point energy of the phonons). Finally, by summing these two equations, one obtains an equation that describes the change in the total energy of the system. Conservation of this energy implies that the collision integrals should satisfy Eq. (15.46).

3. Similar considerations for the conservation of the total momentum give:

$$\int \frac{d^d k}{(2\pi)^d} \hbar \mathbf{k} \text{St}_e [f(\mathbf{k})] + \sum_{\alpha} \int \frac{d^d q}{(2\pi)^d} \hbar \mathbf{q} \text{St}_{\text{ph}} [n_{\alpha}(\mathbf{q})] = 0. \quad (15.47)$$

15.4 Thermodynamic equilibrium

In this section, we seek to find solutions of the kinetic equations (15.39) & (15.42) that describe a steady-state of the system. Namely, solutions that do not contain time dependence and hence should satisfy the condition that the collision integrals vanish:

$$\text{St}_{\text{ph}} [n_{\text{ss}}(\mathbf{q})] = \text{St}_e [f_{\text{ss}}(\mathbf{k})] = 0 \quad (15.48)$$

Because of our expectation that these solutions should include the equilibrium distribution of electrons and phonons, it is instructive to represent them as follows:

$$f_{ss}(\mathbf{k}) = \frac{1}{1+A(\mathbf{k})} \quad \text{so that} \quad 1-f_{ss}(\mathbf{k}) = \frac{A(\mathbf{k})}{1+A(\mathbf{k})}, \quad (15.49)$$

and

$$n_{ss}(\mathbf{q}) = \frac{1}{B(\mathbf{q})-1} \quad \text{so that} \quad 1+n_{ss}(\mathbf{q}) = \frac{B(\mathbf{q})}{B(\mathbf{q})-1}, \quad (15.50)$$

where $A(\mathbf{k})$ and $B(\mathbf{q})$ are some general functions. Substituting the above formulas in Eq. (15.41) for $\text{St}_e[f_{ss}(\mathbf{k})]$, we obtain that the expression in the curly brackets of the first term is:

$$\begin{aligned} & \{[n_{ss}(\mathbf{q})+1]f_{ss}(\mathbf{k}_1)[1-f_{ss}(\mathbf{k})]-n_{ss}(\mathbf{q})f_{ss}(\mathbf{k})[1-f_{ss}(\mathbf{k}_1)]\} \\ &= \frac{A(\mathbf{k}_1)}{[B(\mathbf{q})-1][1+A(\mathbf{k}_1)][1+A(\mathbf{k})]} [B(\mathbf{q})A(\mathbf{k})A^{-1}(\mathbf{k}_1)-1] \end{aligned} \quad (15.51)$$

To nullify this expression, one can employ the energy and the momentum conservation laws of the scattering process, which are enforced by the δ -functions in Eq. (15.36). In particular, choosing:

$$\begin{aligned} A(\mathbf{k}) &= \exp(\beta[\varepsilon(\mathbf{k})-\hbar\mathbf{v}\cdot\mathbf{k}-\varepsilon_F]) \\ B(\mathbf{q}) &= \exp(\beta\hbar[\omega_\alpha(\mathbf{q})-\mathbf{v}\cdot\mathbf{q}]) \end{aligned} \quad (15.52)$$

where the scalar β (having a dimension of inverse energy) and the vector $\mathbf{v}$ (having dimensions of velocity) are arbitrary constants, we obtain:

$$\begin{aligned} & B(\mathbf{q})A(\mathbf{k})A^{-1}(\mathbf{k}_1)-1 \\ &= \exp\{\beta[\hbar\omega_\alpha(\mathbf{q})+\varepsilon(\mathbf{k})-\varepsilon(\mathbf{k}_1)]-\beta\hbar\mathbf{v}\cdot(\mathbf{q}+\mathbf{k}-\mathbf{k}_1)\}-1=0. \end{aligned} \quad (15.53)$$

One can quickly check that the second term in the collision integral $\text{St}_e[f_{ss}(\mathbf{k})]$, as well as those in $\text{St}_{ph}[n_{ss}(\mathbf{q})]$, vanish for the same reason.

Identifying β as the inverse temperature of the system (by calculating the average energy of a particle), the steady-state solutions describe a system of fermions and bosons in equilibrium that moves at a constant velocity, $\mathbf{v}$, with respect to the laboratory reference system. Such a solution is expected by the Galilean invariance of the system. When $\mathbf{v}=0$ we obtain the Fermi-Dirac and the Plank distributions:

$$f_{eq}[\varepsilon(\mathbf{k})] = \frac{1}{\exp\{\beta[\varepsilon(\mathbf{k})-\varepsilon_F]\}+1}, \quad (15.54)$$

and

$$n_{\text{eq}}[\omega(\mathbf{q})] = \frac{1}{\exp[\beta\hbar\omega(\mathbf{q})] - 1}, \quad (15.55)$$

respectively. Here $\beta = 1/k_B T$ and ε_F is the chemical potential (Fermi energy) of the electrons. Notice that the chemical potential of the phonon is zero. The following argument clarifies this property of the phonon system: The chemical potential, defined as the energy needed to add a phonon to the system, is $\mu = \partial F / \partial N$ where F is the free energy while N is the number of phonons. Since there is no conservation law for the number of phonons, the system minimizes the free energy by choosing some particular number. However, at this number, $\partial F / \partial N = 0$, hence the chemical potential vanishes.

The solution obtained here demonstrates that electron-phonon interaction drives the system into equilibrium distribution. Notice, however, that the system's thermalization may also be obtained by other types of interactions, for instance, electron-electron interactions.

To find the rate at which the system decays to the equilibrium distribution, let us perturb the system such that the electrons move slightly from their equilibrium distribution:

$$f[\varepsilon(\mathbf{k})] = f_{\text{eq}}[\varepsilon(\mathbf{k})] + \delta f(\mathbf{k}), \quad (15.56)$$

but the phonon bath remains at equilibrium. Then expanding the collision integral to first order in $\delta f(\mathbf{k})$ yields:

$$\text{St}_e[f_{\text{eq}}[\varepsilon(\mathbf{k})] + \delta f(\mathbf{k})] = -\frac{\delta f(\mathbf{k})}{\tau_{\text{e-ph}}}, \quad (15.57)$$

where $\tau_{\text{e-ph}}$ is the electron-phonon scattering time, which sets the relaxation time to equilibrium. One can identify its temperature dependence without calculation by taking into account that the relaxation rate is determined only by the phonons that contribute to the process. The electron-phonon scattering contains three factors: the volume element of the collision integral, the matrix element, and the constraint set by momentum conservation:

$$\frac{1}{\tau_{\text{e-ph}}} \sim \underbrace{q^3}_{\substack{\text{integration volume} \\ d^3q}} \times \underbrace{\frac{q^2}{q}}_{\substack{\text{matrix element}}} \times \underbrace{\frac{1}{q}}_{\substack{\delta\text{-function} \\ \text{of momentum}}} \sim q^3 \propto \frac{1}{\beta^3} \propto T^3, \quad (15.58)$$

where we took into account only acoustic phonons, hence $q \propto 1/\beta$ as follows from the condition $\beta\hbar\omega(\mathbf{q}) = \beta\hbar cq \sim 1$.

Example: Calculation of $\tau_{\text{e-ph}}$

Here we calculate the relaxation time due to the interaction of an electron with acoustic phonons. We shall consider a three-dimensional system, assume that phonons are at thermal

equilibrium, and, for simplicity, that the sound velocities of all-acoustic branches of the phonon spectrum are equal to c , i.e. $\omega_\alpha(\mathbf{q}) = \omega_q = cq$. Then from Eqs. (15.57) and (15.41) we obtain:

$$\frac{1}{\tau_{\text{e-ph}}} = \int \frac{d^3 k_1}{(2\pi)^3} \frac{d^3 q}{(2\pi)^3} \times \left\{ w_\alpha(\mathbf{k}_1, \mathbf{k}, \mathbf{q}) \left[f_{\text{eq}}(\varepsilon_1) + n_{\text{eq}}(\omega_q) \right] + w_\alpha(\mathbf{k}, \mathbf{k}_1, \mathbf{q}) \left[n_{\text{eq}}(\omega_q) + 1 - f_{\text{eq}}(\varepsilon_1) \right] \right\}, \quad (15.59)$$

where to shorten the notations we have defined $\varepsilon_1 = \varepsilon(\mathbf{k}_1)$. The above equation is obtained by substituting $f = f_{\text{eq}}(\varepsilon) + \delta f(\mathbf{k})$ in the collision integral (15.41) and expanding it to leading order in $\delta f(\mathbf{k})$ in order to get Eq. (15.57). In principle, there is an additional contribution coming from the expansion of $f(\mathbf{k}_1) = f_{\text{eq}}(\varepsilon_1) + \delta f(\mathbf{k}_1)$. However, this contribution should be calculated at $\mathbf{k}_1 = \mathbf{k}$, i.e., $\mathbf{q} = 0$ where $w_\alpha(\mathbf{k}_1, \mathbf{k}, 0) = 0$, therefore, it vanishes.

Performing the integral over $\mathbf{k}_1$ we have

$$\begin{aligned} \frac{1}{\tau_{\text{e-ph}}} = & \int \frac{d^3 q}{(2\pi)^3} \tilde{w}(\mathbf{q}) \left[f_{\text{eq}} \left[\varepsilon(\mathbf{k}) + \hbar \omega_q \right] + n_{\text{eq}}(\omega_q) \right] \delta \left[\varepsilon(\mathbf{k} + \mathbf{q}) - \varepsilon(\mathbf{k}) - \hbar \omega_q \right] \\ & + \int \frac{d^3 q}{(2\pi)^3} \tilde{w}(\mathbf{q}) \left\{ n_{\text{eq}}(\omega_q) + 1 - f_{\text{eq}} \left[\varepsilon(\mathbf{k}) - \hbar \omega_q \right] \right\} \delta \left[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k} - \mathbf{q}) - \hbar \omega_q \right], \end{aligned} \quad (15.60)$$

where to obtain the arguments of the distribution functions we have used conservation of energy in order to replace $\varepsilon_1 = \varepsilon(\mathbf{k} \pm \mathbf{q})$ by $\varepsilon(\mathbf{k}) \pm \hbar \omega_q$, and defined

$$\tilde{w}(\mathbf{q}) = 2\pi \sum_{\alpha} \frac{|\lambda_{ij} e_i^{(\alpha)} q_j|^2}{2\rho \omega_q}. \quad (15.61)$$

To simplify the calculation, from now on, we assume that the electron-phonon coupling tensor is diagonal $\lambda_{ij} = \lambda \delta_{ij}$. With this assumption, we obtain:

$$\frac{\tilde{w}(\mathbf{q})}{2\pi} = \sum_{\alpha} \frac{|\lambda_{ij} e_i^{(\alpha)} q_j|^2}{2\rho \omega_q} = \sum_{\alpha} \frac{\lambda^2 \delta_{ij} e_i^{(\alpha)} q_j \delta_{kl} e_k^{(\alpha)} q_l}{2\rho \omega_q} = \frac{\lambda^2 \delta_{ij} q_j \delta_{kl} q_l \delta_{ik}}{2\rho \omega_q} = \frac{\lambda^2 q^2}{2\rho \omega_q} = \frac{\lambda^2 q}{2\rho c}. \quad (15.62)$$

The third equality in the above formula follows from the completeness relation of the polarization vectors expressed in Eq. (11.46), while for the last equality, we substitute $\omega_q = cq$.

Now to perform the integral (15.60) over $\mathbf{q}$, notice that under our assumptions, the only angular dependence of this vector appears in the argument of the δ -function that secures energy conservation in Eq. (15.60). Hence if we consider the typical situation in metals, where $q \ll k_F$, we may use the approximation

$$\varepsilon(\mathbf{k} \pm \mathbf{q}) - \varepsilon(\mathbf{k}) \mp \hbar \omega_q \simeq \pm (\hbar \mathbf{v}_F \cdot \mathbf{q} - \hbar c q). \quad (15.63)$$

Here we have neglected terms of order q^2 and for the first order expansion in $\mathbf{q}$ used the approximation:

$$\left. \frac{\partial \varepsilon(\mathbf{k})}{\partial \mathbf{k}} \simeq \frac{\partial \varepsilon(\mathbf{k})}{\partial \mathbf{k}} \right|_{\mathbf{k}=\mathbf{k}_F} = \hbar \mathbf{v}_F. \quad (15.64)$$

Namely, we approximate the velocity by its value at the Fermi energy. This approximation is justified by the fact that deviations from the equilibrium distribution function of the electrons usually take place only near the Fermi energy.

With approximation (15.63), the angular integration over the energy δ -functions in Eq. (15.60) yields

$$\int_0^{2\pi} d\varphi \int_0^\pi d\theta \sin \theta \delta[\pm (\hbar v_F q \cos \theta - \hbar c q)] = \frac{2\pi}{\hbar v_F q} \int_{-1}^1 dx \delta\left[x - \frac{c}{v_F}\right] = \frac{2\pi}{\hbar v_F q}, \quad (15.65)$$

where we took into account that the sound velocity is much smaller than the Fermi velocity, $c \ll v_F$ (because electrons are light, and ions are heavy). Using this result for the integral (15.60) and rearranging the terms we obtain

$$\frac{1}{\tau_{\text{e-ph}}} = \int \frac{q^2 dq}{(2\pi)^2} \frac{\tilde{w}(\mathbf{q})}{\hbar v_F q} \left\{ 2n_{\text{eq}}(\mathbf{q}) + 1 + f_{\text{eq}}[\varepsilon(\mathbf{k}) + \hbar \omega_q] - f_{\text{eq}}[\varepsilon(\mathbf{k}) - \hbar \omega_q] \right\}. \quad (15.66)$$

To approximate the expression in the curly brackets, we use the property that deviations from equilibrium are usually very close to the Fermi energy, thus, we may set $\varepsilon(\mathbf{k}) \simeq \varepsilon_F$. With this approximation, one can verify that

$$\begin{aligned} & 2n_{\text{eq}}(\mathbf{q}) + 1 + f_{\text{eq}}[\varepsilon(\mathbf{k}) + \hbar \omega_q] - f_{\text{eq}}[\varepsilon(\mathbf{k}) - \hbar \omega_q] \\ & \simeq \frac{2}{\exp[\beta \hbar \omega_q] - 1} + \frac{2}{\exp[\beta \hbar \omega_q] + 1}. \end{aligned} \quad (15.67)$$

Substituting this result and Eq. (15.62) in (15.66), we obtain

$$\begin{aligned} \frac{1}{\tau_{\text{e-ph}}} &= \int_0^\infty \frac{q^2 dq}{2\pi} \frac{\lambda^2}{\hbar v_F \rho c} \left[\frac{1}{\exp(\beta \hbar c q) - 1} + \frac{1}{\exp(\beta \hbar c q) + 1} \right] \\ &= \frac{\lambda^2}{\hbar^4 c^4 v_F \rho \beta^3} \int_0^\infty \frac{y^2 dy}{2\pi} \left[\frac{1}{\exp(y) - 1} + \frac{1}{\exp(y) + 1} \right] \\ &= \frac{7\zeta(3)}{4\pi} \frac{\lambda^2 (k_B T)^3}{\hbar^4 c^4 v_F \rho}, \end{aligned} \quad (15.68)$$

where

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad (15.69)$$

is the Reimann zeta function, and in particular $\zeta(3) \simeq 1.202$.

15.5 The thermodynamic equilibrium from entropy considerations

Equations (15.48), from which we deduced the equilibrium distribution of the system, are nonlinear integral equations. Being nonlinear, it is unclear whether (15.54) and (15.55) are indeed the correct solutions into which the system relaxes. To prove that, we shall employ entropy considerations.

Let us define the entropy density (entropy per unit volume) of the electrons and the phonons by:

$$S_e = -k_B \int \frac{d^d k}{(2\pi)^d} \left\{ f(\mathbf{k}) \ln f(\mathbf{k}) + [1 - f(\mathbf{k})] \ln [1 - f(\mathbf{k})] \right\}, \quad (15.70)$$

and

$$S_{ph} = -k_B \int \frac{d^d q}{(2\pi)^d} \left\{ n(\mathbf{q}) \ln n(\mathbf{q}) - [1 + n(\mathbf{q})] \ln [1 + n(\mathbf{q})] \right\}, \quad (15.71)$$

respectively. Using the kinetic equations (15.39) and (15.42), one can prove (see Ex. 1) that these quantities satisfy the inequality

$$\frac{d}{dt} [S_e + S_{ph}] \geq 0. \quad (15.72)$$

This inequality is a generalization of Boltzmann's H-theorem for electrons and phonons in a lattice. It implies that an equilibrium state is a state with maximal total entropy. Hence to identify this state, one should maximize the entropy density under the constraints of a fixed number of electrons, and fixed energy, namely:

$$\frac{\delta}{\delta f(\mathbf{k})} \left\{ S_e + S_{ph} - \frac{1}{T} \left[\int \frac{d^d k}{(2\pi)^d} [\varepsilon(\mathbf{k}) - \varepsilon_F] f(\mathbf{k}) \right] \right\} = 0, \quad (15.73)$$

where T and ε_F are Lagrange multipliers that impose the constraints mentioned above. From this equation, we obtain:

$$\ln \left(\frac{f(\mathbf{k})}{1 - f(\mathbf{k})} \right) = -\frac{1}{k_B T} [\varepsilon(\mathbf{k}) - \varepsilon_F], \quad (15.74)$$

which leads to formula (15.54). Similar considerations show that the equilibrium distribution of the phonons is given by Eq. (15.55).

Example: Relation to Shannon's entropy

Although the above definitions for the entropy densities, (15.70) and (15.71) need no justification (because it is enough to prove that they satisfy the inequality (15.72)), it is instructive to show that they are consistent with the definition of the *information entropy*,

i.e., *Shannon's entropy* (Shannon, 1948). The information entropy is defined as follows: Let P_i be the probability to find the system in a state i , then the information entropy is

$$S = -\sum_i P_i \ln P_i. \quad (15.75)$$

In this example, we show that this definition is consistent with the electronic entropy density defined in Eq. (15.70). The case of phonon entropy is given as an exercise.

Since we neglect electron-electron interaction in our treatment, the probability of finding a configuration with N_i electrons in state i is independent of the occupation of other states of the system. Thus, the probability of finding N_1 electrons in state 1, N_2 in state 2, etc. is given by a product of probabilities:

$$P(N_1, N_2, \dots) = P_1(N_1) P_2(N_2) \dots \quad (15.76)$$

Now, let us assume that the probability of occupying the state j by $N_j = 0, 1$ electrons is $P_j(N_j) \propto q_j^{N_j}$ where $0 \leq q_j \leq 1$ is an unknown quantity (which depends, for instance, on the state that we have prepared the system). From the requirement of normalization of the probability, we obtain that:

$$P_j(N_j) = \frac{1}{1+q_j} q_j^{N_j}. \quad (15.77)$$

With this probability, the average occupation of the j -th state is given by:

$$f_j = \sum_{N_j=0,1} N_j \frac{q_j^{N_j}}{1+q_j} = \frac{q_j}{1+q_j}. \quad (15.78)$$

Now let us substitute (15.76) and (15.77) in definition (15.75) and use (15.78):

$$\begin{aligned} S &= -\sum_{N_1=0}^1 \dots \sum_{N_j=0}^1 \dots P(N_1, N_2, \dots) \ln P(N_1, N_2, \dots) \\ &= -\sum_j \sum_{N_j=0}^1 \frac{q_j^{N_j}}{1+q_j} \ln \left[\frac{q_j^{N_j}}{1+q_j} \right] \\ &= -\sum_j \frac{1}{1+q_j} \ln \left[\frac{1}{1+q_j} \right] + \frac{q_j}{1+q_j} \ln \left[\frac{q_j}{1+q_j} \right] \\ &= -\sum_j f_j \ln(f_j) + (1-f_j) \ln(1-f_j). \end{aligned} \quad (15.79)$$

Thus, apart from some physical constants, this formula is precisely the same as the electronic entropy density defined by Eq. (15.70).

15.6 Piezoelectric interaction

As we have learned in chapters 13 & 14, deformations in piezoelectric crystals generate a polarization vector. The polarization, in turn, induces an electric field that acts on the electrons. This mechanism is the primary mechanism for electron-phonon interaction in piezoelectric crystals. The Hamiltonian describing this interaction (see Eq. (14.2)) is:

$$H_{\text{e-ph,piezo}} = \int d^3r' \frac{e}{4\pi\epsilon_0\epsilon_r |\mathbf{r} - \mathbf{r}'|} \gamma_{i;kl} \frac{\partial u_{kl}(\mathbf{r}')}{\partial r'_i}. \quad (15.80)$$

Substituting Eq. (15.7) for the strain tensor we obtain:

$$H_{\text{e-ph,piezo}} = \int d^3r' \frac{-e\gamma_{i;lk}}{4\pi\epsilon_0\epsilon_r |\mathbf{r} - \mathbf{r}'|} \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q}) \text{Vol}}} e_k^{(\alpha)} q_l q_i \left[\hat{a}_\alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{r}') + \hat{a}_\alpha^\dagger(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r}') \right], \quad (15.81)$$

where we have used that $\gamma_{i;lk}$ is symmetric to the interchange of the indices k and l . The integral over $\mathbf{r}'$ is the Fourier transform of the Coulomb potential in three dimensions, hence

$$H_{\text{e-ph,piezo}} = \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q}) \text{Vol}}} \frac{-e\gamma_{i;lk} e_k^{(\alpha)} q_l q_i}{\epsilon_0\epsilon_r q^2} \left[\hat{a}_\alpha(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{r}) + \hat{a}_\alpha^\dagger(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r}) \right]. \quad (15.82)$$

Finally, calculating the matrix elements of this Hamiltonian between two Bloch's wave functions, we obtain:

$$H_{\text{e-ph,piezo}}^{kk'} = - \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\rho\omega_\alpha(\mathbf{q}) \text{Vol}}} \frac{e\gamma_{i;lk} e_k^{(\alpha)} q_l q_i}{\epsilon_0\epsilon_r q^2} \left[\hat{a}_\alpha(\mathbf{q}) \delta_{\mathbf{k}-\mathbf{k}',-\mathbf{q}} + \hat{a}_\alpha^\dagger(\mathbf{q}) \delta_{\mathbf{k}-\mathbf{k}',\mathbf{q}} \right]. \quad (15.83)$$

Now, following the same steps that lead to formula (15.26) for the density of transition rate in the case of deformation potential, for piezoelectric interaction, we obtain

$$w_\alpha(\mathbf{k}_f, \mathbf{k}_i, \mathbf{q}) = \frac{(2\pi)^{d+1}}{2\rho\omega_\alpha(\mathbf{q})} \left| \frac{e\gamma_{i;lk} e_k^{(\alpha)} q_l q_i}{\epsilon_0\epsilon_r q^2} \right|^2 \delta(\mathbf{k}_f - \mathbf{k}_i - \mathbf{q}) \delta[\varepsilon(\mathbf{k}_f) - \varepsilon(\mathbf{k}_i) - \hbar\omega_\alpha(\mathbf{q})]. \quad (15.84)$$

With this formula, one can deduce the temperature dependence of the relaxation rate using the same arguments that lead to Eq. (15.58). The only difference is that the matrix element is proportional to $1/q$ rather than to q , and since $q \propto T$ one obtains

$$\frac{1}{\tau_{\text{e-ph}}^{\text{piezo}}} \propto T. \quad (15.85)$$

This result applies for lightly doped semiconductors in which the density of charge carriers is small and dictated by the temperature (as will be discussed in Chapter 17), and screening effects are small. A more accurate approximation is obtained when replacing the Coulomb

interaction with screened Coulomb interaction (see Eq. (14.14)), i.e., by changing the denominator in Eq. (15.84) according to $\varepsilon_0 \varepsilon_r q^2 \rightarrow \varepsilon_0 \varepsilon_r (q^2 + q_D^2)$, where $1/q_D$ is *Debye's screening length*. The latter is given by

$$q_D = \sqrt{\frac{e^2 n_*}{\varepsilon_0 \varepsilon_r k_B T}}, \quad (15.86)$$

where n_* is the effective density of the charge carriers. This type of screening adds a logarithmic temperature dependence to Eq. (15.85), see Ex. 2. We will return to discuss Debye's screening (15.86) in chapter 17 (see Example on page 333).

15.7 Fröhlich's polaron

Consider an electron moving in a polarizable material. The electron's electric field will polarize the medium by attracting the positive ions and repelling the negative ions, as illustrated in Fig. 15-2. This lattice deformation lowers the energy of the system. When the coupling between the electron and the lattice deformation is strong, one cannot decouple these two degrees of freedom. The excitation that mixes them is a particle called *polaron*. This section, first, presents the Fröhlich Hamiltonian (Fröhlich, 1950) that provides a simple model for such a system. Then, we use a variational approach to calculate the ground state energy of the polaron and estimate its size.

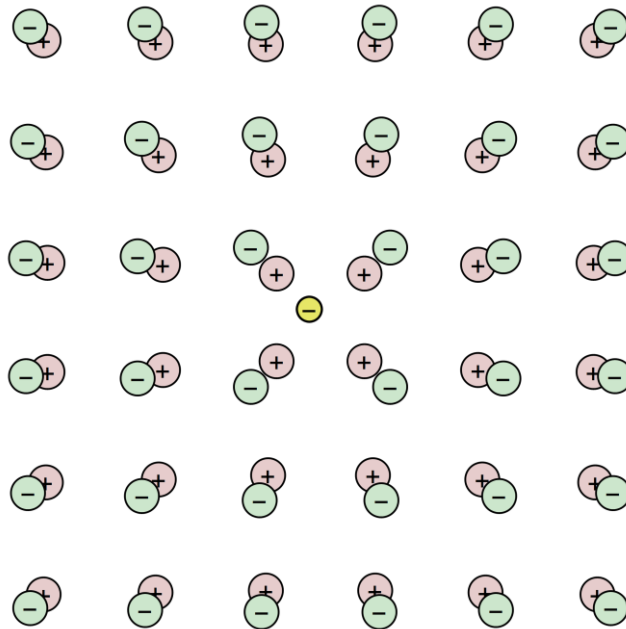


Figure 15-2 An electron in a polarizable material

Recall that deformations in a polarizable crystal generate polarization vector:

$$\mathbf{P} = \rho_p \boldsymbol{\xi}, \quad (15.87)$$

where $\boldsymbol{\xi}$ represents a linear combination of the optical normal modes that create a local deformation of the lattice¹. The polarization charge is minus the divergence of the polarization vector, $-\nabla \cdot \mathbf{P}$, and this charge generates a Coulomb potential that acts on the electron. Thus, the Hamiltonian that describes the coupling between the electron and the optical phonons of a lattice is:

$$H_{\text{e-ph}}(\mathbf{r}) = \int d^d r' \frac{e \rho_p \nabla \cdot \boldsymbol{\xi}(\mathbf{r}')}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|}. \quad (15.88)$$

The general expansion of $\boldsymbol{\xi}(\mathbf{r}')$, in terms of the optical modes of the system, takes the form:

$$\boldsymbol{\xi}(\mathbf{r}') = \sum_{\alpha, \mathbf{q}} \sqrt{\frac{\hbar}{2\text{Vol}\rho\omega_\alpha(\mathbf{q})}} \mathbf{e}_\alpha Q_{\alpha, \mathbf{q}} \exp(-i\mathbf{q} \cdot \mathbf{r}'), \quad (15.89)$$

where ρ is the mass density of the lattice, $\omega_\alpha(\mathbf{q})$ is the frequency of the normal mode with polarization $\mathbf{e}_\alpha$ and wavenumber $\mathbf{q}$, and $Q_{\alpha, \mathbf{q}}$ is the dimensionless amplitude of the normal mode. Notice, however, that the divergence operator leaves only the contribution from the longitudinal modes, $\nabla \cdot \boldsymbol{\xi}(\mathbf{r}') \propto \mathbf{q} \cdot \mathbf{e}_\alpha$. To simplify this Hamiltonian, we assume that all the longitudinal optical modes of the system have the same frequency, ω_0 , independent of the wavelength (Einstein model), thus for our purpose,

$$\boldsymbol{\xi}(\mathbf{r}') = \sum_{\mathbf{q}} \sqrt{\frac{\hbar}{2\text{Vol}\rho\omega_0}} \hat{\mathbf{q}} Q_{\mathbf{q}} \exp(-i\mathbf{q} \cdot \mathbf{r}'), \quad (15.90)$$

where $\hat{\mathbf{q}}$ is a unit vector in the direction of $\mathbf{q}$. Substituting this equation in (15.88) and integrating over $\mathbf{r}'$ yields (for a three-dimensional system):

$$H_{\text{e-ph}}(\mathbf{r}) = \sqrt{\frac{2}{\text{Vol}}} M_0 \sum_{\mathbf{q}} \frac{1}{q} Q_{\mathbf{q}} \exp(-i\mathbf{q} \cdot \mathbf{r}), \quad (15.91)$$

where

$$M_0 = \frac{-ie\rho_p}{2\epsilon_0} \sqrt{\frac{\hbar}{\rho\omega_0}}. \quad (15.92)$$

The Fröhlich Hamiltonian describes a single electron in the environment of optical phonons. It contains three contributions:

¹ Notice that $\boldsymbol{\xi}$ is different from the displacement vector $\mathbf{u}$ because the latter refers to displacement of the whole lattice cell which do not produce polarization.

$$\begin{aligned}
H &= H_e + H_{ph} + H_{e-ph} \\
&= \frac{p^2}{2m} + \frac{\hbar\omega_0}{2} \sum_q (P_q P_q^* + Q_q Q_q^*) + \sqrt{\frac{2}{\text{Vol}}} M_0 \sum_q \frac{1}{q} Q_q \exp(-i\mathbf{q} \cdot \mathbf{r}).
\end{aligned} \tag{15.93}$$

The first term describes a parabolic dispersion of an electron near the bottom of a band with effective mass m . The second term is the contribution from the optical phonons of the system, which is expressed in terms of a sum over harmonic oscillators having an equal frequency, ω_0 . Here P_q is the dimensionless momentum that is conjugated to Q_q , and satisfies the commutation relations:

$$[P_q; Q_{q'}] = -i\delta_{q,q'}. \tag{15.94}$$

Finally, the third term in Eq. (15.93) represents the electron coupling to the phonons.

To find the approximate minimal energy of the above Hamiltonian, we use a variational principle calculation. We choose an electronic wave function describing a localized particle:

$$\psi_e(\mathbf{r}) = \frac{b^{3/2}}{\pi^{3/4}} \exp\left(-\frac{b^2 r^2}{2}\right), \tag{15.95}$$

where b is the variational parameter. Now, we project the Hamiltonian (15.93) on this family of wave functions (parametrized by b) by calculating its expectation value:

$$\langle H \rangle = \langle \psi_e | H | \psi_e \rangle = \frac{3}{4} \frac{\hbar^2 b^2}{m} + \frac{\hbar\omega_0}{2} \sum_q (P_q P_q^* + Q_q Q_q^*) - \frac{1}{2} \sum_q (L_q Q_q^* + L_q^* Q_q), \tag{15.96}$$

where

$$L_q = \sqrt{\frac{2}{\text{Vol}}} \frac{M_0}{q} \exp\left(-\frac{q^2}{4b^2}\right). \tag{15.97}$$

The projected Hamiltonian is quadratic in the variables of the harmonic oscillators, and its minimum can be calculated exactly. The minimum is obtained at $Q_q = Q_q^{\min}$ which satisfies the condition:

$$\frac{\partial}{\partial Q_q^*} \langle H \rangle = \frac{\hbar\omega_0}{2} Q_q - \frac{1}{2} L_q = 0. \tag{15.98}$$

Thus

$$Q_q^{\min} = \frac{L_q}{\hbar\omega_0}. \tag{15.99}$$

Shifting the variables as

$$Q_q \rightarrow Q_q - Q_q^{\min} = Q_q - \frac{L_q}{\hbar\omega_0} \tag{15.100}$$

transforms the projected Hamiltonian to

$$\langle H \rangle = \langle \psi_e | H | \psi_e \rangle = \frac{3}{4} \frac{\hbar^2 b^2}{m} + \frac{\hbar \omega_0}{2} \sum_q (P_q P_q^* + Q_q Q_q^*) - \frac{1}{2\hbar \omega_0} \sum_q |L_q|^2. \quad (15.101)$$

The last term on the right-hand side of the equation is:

$$\begin{aligned} \frac{1}{2\hbar \omega_0} \sum_q |L_q|^2 &= \frac{|M_0|^2}{\hbar \omega_0 \text{Vol}} \sum_q \frac{1}{q^2} \exp\left(-\frac{q^2}{2b^2}\right) \\ &= \frac{|M_0|^2}{\hbar \omega_0} \int \frac{d^3 q}{(2\pi)^3} \frac{1}{q^2} \exp\left(-\frac{q^2}{2b^2}\right) = \frac{|M_0|^2 b}{\hbar \omega_0 (2\pi)^{3/2}}, \end{aligned} \quad (15.102)$$

thus, the projected Hamiltonian takes the form

$$\langle H \rangle = \frac{\hbar \omega_0}{2} \sum_q (P_q P_q^* + Q_q Q_q^*) + \underbrace{\frac{3}{4} \frac{\hbar^2 b^2}{m} - \alpha \left(\frac{\hbar^3 b^2 \omega_0}{\pi m} \right)^{\frac{1}{2}}}_{E(b)}, \quad (15.103)$$

with

$$\alpha = \frac{\sqrt{m} |M_0|^2}{\hbar (2\hbar \omega_0)^{3/2} \pi}. \quad (15.104)$$

In formula (15.103), the first term represents the phonon energy of the system, which is now decoupled from the polaron energy, $E(b)$. The minimization over the parameter b can be straightforwardly calculated:

$$\frac{d\langle H \rangle}{db} = \frac{dE(b)}{db} = \frac{3}{2} \frac{\hbar^2 b}{m} - \alpha \left(\frac{\hbar^3 \omega_0}{\pi m} \right)^{\frac{1}{2}} = 0 \quad \Rightarrow \quad b_{\min} = \frac{2\alpha}{3} \sqrt{\frac{m \omega_0}{\pi \hbar}}, \quad (15.105)$$

and substituting this result in $E(b)$ yields an upper limit on the polaron's ground state energy:

$$E_0 \leq E(b_{\min}) = -\frac{\alpha^2}{3\pi} \hbar \omega_0 = -0.106 \alpha^2 \hbar \omega_0. \quad (15.106)$$

The exact asymptotic expansion of the ground state energy in the limit of strong coupling, $\alpha \rightarrow \infty$ (Miyake, 1975), is given by

$$E_0(\alpha) = -\hbar \omega_0 \left[0.1085 \alpha^2 + 2.836 + O\left(\frac{1}{\alpha^2}\right) \right]. \quad (15.107)$$

Thus, the above variational calculation gives a rather accurate result. Finally, we can also deduce the size of the polaron, l . It equals the inverse of $b_{\min}$ as can be seen from Eq. (15.95).

Thus

$$l \simeq \frac{1}{b_{\min}} \simeq \frac{3}{2\alpha} \sqrt{\frac{\pi\hbar}{m\omega_0}} \sim \frac{42\text{\AA}}{\alpha}. \quad (15.108)$$

Here we have used a typical value of frequency of optical phonons, $\hbar\omega_0 \sim 0.03\text{eV}$, and the mass of a free electron. Typically, $\alpha \sim 5 \div 6$ therefore the size of the polaron is of the order of $10\text{\AA}$.

15.8 Exercises

1. Prove Boltzmann's H-theorem for a system of interacting electron and phonons

Advice: The total entropy of the system is

$$S = -k_B \int \frac{d^d k}{(2\pi)^d} \left\{ f(\mathbf{k}) \ln f(\mathbf{k}) + [1 - f(\mathbf{k})] \ln [1 - f(\mathbf{k})] \right. \\ \left. + n(\mathbf{k}) \ln n(\mathbf{k}) - [1 + n(\mathbf{k})] \ln [1 + n(\mathbf{k})] \right\}, \quad (15.109)$$

where $f(\mathbf{k})$ and $n(\mathbf{k})$ are the electrons and the phonons distribution functions, respectively. Take the time derivative of this entropy and use the kinetic equations (15.39) and (15.42) to show that

$$\frac{\partial S}{\partial t} = k_B \int d^d k d^d k_1 d^d q w_\alpha(\mathbf{k}_1, \mathbf{k}, \mathbf{q}) [1 - f(\mathbf{k})] f(\mathbf{k}_1) [1 + n(\mathbf{q})] (x - 1) \ln(x), \quad (15.110)$$

where

$$x = \frac{f(\mathbf{k})}{1 - f(\mathbf{k})} \frac{n(\mathbf{k})}{1 + n(\mathbf{k})} \frac{1 - f(\mathbf{k}_1)}{f(\mathbf{k}_1)}. \quad (15.111)$$

Finally, show that maximum entropy is obtained at $x = 1$, and this condition is satisfied by the equilibrium distribution functions of the electrons and the phonons.

2. Calculate the relaxation rate of electrons due to piezoelectric interaction for the simplest case where only one parameter controls the interaction:

$$\gamma_{i,jk} = \gamma \mathcal{E}_{ijk}^{\text{sym}}, \quad (15.112)$$

where γ is a constant (in the literature, it is usually denoted by d_{14}) and $\mathcal{E}_{ijk}^{\text{sym}}$ is a symmetric tensor which equals one when all its indices are different and equal zero if any two of them are equal. Also, assume acoustic branches of the phonon spectrum can be approximated by the spherical approximation, namely the sound velocities are independent of the

propagation direction of the wave but only on its polarization: The velocities of the two transverse waves are degenerate and equal $c_{\perp}$ while the velocity of the longitudinal wave is $c_{\parallel}$. Go along the following steps:

(a) Show that the relaxation rate is given by

$$\begin{aligned} \frac{1}{\tau_{\text{e-ph}}^{\text{piezo}}} = & \sum_{\alpha=\perp,\parallel} \int \frac{d^3q}{(2\pi)^3} \tilde{w}_{\alpha}(\mathbf{q}) \left[f_{\text{eq}}[\varepsilon(\mathbf{k}) + \hbar\omega_{\alpha}] + n_{\text{eq}}(\omega_{\alpha}) \right] \delta[\varepsilon(\mathbf{k} + \mathbf{q}) - \varepsilon(\mathbf{k}) - \hbar\omega_{\alpha}] \\ & + \sum_{\alpha=\perp,\parallel} \int \frac{d^3q}{(2\pi)^3} \tilde{w}_{\alpha}(\mathbf{q}) \left\{ n_{\text{eq}}(\omega_{\alpha}) + 1 - f_{\text{eq}}[\varepsilon(\mathbf{k}) - \hbar\omega_{\alpha}] \right\} \delta[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k} - \mathbf{q}) - \hbar\omega_{\alpha}] \end{aligned} \quad (15.113)$$

with

$$\tilde{w}_{\perp}(\mathbf{q}) = \frac{\pi}{\rho} \sum_{\alpha=\perp,2} \frac{1}{\omega_{\alpha}(\mathbf{q})} \left| \frac{e\gamma_{i,jk} e_j^{(\alpha)} q_i q_k}{\varepsilon_{\infty} q^2} \right|^2 = \frac{\gamma^2 e^2 \pi}{\rho c_{\perp} q \varepsilon_{\infty}^2} \left[2 \left(1 - \sum_k \hat{q}_k^4 \right) - \left(\gamma_{i,jk} \hat{q}_i \hat{q}_j \hat{q}_k \right)^2 \right] \quad (15.114)$$

and

$$\tilde{w}_{\parallel}(\mathbf{q}) = \frac{\pi}{\rho} \frac{1}{\omega_{\parallel}(\mathbf{q})} \left| \frac{e\gamma_{i,jk} e_j^{(\parallel)} q_i q_k}{\varepsilon_{\infty} q^2} \right|^2 = \frac{\gamma^2 e^2 \pi}{\rho c_{\parallel} q \varepsilon_{\infty}^2} \left(\varepsilon_{ijk}^{\text{sym}} \hat{q}_i \hat{q}_j \hat{q}_k \right)^2 \quad (15.115)$$

where $\hat{\mathbf{q}}$ is a unit vector in the direction of $\mathbf{q}$, $\varepsilon_{\infty} = \varepsilon_0 \varepsilon_r$ is the material's dielectric constant, and, for the time being, we ignore screening effects.

(b) Ignore the angular dependence of $\varepsilon(\mathbf{k})$ (because in what follows we shall assume $\varepsilon(\mathbf{k}) \simeq \varepsilon_F$), and use the same approximation as in Eq. (15.63) in order to average the δ -function that ensures momentum conservation over the directions of the electron velocity, i.e., over $\mathbf{v}_F$. Show that with this averaging Eq. (15.114) reduces to

$$\frac{1}{\tau_{\text{e-ph}}^{\text{piezo}}} = \sum_{\alpha} \int \frac{d^3q}{(2\pi)^3} \frac{\tilde{w}_{\alpha}(\mathbf{q})}{2\hbar v_F q} \left\{ 2n_{\text{eq}}(\mathbf{q}) + 1 + f_{\text{eq}}[\varepsilon(\mathbf{k}) + \hbar\omega_{\alpha}] - f_{\text{eq}}[\varepsilon(\mathbf{k}) - \hbar\omega_{\alpha}] \right\}. \quad (15.116)$$

(c) Integrate over the directions of $\mathbf{q}$ and show that

$$\frac{1}{\tau_{\text{e-ph}}^{\text{piezo}}} = \sum_{\alpha} \int \frac{dq}{(2\pi)^2} \frac{\langle q \tilde{w}_{\alpha}(\mathbf{q}) \rangle}{\hbar v_F} \left\{ 2n_{\text{eq}}(\mathbf{q}) + 1 + f_{\text{eq}}[\varepsilon(\mathbf{k}) + \hbar\omega_{\alpha}] - f_{\text{eq}}[\varepsilon(\mathbf{k}) - \hbar\omega_{\alpha}] \right\}, \quad (15.117)$$

where

$$\langle q \tilde{w}_{\perp}(\mathbf{q}) \rangle = \frac{16}{35} \frac{\gamma^2 e^2 \pi}{\rho c_{\perp} \varepsilon_{\infty}^2} \quad \text{and} \quad \langle q \tilde{w}_{\parallel}(\mathbf{q}) \rangle = \frac{12}{35} \frac{\gamma^2 e^2 \pi}{\rho c_{\parallel} \varepsilon_{\infty}^2}. \quad (15.118)$$

(d) Assume that $\varepsilon(\mathbf{k}) \simeq \varepsilon_F$ in order to show that

$$\frac{1}{\tau_{\text{e-ph}}^{\text{piezo}}} = \frac{\gamma^2 e^2}{\pi \hbar \rho v_F \varepsilon_\infty^2} \sum_{\alpha=\perp, \parallel} \frac{B_\alpha}{c_\alpha} \int dq \left[\frac{1}{\exp(\beta \hbar c_\alpha q) - 1} + \frac{1}{\exp(\beta \hbar c_\alpha q) + 1} \right], \quad (15.119)$$

with

$$B_\perp = \frac{8}{35}, \quad \text{and} \quad B_\parallel = \frac{6}{35}. \quad (15.120)$$

The above integral diverges logarithmically at $q=0$. In particular, its asymptotic behavior is:

$$\lim_{y_{\min} \rightarrow 0} \int_{y_{\min}}^{\infty} dy \left[\frac{1}{\exp(y) - 1} + \frac{1}{\exp(y) + 1} \right] = \ln \left(\frac{2}{y_{\min}} \right). \quad (15.121)$$

Now, take into account screening effects (with screening wavenumber q_D given by (15.86)) in order to cut off this logarithmic divergence and show that

$$\frac{1}{\tau_{\text{e-ph}}^{\text{piezo}}} = \frac{2\gamma^2 e^2 k_B T}{35\pi \hbar^2 \rho v_F \varepsilon_\infty^2} \left[\frac{4}{c_\perp^2} \ln \left(\frac{2k_B T}{\hbar c_\perp q_D} \right) + \frac{3}{c_\parallel^2} \ln \left(\frac{2k_B T}{\hbar c_\parallel q_D} \right) \right] \quad (15.122)$$

What is the temperature range where this formula applies?

3. Prove that Eq. (15.71) for the entropy of phonons is obtained from Shannon's entropy (15.75).
4. Calculate the approximate energy of the first excited state of a polaron by choosing a variational wave function in the form of a $2p$ state:

$$\psi = \frac{\sqrt{2} b^{5/2}}{\pi^{3/4}} z \exp \left(-\frac{b^2 r^2}{2} \right). \quad (15.123)$$

What is the size of the polaron in this state?

16 Umklapp processes

In the previous chapter, we consider electron-phonon interaction in which the momentum transfer from the lattice, either by acoustic or by optical phonons, is small, $q \rightarrow 0$. This chapter focuses on the opposite limit, i.e., when the phonon momentum is large enough to transfer an electron between different Brillouin zones. As we have seen, see Eq. (15.13), momentum conservation in the scattering process dictates that

$$\mathbf{k} - \mathbf{k}' = \pm \mathbf{q} + \mathbf{b}. \quad (16.1)$$

Here $\mathbf{k}$ and $\mathbf{k}'$ are, respectively, the wavenumbers of the electron before and after scattering, $\mathbf{q}$ is the wavenumber of either absorbed or emitted phonon and $\mathbf{b}$ is an arbitrary lattice vector of the reciprocal lattice. Scattering processes where $\mathbf{b} = 0$ are called “normal”, while those with $\mathbf{b} \neq 0$ are called *umklapp processes* (derived from the German word umklappen – ‘to turn over’). In normal processes, the momentum transferred by the phonon is usually very small, while in umklapp scattering the momentum, $\hbar \mathbf{b}$, is sufficiently large to transfer the electron from one Brillouin zone to another.

In this chapter, we present the physics of umklapp processes using the example of graphene. In particular, we focus our attention on scattering processes that transfer an electron from the K -point to the K' -point (or vice versa),

$$|e, \mathbf{k}_K\rangle + |\text{ph}, \mathbf{q}_K\rangle \longrightarrow |e, \mathbf{k}_{K'}\rangle, \quad (16.2)$$

as illustrated in Fig. 16-1.

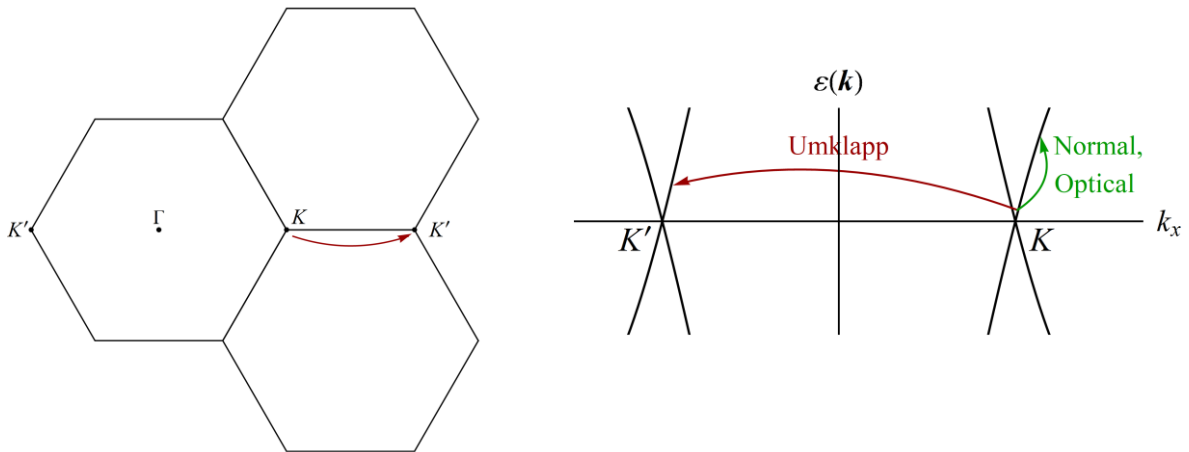


Figure 16-1 Umklapp scattering in graphene

The left panel of the figure shows the Brillouin zones of the system. Here it can be seen that when an electron at K -point changes its momentum by $\hbar \mathbf{k}_K$, it moves to the K' -point in the

adjacent Brillouin zone, namely it appears on the other side of the first Brillouin zone. The right panel shows the normal and the umklapp processes in energy space.

To build a physical picture for the umklapp scattering in graphene, we begin with a qualitative explanation of the process based on the structure of the phonon's vibrational modes. Next, we employ group-theoretical considerations to construct the umklapp Hamiltonian. Finally, we rederive the result in the framework of the $k \cdot p$ approximation for a tight-binding model where bond lengths (and hence hopping matrix elements) are modulated in space and time.

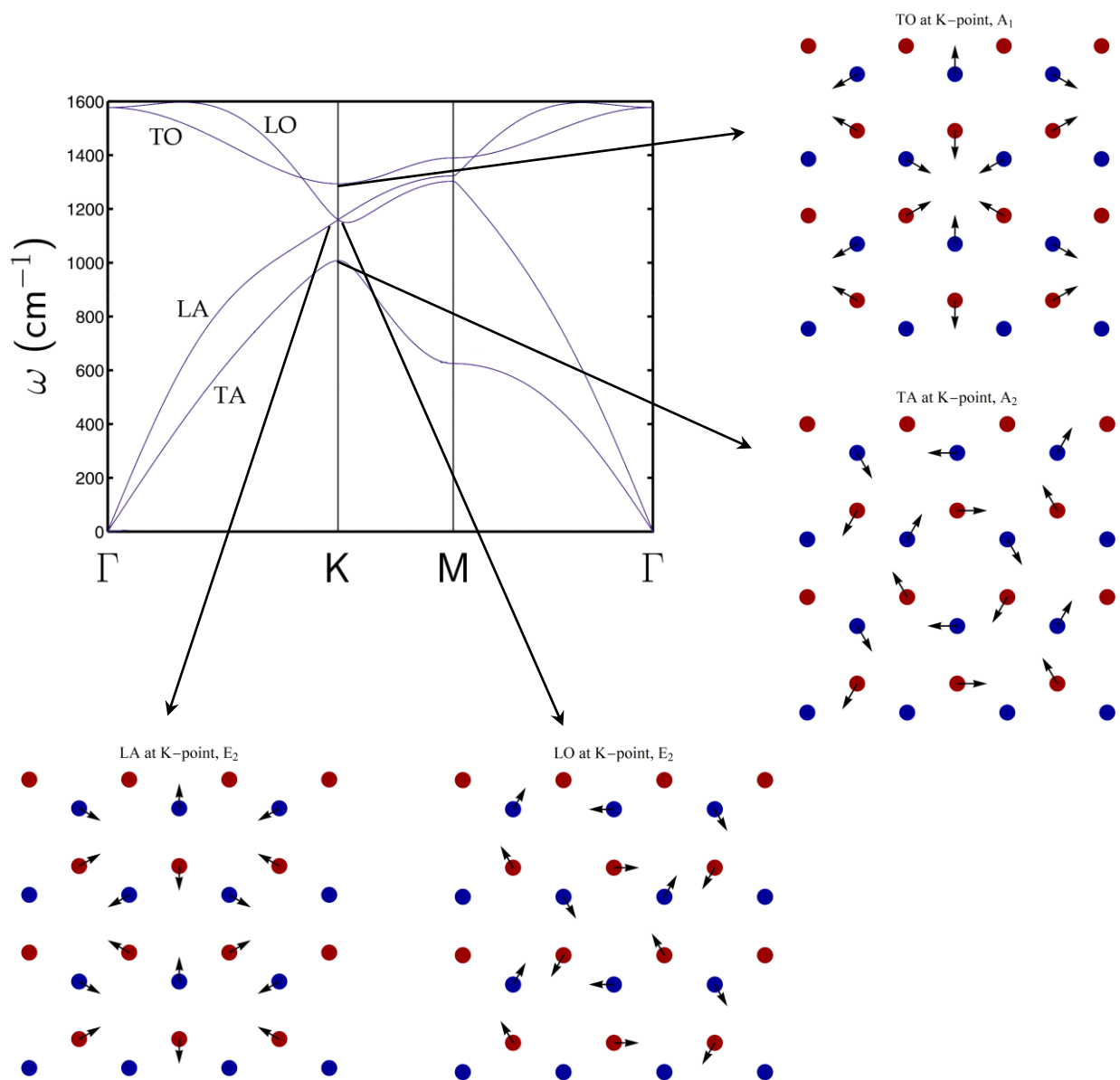


Figure 16-2 The phonon spectrum of graphene and the vibrational modes at the K -point

16.1 Qualitative discussion

In understanding the nature of electron-phonon interaction, one should keep in mind that electrons are light and fast while ions are heavy and slow. For instance, the sound velocity of acoustic phonons is $12.9 \cdot 10^3 \text{ m/s}$ for transverse waves and $19.9 \cdot 10^3 \text{ m/s}$ for longitudinal waves, while the Fermi velocity of the electrons is larger by more than two orders of magnitude, $v_F \simeq 3 \cdot 10^6 \text{ m/s}$. It means that one may consider the lattice deformation due to phonon modes to be, essentially, stationary in time and study the scattering of electrons on a deformed lattice. In order to employ this approach, let us, first, identify the phonon modes that induce the umklapp processes, i.e., phonons with momentum $\pm \hbar \mathbf{k}_K$, that provide the momentum needed for the transfer of an electron between the K and the K' -points.

In Fig. 16-2, we depict the phonon spectrum of graphene (associated only with in-plane vibrations) and the structure of the vibrational modes at the K -point. The symbols used in this figure are 'T' and 'L' for transverse and longitudinal waves, respectively, while 'A' and 'O' for acoustic and optical branches (although this classification is problematic at the K -point). Each normal mode is denoted by the irreducible representation of the C_{6v} group it belongs to. Notice, the phonon spectrum exhibits a degeneracy in the phonon frequency spectrum at the K -point, similar to the electronic spectrum.

As we are interested in scattering process that mixes the K and the K' -points, the suitable group is the extended group,

$$C''_{6v} = C_{6v} \otimes (E \oplus t \oplus t^2), \quad (16.3)$$

where $t = T_{a_1}$ and $t^2 = T_{a_2}$ are the translation operations in the primitive lattice vectors of the system. As we saw in Chapter 6, these translation vectors form the cyclic group C_3 . The vibrational modes at the K -point, being basis functions for the irreducible representation of C''_{6v} , should exhibit this C_3 property of the translation group. Namely, translation by one lattice constant rotates the displacement vectors associated with the vibrational modes by $\pm 120^\circ$ where the sign depends on the sublattice. Indeed, the normal modes that are shown in Fig. 15-4 exhibit this property.

To be concrete, let $\mathbf{u}_A(\mathbf{r}_A)$ and $\mathbf{u}_B(\mathbf{r}_B)$ be the displacement vectors on the points $\mathbf{r}_A$ and $\mathbf{r}_B$ located on sublattice A and B , respectively. Then the phonons normal modes (belonging to A_1 and A_2 irreducible representations) can be written in the form:

$$\mathbf{u}_A(\mathbf{r}_A) = \begin{pmatrix} \text{Re } u_A e^{-i\mathbf{k}_K \cdot \mathbf{r}_A} \\ \text{Im } u_A e^{-i\mathbf{k}_K \cdot \mathbf{r}_A} \end{pmatrix}, \quad \mathbf{u}_B(\mathbf{r}_B) = \begin{pmatrix} \text{Re } u_B e^{i\mathbf{k}_K \cdot \mathbf{r}_B} \\ \text{Im } u_B e^{i\mathbf{k}_K \cdot \mathbf{r}_B} \end{pmatrix}, \quad (16.4)$$

where the upper and lower components of these vectors are the x and the y components of the displacement vector, while u_A and u_B are complex numbers associated with each one of the modes. In particular, choosing $u_A = -u_B = i\xi_{A_1}$, where ξ_{A_1} is real, gives the so-called *Kekulé¹ vibrational mode* associated with the irreducible identity representation A_1 , while $u_A = -u_B = \xi_{A_2}$, with real ξ_{A_2} , describes the vibrational mode that belongs to the A_2 irreducible representation. A similar description applies to the E_2 vibrational modes (see Ex.1). More generally u_A and u_B are slowly changing functions in space (and time), i.e. over distances that are much larger than the lattice constant.

The important property of the vibrational modes at the K -point is that they triple the unit cell's size (i.e., its area). It is most easily observed in the Kekulé vibrational mode. In the left panel of Fig. 16-3, we present the unit cell of an undeformed honeycomb lattice. The unit cell contains two atoms belonging to the different sublattices of the system. On the right panel, we illustrate an instantaneous state of the deformed lattice due to Kekulé's vibrational mode. The different colors highlight the fact that there are three different types of cells in the deformed lattice; hence the new unit cell contains three cells (one from each color), i.e., six lattice points. The shaded parallelogram represents this unit cell. Notice that its side is larger by a factor of $\sqrt{3}$ compared to the original unit cell, and it is also rotated by 30° .

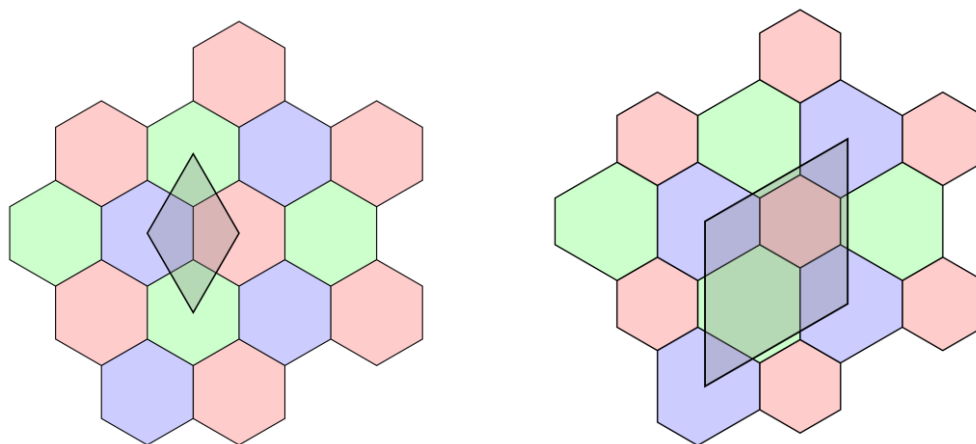


Figure 16-3 The rhombic unit cells of undeformed (left) and deformed (right) lattices of graphene

The increased size of the unit cell of the deformed lattice, by a linear factor of $\sqrt{3}$ compared to that of the original lattice, implies that the same factor reduces the Brillouin zone of the deformed lattice. The rotation of the unit cell induces a similar rotation of the Brillouin zone, as illustrated in Fig. 16-4. As we know, the energy spectrum in the new Brillouin zone is

¹ August Kekulé (1829-1896) was a German theoretical chemist. His famous discovery is the structure of benzene molecule, C_6H_6 , allegedly, after having a day-dream of a snake seizing its own tail.

obtained by folding the spectrum of the original (undeformed) system. This folding shows that both K and K' points move into the Γ point, as indicated by the arrows in Fig. 16-4.

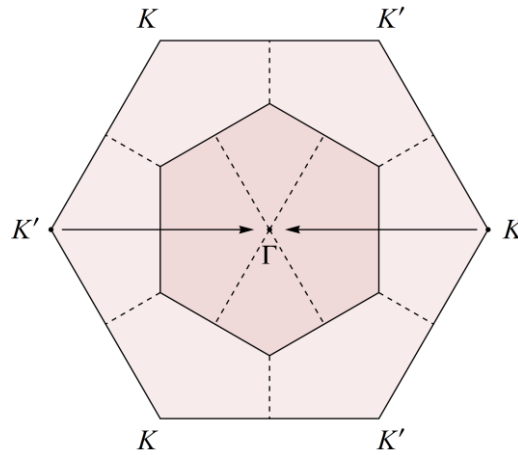


Figure 16-4 The Brillouin zones of the lattices shown in Fig. 16-3

The same result applies to the other phonon modes shown in Fig. 16-2. All these modes bring together the K and the K' points such that for a sufficiently long time, compared to the typical time scale of the electron scattering, they are essentially the same. In this way, the phonon induces the umklapp transition.

16.2 The umklapp Hamiltonian: Group theory approach

The umklapp Hamiltonian, H_{umklapp} , mixes the K and the K' -point; therefore, ignoring spin effects, it acts on an electronic wave function having four components as presented in Eq. (6.17). This form of the wave function implies that the Hamiltonian should be expressed in terms of direct products of Pauli matrices, $\tau_i^{KK'} \otimes \tau_j^{AB}$ where in general, $i, j = x, y, z$. However, the Pauli matrix $\tau_z^{KK'}$ does not describe a transition between the K and the K' -points, thus we are left with the matrices: $\tau_x^{KK'}$, $\tau_y^{KK'}$ and the three Pauli matrices that act on the sublattice space, τ_j^{AB} with $j = x, y, z$. The matrix τ_z^{AB} is associated with electronic transitions between sites on the same sublattice. If we are interested in the largest contribution to H_{umklapp} , it should be also be left aside because the transition between nearest neighbors sites, (i.e. different sublattices) is usually much stronger. Thus H_{umklapp} is constructed from products of the form $\tau_i^{KK'} \otimes \tau_j^{AB}$ with $i, j = x, y$. It should also be linear in the phonon's normal modes and invariant under all symmetry operations of the system.

All elements of the point group of the problem can be generated from c_3, σ_x, σ_d and c_2 . The reflection operation through the horizontal axis, σ_x , replaces A and B sublattices but does not affect the K and the K' points; hence it is given by:

$$\Gamma(\sigma_x) = I^{KK'} \otimes \tau_x^{AB} = \begin{pmatrix} \tau_x^{AB} & 0 \\ 0 & \tau_x^{AB} \end{pmatrix}. \quad (16.5)$$

The rotation by 180° interchanges both A and B sublattices as well as the K and K' points; hence it is described by

$$\Gamma(c_2) = \tau_x^{KK'} \otimes \tau_x^{AB} = \begin{pmatrix} 0 & \tau_x^{AB} \\ \tau_x^{AB} & 0 \end{pmatrix}. \quad (16.6)$$

Reflection σ_d through a line that connects two opposite lattice points does not interchange the A and B sublattices but interchanges the K and K' points; thus it is represented by

$$\Gamma(\sigma_d) = \tau_x^{KK'} \otimes I^{AB} = \begin{pmatrix} 0 & I^{AB} \\ I^{AB} & 0 \end{pmatrix}. \quad (16.7)$$

Finally, as we show in chapter 5 (see Eq. 5.5), the c_3 rotation operation on the wave function associated with K -point is given by $\exp(-i2\pi\tau_z^{AB}/3)$, and one can check that the corresponding operation on the K' -point is obtained by the complex conjugate operator; hence:

$$\Gamma(c_3) = \begin{pmatrix} \exp\left(-i\frac{2\pi}{3}\tau_z^{AB}\right) & 0 \\ 0 & \exp\left(i\frac{2\pi}{3}\tau_z^{AB}\right) \end{pmatrix}. \quad (16.8)$$

Now, let us construct the contribution to the umklapp Hamiltonian. Since the Hamiltonian belongs to the A_1 irreducible representation, we need to identify the product $\tau_i^{KK'} \otimes \tau_j^{AB}$ (with $i, j = x, y$) which is a singlet. One can check that all these matrices are invariant with respect to the c_3 operation, but only $\tau_x^{KK'} \otimes \tau_x^{AB}$ and $\tau_y^{KK'} \otimes \tau_x^{AB}$ are invariant under reflection, σ_x . Out of these only $\tau_x^{KK'} \otimes \tau_x^{AB}$ is invariant to both c_2 and σ_y operations. Thus, the umklapp Hamiltonian is

$$H_{\text{umklapp}} = \lambda \xi_{A_1} \tau_x^{KK'} \otimes \tau_x^{AB}, \quad (16.9)$$

where λ is a constant that depends on the details of the model.

16.3 The umklapp Hamiltonian from the modulated hopping approach

In this section, we rederive Eq. (16.9) from a tight-binding model in which the hopping lengths between nearest neighbors sites change due to spatial modulation of the bond length according to Kekulé's vibrational mode, as illustrated in the right panel of Fig. 16-3. Our strategy is in line with the $k \cdot p$ approximation. Namely, we start from an electronic wave function in the form

$$\Psi = \begin{pmatrix} c_A^K(\mathbf{r})\psi_A^K(\mathbf{r}) \\ c_B^K(\mathbf{r})\psi_B^K(\mathbf{r}) \\ c_A^{K'}(\mathbf{r})\psi_A^{K'}(\mathbf{r}) \\ c_B^{K'}(\mathbf{r})\psi_B^{K'}(\mathbf{r}) \end{pmatrix}, \quad (16.10)$$

where $\psi_{A(B)}^{K(K')}(\mathbf{r})$ is the exact Bloch wave function on sublattice $A(B)$ at the $K(K')$ point, and $c_{A(B)}^{K(K')}(\mathbf{r})$ are functions that change slowly in space, and then construct the effective Hamiltonian acting on the coefficients $c_{A(B)}^{K(K')}(\mathbf{r})$. This construction is obtained by calculating the matrix elements of the full Hamiltonian of the system, H , and averaging over the fast degrees of freedom, treating $c_{A(B)}^{K(K')}(\mathbf{r})$ as constants. This averaging results in the following expression:

$$\langle\langle\Psi|H|\Psi\rangle\rangle_{\text{fast}} = \begin{pmatrix} c_A^K(\mathbf{r}) & c_B^K(\mathbf{r}) & c_A^{K'}(\mathbf{r}) & c_B^{K'}(\mathbf{r}) \end{pmatrix} \begin{pmatrix} H_{AA}^{KK} & H_{AB}^{KK} & H_{AA}^{KK'} & H_{AB}^{KK'} \\ H_{BA}^{KK} & H_{BB}^{KK} & H_{BA}^{KK'} & H_{BB}^{KK'} \\ H_{AA}^{K'K} & H_{AB}^{K'K} & H_{AA}^{K'K'} & H_{AB}^{K'K'} \\ H_{BA}^{K'K} & H_{BB}^{K'K} & H_{BA}^{K'K'} & H_{BB}^{K'K'} \end{pmatrix} \begin{pmatrix} c_A^K(\mathbf{r}) \\ c_B^K(\mathbf{r}) \\ c_A^{K'}(\mathbf{r}) \\ c_B^{K'}(\mathbf{r}) \end{pmatrix}. \quad (16.11)$$

The 4×4 matrix on the right-hand side of this equation is the Hamiltonian in the generalized $k \cdot p$ approximation, which includes umklapp transitions between the K and K' -points. In what follows, we shall calculate its matrix elements, but to keep the calculation simple, we consider only hopping between nearest neighbors sites and set all onsite energies to be zero so that $H_{k \cdot p} = H_0 + H_{\text{umklapp}}$ with

$$H_0 = \begin{pmatrix} 0 & H_{AB}^{KK} & 0 & 0 \\ H_{BA}^{KK} & 0 & 0 & 0 \\ 0 & 0 & 0 & H_{AB}^{K'K'} \\ 0 & 0 & H_{BA}^{K'K'} & 0 \end{pmatrix}, \text{ and } H_{\text{umklapp}} = \begin{pmatrix} 0 & 0 & 0 & H_{AB}^{KK'} \\ 0 & 0 & H_{BA}^{KK'} & 0 \\ 0 & H_{AB}^{K'K} & 0 & 0 \\ H_{BA}^{K'K} & 0 & 0 & 0 \end{pmatrix}. \quad (16.12)$$

To illustrate the averaging procedure, let us start by calculating the term

$$\langle c_A^K(\mathbf{r}) | H_{AB}^{KK} | c_B^K(\mathbf{r}) \rangle = \langle c_A^K(\mathbf{r}) \psi_A^K(\mathbf{r}) | H | c_B^K(\mathbf{r}) \psi_B^K(\mathbf{r}) \rangle_{\text{fast}}. \quad (16.13)$$

For this purpose, we represent the Bloch wave function as a sum over the Wannier functions, $w_{\mathbf{r}_{A(B)}}(\mathbf{r})$, i.e.

$$\psi_{A(B)}^{K(K')}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{r}_{A(B)}} \exp(i\mathbf{k}_{K(K')} \cdot \mathbf{r}_{A(B)}) w_{\mathbf{r}_{A(B)}}(\mathbf{r}), \quad (16.13)$$

where the sum runs over all points $\mathbf{r}_{A(B)}$ of sublattice $A(B)$ and N is the number of unit cells. Substituting the expansion in (16.13) we obtain:

$$\begin{aligned}
& \langle c_A^K(\mathbf{r}) | H_{AB}^{KK} | c_B^K(\mathbf{r}) \rangle \\
&= \frac{1}{N} \sum_{\mathbf{r}_A, \mathbf{r}_B} \langle c_A^{K*}(\mathbf{r}) \exp(-i\mathbf{k}_K \cdot \mathbf{r}_A) w_{r_A}^*(\mathbf{r}) H c_B^K(\mathbf{r}) \exp(i\mathbf{k}_K \cdot \mathbf{r}_B) w_{r_B}(\mathbf{r}) \rangle_{\text{fast}} \\
&= \frac{1}{N} \sum_{\mathbf{r}_A, \mathbf{a}_i} \langle c_A^{K*}(\mathbf{r}) w_{r_A}^*(\mathbf{r}) H w_{r_A+\mathbf{a}_i}(\mathbf{r}) c_B^K(\mathbf{r} + \mathbf{a}_i) \exp(i\mathbf{k}_K \cdot \mathbf{a}_i) \rangle_{\text{fast}} \\
&= \frac{1}{N} \sum_{\mathbf{r}_A, \mathbf{a}_i} \langle w_{r_A}^*(\mathbf{r}) | H | w_{r_A+\mathbf{a}_i}(\mathbf{r}) \rangle \exp(i\mathbf{k}_K \cdot \mathbf{a}_i) c_A^{K*}(\mathbf{r}) c_B^K(\mathbf{r} + \mathbf{a}_i).
\end{aligned} \tag{16.14}$$

The second line of the equation is obtained under the assumption that the Hamiltonian allows only hopping between nearest neighbors sites, namely that $\mathbf{r}_B = \mathbf{r}_A + \mathbf{a}_i$ where $\mathbf{a}_i$ ($i=1,2,3$) are the vectors shown in Fig. 16-4. These are given by

$$\begin{aligned}
\mathbf{a}_1 &= \left(0, \frac{1}{\sqrt{3}}\right) a, & \mathbf{a}_2 &= \left(\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right) a, \\
\mathbf{a}_3 &= \left(-\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right) a,
\end{aligned} \tag{16.15}$$

where a is the lattice constant.

Also observe that the argument of $c_B^K(\mathbf{r})$ is replaced by $c_B^K(\mathbf{r} + \mathbf{a}_i)$ because the reference point of the expansion is $\mathbf{r}_A$ and although $c_B^K(\mathbf{r})$ changes slowly in space, as we shall see below, this small shift yields the leading order contribution. The third line of Eq. (16.14) is obtained by treating $c_{A(B)}^{K*}(\mathbf{r})$ as constants compared to the Wannier functions, which change over the scale of a lattice constant. Defining the hopping matrix element to be

$$\langle w_{r_A}^*(\mathbf{r}) | H | w_{r_A+\mathbf{a}_i}(\mathbf{r}) \rangle = -t, \tag{16.16}$$

we obtain

$$\langle c_A^K(\mathbf{r}) | H_{AB}^{KK} | c_B^K(\mathbf{r}) \rangle = -t \sum_{\mathbf{a}_i} \exp(i\mathbf{k}_K \cdot \mathbf{a}_i) c_A^{K*}(\mathbf{r}) c_B^K(\mathbf{r} + \mathbf{a}_i). \tag{16.17}$$

Now, the assumption that $c_B^K(\mathbf{r} + \mathbf{a}_i)$ changes slowly in space allows one to expand it as

$$c_B^K(\mathbf{r} + \mathbf{a}_i) = c_B^K(\mathbf{r}) + \mathbf{a}_i \cdot \nabla c_B^K(\mathbf{r}) + \dots \tag{16.18}$$

Substituting this expansion in Eq. (16.17), we see that the leading term vanishes because $\sum_{\mathbf{a}_i} \exp(i\mathbf{k}_K \cdot \mathbf{a}_i) = 0$, however, the next to the leading term gives the familiar Dirac Hamiltonian:

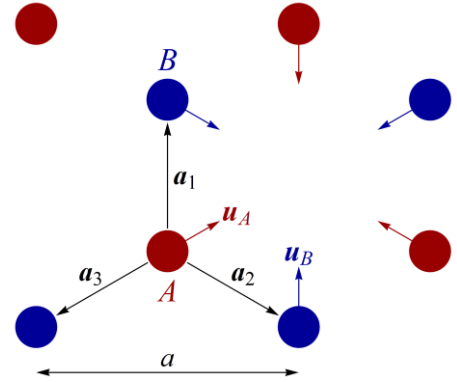


Figure 16-4 Definition of the vectors $\mathbf{a}_i$. The displacement vectors show the modulation due to Kekulé vibrational mode

$$H_{AB}^{KK} = t \frac{\sqrt{3}a}{2\hbar} (-i\hbar) \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right). \quad (16.19)$$

From the above derivation, it is clear that $H_{AB}^{K'K'} = H_{AB}^{KK*}$ and Hermiticity of the Hamiltonian implies

$$H_0 = t \frac{\sqrt{3}a}{2\hbar} \begin{pmatrix} \boldsymbol{\tau}^{AB} \cdot (-i\hbar \nabla) & 0 \\ 0 & -\boldsymbol{\tau}^{AB*} \cdot (-i\hbar \nabla) \end{pmatrix}. \quad (16.20)$$

We turn now to calculate the terms that couple the two valleys. Ignoring the modulation of the hopping matrix elements due to phonons we have:

$$\begin{aligned} \langle c_A^{K'}(\mathbf{r}) | H_{AB}^{KK} | c_B^K(\mathbf{r}) \rangle &= \langle c_A^{K'}(\mathbf{r}) \psi_A^{K'}(\mathbf{r}) | H | c_B^K(\mathbf{r}) \psi_B^K(\mathbf{r}) \rangle_{\text{fast}} \\ &= \frac{1}{N} \sum_{\mathbf{r}_A, \mathbf{r}_B} \langle c_A^{K'*}(\mathbf{r}) \exp(-i\mathbf{k}_{K'} \cdot \mathbf{r}_A) w_{\mathbf{r}_A}^*(\mathbf{r}) H w_{\mathbf{r}_B}(\mathbf{r}) c_B^K(\mathbf{r}) \exp(i\mathbf{k}_K \cdot \mathbf{r}_B) \rangle_{\text{fast}} \\ &= \frac{1}{N} \sum_{\mathbf{r}_A, \mathbf{a}_i} \langle c_A^{K'*}(\mathbf{r}) \exp(-i\mathbf{k}_{K'} \cdot \mathbf{r}_A) w_{\mathbf{r}_A}^*(\mathbf{r}) H w_{\mathbf{r}_A + \mathbf{a}_i}(\mathbf{r}) c_B^K(\mathbf{r}) \exp[i\mathbf{k}_K \cdot (\mathbf{r}_A + \mathbf{a}_i)] \rangle_{\text{fast}} \\ &= -\frac{1}{N} \sum_{\mathbf{r}_A, \mathbf{a}_i} t c_A^{K'*}(\mathbf{r}) c_B^K(\mathbf{r} + \mathbf{a}_i) \exp[i\mathbf{k}_K \cdot (\mathbf{r}_A + \mathbf{a}_i) - i\mathbf{k}_{K'} \cdot \mathbf{r}_A] = 0. \end{aligned} \quad (16.21)$$

Notice that this term vanishes due to the sum over all lattice points, $\mathbf{r}_A$. Since $\mathbf{k}_{K'} = -\mathbf{k}_K$ this sum is a sum over a rapidly oscillating function that averages out to zero when the system is large enough:

$$\frac{1}{N} \sum_{\mathbf{r}_A} \exp(i2\mathbf{k}_K \cdot \mathbf{r}_A) \xrightarrow{N \rightarrow \infty} 0. \quad (16.22)$$

Now let us take into account the modulation in the hopping matrix elements due to a phonon mode that displaces the atoms from their sites, as shown in Fig. 16-2. The distance between two neighboring atoms is

$$l_{AB} = l_0 + \hat{\mathbf{a}}_i \cdot [\mathbf{u}_B(\mathbf{r}_A + \mathbf{a}_i) - \mathbf{u}_A(\mathbf{r}_A)], \quad (16.23)$$

where l_0 is the equilibrium bond length, $\hat{\mathbf{a}}_i$ is a unit vector in the direction that connects the neighboring sites, while $\mathbf{u}_A(\mathbf{r}_A)$ and $\mathbf{u}_B(\mathbf{r}_A + \mathbf{a}_i)$ are the displacement vectors at these sites. Since the hopping matrix element depends on the distance between the atoms, assuming the displacement to be very small, we can expand the hopping matrix element in the form:

$$t(l_{AB}) \simeq t(l_0) + \frac{dt}{dl} \hat{\mathbf{a}}_i \cdot [\mathbf{u}_B(\mathbf{r}_A + \mathbf{a}_i) - \mathbf{u}_A(\mathbf{r}_A)] \quad (16.24)$$

The second term on the right-hand side of this equation, hereinafter denoted by δt , is responsible for the umklapp processes. Using Eqs. (16.4) it can be represented in the form:

$$\delta t = \frac{dt}{dl} \hat{\mathbf{a}}_i \cdot \Delta \mathbf{u} \quad (16.25)$$

where $\Delta \mathbf{u} = (\Delta u_x, \Delta u_y)$ and using (6.4), the components of this vector are given by

$$\begin{aligned} \Delta u_x &= \frac{1}{2} \left[u_B e^{i\mathbf{k}_K \cdot (\mathbf{r}_A + \mathbf{a}_i)} + u_B^* e^{-i\mathbf{k}_K \cdot (\mathbf{r}_A + \mathbf{a}_i)} - u_A e^{-i\mathbf{k}_K \cdot \mathbf{r}_A} - u_A^* e^{i\mathbf{k}_K \cdot \mathbf{r}_A} \right] \\ \Delta u_y &= \frac{1}{2i} \left[u_B e^{i\mathbf{k}_K \cdot (\mathbf{r}_A + \mathbf{a}_i)} - u_B^* e^{-i\mathbf{k}_K \cdot (\mathbf{r}_A + \mathbf{a}_i)} - u_A e^{-i\mathbf{k}_K \cdot \mathbf{r}_A} + u_A^* e^{i\mathbf{k}_K \cdot \mathbf{r}_A} \right] \end{aligned} \quad (16.26)$$

Now let us use the above expansion to calculate the matrix element (16.21) where the hopping matrix element, t , is now replaced by δt , i.e.

$$\langle c_A^{K'}(\mathbf{r}) | H_{AB}^{KK} | c_B^K(\mathbf{r}) \rangle = -\frac{1}{N} \sum_{\mathbf{r}_A, \mathbf{a}_i} \delta t c_A^{K'*}(\mathbf{r}) c_B^K(\mathbf{r} + \mathbf{a}_i) \exp[i\mathbf{k}_K \cdot (\mathbf{r}_A + \mathbf{a}_i) - i\mathbf{k}_{K'} \cdot \mathbf{r}_A]. \quad (16.27)$$

Recall that when we calculated this matrix element without modulation, it vanished due to the sum of the oscillating function, $\exp(i2\mathbf{k}_K \cdot \mathbf{r}_A)$, over the lattice points $\mathbf{r}_A$, see Eq. (6.22). However, now, δt contributes an additional oscillating factor. In particular, the exponents that multiply u_A and u_B^* contain the term $\exp(i\mathbf{k}_K \cdot \mathbf{r}_A)$ that gives a sum of the form $\sum_{\mathbf{r}_A} \exp(i3\mathbf{k}_K \cdot \mathbf{r}_A)$. This sum does not vanish. To why, we substitute $\mathbf{k}_K = (4\pi/3a, 0)$, and notice that the x component of the lattice point positions, $\mathbf{r}_A$ are multiples of half the lattice constant, $a/2$, see Fig. 16-4. Hence $3\mathbf{k}_K \cdot \mathbf{r}_A = 2\pi j$, where j is an integer. Thus, taking the factor $1/N$ in Eq. (16.27), the sum of these terms over $\mathbf{r}_A$ gives one. One can quickly check that the other terms in (16.26) do not have this property and vanish when summing over the lattice points.

Thus substituting Eq. (16.25) with (16.26) in (16.27), we obtain

$$\begin{aligned} H_{AB}^{KK'} &= -\frac{dt}{dl} \sum_{\mathbf{a}_i} \left\{ \hat{a}_{i,x} \frac{1}{2} [u_B \exp(i2\mathbf{k}_K \cdot \mathbf{a}_i) - u_A^* \exp(i\mathbf{k}_K \cdot \mathbf{a}_i)] \right. \\ &\quad \left. + \sum_{\mathbf{a}_i} \hat{a}_{i,y} \frac{1}{2i} [u_B \exp(i2\mathbf{k}_K \cdot \mathbf{a}_i) + u_A^* \exp(i\mathbf{k}_K \cdot \mathbf{a}_i)] \right\} \\ &= \frac{i\sqrt{3}}{2} \frac{dt}{dl} (u_B + u_A^*) \end{aligned} \quad (16.28)$$

For the Kekulé vibrational mode, $u_A = -u_B = i\xi_{A_1}$, with real ξ_{A_1} one obtains:

$$H_{AB}^{KK'} = \sqrt{3} \frac{dt}{dl} \xi_{A_1}, \quad (16.29)$$

while for the A_2 mode $u_A = -u_B = \xi_{A_2}$ with real ξ_{A_2} hence $H_{AB}^{KK'} = 0$. The contribution from the E_2 vibrational modes also vanishes (see Ex. 1); thus A_1 vibrational mode is the only one that contributes to the umklapp Hamiltonian. Employing time-reversal and inversion symmetries, one deduces that $H_{AB}^{KK'} = H_{AB}^{K'K} = H_{BA}^{KK'} = H_{BA}^{K'K}$, hence

$$H_{\text{umklapp}} = \sqrt{3} \frac{dt}{dl} \xi_{A_1} \tau_x^{KK'} \otimes \tau_x^{AB}. \quad (16.30)$$

This formula is the same one obtained in Eq. (16.9) from symmetry considerations.

Finally, we comment that diagonalization of the Hamiltonian $H = H_0 + H_{\text{umklapp}}$, with H_0 and H_{umklapp} given by (16.20) and (16.31) respectively, give:

$$\varepsilon_{\pm}(\delta \mathbf{k}) = \pm \frac{\sqrt{3}}{2} t \sqrt{(\delta k a)^2 + \left(\frac{2}{t} \frac{dt}{dl} \xi_{A_1} \right)^2}. \quad (16.31)$$

Thus, the umklapp process opens a gap in the spectrum of graphene.

16.4 Exercises

1. Show that the E_2 vibrational modes do not contribute to the umklapp Hamiltonian within the framework of the modulated hopping to nearest neighbors' sites.

First, verify that these modes are described by formulas that are similar to (16.4) except for different signs in the exponents:

$$\mathbf{u}_A(\mathbf{r}_A) = \begin{pmatrix} \text{Re} u_A e^{i\mathbf{k}_K \cdot \mathbf{r}_A} \\ \text{Im} u_A e^{i\mathbf{k}_K \cdot \mathbf{r}_A} \end{pmatrix}, \quad \mathbf{u}_B(\mathbf{r}_B) = \begin{pmatrix} \text{Re} u_B e^{-i\mathbf{k}_K \cdot \mathbf{r}_B} \\ \text{Im} u_B e^{-i\mathbf{k}_K \cdot \mathbf{r}_B} \end{pmatrix}. \quad (16.32)$$

Next, repeat the calculation presented in Sec. 16.3 to show that their contribution vanishes.

17 Semiconductors and disordered crystals

Until now, except for a short discussion on dislocations, we assumed our crystals to be perfect. This assumption is far from reality. In realistic systems, there is always some disorder that comes, for example, from domain walls between regions where the crystal structure is in different orientation, dislocation, disclinations, vacancies, interstitials, and impurities. Much of the advances in understanding the role of disorder in crystals came with the technological development of semiconductors that are extensively used for electronic devices. In these crystals, doping by different types of atoms is achieved in a controlled manner. This doping is a source of disorder and, at the same time, the way of introducing charge carriers (electrons or holes) into the system. In this chapter, we present a few models for disordered crystals based on semiconductors. Then we discuss scattering from impurities and the kinetic equations that describe the dynamics of electrons in such systems.

17.1 Impurities and defects in crystals

In order to describe how impurities are combined into a semiconductor, let us first recall the basic physics of chemical bonding. A chemical bond between two identical atoms can be understood as coupling between two potential wells, as illustrated in Fig. 17-1. Each potential well represents the potential seen by an electron due to the atom's nucleus and the other electrons of the atom. When the atoms are far away from each other, the potential wells are decoupled, and the system's ground state is fourfold degenerate (twice due to the two wells and another factor of two due to spin), as illustrated on the left panel of Fig. 17-1. When the two atoms become close to each other, the potential wells are coupled by tunneling, and the two degenerate levels split into upper and lower levels, as illustrated in the right panel of the figure. Occupying the lower energy level by two electrons with opposite spins lowers the system's total energy and creates chemical bonding.

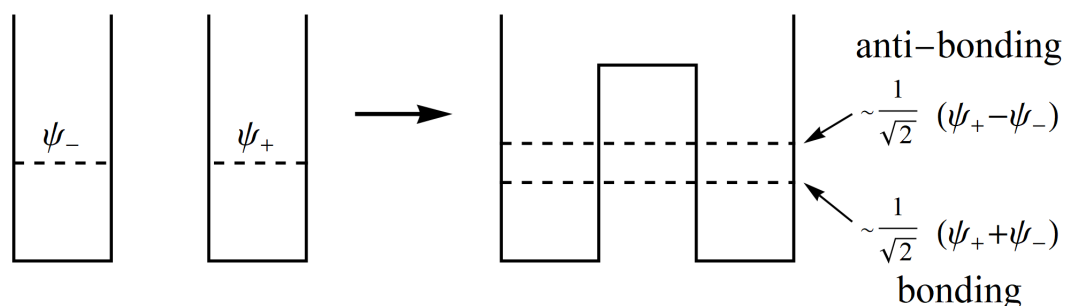


Figure 17-1 An illustration of chemical bonding as coupling between potential wells.

In the limit of weak coupling between the potential wells, the ground state wave function is approximately given by the singlet:

$$\psi_{\text{singlet}} \approx \frac{1}{2}(\psi_+ + \psi_-)[|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle], \quad (17.1)$$

where $\psi_{\pm}$ are the ground state wave functions of the uncoupled wells. The orbital part of this wave function is symmetric so that the spin part secures the antisymmetric property of the fermionic wave function. This wavefunction is associated with chemical bonding.

The excited state of the system is an anti-bonding state described by the triplet :

$$\psi_{\text{triplet}} \approx \frac{1}{\sqrt{2}}(\psi_+ - \psi_-) \begin{cases} |\uparrow\uparrow\rangle \\ \frac{1}{\sqrt{2}}[|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle] \\ |\downarrow\downarrow\rangle \end{cases} \quad (17.2)$$

When the atoms are weakly coupled, their electronic orbitals can approximate the wave functions $\psi_{\pm}$. For instance, for the hydrogen molecule, H_2 , these orbitals are the $1s$ states of each atom. However, there can be more complicated situations like the methane molecule, CH_4 , shown in Fig. 17-2. In this molecule, the carbon creates four symmetric bonding with the hydrogen atoms resulting in a tetrahedral shape. However, this structure is puzzling because the electronic configuration of a carbon atom, $1s^2 2s^2 2p^2$, apparently, allows for only two electronic orbitals to form chemical bonds, say $2p_x$ and $2p_y$, as shown in the left panel of Fig. 17-3. Thus, naively one would expect carbon and hydrogen atoms to form a CH_2 molecule similar to a water molecule. However, it is not what one finds in nature.

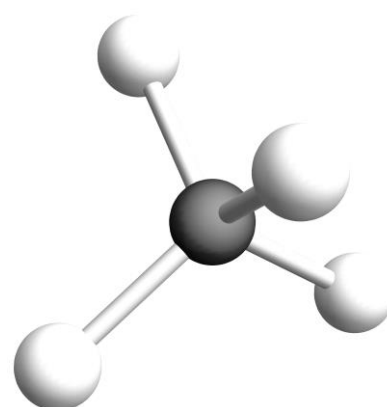


Figure 17-2 Methane molecule

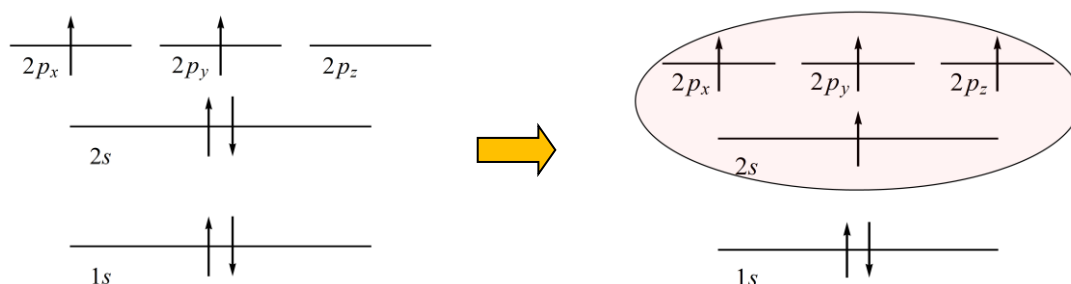


Figure 17-3 Promotion of an electron from $2s$ to $2p$ energy level to form four bonding orbitals

The explanation is as follows: To obtain four orbitals available for chemical bonding, we promote an electron from the $2s$ orbital to $2p$ orbital, as shown in the right panel of Fig. 17-3. This cost some energy, however, it will be compensated, and much more, by sharing

these electrons with the hydrogen orbitals. The molecular orbitals that create the strong covalent bonding with the four hydrogen atoms are presented in Fig. 17-4. These molecular orbitals are classified by the irreducible representation of the tetrahedral group T_d . The upper orbital is associated with the A_1 irreducible representation emerging from the $2s$ orbital, while the lower orbitals are three degenerate molecular orbitals associated with the F_2 irreducible representation that emerge from the $2p$ orbitals.

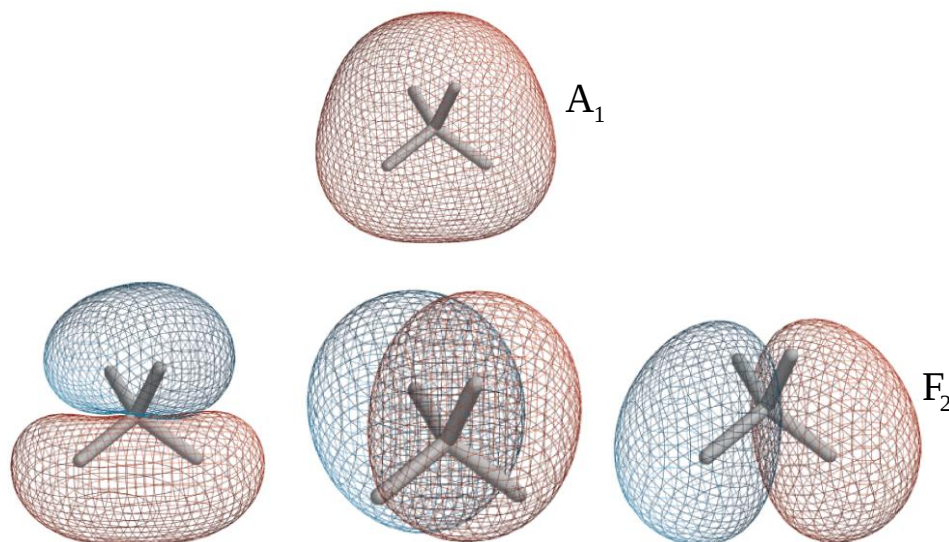


Figure 17-4 The molecular orbitals of Methane. The surfaces represent a constant value of the wave function where blue and red represent opposite signs.

Comment: In the *valence bond theory* approach to covalent bonding (initiated by Linus Pauling), the four orbitals, $2s$, $2p_x$, $2p_y$ and $2p_z$ are assumed to hybridize together to form new equivalent four orbitals denoted by sp^3 . These are and given by

$$\begin{aligned}
 |1\rangle &= \frac{1}{2}(|2s\rangle + |2p_x\rangle + |2p_y\rangle + |2p_z\rangle), \\
 |2\rangle &= \frac{1}{2}(|2s\rangle - |2p_x\rangle - |2p_y\rangle + |2p_z\rangle), \\
 |3\rangle &= \frac{1}{2}(|2s\rangle + |2p_x\rangle - |2p_y\rangle - |2p_z\rangle), \\
 |4\rangle &= \frac{1}{2}(|2s\rangle - |2p_x\rangle + |2p_y\rangle - |2p_z\rangle).
 \end{aligned}
 \tag{17.3}$$

One can check that the wave functions associated with these orbitals are indeed oriented toward the four corners of a regular tetrahedron. However, this approach is inaccurate. For instance, measurements of the absorption of an electromagnetic field in methane molecule show that there are two different absorption lines associated with energy levels of A_1 and F_2 irreducible representations. This finding contradicts the single absorption line predicted by the valence bond approach.

In a silicon crystal, the situation is similar. Silicon has 14 electrons; hence its orbital configuration is $1s^2 2s^2 2p^6 3s^2 3p^2$, namely, there are only two electrons in the $3p$ level. However, promoting one electron from $3s$ to $3p$ enables the formation of four covalent bonds with four neighboring silicon atoms creating the crystal shown in Fig. 17-5.

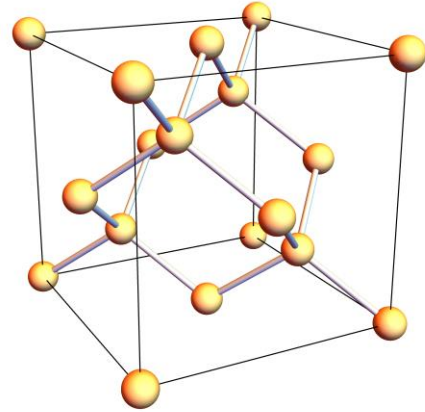


Figure 17-5 A silicon crystal

Consider now a situation in which the silicon crystal is lightly doped with phosphorus atoms, P, that substitute some silicon atoms. Let us also assume that the phosphorous atoms are sufficiently far apart so that we may ignore interactions between them and focus on a single site. The atomic number of a phosphorus atom is 15. Namely, it has an additional electron compared to the silicon atom. Four of its electrons create four covalent bonds with the neighboring silicon atoms, leaving one electron weakly tight to the positive phosphorous ion. Thus, the additional electron behaves as a hydrogen-like atom system. The relative dielectric constant of the silicon crystal is 11.7, and the effective mass of the electron is half that of the bare electron. Hence if we use the familiar formulas for the ground state energy and Bohr radius of hydrogen atom:

$$E_0 = -\frac{m_e e^4}{32\pi^2 \epsilon_0 \hbar^2}, \quad a_B = \frac{4\pi\epsilon_0 \hbar^2}{m_e e^2}, \quad (17.4)$$

and substitute $\epsilon_0 \rightarrow 11.7\epsilon_0$ and $m_e \rightarrow m_e/2$, we obtain that the ground state energy of the electron is $E_0 \rightarrow \epsilon_d = -0.045\text{eV}$ (the subscript d stands for ‘donors’ because the phosphorus atoms donate electrons to the system), while the effective Bohr radius is about 20 times larger than the radius of hydrogen atom (this large radius justifies tearing the crystal as an isotropic material).

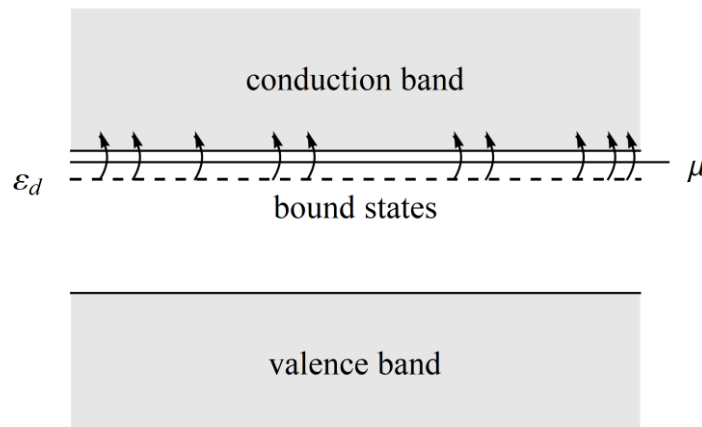


Figure 17-6 Band diagram of n-type semiconductor

The ground state energy of the electron is approximately (minus) the ionization energy. Since ionization means that the electron moves into the conduction band, the energy level of the additional electron of the phosphorous is 0.045eV below the conduction band, as illustrated in Fig. 17-6. In temperature units, bonding energy of 0.045eV corresponds to $450^{\circ}K$, therefore, at room temperature, many of the phosphorus electrons move into the conduction band. A semiconductor in which the charge carriers are electrons is called an *n-type semiconductor*.

The density of charge carriers in the system depends on the doping level and the temperature. In particular, it depends on the probability that a given phosphorus atom is ionized. To calculate this probability, let N_D be the density of phosphorous atoms (assumed to be uniform in space), and denote by N_D^+ and N_D^0 the densities of the ionized and the neutral phosphorous atoms, respectively. In equilibrium, these are given by:

$$N_D^+ = bN_D [1 - f(\varepsilon_d)] \quad \text{and} \quad N_D^0 = bgN_D f(\varepsilon_d) \quad (17.5)$$

where $f(\varepsilon_d)$ is the Fermi-Dirac distribution function at the binding energy ε_d , and g is the degeneracy of the occupied state. Here $g = 2$ due to the spin degeneracy of the occupied state. The unoccupied state, on the other hand, is non-degenerate. The normalization constant b is determined by the condition $N_D = N_D^0 + N_D^+$. This normalization is required because the number of states does not equal the number of impurities. From these considerations, one concludes that the probability that a randomly chosen phosphorus atom is neutral is given by:

$$P_D = \frac{N_D^0}{N_D} = \frac{N_D^0}{N_D^0 + N_D^+} = \frac{1}{1 + \frac{1}{2} \exp\left(\frac{\varepsilon_d - \mu}{k_B T}\right)} \quad (17.6)$$

Notice, this is not Fermi-Dirac distribution because the spin degeneracy of the occupied state increases its probability.

Now consider a situation where the silicon crystal is lightly doped with aluminum (atomic number 13), i.e., when some silicon atoms are substituted by aluminum. Now there are only 3 electrons available for covalent bonds with the neighboring silicon atoms; namely, there is a deficit of one electron, or in other words, a hole. In a binding state of the hole, one of the four covalent bonds of the aluminum is missing; therefore, this state is fourfold degenerate (here, we assume that temperature is sufficiently high such that one can neglect the energy difference between the A_1 and the F_2 binding orbitals). In an ionized state, an electron from one of the other covalent bonds takes this place, i.e., the hole moves to some other location in the crystal. Such semiconductor is called *p-type*, and its band diagram is presented in Fig. 17-7, where ε_a is the binding energy of a hole (the subscript a stands for acceptor).

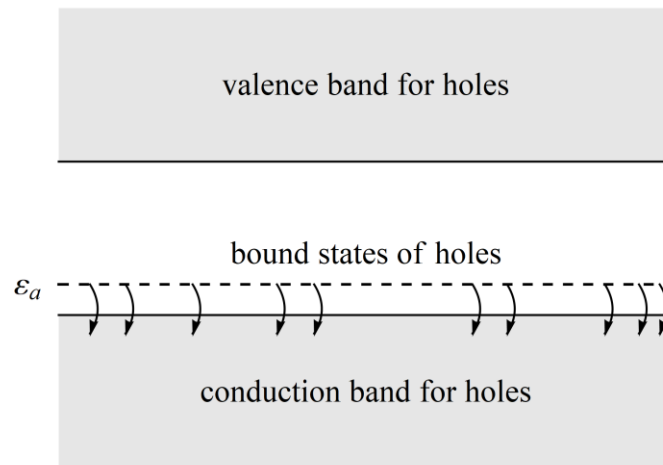


Figure 17-7 Band diagram of p-type semiconductor

Repeating the same arguments presented above, one finds that the probability of finding an aluminum atom in a neutral state is:

$$P_A = \frac{1}{1 + \frac{1}{4} \exp\left(\frac{\mu - \epsilon_a}{k_B T}\right)}, \quad (17.7)$$

where $\mu < \epsilon_a$ is the chemical potential set to ensure that the number of ionized acceptors plus the number of neutral ones equals the total number of acceptors.

Finally, compensated semiconductors are obtained when doping is with both donors and acceptors. In this case, some of the electrons occupy some of the hole states, as illustrated in Fig. 17-8. This recombination generates local electric dipoles oriented in random directions. These dipoles create a random-like potential acting on the uncompensated charge carriers (obtained, e.g., when the number of phosphorous atoms is larger than that of the aluminum atoms). In this way, one can control the amount of disorder in the system.

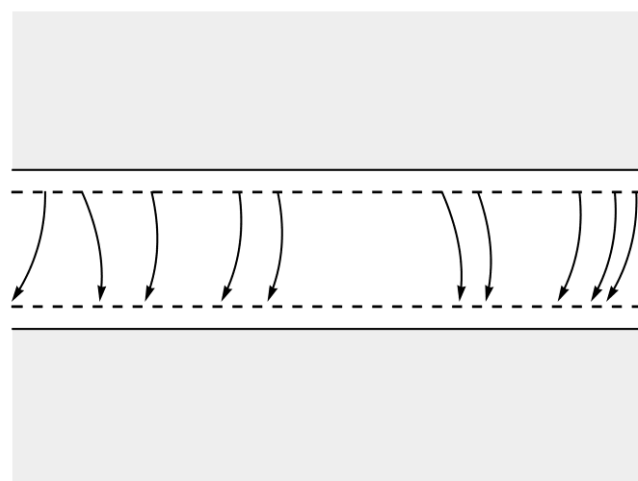


Figure 17-8 Compensated semiconductors

Additional common sources of disorder in crystals are associated with vacancies and interstitials. In many cases, there are correlations between two such defects. For instance, the *Schottky defect* in ionic crystals, such as sodium chloride salt, is realized when vacancies of the negative and the positive ions are located close to each other to reduce the Coulomb energy (see left panel of Fig. 17-9). Another example is the *Frenkel defect* (also in ionic crystals) realized when an ion moves from its lattice position into a nearby point not on the lattice, as illustrated on the right panel of Fig. 17-9. Here, Coulomb attraction also results in a tendency of the vacancy to be near the interstitial.

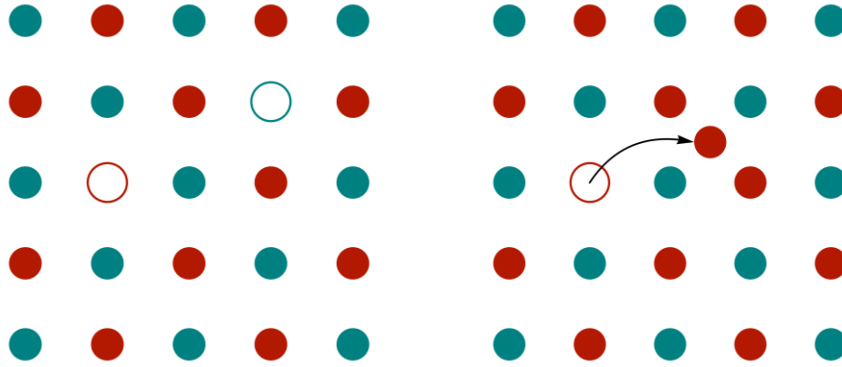


Figure 17-9 An illustration of Schottky (left) and Frenkel (right) defects in a lattice

Example: The law of mass action in semiconductors

According to the *law of mass action*, the geometric mean of the hole and the electron concentrations in a semiconductor at equilibrium is a constant independent of the doping level. To derive this property, let ε_C be the lowest energy of the conduction band in the semiconductor, and ν_C the density of state in the conduction band assumed to be constant for simplicity. Then the density of electrons is given by:

$$n_e = \int_{\varepsilon_C}^{\infty} \nu_C d\varepsilon \frac{1}{1 + \exp\left(\frac{\varepsilon - \mu}{k_B T}\right)} \simeq \nu_C k_B T \exp\left(-\frac{\varepsilon_C - \mu}{k_B T}\right), \quad (17.8)$$

where μ is the chemical potential (that depends on the doping level). To obtain this result we assumed that $\varepsilon_C - \mu > k_B T$, hence Boltzmann's distribution could approximate the Fermi-Dirac distribution. We also assume the width of the conduction band to be sufficiently large so that one may replace the upper limit of the integral by infinity.

A similar calculation gives the density of holes:

$$n_h = \int_{-\infty}^{\varepsilon_V} \nu_V d\varepsilon \frac{1}{1 + \exp\left(\frac{\mu - \varepsilon}{k_B T}\right)} \simeq \nu_V k_B T \exp\left(-\frac{\mu - \varepsilon_V}{k_B T}\right) \quad (17.9)$$

where ν_V is the density of state in the valance band, while ε_V is the highest energy of the valance band. From the above equations, we obtain:

$$n_h n_e \simeq n_i^2 \quad (17.10)$$

with

$$n_i^2 \simeq \nu_C \nu_V (k_B T)^2 \exp\left(-\frac{\varepsilon_{\text{gap}}}{k_B T}\right). \quad (17.11)$$

Here $\varepsilon_{\text{gap}} = \varepsilon_C - \varepsilon_V$ is the gap energy between the conduction and the valance band of the intrinsic semiconductors. This result implies that n_i is independent of the chemical potential and consequently independent of the type of doping and its concentration. Hence, n_i is called the intrinsic density of the charge carriers in the semiconductor.

Example: Debye's screening length in semiconductors

As explained above, in a semiconductor, the density of the charge carriers depends on the temperature. Therefore, the screening length in lightly doped semiconductors is expected to be temperature-dependent. Namely, it is not described by the Thomas-Fermi theory (presented in section 14.2). This example is devoted to the calculation of the screening length in semiconductors. To be concrete, we consider the case of a semiconductor doped only by donors, but the result is more general and applies to semiconductors doped by acceptors.

Let N_D denote the density of donors in the semiconductor, which is assumed to be homogenous throughout the system. The effective density of charge carriers (see Eq. (17.5)) is given by

$$n_* = N_D [1 - f(\varepsilon_d)], \quad (17.12)$$

where $f(\varepsilon)$ is the Fermi-Dirac distribution. An external charge density, $\rho_{\text{ext}}(\mathbf{r})$, introduced to the system creates an electric potential $\varphi(\mathbf{r})$ seen by the electrons. Here we assume this potential to change slowly over a distance of the order of the typical electron wavelength to employ a semiclassical approach for its calculation. Namely, we use Gauss law,

$$-\varepsilon_0 \varepsilon_r \nabla^2 \varphi(\mathbf{r}) = \rho_+ + \rho_- + \rho_{\text{ext}}(\mathbf{r}), \quad (17.13)$$

where ε_r is the relative dielectric constant of the crystal, $\rho_+ = en_*$ is the positive charge density due to the ions in the system, and

$$\rho_- = -eN_D \left\{ 1 - f \left[\varepsilon_d + e\varphi(\mathbf{r}) \right] \right\} \quad (17.14)$$

is the charge density of the electrons in the presence of the external charge, which shifts the ground state energy of the donors. Expanding this function to linear order in φ , assuming $|e\varphi| \ll k_B T$, and using the formula

$$\frac{df(\varepsilon)}{d\varepsilon} = \frac{1}{k_B T} f(\varepsilon) [1 - f(\varepsilon)] \quad (17.15)$$

give:

$$\rho_- = -n_* e - b \frac{n_* e}{k_B T} \varphi(\mathbf{r}), \quad (17.16)$$

where $b = f(\varepsilon_d)$. In the high-temperature regime, $k_B T \gg \mu - \varepsilon_d$, b is approximately one-half, while in the opposite limit, $k_B T \ll \mu - \varepsilon_d$ it is approximately one. Therefore, this constant does not play an important role, and from now on, we assume $b \simeq 1$. These considerations imply that Eq. (17.13) can be approximated by

$$-\nabla^2 \varphi(\mathbf{r}) = -\frac{n_* e^2}{\varepsilon_0 \varepsilon_r k_B T} \varphi(\mathbf{r}) + \frac{\rho_{\text{ext}}(\mathbf{r})}{\varepsilon_0 \varepsilon_r}. \quad (17.17)$$

This equation has the same structure as the equation obtained from the Thomas-Fermi approximation of metals, see Eq. (14.14). Hence, by similar considerations, one can identify the screening wavenumber as

$$q_D = \sqrt{\frac{n_* e^2}{\varepsilon_0 \varepsilon_r k_B T}}. \quad (17.18)$$

This quantity is the inverse of *Debye's screening length*.

17.2 The Jahn-Teller effect of an impurity in a crystal

Occasionally, crystal defects are accompanied by local lattice deformations – i.e., local phonons. These appear when the impurity's electrons occupy degenerate energy levels. The bound state of an electron on such impurity is characterized by one of the irreducible representations of the lattice. However, a local deformation may change the symmetry and open a gap in the degenerate electronic levels associated with the impurity. In this event, the electron will occupy the lower energy and reduce the system's total energy. This effect is called the *Jahn-Teller effect* (1937).

Let ξ_i denote the coordinates that represent a local deformation of the lattice. The local Hamiltonian that describes the deformation and its coupling to the electronic spectrum is:

$$\delta H = \sum_i \eta_i \xi_i + \frac{1}{2} K_i \xi_i^2, \quad (17.19)$$

where $K_i \xi_i$ represents the restoring force due to the deformation while η_i characterizes the coupling between the deformation and the electronic energy levels. As the change in the electron energy levels is linear in $|\xi|$, while the deformation is quadratic in ξ , there will be a spontaneous symmetry breaking with a new minimum at non zero value of ξ , as illustrated in Fig. 17-10.

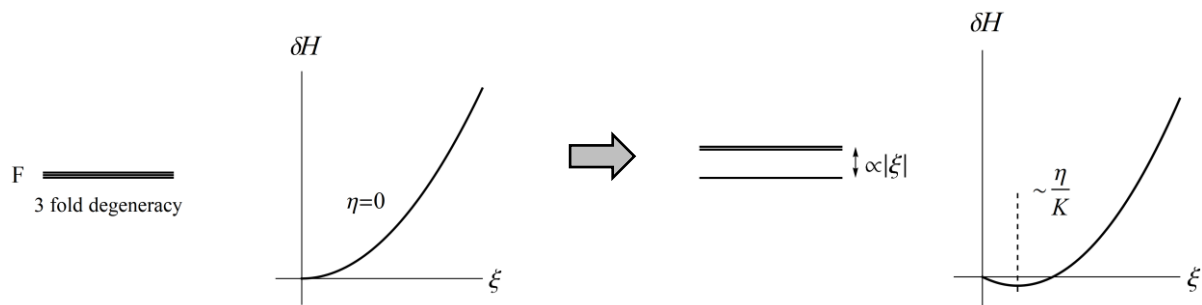


Figure 17-10 An illustration of the Jahn-Teller effect

The above picture rests on the assumption that $\eta_i \neq 0$. To clarify the conditions for this assumption, let us expand the Hamiltonian of the impurity in the normal coordinates of the local deformation:

$$H(\mathbf{r}, \xi) \approx H(\mathbf{r}, 0) + \left. \frac{\partial H(\mathbf{r}, \xi)}{\partial \xi_k} \right|_{\xi=0} \xi_k, \quad (17.20)$$

where $\mathbf{r}$ is the electron coordinate on the impurity, and $H(\mathbf{r}, 0)$ is its Hamiltonian in the absence of deformation. We shall assume that the electron energy levels are degenerate and described by the wave function $\psi_j^{(\alpha)}$, which belongs to an irreducible representation α whose dimension is ℓ_α .

The corrections to the energy of the electron due to the local deformation is given by degenerate perturbation theory; namely, one should diagonalize the matrix $\xi_k V_{k;ij}$, where

$$V_{k;ij} = \left\langle \psi_i^{(\alpha)} \left| \frac{\partial H(\mathbf{r}, \xi)}{\partial \xi_k} \right|_{\xi=0} \right| \psi_j^{(\alpha)} \rangle. \quad (17.21)$$

If one of the elements of this tensor is nonzero, then the correction to the energy levels will be linear in at least one of the components of ξ . Then assuming a single electron occupies

these energy levels, the system becomes unstable to local deformation since ξ can be chosen in such a way that the electron energy reduces.

Consider, for example, an impurity in a lattice with tetrahedral symmetry and that the electronic state belongs to either to F_1 or to F_2 irreducible representations. The representation obtained from a product of two wave functions, $\psi_i^{(F)*} \psi_j^{(F)}$, belonging to this irreducible representation (which we denote by F) is $F \otimes F = A_1 + E + F_1 + F_2$. On the other hand, the operator $\partial H(\mathbf{r}, \xi)/\partial \xi$ has the same symmetry as ξ , hence $\xi \partial H(\mathbf{r}, \xi)/\partial \xi$ generically contains at least one component that belongs to one of the irreducible representations in the product $F \otimes F$ (the only irreducible representation of T_d that does not appear in this product is A_2). From here it follows that $\xi_k V_{k;ij}$ has a singlet. Thus, the coupling constant, η_i , is generically nonzero.

Example: Cr^{2+} impurities in a CdSe crystal

Cadmium selenide crystal is a wurtzite hexagonal crystal where each cadmium atom is surrounded by four selenium atoms (and vice versa), forming a regular tetrahedral shape. When doped with chromium, ions of Cr^{2+} replace some of the cadmium atoms, as demonstrated in Fig. 17-11.

Let us first identify the atomic term of Cr^{2+} . Chromium contains 24 electrons; therefore, the electronic configuration of Cr^{2+} is $1s^2 2s^2 2p^6 3s^2 3p^6 3d^4$. According to Hund's rules the electrons occupy 4 of the d-orbitals such that all spins are parallel and the total orbital angular momentum is maximal. For instance, the electrons may occupy the states with $m_l = 2, 1, 0, -1$.

Thus, the total orbital angular momentum is $L = 2$, and the total spin is also $S = 2$. As we are dealing with less than a half-filled state, the total angular momentum is $J = L - S = 0$. Thus, the term of the ground state of free Cr^{2+} is 1D_4 , and the orbital momentum is fivefold degenerate.

However, when the chromium ion is placed in a crystal, the crystal field breaks the full rotational symmetry of the system and reduces the degeneracy of the ground state. In exercises 6 and 7 of chapter 4 it was shown that in a crystal with tetrahedral symmetry, T_d , the $L = 2$ states split into $E \oplus F_2$ irreducible representations. It turns out that the ground state is associated with the triplet F_2 , while the excited state is associated with the doublet E .

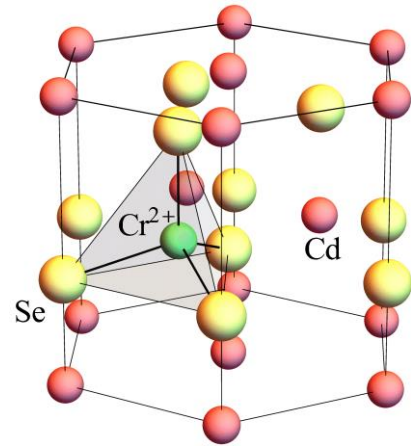


Figure 17-11 Cadmium selenide doped by chromium

Now let us identify the possible deformations of the tetrahedron surrounding a Cr^{2+} ion. As we are not interested in rotations or translations of this tetrahedron (which do not change the crystal field around the ion), all possible deformations are described by the normal modes of a single tetrahedron. Since this system has 15 degrees of freedom, it has nine vibrational modes. In exercise 4 of Chapter 4, we saw that the composition of the normal modes of a methane molecule (that has the same structure) is $A_1 \oplus E \oplus F_2 \oplus F_2$. The corresponding normal modes are shown in Fig. 17-12 below.

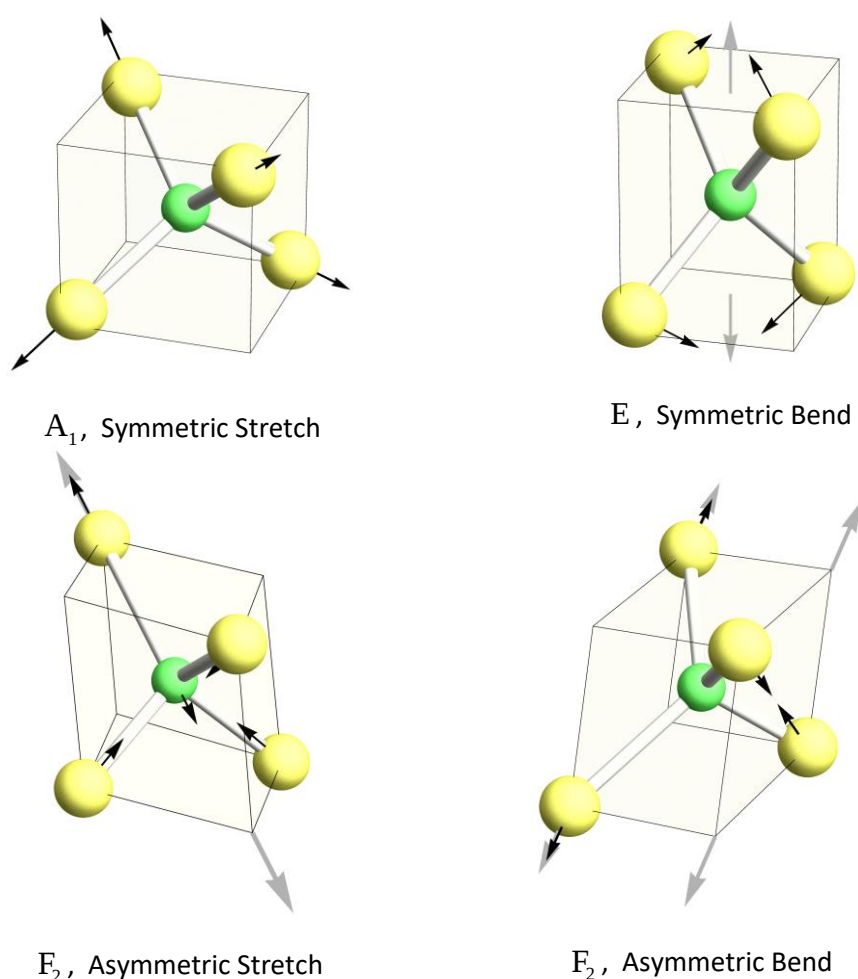


Figure 17-12 Local deformation of the octahedron surrounding the chromium impurity

From this figure, it is clear that the symmetric stretch is a basis function of the identity irreducible representation; therefore, it shifts all three levels of the F_2 electronic levels in the same amount. The symmetric bend, on the other hand, has the symmetry of D_{2d} group whose largest irreducible representation is two-dimensional and therefore opens a gap in the electronic spectrum, as illustrated in Fig. 17-10. The asymmetric stretch has C_{3v} symmetry

implying that it also opens a gap in the spectrum. Finally, the asymmetric bend mode has the lowest symmetry: C_{2v} . This group has only one-dimensional irreducible representations; therefore, such deformation will split the F_2 degenerate electronic levels into three different levels. These considerations show that some of the coefficients η_i in Eq. (17.19) are nonzero.

Consider now the quadratic term in Eq. (17.19), which accounts for the elastic energy of the local deformation. Each one of the normal modes shown in Fig. 17-12 can be associated with a deformation of a cube defined by the tetrahedron, as shown in the figure. This local deformation serves as a boundary condition for the local phonon mode obtained by solving the elasticity problem. In what follows, we will illustrate the calculation of this mode and its energy for the symmetric stretch. For simplicity, we assume that the elastic modulus tensor can be approximated by its spherically symmetric component (see Eq. (11.48)), i.e.

$$\Xi_{ij;kl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}), \quad (17.22)$$

where λ and μ are the *Lamé parameters*. The corresponding elastic energy is given by

$$E_{\text{elastic}} = \frac{1}{2} \int d^3r \left[\lambda (u_{ii})^2 + 2\mu u_{ij} u_{ji} \right], \quad (17.23)$$

where u_{ij} is the strain tensor. We shall also assume that the boundary conditions set by local deformation of the symmetric stretch mode can be approximated by setting the value of the displacement vector on a small sphere of radius a to be radial and constant:

$$\mathbf{u}|_{r=a} = \xi \hat{\mathbf{r}}. \quad (17.24)$$

Here ξ measures for the amount of deformation while a is of the order of the size of the cube shown in the upper left panel of Fig. 17-12.

The equilibrium condition of an elastic system is given by the equation

$$\frac{\sigma_{kl}}{\partial r_k} = 0 \quad (17.25)$$

where

$$\sigma_{kl} = \Xi_{ij;kl} u_{ij} = \lambda u_{ii} \delta_{kl} + 2\mu u_{kl} \quad (17.26)$$

is the strain tensor, see Eqs. (11.37) and (11.38). Expressing the strain tensor as derivative of the displacement vector (see Eq. (11.26)), we obtain that

$$\frac{\sigma_{kl}}{\partial r_k} = \lambda \frac{\partial}{\partial r_l} \nabla \cdot \mathbf{u} + \mu \left(\frac{\partial}{\partial r_l} \nabla \cdot \mathbf{u} + \nabla^2 u_l \right). \quad (17.27)$$

Hence the equilibrium condition (17.25) is:

$$(\lambda + \mu) \nabla (\nabla \cdot \mathbf{u}) + \mu \nabla^2 \mathbf{u} = 0. \quad (17.28)$$

Now, using the identity $\nabla^2 \mathbf{u} = \nabla(\nabla \cdot \mathbf{u}) - \nabla \times (\nabla \times \mathbf{u})$ and taking the divergence of the resulting equation, one concludes that

$$\nabla^2 (\nabla \cdot \mathbf{u}) = 0. \quad (17.29)$$

Thus, the divergence of the displacement vector satisfies the Laplace equation. Finally, operating with the Laplacian on Eq. (17.28) and using (17.29) we obtain that the displacement vector satisfies the biharmonic equation:

$$\nabla^2 \nabla^2 \mathbf{u} = 0. \quad (17.30)$$

The spherical symmetry of the problem implies that we should seek for a solution of the form $\mathbf{u} = f(r) \hat{\mathbf{r}}$. From here we obtain the equation

$$\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} \left(\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial f}{\partial r} \right) = \frac{4}{r} \frac{d^3 f}{dr^3} + \frac{d^4 f}{dr^4} = 0 \quad (17.31)$$

whose general solution is

$$f(r) = \frac{b_1}{r} + b_2 + b_3 r + b_4 r^2, \quad (17.32)$$

where b_i are constants. However, being interested in a local deformation that decays far from the origin, the only relevant term in this solution is the first one. Imposing the boundary conditions (17.24), we obtain:

$$\mathbf{u}(r) = \frac{a\xi}{r} \hat{\mathbf{r}}, \quad \text{for } r \geq a. \quad (17.33)$$

Thus, the displacement vector decays slowly as $1/r$. The corresponding strain tensor contains only one component:

$$u_{rr} = -\frac{a\xi}{r^2}, \quad \text{for } r > a. \quad (17.34)$$

Substituting it in Eq. (17.23) we obtain the energy of the local deformation:

$$E_{\text{elastic}} = \left(\frac{\lambda}{2} + \mu \right) a^2 \xi^2 \int_{r>a} \frac{d^3 r}{r^4} = \frac{1}{2} 4\pi (\lambda + 2\mu) \xi^2 a. \quad (17.35)$$

Comparing this energy with the second term in Eq. (17.19), one can identify the constant K_i as $4\pi(\lambda + 2\mu)a$.

Notice that although the displacement vector decays slowly, the energy density behaves as $1/r^4$, namely, it is concentrated within a region near the impurity. Deformations that are volume-preserving behave as quadrupoles rather than monopoles and result in a much faster decay of the local phonon mode, $\mathbf{u} \sim 1/r^3$, hence the energy density decays as $1/r^8$.

17.3 Elastic scattering and the optical theorem

To describe the basic ideas of elastic scattering (i.e., scattering which preserves the particle's energy), we begin with the scattering of a free electron from a single impurity. The impurity is represented by a scattering potential localized within a small domain in space, as illustrated in Fig. 17-13.

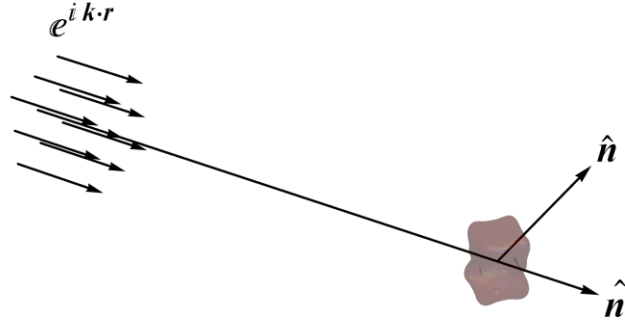


Figure 17-13 Scattering of a plane wave from a finite range potential

Far from the impurity, the electron satisfies the free-particle Schrödinger equation:

$$-\frac{\hbar^2 \nabla^2}{2m} \psi = \varepsilon \psi, \quad (17.36)$$

where m is the particle's (effective) mass and ε is its energy. Thus, far from the scatterer, the wave function is of the form:

$$\psi(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) + \frac{a(\hat{\mathbf{n}}', \hat{\mathbf{n}})}{r} \exp(ikr). \quad (17.37)$$

Here we assume that the scatterer is located at the origin, $\mathbf{k}$ is the wavenumber of the incoming wave, and $a(\hat{\mathbf{n}}', \hat{\mathbf{n}})$ is the scattering amplitude in which

$$\hat{\mathbf{n}} = \frac{\mathbf{k}}{|\mathbf{k}|}, \quad \text{and} \quad \hat{\mathbf{n}}' = \frac{\mathbf{r}}{|\mathbf{r}|} \quad (17.38)$$

are unit vectors in the directions of the incoming and outgoing waves, respectively. The scattering amplitude is obtained from the solution of the Schrödinger equation in the presence of the scattering potential. If this potential is spherically symmetric, $a(\hat{\mathbf{n}}', \hat{\mathbf{n}})$ depends only on the angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}'$.

The generalization of Eq. (17.33) to a particle in a lattice, in the framework of the $k \cdot p$ approximation, is obtained by taking the envelope of Bloch's wave function to be

$$\psi(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) + \sqrt{\frac{v(\hat{\mathbf{n}})}{v(\hat{\mathbf{n}}')}} \frac{a(\hat{\mathbf{n}}', \hat{\mathbf{n}})}{r} \exp(ikr), \quad (17.39)$$

where $v(\hat{\mathbf{n}})$ is the velocity of a particle moving in direction $\hat{\mathbf{n}}$. This formula is valid when assuming the dynamics to take place within a single band. A generalization for several bands is obtained when replacing the scattering amplitude $a(\hat{\mathbf{n}}', \hat{\mathbf{n}})$ with a matrix in the band space. In what follows, we shall consider only the simple case described by Eq. (17.37).

The current density obtained from the wave function (17.37) is

$$\begin{aligned} \mathbf{j} &= \frac{\hbar}{2mi} [\psi^* \nabla \psi - \psi \nabla \psi^*] \\ &= \left[\hat{\mathbf{n}} + \frac{\hat{\mathbf{n}}'}{r^2} |a(\hat{\mathbf{n}}', \hat{\mathbf{n}})|^2 \right] v + \frac{(\hat{\mathbf{n}} + \hat{\mathbf{n}}')v}{2r} [\exp(i\mathbf{k} \cdot \mathbf{r} - ikr) a^*(\hat{\mathbf{n}}', \hat{\mathbf{n}}) + c.c.], \end{aligned} \quad (17.40)$$

where $v = \hbar k/m$. To obtain this result, we kept only leading order terms in $1/kr$, as we are interested in the limit where r is much larger than the particle wavelength. In particular, we have neglected the gradient of the preexponential factor in Eq. (17.33) but kept gradients of the exponential term.

Let us now calculate the total current passing through a distant spherical shell surrounding the scatterer:

$$\begin{aligned} \oint r^2 d\hat{\mathbf{n}} \cdot \mathbf{j} &= v \oint d^2 \hat{\mathbf{n}}' |a(\hat{\mathbf{n}}', \hat{\mathbf{n}})|^2 \\ &+ \frac{vr}{2} \oint d\hat{\mathbf{n}}' \cdot (\hat{\mathbf{n}} + \hat{\mathbf{n}}') [\exp(i\mathbf{k} \cdot \mathbf{r} - ikr) a^*(\hat{\mathbf{n}}', \hat{\mathbf{n}}) + c.c.], \end{aligned} \quad (17.41)$$

where $d^2 \hat{\mathbf{n}} = d\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}$ is an infinitesimal element of the solid angle. The first contribution to the integral,

$$J_1 = v \oint d^2 \hat{\mathbf{n}}' |a(\hat{\mathbf{n}}', \hat{\mathbf{n}})|^2, \quad (17.42)$$

is manifestly independent of the position of the spherical shell. However, apparently, the second contribution to the total current (given by the second line in Eq. (17.41)) contains r dependence. On the other hand, since the total current passing through the spherical shell must vanish, there should be no such dependence when $r \rightarrow \infty$.

To calculate the second contribution to the total current, it is convenient to express the argument of the exponential function in (17.41) in the form $\mathbf{k} \cdot \mathbf{r} - kr = (\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' - 1)kr$ so that

$$J_2 = \frac{vr}{2} \oint d\hat{\mathbf{n}}' \cdot (\hat{\mathbf{n}} + \hat{\mathbf{n}}') \left\{ \exp[-ikr(1 - \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}')] a^*(\hat{\mathbf{n}}', \hat{\mathbf{n}}) + c.c. \right\}. \quad (17.43)$$

In the limit $r \rightarrow \infty$, the integrand is a rapidly oscillating function that can be calculated by the stationary phase approximation. The stationary phase points are $\hat{\mathbf{n}}' = \pm \hat{\mathbf{n}}$. It is easy to derive

this result by writing the phase in terms of the angle, θ , between $\hat{n}$ and $\hat{n}'$. In this form, the stationary phase condition,

$$\frac{d}{d\theta} kr(1 - \cos \theta) = kr \sin \theta = 0, \quad (17.44)$$

yields $\theta = 0, \pi$ which imply that $\hat{n}' = \pm \hat{n}$. The contribution from the stationary point $\hat{n}' = -\hat{n}$ can be neglected due to the preexponential factor $(\hat{n} + \hat{n}')$. To calculate the contribution from the second stationary point, $\hat{n}' = \hat{n}$, let us define $\delta \hat{n}$ to be the deviation from the stationary point, i.e.

$$\hat{n}' = \hat{n} + \delta \hat{n}. \quad (17.45)$$

The normalization condition of this unit vector imposes the constraint

$$2\hat{n} \cdot \delta \hat{n} + \delta n^2 = 0 \quad (17.46)$$

Substituting (17.45) in (17.43) using Eq. (17.46) shows that within the stationary phase approximation (in which the pre-exponential factor is approximated by its value at the stationary point):

$$\begin{aligned} J_2 &= vr \oint d^2 \delta n \cdot \left\{ \exp \left[i \frac{kr}{2} \delta n^2 \right] a^*(\hat{n}, \hat{n}) + c.c. \right\} \\ &= \frac{2\pi v}{ik} a^*(\hat{n}, \hat{n}) + c.c. = -\frac{4\pi v}{k} \text{Im} a(\hat{n}, \hat{n}). \end{aligned} \quad (17.47)$$

This formula becomes exact in the limit $r \rightarrow \infty$. Now, conservation of the total number of particles dictates that $J_1 + J_2 = 0$, hence

$$\text{Im} a(\hat{n}, \hat{n}) = \frac{k}{4\pi} \oint d^2 \hat{n}' |a(\hat{n}', \hat{n})|^2. \quad (17.48)$$

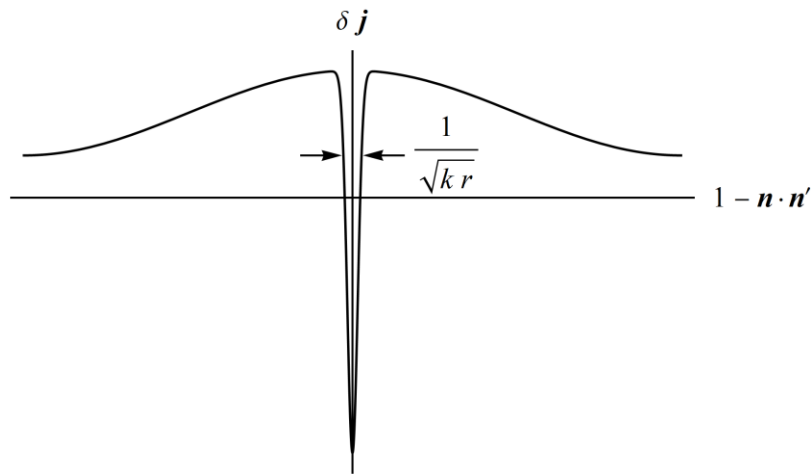


Figure 17-14 The angular dependence of the scattered particle current

This relation is known as the “*optical theorem*” of scattering theory. Its physical interpretation is that the total scattered particle current in directions that are different from the incoming wave, $\hat{n}$, must come from a reduction in the particle current moving in the original direction. In other words, if $\delta \mathbf{j} = \mathbf{j} - v\mathbf{n}$ is the scattered current, its integral over all directions vanishes, and the manner by which the positive and negative contributions add up to give zero is demonstrated schematically in Fig. 17-14.

Example: Scattering from a spherical potential

The large-distance asymptotic form of a wave function describing scattering from a finite range potential with spherical symmetry (located at the origin) is:

$$\psi = \exp(ikz) + a(\theta) \frac{\exp(ikr)}{r}. \quad (17.49)$$

Here $\exp(ikz)$ is the incoming waves describing a plane wave moving in the z direction, and $a(\theta)$ is the scattering amplitude that depends only on the angle between the direction scattered wave and the z axis. This function is a particular form of the general wave function of a free particle having a cylindrical symmetry:

$$\psi = \sqrt{4\pi} \sum_{l=0}^{\infty} \sqrt{2l+1} i^l Y_{l,0}(\theta) A_l \left[h_l^{(1)}(kr) e^{i2\delta_l} + h_l^{(2)}(kr) \right]. \quad (17.50)$$

Here $Y_{l,0}(\theta)$ are the spherical harmonic functions, while $h_l^{(1)}(x)$ and $h_l^{(2)}(x)$ are the spherical Hankel functions which have the following asymptotic behavior at large argument:

$$\begin{aligned} h_l^{(1)}(x) &\xrightarrow{x \rightarrow \infty} \frac{1}{i} \frac{\exp\left(ix - i\frac{\pi}{2}l\right)}{x}, \\ h_l^{(2)}(x) &\xrightarrow{x \rightarrow \infty} -\frac{1}{i} \frac{\exp\left(-ix + i\frac{\pi}{2}l\right)}{x}. \end{aligned} \quad (17.51)$$

Thus, the functions $h_l^{(1)}(x)$ are associated with radially outgoing waves, while $h_l^{(2)}(x)$ describe the incoming waves. The parameters δ_l are called the *scattering phase shift*. These parameters must be real in order to ensure the conservation of probability. Namely, that the flux of the incoming wave, with a given angular momentum, equals that of the outgoing wave with the same angular momentum. Finally, the coefficients A_l are arbitrary.

To identify the relation between $a(\theta)$ and the scattering phase shifts, we use the expansion of the plane wave in terms of the spherical Hankel functions:

$$\exp(ikz) = \sqrt{4\pi} \sum_{l=0}^{\infty} \sqrt{2l+1} i^l Y_{l,0}(\theta) \frac{h_l^{(1)}(kr) + h_l^{(2)}(kr)}{2}. \quad (17.52)$$

Equating the general solution (17.50) to (17.49) using (17.52) and substituting the asymptotic form of the Hankel functions (17.51), we have

$$\begin{aligned} & \frac{\sqrt{4\pi}}{k} \sum_{l=0}^{\infty} \sqrt{2l+1} i^l Y_{l,0}(\theta) \frac{1}{2i} \left[\frac{e^{ikr-i\frac{\pi}{2}l}}{r} e^{i2\delta_l} - \frac{e^{-ikr+i\frac{\pi}{2}l}}{r} \right] \\ &= \sqrt{4\pi} \sum_{l=0}^{\infty} \sqrt{2l+1} i^l Y_{l,0}(\theta) \frac{1}{2i} \left[\frac{e^{ikr-i\frac{\pi}{2}l}}{r} - \frac{e^{-ikr+i\frac{\pi}{2}l}}{r} \right] + a(\theta) \frac{\exp(ikr)}{r}, \end{aligned} \quad (17.53)$$

where we choose $A_k = 1/2$ to ensure that the incoming component of the wave function, on both sides of the equation, are equal. Solving for $a(\theta)$ we obtain:

$$\begin{aligned} a(\theta) &= \frac{\sqrt{4\pi}}{k} \sum_{l=0}^{\infty} \sqrt{2l+1} i^l Y_{l,0}(\theta) \frac{1}{2i} e^{-i\frac{\pi}{2}l} (e^{i2\delta_l} - 1) \\ &= \frac{\sqrt{4\pi}}{k} \sum_{l=0}^{\infty} \sqrt{2l+1} Y_{l,0}(\theta) e^{i\delta_l} \sin \delta_l. \end{aligned} \quad (17.54)$$

Now we see that

$$\begin{aligned} \int d\Omega |a(\theta)|^2 &= \frac{4\pi}{k^2} \sum_{l,m} \sqrt{2l+1} \sqrt{2m+1} e^{-i\delta_m} \sin \delta_m e^{i\delta_l} \sin \delta_l \underbrace{\int d\Omega Y_{m,0}^*(\theta) Y_{l,0}(\theta)}_{\delta_{lm}} \\ &= \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l, \end{aligned} \quad (17.55)$$

and using the property

$$Y_{l,0}(0) = \sqrt{\frac{2l+1}{4\pi}} \quad (17.56)$$

we obtain

$$\begin{aligned} \text{Im } a(0) &= \text{Im} \frac{\sqrt{4\pi}}{k} \sum_{l=0}^{\infty} \sqrt{2l+1} Y_{l,0}(0) e^{i\delta_l} \sin \delta_l \\ &= \text{Im} \frac{\sqrt{4\pi}}{k} \sum_{l=0}^{\infty} \sqrt{2l+1} \sqrt{\frac{2l+1}{4\pi}} e^{i\delta_l} \sin \delta_l = \frac{1}{k} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l. \end{aligned} \quad (17.57)$$

From Eqs. (17.55) and (17.57) follows the optical theorem:

$$\text{Im } a(0) = \frac{k}{4\pi} \int d\Omega |a(\theta)|^2. \quad (17.58)$$

Example: Scattering from a “hard” (i.e., impenetrable) sphere

Consider the case in which the scatterer is an impenetrable sphere of radius R . The wave function, ψ , in this case, satisfies Dirichlet boundary conditions on the surface of the sphere:

$$\psi(\mathbf{r})|_{r=R} = 0. \quad (17.59)$$

Imposing this condition on the wave function (17.50) yields the equations,

$$h_l^{(1)}(kR)e^{i2\delta_l} + h_l^{(2)}(kR) = 0, \quad (17.60)$$

that determine the phase shifts to be

$$e^{-i2\delta_l} = -\frac{h_l^{(1)}(kR)}{h_l^{(2)}(kR)} \quad (17.61)$$

Notice the absolute value of the ratio on the right-hand side of this equation is one because $h_l^{(2)}(z) = h_l^{(1)*}(z)$ for real z . The latter condition follows from the fact that the incoming and outgoing waves are related by time-reversal symmetry; hence one component is the complex conjugate of the other.

It is convenient to express the spherical Hankel functions in terms of the spherical Bessel and Neuman functions:

$$\begin{aligned} h_l^{(1)}(x) &= j_l(x) + in_l(x), \\ h_l^{(2)}(x) &= j_l(x) - in_l(x). \end{aligned} \quad (17.62)$$

With these definition,

$$e^{-i2\delta_l} = -\frac{j_l(kR) + in_l(kR)}{j_l(kR) - in_l(kR)} = \frac{n_l(kR) - ij_l(kR)}{n_l(kR) + ij_l(kR)}, \quad (17.63)$$

hence

$$\tan \delta_l = \frac{j_l(kR)}{n_l(kR)}, \quad (17.64)$$

and

$$\sin^2 \delta_l = \frac{\tan^2 \delta_l}{1 + \tan^2 \delta_l} = \frac{j_l^2(kR)}{j_l^2(kR) + n_l^2(kR)}. \quad (17.65)$$

We shall use this formula to calculate the total scattering cross-section in the limits where the sphere's radius is much smaller ($kR \ll 1$) or much larger ($kR \gg 1$) than the particle wavelength.

Recall that the scattering cross-section is defined as the number of scattered particles per unit time divided by the incoming particles' flux (number of particles per unit time per unit area).

In our calculation, we choose the amplitude of the incoming wave to be one; hence the cross-section is:

$$\sigma_{\text{scat}} = \int d\Omega |a(\theta)|^2 = \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l \quad (17.66)$$

To calculate this sum, we need the asymptotic behavior of the spherical Bessel and Neuman functions:

$$j_l(x) \sim \begin{cases} \frac{2^l l!}{(2l+1)!} x^l & x \rightarrow 0 \\ \frac{1}{x} \sin\left(x - l \frac{\pi}{2}\right) & x > l \end{cases} \quad n_l(x) \sim \begin{cases} -\frac{(2l)!}{2^l l!} \frac{1}{x^{l+1}} & x \rightarrow 0 \\ -\frac{1}{x} \cos\left(x - l \frac{\pi}{2}\right) & x > l \end{cases} \quad (17.67)$$

Consider first the limit $kR \ll 1$. Using Eqs. (17.65) and (17.67) we obtain

$$\lim_{kR \rightarrow 0} \sin^2 \delta_l = \frac{\left(\frac{2^l l!}{(2l+1)!} x^l \right)^2}{\left(\frac{(2l)!}{2^l l!} \frac{1}{x^{l+1}} \right)^2} \bigg|_{x=kR} = \left(\frac{4^l (l!)^2}{(2l+1)!(2l)!} \right)^2 (kR)^{4l+2}. \quad (17.68)$$

From here, it follows that the main contribution comes from $l=0$, where $\sin^2 \delta_0 = (kR)^2$. Substituting this result in (17.66), we obtain:

$$\sigma_{\text{scat}} = 4\pi R^2, \quad \text{for} \quad kR \ll 1. \quad (17.69)$$

Notice that this scattering cross-section equals the full surface of the sphere rather than its cross-section area, πR^2 . This is not surprising because this result applies to the ultra-quantum limit, where the size of the sphere is much smaller than the particle wavelength.

Consider now the opposite limit, $kR \gg 1$. Using Eqs. (17.65) and (17.67) give that for $kR > l$

$$\sin^2 \delta_l \approx \sin^2 \left(kR - l \frac{\pi}{2} \right) = \frac{1}{2} [1 - \cos(2kR - l\pi)] = \frac{1}{2} [1 - (-1)^l \cos(2kR)] \quad (17.70)$$

Substituting this result in Eq. (17.66) and taking into account that the contribution from the second term in (17.70) is negligible because of the alternating signs, and that sum over l extends up to kR where the asymptotic formula (17.67) for large argument applies (at higher values δ_l decays exponentially with l), we obtain:

$$\sigma_{\text{scat}} = \frac{4\pi}{k^2} \sum_{l=0}^{kR} (2l+1) \frac{1}{2} = 2\pi R^2, \quad \text{for} \quad kR \gg 1. \quad (17.71)$$

Thus, we obtain a cross-section that is precisely twice the classical value (the area of the sphere cross-section). This result is surprising because, from the correspondence principle, we expect that when the wavelength approaches zero, one should recover the classical result.

This discrepancy between the classical value of the cross-section and its quantum value in the classical limit is called the *extinction paradox*.

The reason for paradoxes is usually rooted in some hidden assumptions. Here it is associated with the orders of limit in which the cross-section is calculated. To reveal this assumption, let us recall the problem of diffraction from a slit whose width, w , is much larger than the wavelength λ . In this problem, one identifies two regimes that depend on the distance, r , of the detector from the slit. In the near-field regime, $\lambda \ll r \ll \sqrt{\lambda w}$, diffraction effects due to the edges of the slit are small so that rays move, essentially, along the classical trajectories.

On the other hand, in the far-field regime, $r \gg \sqrt{\lambda w}$, essentially all rays detract and generate the distinctive Fraunhofer interference pattern. This example shows that although one may assume λ to be the smallest length scale in the problem in both cases, the order of limits, $\lambda \rightarrow 0$ and $r \rightarrow \infty$, is important. The classical result is obtained when first $\lambda \rightarrow 0$ and only then $r \rightarrow \infty$, while strong diffraction is revealed in the opposite order, first $r \rightarrow \infty$ and only then $\lambda \rightarrow 0$.

The calculation that led to Eq. (17.71) has been performed in the orders of limits where quantum diffractions effects are always significant (i.e., first $r \rightarrow \infty$ and then $\lambda \rightarrow 0$). To understand the reason for the factor of two between the quantum and the classical result, let us consider the problem of diffraction from a thin circular mirror of area A , as illustrated in Fig. 17-15. Clearly, all rays that hit the mirror will be reflected and contribute an area A to the total cross-section. However, diffraction takes place also for rays that do not hit the mirror, as illustrated in the figure.

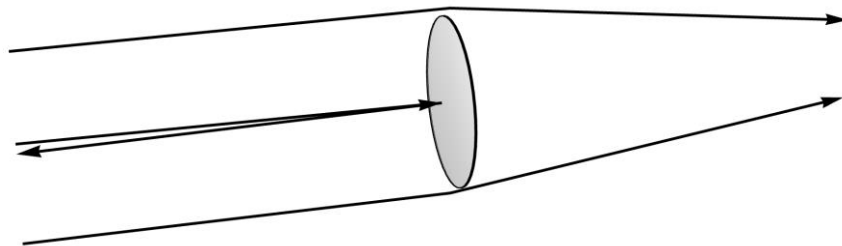


Figure 17-15 Diffraction from a thin circular mirror

To find the contribution of these rays to the total cross-section, we use Babinet's principle according to which, in the far-field regime, the diffraction pattern from an impenetrable screen is the same as from a slit with the same shape (in any direction which is different from that of the incoming plane wave). This principle implies that the effective area of the diffractive rays (which do not hit the mirror directly) is also A , hence the total cross-section is $2A$.

17.4 The kinetic (Boltzmann) equation for elastic scattering

Impurities and lattice defects scatter the electrons from one Bloch state to another. In this section, we construct the kinetic equations that describe this process. This description is semiclassical in nature: It constitutes an equation for the probability density, $f(\mathbf{r}, \mathbf{p}, t)$, of finding a particle at a point $(\mathbf{r}, \mathbf{p})$ in phase space at a time t . It is based on the assumption that the elastic mean free path, that characterizes the distance between successive scattering events, is much larger than the particle's wavelength and neglects interference effects. Solution of the kinetic equation allows one to calculate the particle's density,

$$\rho = \int d^3 p f(\mathbf{r}, \mathbf{p}, t), \quad (17.72)$$

and the current density:

$$\mathbf{j} = \int d^3 p f(\mathbf{r}, \mathbf{p}, t) \frac{\mathbf{p}}{m}. \quad (17.73)$$

If there is no scattering, the electron distribution function satisfies the Liouville equation, which in the absence of external forces reduces to

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} = 0, \quad (17.74)$$

where $\mathbf{v} = \mathbf{p}/m$ is the particle's velocity. A solution of this equation, analogous to a plane wave solution of the Schrödinger equation of a free particle, is:

$$f(\mathbf{r}, \mathbf{p}, t) = \rho \delta(\mathbf{p} - \mathbf{p}_0), \quad (17.75)$$

where ρ is the density, which is constant and homogeneous in space, while $\mathbf{p}_0$ is the initial momentum. The particle current obtained from this distribution function is

$$\mathbf{j} = \rho \frac{\mathbf{p}_0}{m}. \quad (17.76)$$

Since elastic scattering conserves energy, it is convenient to express the conservation of momentum, as conservation of its absolute value and direction, $\hat{\mathbf{n}} = \mathbf{p}/p$. For parabolic energy spectrum, conservation of the absolute value of the momentum is equivalent to conservation of the energy, $\varepsilon_0 = p_0^2/2m$, hence representing the probability density function in the form $f(\mathbf{r}, \mathbf{p}, t) = f(\varepsilon, \hat{\mathbf{n}})$, the solution (17.71) may be written in the form:

$$f(\varepsilon, \hat{\mathbf{n}}) = \frac{\rho v}{8\pi m \varepsilon} \delta(\varepsilon - \varepsilon_0) \delta(\hat{\mathbf{n}} - \hat{\mathbf{n}}_0), \quad (17.77)$$

where $\hat{\mathbf{n}}_0 = \mathbf{p}_0/p_0$ is the initial direction of the particle's momentum. From now on, we shall suppress the energy dependence of the distribution function because it is conserved.

To take into account scattering by disorder, one has to add to the right-hand side of the Liouville equation (17.70) a term that describes the change of the distribution in time due to scattering:

$$\frac{\partial f(\hat{n}, \mathbf{r}, t)}{\partial t} + \mathbf{v} \cdot \frac{\partial f(\hat{n}, \mathbf{r}, t)}{\partial \mathbf{r}} = \text{St}_{\text{el}}[f(\hat{n}, \mathbf{r}, t)], \quad (17.78)$$

with the collision integral

$$\text{St}_{\text{el}}[f] = \sum_i \delta(\mathbf{r} - \mathbf{R}_i) \iint d^2 \hat{n}' v \left[|a(\hat{n}, \hat{n}')|^2 f(\hat{n}', \mathbf{r}, t) - |a(\hat{n}', \hat{n})|^2 f(\hat{n}, \mathbf{r}, t) \right]. \quad (17.79)$$

Here, for simplicity, we assume that the lattice disorder comprises identical scatterers located at points $\mathbf{R}_i$, with a differential cross-section of each scatterer given by $|a(\hat{n}', \hat{n})|^2$ (for scattering from direction $\mathbf{n}$ to $\mathbf{n}'$). The first term of Eq. (17.79) describes the increase in the probability density to find a particle in direction $\mathbf{n}$, due to scattering of a particle from a different direction $\mathbf{n}'$ into $\mathbf{n}$. The second term represents the decrease in the probability distribution as a result of scattering from $\mathbf{n}$ to $\mathbf{n}'$.

Notice that except for the scattering amplitude, which may take into account quantum properties of the system, the above equation is essentially classical. Expressed only in terms of the probability for scattering from individual scatterers, it neglects possible interference effects. To explain this point, consider two trajectories of a particle passing from point A to point B , via several scattering events, as illustrated in Fig. 17-16. Denoting by w_1 and w_2 the probabilities of passing through the upper and lower trajectories, respectively, the classical probability of passing from point A to point B , is given by the sum $w = w_1 + w_2$.

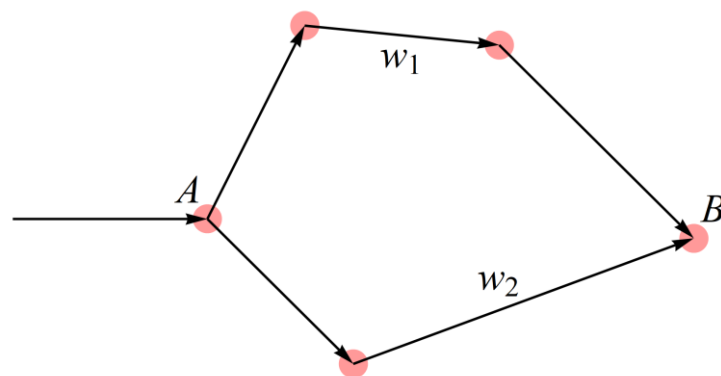


Figure 17-16 A particle passing from point A to point B experiencing several scattering events

On the other hand, when treating the problem quantum mechanically, one should sum probability amplitudes rather than probabilities. The probability amplitudes are given by

$\sqrt{w_i} \exp(i\phi_i)$, where ϕ_1 and ϕ_2 are the phases accumulated along the upper and lower trajectories, respectively. Thus, the probability of transition is

$$w = \left| \sqrt{w_1} \exp(i\phi_1) + \sqrt{w_2} \exp(i\phi_2) \right|^2 = w_1 + w_2 + 2\sqrt{w_1 w_2} \cos(\phi_1 - \phi_2) \quad (17.80)$$

The last term on the right-hand side of this equation is the interference contribution. This term is not taken into account by the kinetic equation (17.73). Interference might become an important factor leading to various physical phenomena such as localization, weak localization, and mesoscopic fluctuations. These subjects are beyond the scope of this course. The interference effects can be neglected when scattering is sufficiently weak or in the presence of processes that destroy phase coherence (such as electron-electron interaction at sufficiently high temperature).

Usually (even when taking interference effects into account), the microscopic description of elastic scattering from individual scatterers contains too much information, which is not required to characterize the behavior of a large system. Instead, one may characterize the scattering process by the transition rate between different momentum states (or momentum directions). To obtain such a description, we define the density of scatterers as

$$N_{\text{imp}} = \left\langle \sum_i \delta(\mathbf{r} - \mathbf{R}_i) \right\rangle, \quad (17.81)$$

where $\langle \dots \rangle$ represents a statistical average over the positions of the scatterers. Here we assume that the density of scatterers is uniform in space. In this case, the collision integral (17.75) reduces to

$$\text{St}_{\text{el}}[f] = \iint \frac{d^2 \hat{n}'}{4\pi} \left[\frac{f(\hat{n}', \mathbf{r}, t)}{\tau(\hat{n}, \hat{n}')} - \frac{f(\hat{n}, \mathbf{r}, t)}{\tau(\hat{n}', \hat{n})} \right] \quad (17.82)$$

where

$$\frac{1}{\tau(\hat{n}, \hat{n}')} = 4\pi N_{\text{imp}} v |a(\hat{n}, \hat{n}')|^2 \quad (17.83)$$

is the scattering rate from a state where the particle moves in direction $\hat{n}'$ to a state where it moves in direction $\hat{n}$. In the absence of a magnetic field, time-reversal symmetry implies that

$$\tau(\hat{n}, \hat{n}') = \tau(-\hat{n}, -\hat{n}'). \quad (17.84)$$

On the other hand, in general, one cannot replace the roles of the incoming and outgoing momenta during the scattering events, namely

$$\tau(\hat{n}, \hat{n}') \neq \tau(\hat{n}', \hat{n}). \quad (17.85)$$

This feature is explained in Fig. 17-17. Nevertheless, when the scattering potential is weak, and the transition rate can be calculated perturbatively (by Fermi's golden rule), one obtains

that $\tau(\hat{n}, \hat{n}') = \tau(\hat{n}', \hat{n})$, see Ex. 1. This equality is called the “*principle of microscopic reversibility*”. Assuming this property, the kinetic equation (17.78) reduces to:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} = \iint \frac{d^2 \hat{n}'}{4\pi} \frac{1}{\tau(\hat{n}, \hat{n}')} [f(\hat{n}', \mathbf{r}, t) - f(\hat{n}, \mathbf{r}, t)]. \quad (17.86)$$

Finally, we comment that elastic scattering does not take the system into an equilibrium state because the particle energy remains intact. On the other hand, the momentum relaxes and leads to diffusive behavior, as discussed in the following chapter.

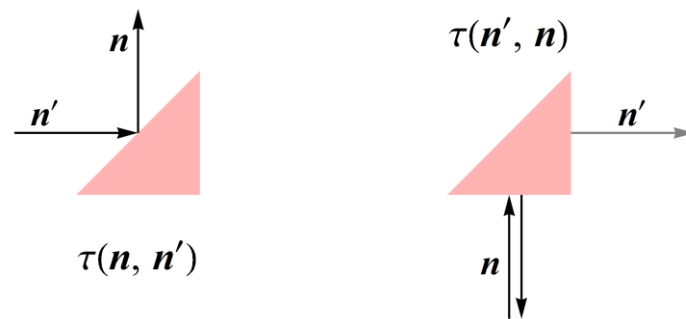


Figure 17-17 Scattering from an impenetrable triangle illustrating the inequality (17.80).

17.5 Exercises

1. Prove the principle of microscopic reversibility with the framework of Fermi's golden rule.
2. For an electron in a two-dimensional system, calculate the total cross-section from an impenetrable circular potential in the following limits:
 - (a) When the circle radius is much smaller than the electron's wavelength. Here, show that the cross-section is determined mainly by the electron wavelength rather than the size of the scatterer.
 - (b) When the wavelength is much smaller than the circle radius. Show that the cross-section obtained in this limit is twice the classical one.

Advice: Use the solution of the two dimensional Schrödinger equation of a free particle in polar coordinates (r, θ) :

$$\psi = \sum_{n=-\infty}^{\infty} i^n \exp(in\theta) A_n \left[H_n^{(2)}(kr) e^{i2\delta_n} + H_n^{(1)}(kr) \right] \quad (17.87)$$

where

$$\begin{aligned} H_n^{(1)}(x) &= J_n(x) + iN_n(x) \\ H_n^{(2)}(x) &= J_n(x) - iN_n(x) \end{aligned} \quad (17.88)$$

are the Hankel functions of the first and the second kind expressed in terms of the Bessel and the Neuman functions which have the following asymptotic behavior:

$$J_n(x) \sim \begin{cases} \frac{1}{n!} \left(\frac{x}{2} \right)^n & x \rightarrow 0 \\ \sqrt{\frac{2}{\pi x}} \cos \left(x - n \frac{\pi}{2} - \frac{\pi}{4} \right) & x \gg n \end{cases} \quad (17.89)$$

$$N_n(x) \sim \begin{cases} \frac{2}{\pi} \ln \frac{x}{2} & x \rightarrow 0, n=0 \\ -\frac{(n-1)!}{\pi} \left(\frac{2}{x} \right)^n & x \rightarrow 0, n>0 \\ \sqrt{\frac{2}{\pi x}} \sin \left(x - l \frac{\pi}{2} - \frac{\pi}{4} \right) & x \gg n \end{cases} \quad (17.90)$$

Use also the expansion:

$$\exp(ikx) = \exp(ikr \cos \theta) = \sum_{n=-\infty}^{\infty} i^n J_n(kr) \exp(in\theta). \quad (17.91)$$

18 Transport coefficients & thermoelectric effects

In the final chapter of this course, we consider the nonequilibrium behavior of electrons in disordered crystals in response to spatial changes in the electron's density, ρ , the temperature gradient, ∇T , and the application of an external electric field, $\mathbf{E}$. We focus our attention on the regime of linear response where the electric current is given by

$$\mathbf{j} = \sigma \mathbf{E} - D \nabla \rho - \beta \nabla T, \quad (18.1)$$

and discuss the coefficients that characterize the response and the relations among them. These are the conductivity tensor σ , the diffusion constant, D , and the thermopower coefficient, β . We conclude by presenting thermoelectric effects - the Seebeck, Peltier, and the Thomson effects - and the relations among the thermoelectric coefficients associated with them.

18.1 Diffusion

Diffusion is a dynamical process by which a particle scatters and changes its direction rapidly, but, similar to random walk, its position changes slowly in space such that the square of the distance from the origin is proportional to time:

$$\langle r^2 \rangle = 2Dt, \quad (18.2)$$

where D is the diffusion constant. Here we derive this equation and identify the diffusion constant by solving the kinetic (Boltzmann) equation (17.74) with the collision integral (17.78) for which we assume the principle of microscopic reversibility, $\tau(\hat{\mathbf{n}}, \hat{\mathbf{n}}') = \tau(\hat{\mathbf{n}}', \hat{\mathbf{n}})$, i.e.,

$$\frac{\partial f}{\partial t} + v \hat{\mathbf{n}} \cdot \frac{\partial f}{\partial \mathbf{r}} = \iint \frac{d^2 \hat{\mathbf{n}}'}{4\pi} \frac{1}{\tau(\hat{\mathbf{n}}, \hat{\mathbf{n}}')} [f(\hat{\mathbf{n}}', \mathbf{r}, t) - f(\hat{\mathbf{n}}, \mathbf{r}, t)]. \quad (18.3)$$

We also assume the absolute value of the velocity to be approximately constant and express it in the form $\mathbf{v} = v \hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is a unit vector.

Averaging Eq. (18.3) over the directions of $\hat{\mathbf{n}}$ yields the continuity equation:

$$\frac{\partial \langle f \rangle}{\partial t} + \nabla \cdot \langle v \hat{\mathbf{n}} f \rangle = 0, \quad (18.4)$$

where

$$\langle \dots \rangle = \int \frac{d^2 \hat{\mathbf{n}}}{4\pi} (\dots). \quad (18.5)$$

It is clear that Eq. (18.3) is satisfied by a constant function. Such a solution is expected to be reached after a long time when momentum relaxes and the particle probability density spreads uniformly in space. This property suggests looking for a solution in the form of a gradient expansion,

$$f(\hat{\mathbf{n}}, \mathbf{r}, t) = f_0(\mathbf{r}, t) + \hat{\mathbf{n}} \cdot \nabla f_1(\mathbf{r}, t) + \hat{n}_i \hat{n}_j \frac{\partial}{\partial r_i} \frac{\partial}{\partial r_j} f_2(\mathbf{r}, t) + \dots, \quad (18.6)$$

where $f_i(\mathbf{r}, t)$ changes slowly in space and time, i.e., over distance (and time) that is much larger than the typical distance (or time) between successive scattering events. In what follows, we consider only the first two terms of the above expansion. Substituting (18.6) in the continuity equation (18.4),

$$\frac{\partial}{\partial t} \langle f_0 + \hat{\mathbf{n}} \cdot \nabla f_1 \rangle + \nabla \cdot \langle v \hat{\mathbf{n}} (f_0 + \hat{\mathbf{n}} \cdot \nabla f_1) \rangle = 0 \quad (18.7)$$

we obtain

$$\frac{\partial f_0(\mathbf{r}, t)}{\partial t} + \frac{v}{d} \nabla^2 f_1(\mathbf{r}, t) = 0. \quad (18.8)$$

The terms linear in $\hat{\mathbf{n}}$ vanish upon averaging, while those quadratic in $\hat{\mathbf{n}}$ are calculated as follows:

$$\nabla \cdot \langle v \hat{\mathbf{n}} f \rangle = \nabla \cdot \langle v \hat{\mathbf{n}} [f_0 + \hat{\mathbf{n}} \cdot \nabla f_1] \rangle = v \langle \hat{n}_i \hat{n}_j \rangle \frac{\partial}{\partial r_i} \frac{\partial}{\partial r_j} f_1. \quad (18.9)$$

Now, it is clear that for $i \neq j$ $\langle \hat{n}_i \hat{n}_j \rangle = 0$, while for $i = j$ the average $\langle \hat{n}_i^2 \rangle$, is, by symmetry, independent of i ; hence

$$\langle \hat{n}_i \hat{n}_j \rangle = \delta_{ij} \langle \hat{n}_i^2 \rangle = \frac{1}{3} \left\langle \sum_{i=1}^3 \hat{n}_i^2 \right\rangle \delta_{ij} = \frac{1}{3} \delta_{ij}. \quad (18.10)$$

Notice that in this equation, we do not sum over repeated indices. Substituting this equation in (18.9) gives the second term in (18.8).

To obtain an additional relation between f_0 and f_1 , we substitute the expansion (18.6) in (18.3), multiply by $\hat{\mathbf{n}}$, and average over the directions of $\hat{\mathbf{n}}$. Keeping those terms that do not vanish upon averaging, the resulting equation takes the form:

$$\frac{\partial}{\partial t} \langle \hat{\mathbf{n}} \hat{n}_i \rangle \frac{\partial}{\partial r_i} f_1(\mathbf{r}, t) + v \langle \hat{\mathbf{n}} \hat{n}_i \rangle \frac{\partial f_0(\mathbf{r}, t)}{\partial r_i} = \int \frac{d^2 \hat{\mathbf{n}}}{4\pi} \int \frac{d^2 \hat{\mathbf{n}}'}{4\pi} \frac{1}{\tau(\hat{\mathbf{n}}', \hat{\mathbf{n}})} \hat{\mathbf{n}} (\hat{n}'_i - \hat{n}_i) \frac{\partial f_1}{\partial r_i}. \quad (18.11)$$

From now on, we assume that the scattering time depends only on the angle between the incoming and outgoing rays, θ , i.e. $\hat{\mathbf{n}}' \cdot \hat{\mathbf{n}} = \cos \theta$, and denote $\tau(\hat{\mathbf{n}}', \hat{\mathbf{n}}) = \tau(\theta)$. With this assumption, the double integral on the right-hand side of Eq. (18.11) reduces to a single integral over the angle θ . Thus,

$$\int \frac{d^2 \hat{n}}{4\pi} \int \frac{d^2 \hat{n}'}{4\pi} \frac{1}{\tau(\hat{n}', \hat{n})} \hat{n}(\hat{n}' - \hat{n}) \frac{\partial f_1}{\partial r_i} = -\frac{1}{\tau_{tr}} \nabla f_1, \quad (18.12)$$

with

$$\frac{1}{\tau_{tr}} = \frac{1}{2} \int_0^\pi d\theta \sin \theta \frac{1}{\tau(\theta)} (1 - \cos \theta). \quad (18.13)$$

The time τ_{tr} is called the *transport mean free time*. It is the typical time when the particle loses memory of its initial direction. When scattering is by small angles (i.e., when $\tau(\theta)$ is peaked at $\theta = 0$), the transport mean free time can be much larger than the *elastic mean free time*,

$$\frac{1}{\tau} = \frac{1}{2} \int_0^\pi d\theta \sin \theta \frac{1}{\tau(\theta)}, \quad (18.14)$$

because many scattering events are required to reverse the particle's direction.

Using Eq. (18.10) in order to evaluate the left-hand side of Eq. (18.11), and the Eq. (18.12) for the right-hand side, we obtain

$$\frac{1}{3} \frac{\partial}{\partial t} \nabla f_1(\mathbf{r}, t) + \frac{1}{3} v \nabla f_0(\mathbf{r}, t) = -\frac{1}{3\tau_{tr}} \nabla f_1(\mathbf{r}, t). \quad (18.15)$$

However, the first term on the left-hand side of the equation is negligible because time derivative is equivalent to higher spatial derivative (as follows from Eq. (18.8)), hence

$$f_1(\mathbf{r}, t) = -v\tau_{tr} \nabla f_0(\mathbf{r}, t). \quad (18.16)$$

Substituting (18.16) in (18.8), we obtain the diffusion equation:

$$\frac{\partial}{\partial t} f_0(\mathbf{r}, t) = \nabla \cdot [D \nabla f_0(\mathbf{r}, t)], \quad (18.17)$$

where

$$D = \frac{v^2 \tau_{tr}}{3} \quad (18.18)$$

is the diffusion constant. The above result is derived for a three-dimensional system. In a d -dimensional system, a similar calculation gives Eq. (18.17) with the diffusion constant, $D = v^2 \tau_{tr} / d$.

Assuming D to be uniform in space, the solution of the diffusion equation (18.17) (assuming initial conditions where the particle is located at the origin) is:

$$f_0(r, t) = \frac{1}{(4\pi Dt)^{\frac{d}{2}}} \exp\left(-\frac{r^2}{4Dt}\right). \quad (18.19)$$

Using this solution to calculate the expectation value of r^2 leads to Eq. (18.2).

To summarize, the above results imply that the diffusion equation (for charged particles) is obtained from the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0, \quad (18.20)$$

where ρ and $\mathbf{j}$ are the charge density and the current density, respectively. Fick's law determines the latter,

$$\mathbf{j} = -D \nabla \rho. \quad (18.21)$$

This diffusion current is only one of the contributions to the total current. Additional contributions may come from the temperature gradient and the electric field, see Eq. (18.1).

18.2 The electric conductivity and Einstein's relation

Consider a system of electrons in thermal equilibrium without a magnetic field. The electric current in the system,

$$\mathbf{j} = -e \int \frac{d^d p}{(2\pi)^d} \mathbf{v} f(\mathbf{r}, \mathbf{p}, t), \quad (18.22)$$

vanishes due to time-reversal symmetry. Mathematically it can be verified by taking into account that the equilibrium distribution function is only a function of the energy $f(\mathbf{r}, \mathbf{p}, t) = f(\varepsilon)$ (so that its Poisson bracket with the Hamiltonian vanishes). Then representing the distribution function in the form $f = \partial F / \partial \varepsilon$, and taking into account that the velocity is the derivative of the energy with respect to the momentum, $\mathbf{v} = \partial \varepsilon(\mathbf{p}) / \partial \mathbf{p}$, the current can be expressed as an integral over a full derivative and therefore vanishes:

$$\mathbf{j} = -e \int \frac{d^d p}{(2\pi)^d} \frac{\partial \varepsilon}{\partial \mathbf{p}} f(\varepsilon) = -e \int \frac{d^d p}{(2\pi)^d} \frac{\partial F}{\partial \mathbf{p}} = 0. \quad (18.23)$$

Now consider an isolated system subjected to an electric field, $\mathbf{E}$, but in an equilibrium state so that there is no electric current flowing in the system. For simplicity, we assume that the system is isotropic and homogeneous; hence the conductivity tensor can be replaced by a scalar, σ . In this situation, the electric current due to the electric field must be compensated by the diffusion current due to the gradient of the charge density, ρ , i.e.,

$$\mathbf{j} = \sigma \mathbf{E} - D \nabla \rho = 0. \quad (18.24)$$

This equation suggests that the conductivity and the diffusion constant are related. To identify this relation, let $\varphi(\mathbf{r})$ be the electric scalar potential that determines the electric field,

$$\mathbf{E} = -\nabla \varphi(\mathbf{r}). \quad (18.25)$$

The Hamiltonian of the system is then given by

$$H = \varepsilon_0(\mathbf{p}) - e\varphi(\mathbf{r}), \quad (18.26)$$

where $\varepsilon_0(\mathbf{p})$ is the electronic spectrum of the system in the absence of an electric field. To proceed, we adopted a semiclassical approximation where the Fermi wavelength is much smaller than any other length scale in the problem. In particular, the electric field is assumed to be sufficiently weak so that $\varphi(\mathbf{r})$ changes very slowly over a distance of the order of the Fermi wavelength. With this assumption, the equilibrium electronic distribution function can be approximated by

$$f_{\text{eq}}(\mathbf{r}, \mathbf{p}) \simeq \theta(\varepsilon_F - H) = \theta[\varepsilon_F + e\varphi(\mathbf{r}) - \varepsilon_0(\mathbf{p})] \quad (18.27)$$

where we assumed the Fermi energy, ε_F , to be much larger than the temperature, $k_B T \ll \varepsilon_F$, therefore the Fermi distribution function can be replaced by a unit step function (this is the typical situation in good metals). Within this semiclassical approach, each electron occupies a volume of $(2\pi\hbar)^d$ in phase space; hence the charge density is obtained by the following integral over the momentum:

$$\rho = -e \int \frac{d^d p}{(2\pi\hbar)^d} \theta[\varepsilon_F + e\varphi(\mathbf{r}) - \varepsilon_0(\mathbf{p})]. \quad (18.29)$$

It follows that to the leading order in $e\varphi(\mathbf{r})/\varepsilon_F$,

$$\nabla \rho \simeq -e^2 \nu \nabla \varphi(\mathbf{r}), \quad (18.30)$$

where

$$\nu = \int \frac{d^d p}{(2\pi\hbar)^d} \delta[\varepsilon_F - \varepsilon_0(\mathbf{p})] \quad (18.31)$$

is the density of states at the Fermi energy. Substituting (18.30) and (18.25) in the equilibrium condition (18.24) yields *Einstein's relation*,

$$\sigma = e^2 \nu D, \quad (18.32)$$

that relates the conductivity to the diffusion constant.

Example: PN junction (a diode)

A PN-junction is obtained when a *p*-type semiconductor is brought to contact with *n*-type semiconductor. Near the interface, electrons from the *n*-type semiconductor diffuse into the *p*-type semiconductor and compensate the acceptors, creating a depletion layer with a low

concentration of free charges, as illustrated in the upper panel of Fig. 18-1. At an equilibrium state, the diffusion current, due to the gradient of concentrations of holes and electrons, is compensated by the electric current induced by the electric field. This electric field is created by the charge carriers that have passed from one side of the junction to its other side.

An electric field's buildup is required to equate chemical potential on both sides of the junction. It leads to a voltage drop, ΔV , as illustrated in the lower panel of Fig. 18-1.

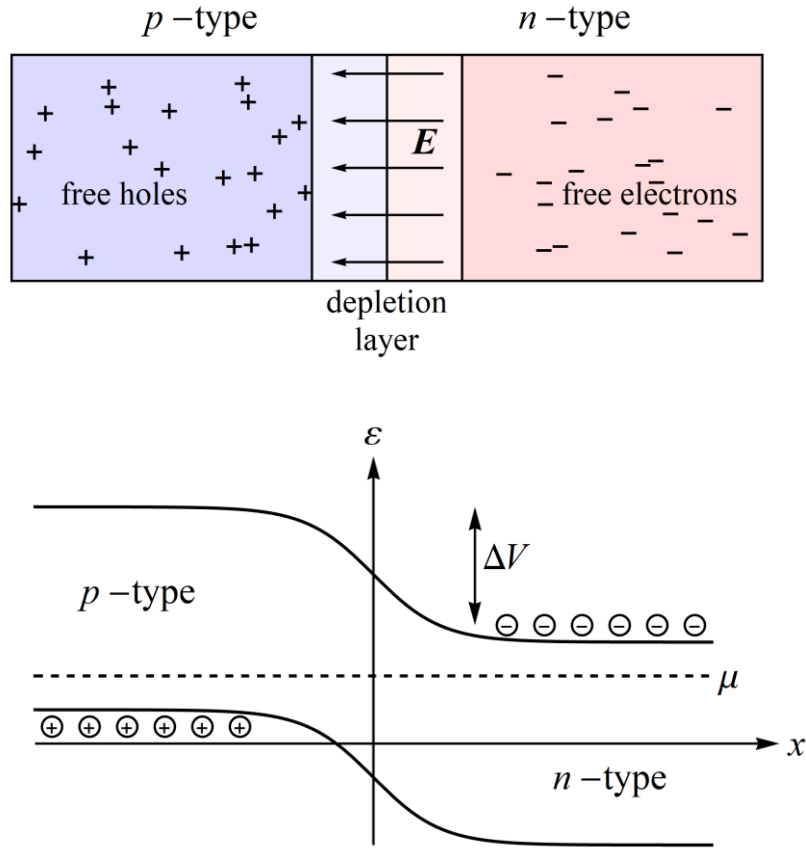


Figure 18-1 A PN junction (diode) at equilibrium. Upper panel: The depletion layer and the buildup of an electric field. Lower panel: The band structure of electrons.

To calculate the potential drop across the PN junction, we consider, for simplicity, the case in which the intrinsic semiconductors on both sides of the junction are identical and construct the equation for a zero current flow, analogous to (18.24). Here, however, one has to take into account that the electric field and the density of carriers are not uniform in space. Focusing on the current due to holes, and assuming one-dimensional geometry, we have

$$j_x = e \left[\mu_h n_h(x) E(x) - D_h \frac{\partial}{\partial x} n_h(x) \right] = 0, \quad (18.33)$$

where $n_h(x)$ is the density of holes at point x , $\mu_h = e\tau_{tr}/m_h$ is the mobility of the holes (see definition below Eq.(1.3)) in which m_h is the effective mass of the hole, and D_h is the hole diffusion constant. Notice that in the case of lightly dope semiconductors, one cannot use relation (18.32) that was derived for a metal. Instead, assuming a temperature range where the Boltzmann distribution approximates the hole distribution function, one can calculate the diffusion constant using the equipartition law:

$$D_h = \frac{v^2 \tau_{tr}}{3} = \frac{m_h v^2}{2} \frac{2\tau_{tr}}{3m_h} = \frac{3k_B T}{2} \frac{2\tau_{tr}}{3m_h} = \frac{k_B T \tau_{tr}}{m_h}. \quad (18.34)$$

This result may also be written in the form $D_h = k_B T \mu_h / e$, which constitutes the Einstein relation for a Brownian particle. Using this result, Eq. (18.33) reduces to:

$$\frac{1}{n_h(x)} \frac{\partial n_h(x)}{\partial x} = \frac{\mu_h}{D_h} E(x) = -\frac{e}{k_B T} \frac{\partial \varphi(x)}{\partial x}, \quad (18.35)$$

where $\varphi(x)$ is the electric potential. Integrating the above equation across the depletion layer from a point located deep in the p -type semiconductor to a point located deep in the n -type semiconductor, we obtain:

$$\ln \left(\frac{n_h^{n\text{-type}}}{n_h^{p\text{-type}}} \right) = -\frac{e}{k_B T} \Delta V, \quad (18.36)$$

where ΔV is the voltage across the junction (i.e. the difference between the voltage in the n -type side of the junction, V_N , and the voltage on the p -type side, V_p), while, $n_h^{p\text{-type}}$ and $n_h^{n\text{-type}}$ are the densities of the holes at the p -type and at the n -type regions of the junction, respectively. In particular, the density of holes in the p -type semiconductor can be approximated by the density of acceptors, $n_h^{p\text{-type}} \simeq N_A$, and using the law of mass action (17.10), we deduce that the density of holes in the n -type semiconductor is $n_h^{n\text{-type}} \simeq n_i^2 / N_D$, where n_i is the intrinsic density of charge carriers in the semiconductor, while N_D is the density of donors (which approximates the density of electrons). From here, we conclude that

$$\Delta V = \frac{k_B T}{e} \ln \left(\frac{N_A N_D}{n_i^2} \right). \quad (18.37)$$

18.3 The second law of thermodynamics in electronic systems

Consider the Boltzmann (kinetic) equation for the distribution function of electrons subjected to an external electric field,

$$\frac{\partial f}{\partial t} + \frac{\partial \varepsilon}{\partial \mathbf{p}} \cdot \frac{\partial f}{\partial \mathbf{r}} - e \mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} = \text{St}[f], \quad (18.38)$$

where the collision integral accounts, both for the scattering of the electrons by impurities and the collisions among electrons¹. These conditions imply that the scattering processes conserve the number of particles and energy. Thus, multiplication of Eq. (18.38) by $-e$ and integration over the momentum yields the conservation law of electrical current (18.20), where the charge density and the current density are given by

$$\begin{pmatrix} \rho \\ \mathbf{j} \end{pmatrix} = \int \frac{d^d p}{(2\pi\hbar)^d} f(\mathbf{p}) \begin{pmatrix} -e \\ -e \frac{\partial \varepsilon}{\partial \mathbf{p}} \end{pmatrix}. \quad (18.39)$$

Notice that this equation follows from two properties: (a) the conservation of the number of particles by collisions, and (b) that the integral over the force term in the Boltzmann equation (18.38), $-e \mathbf{E} \cdot \partial f / \partial \mathbf{p}$, vanish because it is a full derivative.

Similarly, one can construct the dynamical equation for the energy flow in the system,

$$\frac{\partial \rho_\varepsilon}{\partial t} + \nabla \cdot \mathbf{j}_\varepsilon = \mathbf{j} \cdot \mathbf{E}, \quad (18.40)$$

where ρ_ε is the energy density, and $\mathbf{j}_\varepsilon$ is the energy current, defined by the integrals:

$$\begin{pmatrix} \rho_\varepsilon \\ \mathbf{j}_\varepsilon \end{pmatrix} = \int \frac{d^d p}{(2\pi\hbar)^d} f(\mathbf{p}) \begin{pmatrix} \varepsilon \\ \varepsilon \frac{\partial \varepsilon}{\partial \mathbf{p}} \end{pmatrix}. \quad (18.41)$$

Here, the contribution from collision integral vanishes; however, the force term on the left-hand side of the Boltzmann equation (18.38) gives a finite contribution to the energy balance equation (18.40). This contribution can be calculated using integration by parts:

$$-e \int \frac{d^d p}{(2\pi\hbar)^d} \varepsilon \mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} = e \int \frac{d^d p}{(2\pi\hbar)^d} \frac{\partial \varepsilon}{\partial \mathbf{p}} \cdot \mathbf{E} f = -\mathbf{j} \cdot \mathbf{E}. \quad (18.42)$$

It is called *Joule heating*. It describes the property that a moving electron changes its energy due to the work performed against the electric field. The above equations can be straightforwardly generalized to include the magnetic field, and the Lorentz force that acts on the charge carriers.

¹ Similar to collision with phonons, the collision of electrons among themselves bring the system into Fermi-Dirac distribution (in the absence of external forces). At low temperature, this is the dominant mechanism because the electron-electron collision rate is proportional to T (in three dimensions), while the electron-phonon rate is proportional to T^3 .

There is, however, an additional type of forces called *thermodynamic forces*. These forces originate from spatial or temporal changes in the temperature, $T(\mathbf{r}, t)$ and/or in the chemical potential $\mu(\mathbf{r}, t)$. To identify these forces, let us represent the electron distribution function as a sum of two contributions:

$$f = f_{\text{eq}}[T(\mathbf{r}, t); \mu(\mathbf{r}, t)] + \delta f, \quad (18.43)$$

where the first contribution,

$$f_{\text{eq}}[T(\mathbf{r}, t); \mu(\mathbf{r}, t)] = \frac{1}{1 + \exp\left[\frac{\varepsilon(\mathbf{p}) - \mu(\mathbf{r}, t)}{k_B T(\mathbf{r}, t)}\right]}, \quad (18.44)$$

is a distribution function describing local equilibrium (assuming the temperature and the chemical potential to change slowly in space and time), while the second contribution, δf , represents deviations from local equilibrium. Substituting Eq. (18.43) in Boltzmann's equation (18.38) and linearizing with respect to δf (focusing for simplicity on a steady-state solution), we obtain the equation:

$$\hat{L}\delta f = \frac{\partial f_{\text{eq}}}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial \mathbf{p}} \left\{ e\mathbf{E} + \nabla \mu(\mathbf{r}) + \frac{\varepsilon - \mu(\mathbf{r})}{T(\mathbf{r})} \nabla T(\mathbf{r}) \right\}, \quad (18.45)$$

where $\hat{L}$ is a linear operator representing the action of the linearized Boltzmann equation (including the collision term) on δf . The right-hand side of this equation accounts for the external forces (here, the electric field) and the thermodynamic forces (gradients of the temperature and the chemical potential) acting on the system. To derive these forces, let us first consider the action of the spatial derivative (the second term on the left-hand side of Boltzmann's equation (18.38)) on $f_{\text{eq}}[T(\mathbf{r}); \mu(\mathbf{r})]$:

$$\begin{aligned} \frac{\partial \varepsilon}{\partial \mathbf{p}} \cdot \frac{\partial}{\partial \mathbf{r}} f_{\text{eq}}[T(\mathbf{r}); \mu(\mathbf{r})] &= k_B T(\mathbf{r}) \frac{\partial f_{\text{eq}}}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial \mathbf{p}} \cdot \frac{\partial}{\partial \mathbf{r}} \left(\frac{\varepsilon - \mu(\mathbf{r})}{k_B T(\mathbf{r})} \right) \\ &= \frac{\partial \varepsilon}{\partial \mathbf{p}} \cdot \frac{\partial f_{\text{eq}}}{\partial \varepsilon} \left\{ -\nabla \mu(\mathbf{r}) - \frac{\varepsilon - \mu(\mathbf{r})}{T(\mathbf{r})} \nabla T(\mathbf{r}) \right\}. \end{aligned} \quad (18.46)$$

Next, we make use of the fact that the dependence of $f_{\text{eq}}[T(\mathbf{r}); \mu(\mathbf{r})]$ on the momentum is only through the energy dependence, $\varepsilon = \varepsilon(\mathbf{p})$, hence the force term in the Boltzmann equation (the third term on the left-hand side of (18.38)) is given by:

$$-e\mathbf{E} \cdot \frac{\partial}{\partial \mathbf{p}} f_{\text{eq}}[T(\mathbf{r}); \mu(\mathbf{r})] = -e \frac{\partial f_{\text{eq}}}{\partial \varepsilon} \mathbf{E} \cdot \frac{\partial \varepsilon}{\partial \mathbf{p}}. \quad (18.47)$$

The sum of these two contributions gives the inhomogeneous term on the right-hand side of Eq. (18.45).

To weave the connection to the second law of thermodynamics, first, notice that the contribution to the change in the entropy density from the distribution function (18.43) comes only from δf because the local equilibrium distribution, $f_{\text{eq}}[T(\mathbf{r}); \mu(\mathbf{r})]$, already maximizes the entropy. Hence Eq. (18.45) suggest that the entropy density, s , satisfies an equation of the form:

$$\frac{\partial s}{\partial t} + \nabla \cdot \mathbf{j}_Q = \mathbf{J} \cdot \mathbf{F}, \quad (18.48)$$

where $\mathbf{j}_Q$ is the heat (or entropy) current density. In the right hand side of this equation, $\mathbf{F}$ represents the external and thermodynamic forces acting on the system, while $\mathbf{J}$ stands for the current densities that couple to these forces. The left-hand side of the above equation follows from the left-hand side of Eq. (18.45). It describes reversible processes in which the entropy is conserved. The entropy change comes from the right-hand side of the equation and has the same form of Joule heating.

To identify the current $\mathbf{J}$, we begin from the thermodynamic relation,

$$TdS = dU - \mu dN, \quad (18.49)$$

where S is the total entropy, U is the total energy, and N is the number of particles. Passing from these extensive quantities to their intensive counterparts (i.e. densities) and focusing our attention on their time dependence, we obtain

$$\frac{\partial s}{\partial t} = \frac{1}{T} \frac{\partial \rho_\varepsilon}{\partial t} - \frac{\mu}{eT} \frac{\partial \rho}{\partial t}, \quad (18.50)$$

where we replaced the particle density with the charge density (divided by e). Now, let us substitute Eqs. (18.20) and (18.40) for the time derivative of the energy and the charge densities. Rearranging the terms in the resulting equation gives

$$\begin{aligned} \frac{\partial s}{\partial t} &= \frac{1}{T} (-\nabla \cdot \mathbf{j}_\varepsilon + \mathbf{j} \cdot \mathbf{E}) + \frac{\mu}{eT} \nabla \cdot \mathbf{j} \\ &= \frac{\mathbf{j} \cdot \mathbf{E}}{T} + \nabla \cdot \left(-\frac{1}{T} \mathbf{j}_\varepsilon + \frac{\mu}{eT} \mathbf{j} \right) + \mathbf{j}_\varepsilon \cdot \nabla \frac{1}{T} - \mathbf{j} \cdot \nabla \frac{\mu}{eT} \\ &= -\nabla \cdot \mathbf{j}_Q + \frac{1}{T} \left[\mathbf{j} \cdot \left(\mathbf{E} - \frac{\nabla \mu}{e} \right) + \frac{1}{T} \left(\mathbf{j}_\varepsilon - \mathbf{j} \frac{\mu}{e} \right) (-\nabla T) \right], \end{aligned} \quad (18.51)$$

where the heat current is given by

$$\mathbf{j}_Q = \frac{1}{T} \left(\mathbf{j}_\varepsilon - \frac{\mu}{e} \mathbf{j} \right). \quad (18.52)$$

This equation has the same structure as Eq. (18.48). It may be written in a form that allows one to identify the forces and the currents that couple to them:

$$\frac{\partial s}{\partial t} + \nabla \cdot \mathbf{j}_Q = \frac{1}{T} \begin{pmatrix} \mathbf{j} & \mathbf{j}_Q \end{pmatrix} \cdot \begin{pmatrix} \mathbf{E} - \frac{\nabla \mu}{e} \\ -\nabla T \end{pmatrix}. \quad (18.53)$$

The condition set by the second law of thermodynamics is that the right-hand side of Eq. (18.53) is non-negative. This condition implies that within linear response theory, the matrix that relates the currents to the forces,

$$\begin{pmatrix} \mathbf{j} \\ \mathbf{j}_Q \end{pmatrix} = \begin{pmatrix} \hat{\sigma} & \hat{\beta} \\ \hat{\gamma} & \hat{\zeta} \end{pmatrix} \begin{pmatrix} \mathbf{E} - \frac{\nabla \mu}{e} \\ -\nabla T \end{pmatrix} \quad (18.54)$$

is non-negative. The following Onsager relations set this condition:

$$\begin{aligned} \hat{\sigma}(\mathbf{B}) &= \hat{\sigma}^T(-\mathbf{B}), \\ \hat{\zeta}(\mathbf{B}) &= \hat{\zeta}^T(-\mathbf{B}), \\ \hat{\beta}(\mathbf{B}) &= \hat{\gamma}^T(-\mathbf{B}). \end{aligned} \quad (18.55)$$

These relations are written here for the case where the system is subjected to a magnetic field, $\mathbf{B}$.

Example: The I-V characteristic of a diode and Onsager's relations

In their simplest form, Onsager's relations imply that in the linear response regime of a passive system and in the absence of a magnetic field, the conductivity is independent of the current direction. This property seems to contradict the behavior of a diode that allows an electric current to flow only in one direction. In this example, we calculate the I-V characteristic of a diode and show that the magnitude of the current flowing through a PN junction is indeed independent of the direction of voltage bias, provided it is small enough.

The main assumption that we shall need to calculate the I-V curve of a diode is that the injected current is sufficiently small. Notice, at equilibrium, the current in any part of the diode vanishes because the currents of charge carriers moving in opposite directions cancel out. The typical current associated with one of these components is of the order of one ampere, while the current flow through a biased diode is typically a few milliamperes. Small injected current implies a slight voltage bias that has only a small effect on the depletion layer. Hence we may also assume that the depletion layer is free of charge carriers. In addition, we assume that the resistivity of the depletion layer is much higher than that in the other parts of the diode; hence the voltage drop occurs, essentially, only on the depletion layer.

To calculate the current in the diode, we focus our attention on the minority charge carriers in each side of the PN junction. When the diode is forward biased by a voltage V , the equilibrium drop voltage is lowered from ΔV , to $\Delta V - V$ and holes are injected into the n -type region of the diode at the point W_n , see Fig. 18-2. The holes in this region diffuse and recombine with electrons, thereby generating a current flow.

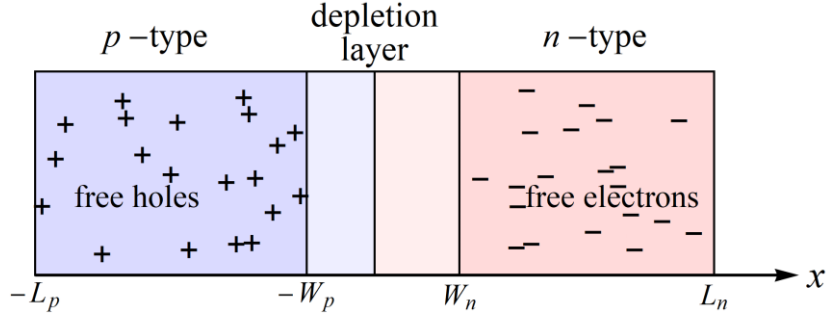


Figure 18-2 The geometry of the PN junction used for calculating the I-V characteristic of a diode

To be specific, let $\delta n_h(x)$ be the deviation of the hole density from its equilibrium value in the n -type region. The time evolution, describing diffusion and recombination process, is given by the equation:

$$\frac{\partial \delta n_h(x)}{\partial t} - D_h \frac{\partial^2 \delta n_h(x)}{\partial x^2} = -\frac{\delta n_h(x)}{\tau_{e-h}}, \quad (18.56)$$

where D_h is the hole diffusion constant, and the right-hand side of the equation accounts for recombination at rate $1/\tau_{e-h}$. This equation implies that the hole may diffuse only a distance of order $l_{h-e} = \sqrt{D_h \tau_{e-h}}$ until it recombines with an electron. Assuming the length of the n -type region to be much larger than this distance, $L_n - W_n \gg l_{h-e}$, a steady-state solution of Eq. (18.56) is:

$$\delta n_h(x) = \delta n_h(W_n) \exp\left(-\frac{x - W_n}{l_{h-e}}\right), \quad \text{for } L_n \geq x \geq W_n. \quad (18.57)$$

Here $\delta n_h(W_n)$ is the deviation of the hole density from its equilibrium value at the edge of the depletion region, $x = W_n$, where holes are injected (see Fig. 18-2). Thus, the electric current density coming from the holes is the diffusion current (recall we assume there is no electric field within the diode apart from the depletion region):

$$j_h = -D_h \left. \frac{\partial \delta n_h(x)}{\partial x} \right|_{x=W_n} = \frac{D_h}{l_{h-e}} \delta n_h(W_n) \quad (18.58)$$

A similar process takes place for electrons in the p -type region and the total current is the sum of both contributions.

Our goal now is to find the voltage dependence of the nonequilibrium density of the holes, $\delta n_h(W_n)$. For this purpose, we integrate Eq. (18.35) across the depletion layer, from W_p to W_n , taking into account that the total voltage change is $\Delta V - V$, thus

$$\frac{n_h(W_n)}{n_h(W_p)} = \exp\left(-\frac{e(\Delta V - V)}{k_B T}\right). \quad (18.56)$$

Taking the hole density at the edge of the p -type region, $n_h(W_p)$, to be approximately the density of acceptors, N_A , and using the equilibrium relation (18.37), we obtain:

$$n_h(W_n) = \frac{n_i^2}{N_D} \exp\left(\frac{eV}{k_B T}\right). \quad (18.57)$$

Subtracting the equilibrium hole density (i.e., the value of $n_h(W_n)$ at $V = 0$) we obtain

$$\delta n_h(W_n) = \frac{n_i^2}{N_D} \left[\exp\left(\frac{eV}{k_B T}\right) - 1 \right]. \quad (18.58)$$

Substituting this result in Eq. (18.58) yields the contribution of the holes to the current density:

$$j_h = \frac{D_h}{l_{h-e}} \frac{n_i^2}{N_D} \left[\exp\left(\frac{eV}{k_B T}\right) - 1 \right]. \quad (18.59)$$

A similar calculation gives the electronic contribution to the current density:

$$j_e = \frac{D_e}{l_{e-h}} \frac{n_i^2}{N_A} \left[\exp\left(\frac{eV}{k_B T}\right) - 1 \right], \quad (18.60)$$

where D_e is the electron diffusion constant and l_{e-h} is the recombination length for electrons in the p -type region of the diode.

Finally, summing the two contributions (18.59) and (18.60) and multiplying by the area of the diode cross-section, we find that the total current flowing through the diode satisfies the equation:

$$I = I_s \left[\exp\left(\frac{eV}{k_B T}\right) - 1 \right], \quad (18.61)$$

where I_s is the saturation current when the diode is reversely biased. In particular, close to equilibrium, $|eV| \ll k_B T$, this equation reduces to the familiar Ohm's law describing linear response:

$$I = \frac{eI_s}{k_B T} V. \quad (18.62)$$

This behavior manifestly satisfies Onsager's relation.

18.4 Thermoelectric effects

Thermoelectric effects involve the conversion of temperature gradients to electric current or voltage and vice versa (in metallic systems). In metals, the chemical potential is usually constant due to quasi-neutrality; hence, gradients of the chemical potential can be neglected. Furthermore, the physical quantities that can be controlled in the experimental systems are the electric current and the temperature gradient; therefore a convenient way of representing Eq. (18.54) is

$$\begin{pmatrix} \mathbf{E} \\ \mathbf{j}_Q \end{pmatrix} = \begin{pmatrix} \hat{\rho} & \hat{Q} \\ \frac{\hat{\Pi}}{T} & -\frac{\hat{\kappa}}{T} \end{pmatrix} \begin{pmatrix} \mathbf{j} \\ \nabla T \end{pmatrix}, \quad (18.63)$$

where $\hat{\rho} = \hat{\sigma}^{-1}$ is the resistivity tensor, $\hat{Q} = \hat{\sigma}^{-1} \hat{\beta}$ is *Seebeck coefficient*, $\hat{\Pi} = T \hat{\gamma} \hat{\sigma}^{-1}$ is *Peltier coefficient*, and $\hat{\kappa} = T \left(\hat{\zeta} - \hat{\gamma} \hat{\sigma}^{-1} \hat{\beta} \right)$ the *heat conductivity coefficient*. All these coefficients are, in general, tensors.

The above equations allow us to explain several thermoelectric effects. For simplicity, we consider systems where ρ , Q , Π , and κ are scalars rather than tensors.

1. Seebeck effect (1821)

Consider an open electric circuit made from two metals with different Seebeck coefficients. Let us assume that the interface between the two metals is held at a temperature that is different from the temperature at the other edges of the metals, as illustrated in Fig. 18-3. Such a device is called a *thermocouple*. As the circuit is opened, the current that flowing through the system vanishes, $\mathbf{j} = 0$, hence from Eq. (18.63), it follows that

$$\mathbf{E} = Q \nabla T \quad (18.64)$$

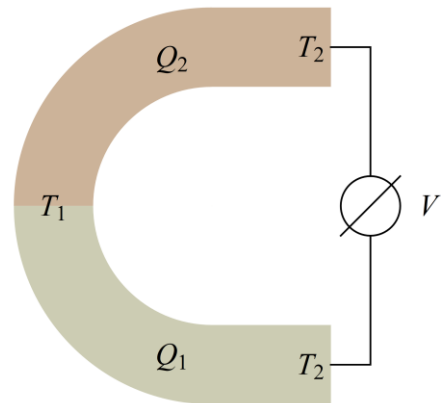


Figure 18-3 A thermocouple

Integrating this equation along a curve that starts at one edge of the thermocouple and ends at the other edge gives the voltage drop between these edges:

$$\begin{aligned} V &= \int \mathbf{E} \cdot d\mathbf{l} = \int Q \nabla T \cdot d\mathbf{l} = \int Q(T) dT \\ &= \int_{T_2}^{T_1} Q_1 dT + \int_{T_1}^{T_2} Q_2 dT = (Q_1 - Q_2)(T_1 - T_2) \end{aligned} \quad (18.65)$$

Thus, the voltage is proportional to the temperature difference and to the difference between the Seebeck coefficient of the two metals.

2. Peltier effect (1834)

Peltier effect can be regarded as the inverse of the Seebeck effect. When current flows through a thermocouple made of two metals with different Peltier coefficients, the junction can emit or absorb heat according to the current direction (see Fig. 18-4).

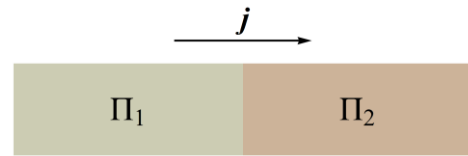


Figure 18-4 A system demonstrating the Peltier effect

Assuming the temperature is constant in the vicinity of the interface between the two metals, the heat current is proportional to the current, $\mathbf{j}_Q = \Pi \mathbf{j} / T$ (see Eq. (18.63)). Since the Peltier coefficient from both sides of the junction are different, the heat current entering (or emitted from) the junction is

$$\Delta \mathbf{j}_Q = \frac{1}{T} (\Pi_1 - \Pi_2) \mathbf{j}. \quad (18.66)$$

Thus, the electric current may heat or cool the junction depending on its direction and the values of Peltier coefficients. The relation between the Seebeck and the Peltier effects is manifested by the relation between the coefficients that characterize these effects:

$$\Pi = TQ. \quad (18.67)$$

3. Thomson effect (1951)

In general, thermoelectric coefficients are temperature dependent. In particular, in the presence of a temperature gradient, the Seebeck coefficient may vary substantially in space. If, in addition, an electric current flows in the system (in the direction or in the opposite direction to the temperature gradient), a continuous variation of the Peltier effect is obtained – this is the *Thomson effect*. To characterize it, let us write Eq. (18.53), assuming constant chemical potential, in the form

$$T \frac{\partial s}{\partial t} = \mathbf{j} \cdot \mathbf{E} - T \nabla \cdot \mathbf{j}_Q - \mathbf{j}_Q \cdot \nabla T = \mathbf{j} \cdot \mathbf{E} - \nabla \cdot (T \mathbf{j}_Q). \quad (18.68)$$

Substituting Eq. (18.63) we obtain

$$T \frac{\partial s}{\partial t} = \rho \mathbf{j}^2 + \mathbf{j} \cdot (Q \nabla T) - \nabla \cdot (Q T \mathbf{j} - \kappa \nabla T), \quad (18.69)$$

where we also used Eq.(18.67).

Next we assume that the Seebeck coefficient, Q , is temperature-dependent, and take into account charge conservation in a steady-state, $\nabla \cdot \mathbf{j} = 0$. These conditions reduce Eq. (18.69) to :

$$T \frac{\partial s}{\partial t} = \rho \mathbf{j}^2 - \mu_T \mathbf{j} \cdot \nabla T + \kappa \nabla^2 T, \quad (18.70)$$

where

$$\mu_T = T \frac{dQ}{dT} \quad (18.71)$$

is the *Thomson coefficient*. The first term on the right-hand side of Eq. (18.70) is independent of the current direction, but the second one does depend on the current direction and accordingly increases or decreases the entropy density in the system. Namely, this effect is manifested by heating or cooling of a current-carrying conductor in the presence of a temperature gradient.

18.5 Relations between the kinetic coefficients

There are relations among the kinetic coefficients presented in Eq. (18.54). This section employs the Boltzmann equation to reveal these relations in simple metallic systems. To begin with, we use Eqs. (18.39), (18.41), and (18.52) to write the electric current and the heat current as integrals over the distribution function of the electrons:

$$\begin{pmatrix} \mathbf{j} \\ \mathbf{j}_Q \end{pmatrix} = \int \frac{d^d p}{(2\pi\hbar)^d} f(\mathbf{p}) \frac{\partial \varepsilon}{\partial \mathbf{p}} \begin{pmatrix} -e \\ \frac{\varepsilon - \mu}{T} \end{pmatrix}. \quad (18.72)$$

Next, we substitute a solution of the form (18.43) with (18.44), i.e. $f = f_{\text{eq}} + \delta f$, where f_{eq} describes a local thermodynamic equilibrium (with chemical potential and temperature that change slowly in space), while δf describes deviations from local equilibrium. When time-reversal symmetry applies, f_{eq} , does not contribute to the currents in the above equation because these currents vanish at equilibrium. Thus $f(\mathbf{p})$, in the above equation, should be replaced by δf .

To simplify the problem, we assume that the electron dynamics is governed by elastic scattering; hence the operator $\hat{L}$ in Eq. (18.45) is:

$$\hat{L} = \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}} - e\mathbf{E} \cdot \frac{\partial}{\partial \mathbf{r}} + \frac{1}{\tau_{tr}}, \quad (18.73)$$

where the collision integral is approximated by the relaxation time approximation. Moreover, we focus our attention on the limit where the transport mean free time is shorter than any other relevant time scale so that $\hat{L} \simeq 1/\tau_{tr}$. With this assumption, the solution of (18.45) is

$$\delta f = \tau_{tr} \frac{\partial f_{eq}}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial \mathbf{p}} \left\{ e\mathbf{E} - \frac{\varepsilon - \mu}{T} (-\nabla T) \right\}, \quad (18.74)$$

where we also assume the chemical potential to be constant in space. Substitution (18.74) in (18.72) yields

$$\begin{pmatrix} j_i \\ j_{Qi} \end{pmatrix} = \int \frac{d^d p}{(2\pi\hbar)^d} \frac{\partial f_{eq}}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial p_j} \left\{ eE_j - \frac{\varepsilon - \mu}{T} (-\nabla T)_j \right\} \frac{\partial \varepsilon}{\partial p_i} \begin{pmatrix} -e \\ \frac{\varepsilon - \mu}{T} \end{pmatrix}, \quad (18.75)$$

and by comparing with Eq. (18.54) we can identify the kinetic coefficients as

$$\begin{pmatrix} \sigma_{ij} & \beta_{ij} \\ \gamma_{ij} & \zeta_{ij} \end{pmatrix} = \int \nu d\varepsilon \int \frac{d\Omega}{4\pi} \tau_{tr} \frac{\partial f_{eq}}{\partial \varepsilon} \left(\frac{\partial \varepsilon}{\partial \mathbf{p}} \right)_i \left(\frac{\partial \varepsilon}{\partial \mathbf{p}} \right)_j \begin{pmatrix} -e^2 & e \frac{\varepsilon - \mu}{T} \\ e \frac{\varepsilon - \mu}{T} & -\left(\frac{\varepsilon - \mu}{T} \right)^2 \end{pmatrix}. \quad (18.76)$$

Here we have replaced the integral over momentum with an integral over the energy, where ν is the density of states (see Eq. (18.31)), and $d\Omega$ is an infinitesimal element of the solid angle.

To calculate the integrals (18.76) we assume, for simplicity, that the system is isotropic, namely that the particle velocity is independent of its direction. This assumption implies that all tensors on the left-hand side of the equation are proportional to the identity matrix. Furthermore, we assume that the temperature is sufficiently low so that Fermi-Dirac distribution can be approximated by a step function and leading order corrections due to the temperature. Namely,

$$f_{eq}(\varepsilon) = \theta(\mu - \varepsilon) - \frac{\pi^2 k_B^2 T^2}{6} \delta'(\varepsilon - \mu) + \dots, \quad (18.77)$$

so that

$$\frac{\partial f_{eq}(\varepsilon)}{\partial \varepsilon} = -\delta(\varepsilon - \mu) - \frac{\pi^2 k_B^2 T^2}{6} \delta''(\varepsilon - \mu) + \dots. \quad (18.78)$$

Consider, first, the conductivity. Observe that when substituting the expansion (18.78) in Eq. (18.76), the leading contribution comes from the delta function. It yields $\sigma = e^2 \nu D$ with $D = v^2 \tau_{tr} / d$, where d is the dimensionality of the system and $v = \partial \varepsilon / \partial \mathbf{p}$ the velocity at Fermi level.

Let us now consider the integral of ζ where the integrand contains the term $(\varepsilon - \mu)^2$. Here it is clear that the contribution from the delta function in (18.78) vanishes, but the second term in this expansion gives a finite contribution. Substituting the second term of (17.78) in (18.76) and integrating twice by parts gives (see Ex. 1):

$$\zeta = \frac{\pi^2 k_B^2}{3e^2} \sigma. \quad (18.79)$$

In both calculations, for σ and ζ , we have neglected the energy dependence of the density of states. On the other hand, if we use the same assumption for calculating β and γ , we would get that these quantities vanish. Hence, in this approximation, the heat conductivity $\kappa = T(\zeta - \gamma \sigma^{-1} \beta)$ (see paragraph below Eq. (18.63)) is $T\zeta$. Thus, the heat conductivity and the electric conductivity are proportional and satisfy the relation:

$$\frac{\kappa}{T\sigma} = \frac{\pi^2 k_B^2}{3e^2}. \quad (18.80)$$

This equation is called *Wiedemann-Franz law* (1853), and the constant on the right-hand side of this equation is the *Lorenz number*.

To calculate the constant β (which equals γ in isotropic and time-reversal symmetric systems), we observe that the following term

$$\int \frac{d\Omega}{4\pi} \nu \tau_{tr} \left(\frac{\partial \varepsilon}{\partial \mathbf{p}} \right)_i \left(\frac{\partial \varepsilon}{\partial \mathbf{p}} \right)_j = \frac{\sigma(\varepsilon)}{e^2}, \quad (18.81)$$

which appears in the integral (18.76), is the conductivity at energy ε divided by the square of the electron charge. Expanding the conductivity to linear order in the deviation from the chemical potential energy,

$$\sigma(\varepsilon) \approx \sigma_0 - \frac{\partial \sigma}{\partial \mu} (\varepsilon - \mu) \quad (18.82)$$

and substituting this expansion for the term (18.81) in the integral (18.76), we see that the leading contribution comes from the second term in (18.78) (similar to the calculation of ζ). Thus, integrating twice by parts (see Ex. 1) gives

$$\beta = \frac{\pi^2 k_B^2 T}{3e} \frac{d\sigma}{d\mu} \quad (18.83)$$

Finally, using the relation $Q = \sigma^{-1} \beta$ (see text below Eq. (18.63)) gives the Seebeck coefficient:

$$Q = \frac{\pi^2 k_B^2 T}{3e} \frac{d \ln \sigma}{d \mu}. \quad (18.84)$$

This relation between the conductivity and Seebeck coefficient is called the *Mott-Cutler formula* (1969).

=

18.6 Exercises

1. Prove formulas (18.79) and (18.84).