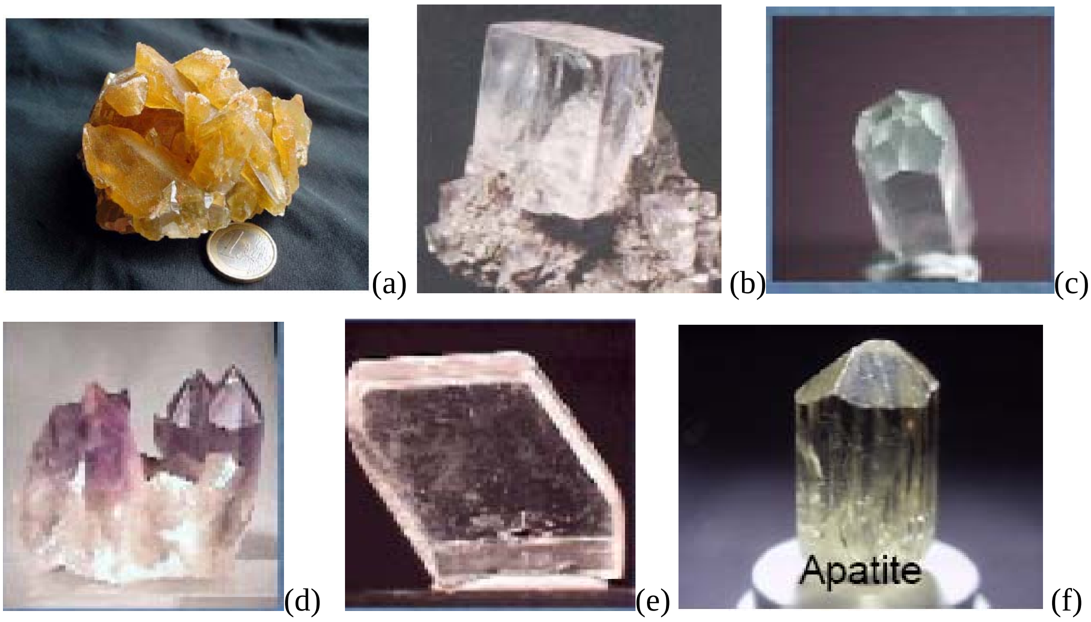
# SOLID STATE PHYSICS

By definition, solid state is that particular aggregation form of matter characterized by strong interaction forces between constituent particles (atoms, ions, or molecules). As a result, a solid state material has an independent geometric form (in contrast to liquids, which take the form of the container) and an invariant volume (in contrast to gases/vapors) in given temperature and pressure conditions. As temperature increases, a solid state material can evolve into another aggregation form (liquid or gas). Solid state physics studies the structural, mechanical, thermodynamic, electrical, magnetic, and optical properties of (poly-)crystalline and non-crystalline solids (for example, amorphous materials, such as glass).

## Crystal structure

The properties of crystalline solids are determined by the symmetry of the crystalline lattice, because both electronic and phononic systems, which determine, respectively, the electric/ magnetic and thermal response of solids, are very sensitive to the regular atomic order of materials and to any (local or non-local) perturbation of it. The crystalline structure can be revealed by the macroscopic form of natural or artificially-grown crystals (see the pictures below), or can be inferred from the resulting debris after cleaving a crystalline material.

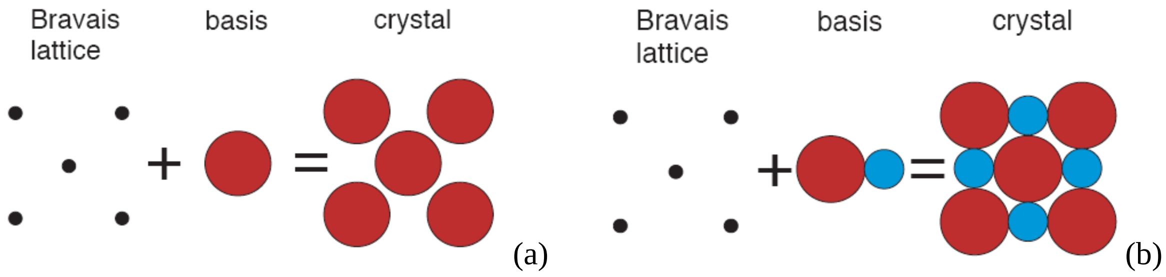


(a) (b) (c) (d) (e) (f)

Crystals of (a) baryt, (b) salt, (c) hexagonal beryl, (d) trigonal quartz, (e) monoclinic gypsum, and apatite (f)

Non-crystalline materials have no long-range order, but at least their optical properties are similar to that of crystalline materials because the wavelength of the incident photons (of the order of 1 μm) is much larger than the lattice constant of crystals and so, photons “see” an effective homogeneous medium. Other properties of non-crystalline materials are derived based on concepts proper to crystalline solids and, therefore, the crystal structure is extremely important in understanding the properties of solid state materials.

The macroscopic, perfect crystal is formed by adding identical building blocks (unit cells) consisting of atoms or groups of atoms. A unit cell is the smallest component of the crystal that, when stacked together with pure translational repetition, reproduces the whole crystal. The periodicity of the crystalline structure that results in this way is confirmed by X-ray diffraction experiments. The figures below illustrate crystals in which the basis consists of (a) one atom and (b) two atoms.



The group of atoms or molecules that forms, by infinite repetition, the macroscopic crystal is called **basis**. The basis is positioned in a set of mathematical/abstract points that form the **lattice** (also called **Bravais lattice**). So, a crystal is a combination of a basis and a lattice. Although usually the basis consists of only few atoms, it can also contain complex organic or inorganic molecules (for example, proteins) of hundreds and even thousands of atoms.

In two dimensions, all Bravais lattice points

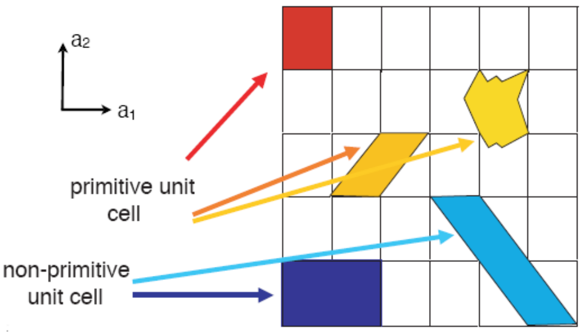
$$\boldsymbol{R}_{mn} = m\boldsymbol{a}_1 + n\boldsymbol{a}_2 \tag{1}$$

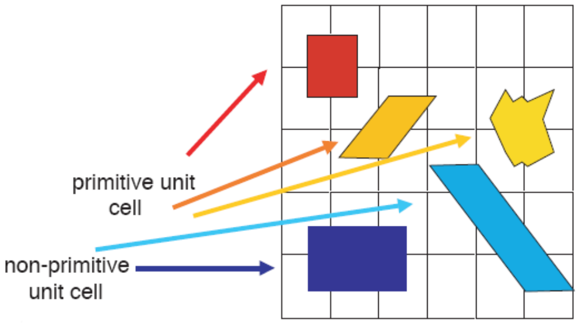
can be obtained as superpositions of integral multiples of two non-collinear vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ ($m$ and $n$ are arbitrary integers). A basis consisting of $s$ atoms is then defined by the set of vectors $\boldsymbol{r}_j = m_j\boldsymbol{a}_1 + n_j\boldsymbol{a}_2$, $j$ = 1,2,…,$s$, that describe the position of the centers of the basis atoms with respect to one point of the Bravais lattice. In general, $0 \le m_j, n_j \le 1$.

Every point of a Bravais lattice is equivalent to every other point, i.e. the arrangement of atoms in the crystal is the same when viewed from different lattice points. The Bravais lattice defined by (1) is invariant under the operation of discrete translation $\boldsymbol{T}_{pq} = p\boldsymbol{a}_1 + q\boldsymbol{a}_2$ along integer multiples $p$ and $q$ of vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, respectively, because

$$\boldsymbol{T}_{pq}(\boldsymbol{R}_{mn}) = \boldsymbol{T}_{pq} + \boldsymbol{R}_{mn} = \boldsymbol{R}_{p+m,q+n} \tag{2}$$

is again a Bravais lattice point. In fact, since the translation operation is additive, i.e. $\boldsymbol{T}_{pq}\boldsymbol{T}_{uv} = \boldsymbol{T}_{p+u,q+v}$, associative, i.e. $\boldsymbol{T}_{pq}(\boldsymbol{T}_{uv}\boldsymbol{T}_{mn}) = (\boldsymbol{T}_{pq}\boldsymbol{T}_{uv})\boldsymbol{T}_{mn}$, commutative, i.e. $\boldsymbol{T}_{pq}\boldsymbol{T}_{uv} = \boldsymbol{T}_{uv}\boldsymbol{T}_{pq}$, and has an inverse, such that $\boldsymbol{T}_{pq}^{-1} = \boldsymbol{T}_{-p,-q}$ and $\boldsymbol{T}_{pq}\boldsymbol{T}_{-p,-q} = \boldsymbol{I}$ with $\boldsymbol{I}$ the identity transformation, it follows that the translations form an abelian (commutative) group. Because condition (2) is satisfied for all Bravais lattice points, $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are called primitive translation vectors, and the unit cell determined by them is called **primitive unit cell**. The modulus of these vectors, $a_1 = |\boldsymbol{a}_1|$ and $a_2 = |\boldsymbol{a}_2|$, are the lattice constants along the respective axes, and the area of the unit cell in two dimensions is $S = |\boldsymbol{a}_1 \times \boldsymbol{a}_2|$. It is important to notice that the set of vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ is not unique (see the figures below), but all primitive unit cells have the same area.

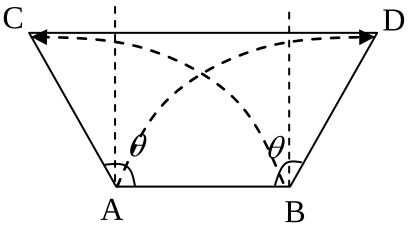




The primitive unit cell covers the whole lattice once, without overlap and without leaving voids, if translated by all lattice vectors. An equivalent definition of the primitive unit cell is a cell with one lattice point per cell (each lattice point in the figures above belong to four adjacent primitive unit cells, so that each primitive unit cell contains 4×(1/4) = 1 lattice point). Non-primitive (or conventional) unit cells are larger than the primitive unit cells, but are sometimes useful since they can exhibit more clearly the symmetry of the Bravais lattice.

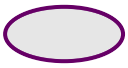
Besides discrete translations, the Bravais lattice is invariant also to the point group operations, which are applied around a point of the lattice that remains unchanged. These operations are:

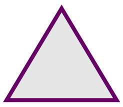
- **Rotations** by an angle $2\pi/n$ about a specific axis, denoted by $C_n$, and its multiples, $C_n^j = (C_n)^j$. Geometric considerations impose that $n$ = 1, 2, 3, 4 and 6, and that repeating the rotation $n$ times one obtains $C_n^n = E$, where $E$ is the identity operation, which acts as $\boldsymbol{r} \to \boldsymbol{r}$. Moreover, $C_1 = 2\pi = E$ does not represent a symmetry element.

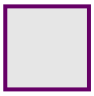


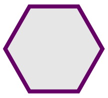
The allowed values of $n$ can be determined assuming that we apply a rotation with an angle $\theta$ around an axis that passes first through a point A and then through an adjacent lattice point B. The points A and B are separated by the lattice constant $a$. If C and D are the resulting points, they should also be separated by an integer multiple of $a$. From the requirement that CD = $a + 2a\sin(\theta - \pi/2) = a - 2a\cos\theta$ = $ma$, or $-1 \le \cos\theta = (1-m)/2 \le 1$, with $m$ integer, it follows that $m$ can only take the values −1, 0, 1, 2, and 3, the corresponding $n = 2\pi/\theta$ taking the values specified above. As for translations, the rotations also form an abelian group.

Examples of two-dimensional figures with different rotation symmetries:


**$C_2$**

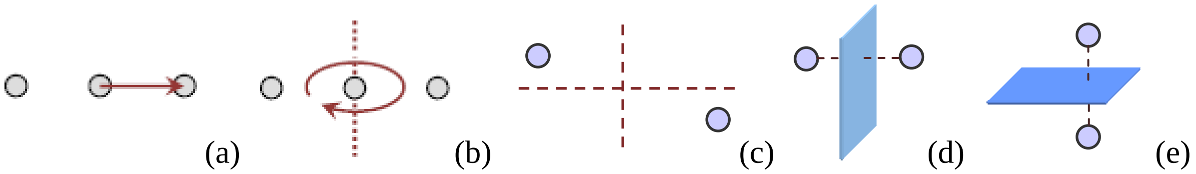

**$C_3$**


**$C_4$**


**$C_6$**

- **Inversion** $I$, which is defined by the operation $\boldsymbol{r} \to -\boldsymbol{r}$ if applied around the origin.
- **Reflection** $\sigma_j$, which can be applied around the horizontal plane ($j = h$), the vertical plane ($j = v$), or the diagonal plane ($j = d$).
- **Improper rotation** $S_n$, which consists of the rotation $C_n$ followed by reflection in the plane normal to the rotation axis. Note that $S_2 \equiv I$.

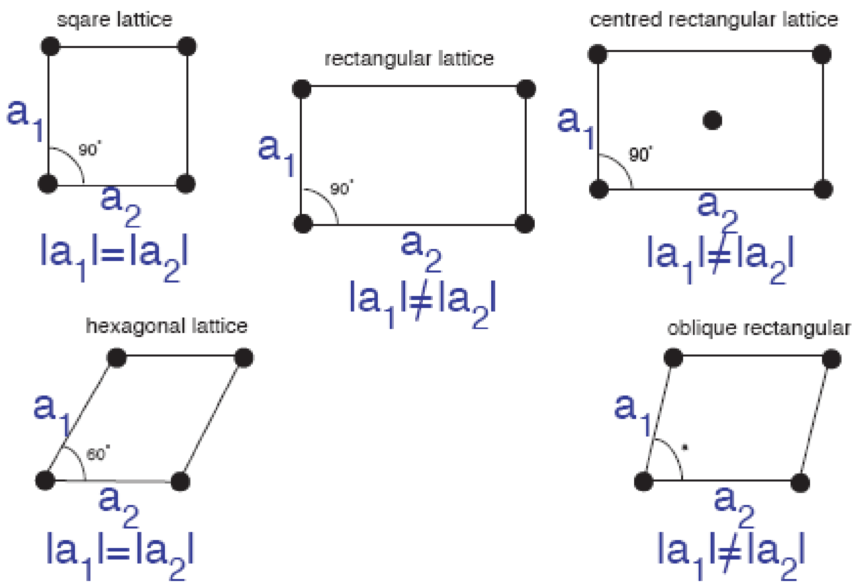
When we combine the point group symmetry with the translational symmetry, we obtain the space-group symmetry. It is important to notice that the basis can introduce additional symmetry elements, such as helicoidal symmetry axes and gliding reflection planes. The figure bellow represents several symmetry operations: (a) translations, (b) rotation, (c) inversion, and reflection with respect to a (d) vertical, and (e) horizontal plane.

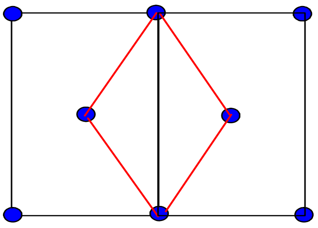


Crystal lattices are classified according to their symmetry properties at point group operations. The five Bravais lattice types in two dimensions are shown in the figure below.

These are:

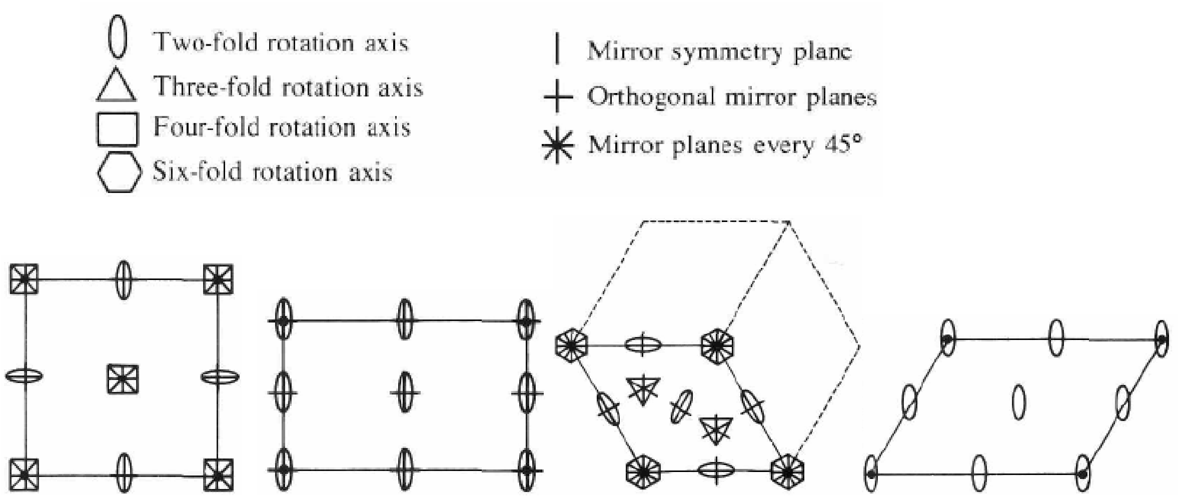
- **square lattice**, for which $|\boldsymbol{a}_1| = |\boldsymbol{a}_2|$, and $\gamma = 90°$, where $\gamma$ is the angle between $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$,
- **rectangular lattice**, for which $|\boldsymbol{a}_1| \neq |\boldsymbol{a}_2|$, and $\gamma = 90°$,
- **centered rectangular lattice**, which is a rectangular lattice with an additional lattice point in the center of the rectangle,
- **hexagonal lattice**, for which $|\boldsymbol{a}_1| = |\boldsymbol{a}_2|$, and $\gamma = 60°$ (or 120° for a different choice of the origin),
- **oblique rectangular lattice** (called also oblique lattice), for which $|\boldsymbol{a}_1| \neq |\boldsymbol{a}_2|$, and $\gamma \neq$ 90°, 60° (or 120°).



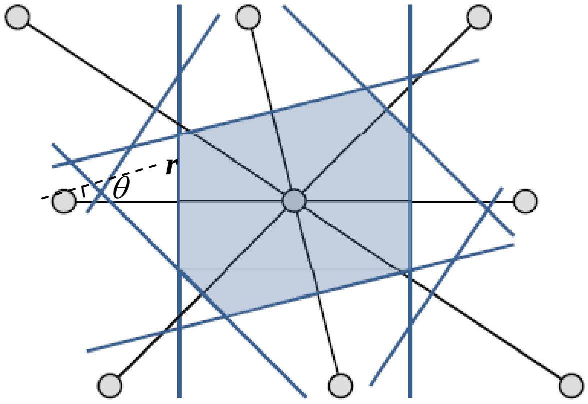


With the exception of the centered rectangular lattice, all unit cells in the figure above are primitive unit cells. The primitive cell for the centered rectangular lattice is a rhombus (see figure at right) and therefore this Bravais lattice is also called rhombic lattice, case in which its primitive unit cell has $|\boldsymbol{a}_1| = |\boldsymbol{a}_2|$, and $\gamma \neq$ 90°, 60° (or 120°).

Each lattice type has a different set of symmetry operations. For all Bravais lattice types in two dimensions, the rotation axes and/or reflection planes occur at lattice points. There are also other locations in the unit cell with comparable or lower degrees of symmetry with respect to rotation and reflection. These locations are indicated in the figure below.



In order to incorporate the information about the point group symmetry in the primitive cell, the **Wigner-Seitz cell** is usually employed. This particular primitive unit cell is constructed by first drawing lines to connect a given lattice point to all nearby lattice points, and then drawing new lines (or planes, in three-dimensional lattices) at the mid point and normal to the first lines. The Wigner-Seitz cell is the smallest area (volume) enclosed by the latter lines (planes). An example of the construction of a Wigner-Seitz cell for a two-dimensional oblique lattice is illustrated in the figure below. For a two-dimensional square lattice the Wigner-Seitz cell is also a square. The Wigner-Seitz cell is always centered on a lattice point and incorporates the volume of space which is closest to that lattice point rather than to any other point.



The faces of the Wigner-Seitz cell satisfy the relation $r\cos\theta = R/2$, where $R$ is the distance to the nearest neighbor and $\theta$ is the angle between $\boldsymbol{r}$ and $\boldsymbol{R}$. This relation can be rewritten as $2(\boldsymbol{r}\cdot\boldsymbol{R}) = \boldsymbol{R}^2$ or, since the equation is equivalent to the replacement of $\boldsymbol{R}$ with $-\boldsymbol{R}$, $2\boldsymbol{r}\cdot\boldsymbol{R}+\boldsymbol{R}^2 = 0$, and finally, $(\boldsymbol{r}+\boldsymbol{R})^2 = \boldsymbol{r}^2$. In other words, the faces of the Wigner-Seitz cell are determined by the intersection between equal-radius spheres centered at the nearest-neighbor points of the Bravais lattice.

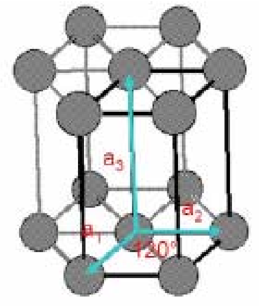
In a similar manner, in three dimensions, all Bravais lattice points

$$\boldsymbol{R}_{mnp} = m\boldsymbol{a}_1 + n\boldsymbol{a}_2 + p\boldsymbol{a}_3 \tag{3}$$

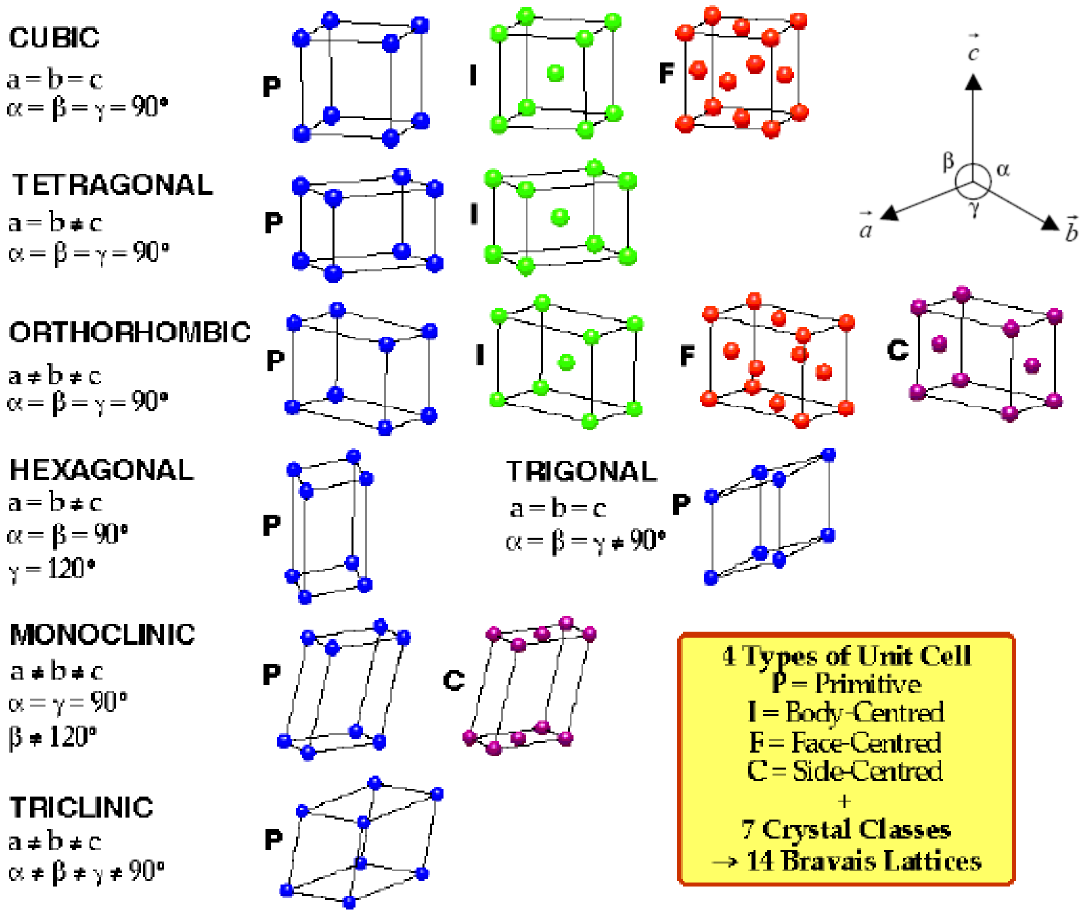
can be obtained as superpositions of integral multiples of three non-coplanar primitive translation vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ ($m$, $n$, and $p$ are arbitrary integers), and the point group operations are defined identically. The volume of the primitive unit cell, which in this case is a parallelepiped, is $\Omega = |(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{a}_3|$.

There are 14 three-dimensional Bravais lattices, which belong to 7 crystal systems, as can be seen from the figure below, where the primitive translation vectors are denoted by $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ (with respective lengths $a$, $b$, and $c$), and $\alpha$, $\beta$, $\gamma$ are the angles between $\boldsymbol{b}$ and $\boldsymbol{c}$, $\boldsymbol{c}$ and $\boldsymbol{a}$, and $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. These crystal systems, which are different point groups endowed with a spherical symmetric basis, are:

- **cubic**, for which $a = b = c$, $\alpha = \beta = \gamma = 90°$. It consists of three non-equivalent space-group lattices: simple cubic, body-centered cubic, and face-centered cubic. This is the crystal system with the highest symmetry and is characterized by the presence of four $C_3$ axes (the diagonals of the cube)
- **tetragonal**, for which $a = b \neq c$, $\alpha = \beta = \gamma = 90°$. It encompasses the simple and body-centered Bravais lattices and contains one $C_4$ symmetry axis.
- **orthorhombic**, for which $a \neq b \neq c$, $\alpha = \beta = \gamma = 90°$. It incorporates the simple, body-centered, face-centered, and side-centered lattices and has more than one $C_2$ symmetry axis or more than one reflection plane (actually, three such axes/planes, perpendicular to each other).
- **hexagonal**, for which $a = b \neq c$, $\alpha = \beta = 90°$, $\gamma = 120°$. It is characterized by the existence of a single $C_6$ symmetry axis. The conventional hexagonal unit cell (see the figure at right) is composed of three primitive cells.

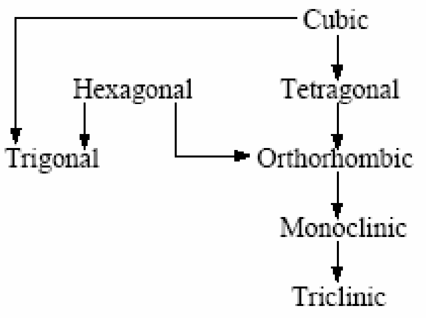


- **trigonal**, for which $a = b = c$, $\alpha = \beta = \gamma \neq 90°$. It contains a single $C_3$ axis.
- **monoclinic**, for which $a \neq b \neq c$, $\alpha = \gamma = 90° \neq \beta$. It includes the simple and side-centered lattices, and has one $C_2$ symmetry axis and/or one reflection plane perpendicular to this axis.
- **triclinic**, for which $a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90°$. This is the crystal system with the lowest symmetry. It is not symmetric with respect to any rotation axis or reflection plane.

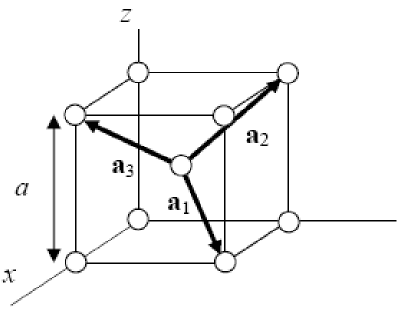


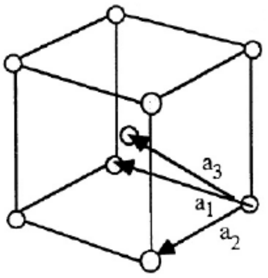
The relations between these lattices can be summarized in the figure at the right.

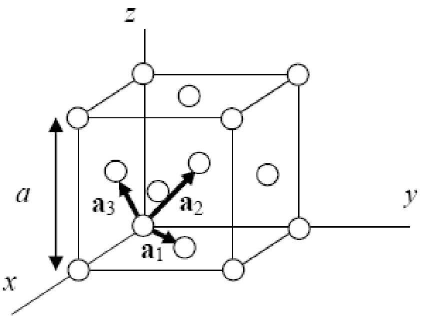
The different crystal systems have different numbers of unit cell types because other possible unit cell types cannot represent new Bravais lattices. For example, both the body-centered and the face-centered monoclinic lattices can be reduced to the side-centered lattice by appropriately choosing the primitive translation vectors.



Examples of two sets of primitive translation vectors for a body-centered cubic (bcc) lattice are represented in the figure below at left and center, while the figure at right displays a set of primitive translation vectors for a face-centered cubic (fcc) lattice.







The primitive translation vectors for the left figure above can be expressed as

$$\boldsymbol{a}_1 = (a/2)(\boldsymbol{x} + \boldsymbol{y} - \boldsymbol{z}),\ \boldsymbol{a}_2 = (a/2)(-\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}),\ \boldsymbol{a}_3 = (a/2)(\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{z}), \tag{4}$$

while those for the right figure are

$$\boldsymbol{a}_1 = (a/2)(\boldsymbol{x} + \boldsymbol{y}),\ \boldsymbol{a}_2 = (a/2)(\boldsymbol{y} + \boldsymbol{z}),\ \boldsymbol{a}_3 = (a/2)(\boldsymbol{z} + \boldsymbol{x}) \tag{5}$$

and the angles between these vectors are 60°.

A simple lattice has lattice points only at the corners, a body-centered lattice has one additional point at the center of the cell, a face-centered lattice has six additional points, one on each side, and a side-centered lattice has two additional points, on two opposite sides. The simple lattices are also primitive lattices and have one lattice point per cell, since the eight sites at the corners are shared by eight adjacent unit cells, so that 8×(1/8) = 1. The non-simple lattices are non-primitive. The volume of the primitive unit cell in these lattices is obtained by dividing the volume of the conventional unit cell by the number of lattice points. In particular, the body-centered lattices have two points per unit cell: the eight at the corners which contribute with 8×(1/8) = 1, and the one in the center, which belongs entirely to the unit cell. The face-centered lattices have 4 lattice points per cell: those in the corners contribute with 8×(1/8) = 1, and those on the faces contribute with 6×(1/2) = 3, since they are shared by two adjacent cells. Finally, the side-centered lattices have two lattice points per cell: the points at the corner contribute with 8×(1/8) = 1, and those on the faces with 2×(1/2) = 1. The characteristics of the cubic lattices with side *a* are summarized in the table below. If each lattice point is expanded into a sphere with a radius equal to half of the distance between nearest neighbors, such that adjacent spheres touch each other, then a packing fraction can be defined as the fraction between the volume of the spheres contained in the conventional unit cell and the volume of the unit cell. Note that in the volume between the spheres one can always insert smaller spheres, which can stand for other atom types.

| | Simple | Body-centered | Face-centered |
|---|---|---|---|
| **Volume of conventional cell** | $a^3$ | $a^3$ | $a^3$ |
| **Lattice points per cell** | 1 | 2 | 4 |
| **Volume of primitive cell** | $a^3$ | $a^3/2$ | $a^3/4$ |
| **Number of nearest neighbors** | 6 | 8 | 12 |
| **Nearest-neighbor distance** | $a$ | $\sqrt{3}a/2$ | $a/\sqrt{2}$ |
| **Number of second neighbors** | 12 | 6 | 6 |
| **Second-neighbor distance** | $\sqrt{2}a$ | $a$ | $a$ |
| **Packing fraction** | $\pi/6 = 0.524$ | $\sqrt{3}\pi/8 = 0.68$ | $\sqrt{2}\pi/6 = 0.74$ |

The 14 Bravais lattices incorporate all possible crystalline structures; they result by taking into consideration the space-group symmetry, i.e. the symmetry at translations and the point group symmetry of the lattice (the symmetry with respect to rotation, reflexion or inversion). When the basis consists of only one atom, the Bravais lattice is identical with the crystalline structure. But when the basis is complex and consists of several atoms, say $s$, the crystalline structure can be seen as formed by the interpenetration of $s$ Bravais lattices. The Bravais lattices have always an inversion center in one of the lattice points, whereas such an inversion center can lack in crystals with complex bases.

By counting the point groups of the possible different crystals (which have bases with different symmetries), one ends with 32 crystalline classes that can be accommodated by the 7 crystal systems. Also, there are 230 space groups that result from the combination of the 32 crystalline structures with the translational symmetry.

## Index system for lattice points, directions and planes

When the origin of the primitive translation vectors is a lattice point, another **lattice point** with a position $\boldsymbol{R}_{mnp} = m\boldsymbol{a}_1 + n\boldsymbol{a}_2 + p\boldsymbol{a}_3$ is simply specified by the set of numbers [[*m*,*n*,*p*]]. A negative integer *m*, *n* or *p* is denoted by a – sign placed on top of it. For example, [[$m\bar{n}p$]] stays for the lattice point specified by the integers *m*, –*n* and *p*, with *m*, *n* and *p* positive numbers. In particular, for the three-dimensional primitive Bravais lattices the coordinates of the lattice point at the origin are [[0,0,0]], the other lattice points differing only through discrete translations along the three coordinate axis. The number of non-equivalent lattice points in a Bravais lattice is given by the number of lattice points per unit cell. In particular, for the body-centered lattice, the position of the lattice point at the center of the cube is denoted by [[1/2,1/2,1/2]], the three additional lattice points in face-centered lattices having coordinates [[0,1/2,1/2]], [[1/2,0,1/2]], [[1/2,1/2, 0]]. In a similar manner, depending on the set of opposite sites they can occupy, the additional site in a face-centered lattice has the coordinates [[0,1/2,1/2]], [[1/2,0,1/2]] or [[1/2,1/2,0]].

A **direction**, by definition, passes through two lattice points. To specify a direction in a crystalline lattice, one uses the symbol [*mnp*], where *m*, *n* and *p* are three integers determined by the following rule: since one can specify a direction by the coordinates [[$m_1, n_1, p_1$]] and [[$m_2, n_2, p_2$]] of two points through which it passes, the indices *m*, *n* and *p* are defined as the smallest integer numbers that satisfy the proportionality relations

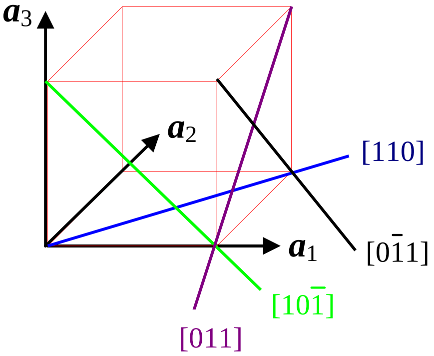
$$\frac{m}{n} = \frac{m_2 - m_1}{n_2 - n_1}, \qquad \frac{n}{p} = \frac{n_2 - n_1}{p_2 - p_1}, \qquad \frac{p}{m} = \frac{p_2 - p_1}{m_2 - m_1}, \tag{6}$$

or

$$m : n : p = (m_2 - m_1) : (n_2 - n_1) : (p_2 - p_1)\,. \tag{7}$$

If one of the integers is negative, the – sign is placed on top of the integer. For example, [$m\bar{n}p$] stays for the direction specified by the integers *m*, –*n* and *p*. If the direction is not considered as an oriented axis but as a simple line, the direction specified by the integers *m*, *n*, and *p* is the same as that specified by –*m*, –*n*, and –*p* (otherwise, the change of all signs means a change of direction of the same line). If there are several **equivalent directions** (equivalent, from the point of view of crystal symmetry), they are denoted as $\langle mnp \rangle$. A particular situation is encountered in the hexagonal lattice, in which lattice directions are labeled by four numbers (this situation is not further discussed in this course).

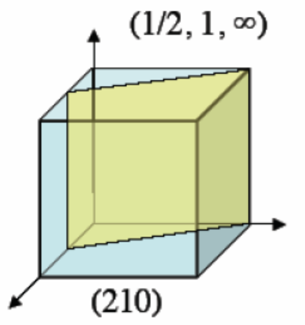
**Examples**: The $\mathbf{a}_1$ axis is the [100] direction. The $-\mathbf{a}_2$ axis is the $[0\bar{1}0]$ direction. Other examples are illustrated in the figure below.

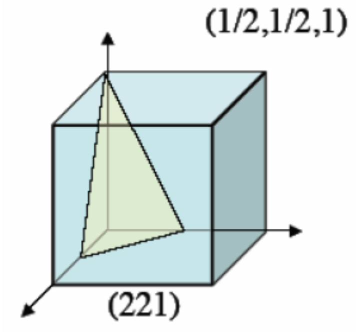


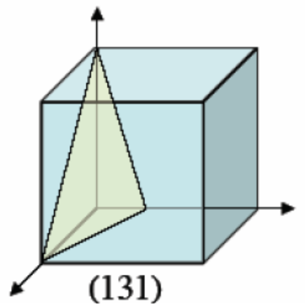
In three-dimensional lattices, the orientation of a **crystal plane** is determined by three non-collinear points in the plane. If each point is situated on a different crystal axis, the plane is specified by the coordinates of the points in terms of the lattice constants $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. Another way to specify the orientation of a plane, which is more useful for structure analysis, involves the determination of three indices, called Miller indices, according to the rule:

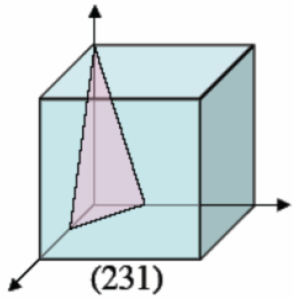
- Find first the intercepts of the plane on the axes in terms of lattice constants $a_1$, $a_2$, and $a_3$, irrespective of the nature (primitive or non-primitive) of the unit cell.
- Take the reciprocal of these numbers.
- If fractional, reduce these numbers to the smallest three integers, say *m*, *n*, *p*, with the same ratio. The result, symbolized by (*mnp*) (or $(m\bar{n}p)$ if the second index, for example, is negative), is the Miller index system of the plane.

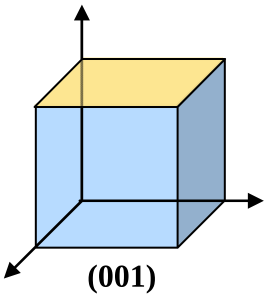
It is obvious that the Miller index for an intercept at infinity is zero. The faces of a cubic crystal, for example, are denoted by (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, and $(00\bar{1})$. Moreover, the plane (200) is parallel to (100), but cuts the $\mathbf{a}_1$ axis at $a/2$. If, from the point of view of crystal symmetry, there is a set of nonparallel **equivalent planes**, they are symbolized as {*mnp*}. For example, the set of faces of a cubic crystal is {100}. Again, for the hexagonal lattice there are four Miller indices instead of three. Examples of Miller indices are given in the figures below.

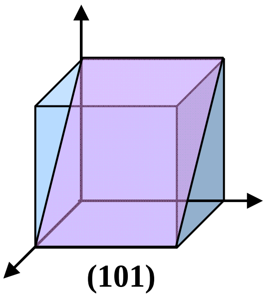


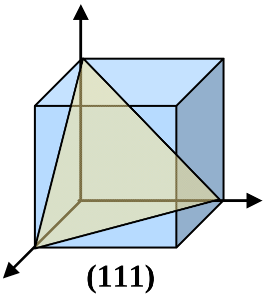








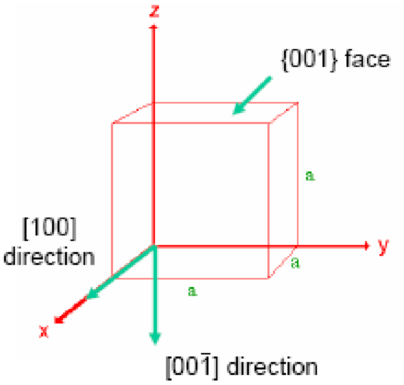




Note that the Miller indices determine not only one plane but a **family of parallel planes**, since there is an infinite number of planes with the same indices, all of which cut the coordinate axes at $s/m$, $s/n$, and $s/p$, with *s* integer. The plane that cuts the axes at $1/m$, $1/n$, and $1/p$ is the closest to the origin from the family of parallel planes.

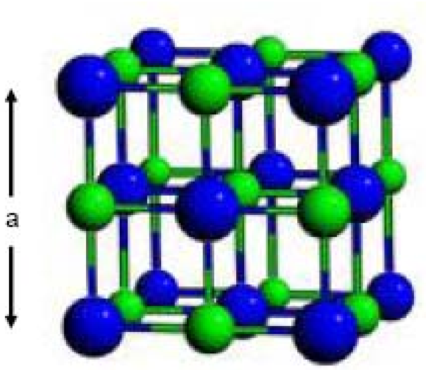
Note also that the planes with Miller indices (*sm*,*sn*,*sp*) are parallel with the plane (*mnp*), but the distance between them is *s* times smaller. For example, the set of planes (222) is parallel to but twice as close as the (111) set of planes.

In cubic crystals, the plane (*mnp*) is perpendicular to the direction [*mnp*] with the same indices, but this result cannot be extended to other crystal systems. An example is given in the figure below.



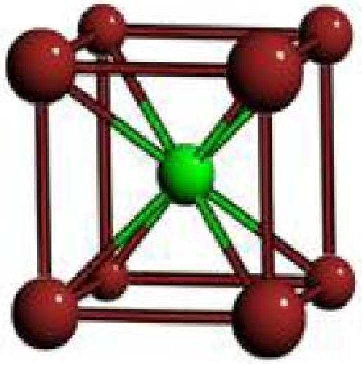
# Simple crystal structures

One of the most simple crystal structures and, at the same time, of general interest, is that of NaCl (sodium chloride). It is illustrated below. The lattice is face-centered cubic, with a basis consisting of one $Cl^-$ ion (blue) at [[000]] and a $Na^+$ ion (green) at [[1/2,1/2,1/2]]. As can be seen from the figure below, a unit cube consists of four NaCl units, with $Na^+$ ions at positions [[1/2,1/2,1/2]], [[0,0,1/2]], [[0,1/2,0]], and [[1/2,0,0]] and $Cl^-$ ions at [[000]], [[1/2,1/2,0]], [[1/2,0,1/2]], and [[0,1/2,1/2]]. Each atom has as nearest neighbors six atoms of opposite kind. Example of crystals with this structure and their lattice constants are given below.



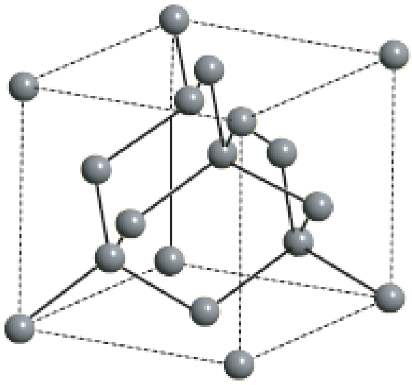
| Crystal | *a*(Å) | Crystal | *a* (Å) | Crystal | *a* (Å) |
|---|---|---|---|---|---|
| LiF | 4.02 | KBr | 6.60 | MgO | 4.21 |
| LiBr | 5.50 | AgBr | 5.77 | MnO | 4.43 |
| NaCl | 5.64 | AgF | 4.92 | MgS | 5.20 |
| NaI | 6.47 | CaSe | 5.91 | PbS | 5.92 |
| KCl | 6.29 | BaO | 5.52 | SrTe | 6.47 |

Another common structure is that of CsCl (other crystals with the same structure are given in the table below). The lattice is in this case simple cubic, with a basis consisting of one $Cs^+$ ion (red) at [[000]], and one $Cl^-$ ion (green) at [[1/2,1/2,1/2]]. The number of nearest neighbors (of opposite kind) is eight.



| Crystal | *a* (Å) | Crystal | *a* (Å) | Crystal | *a* (Å) |
|---|---|---|---|---|---|
| AlNi | 2.88 | CsCl | 4.12 | TlCl | 3.83 |
| CuZn ($\beta$-brass) | 2.94 | CsBr | 4.29 | TlBr | 3.97 |
| AgMg | 3.28 | CsI | 4.57 | TlI | 4.20 |

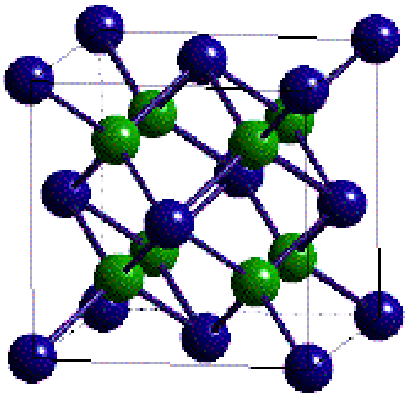
The crystal structure of diamond (and also of Si and Ge semiconductors) is represented below.



| Crystal | *a* (Å) |
|---|---|
| C (diamond) | 3.57 |
| Si | 5.43 |
| Ge | 5.66 |
| $\alpha$-Sn (grey) | 6.49 |

It is a face-centered cubic (fcc) lattice with a basis consisting of two identical atoms, with coordinates [[000]] and [[1/4,1/4,1/4]]. Alternatively, diamond can be viewed as being formed from two interpenetrating fcc lattices, displaced by 1/4 of the volume diagonal. Since the conventional unit cell of the fcc lattice contains 4 lattice points, it follows that the conventional unit cell of diamond has 2×4 = 8 atoms. No primitive cell exists that contains only one atom. In diamond, each atom has 4 nearest neighbors and 12 next nearest neighbors. It is usually encountered in materials where the covalent bonding prevails. Note that, although a fcc lattice, the packing fraction of the diamond structure is only 0.34.

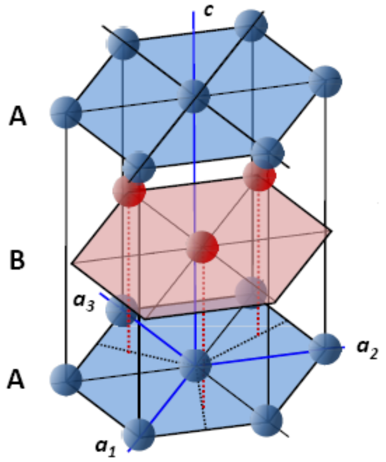
A closely related crystal structure to that of the diamond is the cubic zinc sulfide (zinc blende structure). It differs from diamond in that the two atoms of the basis are different (in this case, Zn and S). The conventional unit cell contains four molecules, the Zn atoms (dark blue in the figure below) being placed at the positions [[000]], [[0,1/2,1/2]], [[1/2,0,1/2]] and [[1/2,1/2,0]], whereas the S atoms (green) occupy the positions [[1/4,1/4,1/4]], [[1/4,3/4,3/4]], [[3/4,1/4,3/4]], and [[3/4,3/4,1/4]]. Each atom is surrounded by four equally distant atoms of the opposite kind, placed in the corners of a regular tetrahedron.



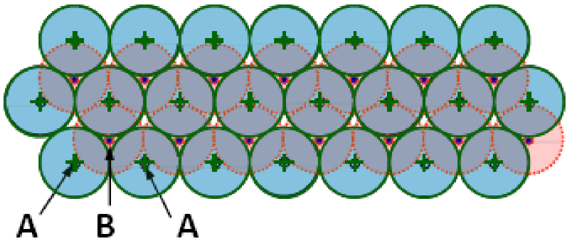
| Crystal | *a* (Å) | Crystal | *a* (Å) | Crystal | *a* (Å) |
|---|---|---|---|---|---|
| SiC | 4.35 | AlP | 5.45 | InAs | 6.04 |
| ZnS | 5.41 | AlAs | 5.66 | InSb | 6.48 |
| ZnSe | 5.67 | GaAs | 5.65 | SiC | 4.35 |
| MnS (red) | 5.60 | GaSb | 6.12 | CuCl | 5.41 |
| CdS | 5.82 | GaP | 5.45 | CuBr | 5.69 |
| CdTe | 6.48 | AgI | 6.47 | HgSe | 6.08 |

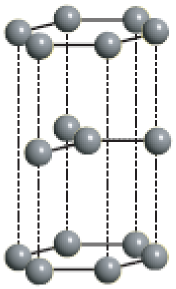
Unlike in the diamond structure, where there is a center of inversion at the midpoint of every line between nearest-neighbor atoms, such inversion centers are absent in the zinc blende structure. This is an example of additional symmetry operations related to the basis of the crystal structure.

The hexagonal close-packed (hcp) crystal structure can be obtained from the hexagonal Bravais lattice if the basis consists of two atoms (blue and red in the figure below, left) and if the atoms in one plane, which touch each other, also touch the atoms in adjacent planes. The packing fraction in this case is 0.74 (as in fcc lattices), and is maximum. This crystal structure is found in the solid state of many elements, as can be seen from the table below. The hcp structure can be viewed as vertical arrangement of two-dimensional hexagonal structures, such as the spherical atoms in the second layer are placed in the depressions left in the center of every other triangle formed by the centers of the spherical atoms in the first layer. The third layer of atoms is then placed exactly above the first, the fourth above the second, and so on. This kind of arrangement is called ABAB… In an ideal hcp structure, the height between the first and the third layers (the height along the *c* axis in the figure below) is $c=\sqrt{8/3}a=1.63a$. Because the symmetry of the hcp lattice is independent of the ratio $c/a$, in real hcp structures this ratio can take values close to, but not exactly identical to the ideal 1.63 value (see the table below).



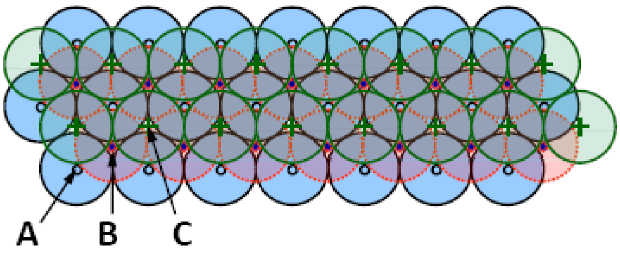
| Crystal | *a* (Å) | *c*/*a* | Crystal | *a* (Å) | *c*/*a* |
|---|---|---|---|---|---|
| He | 3.57 | 1.63 | Mg | 3.21 | 1.62 |
| Be | 2.29 | 1.58 | Ti | 2.95 | 1.58 |
| Nd | 3.66 | 1.61 | Zr | 3.23 | 1.59 |
| Zn | 2.66 | 1.86 | Y | 3.65 | 1.57 |
| Cd | 2.98 | 1.88 | Gd | 3.64 | 1.59 |
| α-Co | 2.61 | 1.62 | Lu | 3.50 | 1.58 |

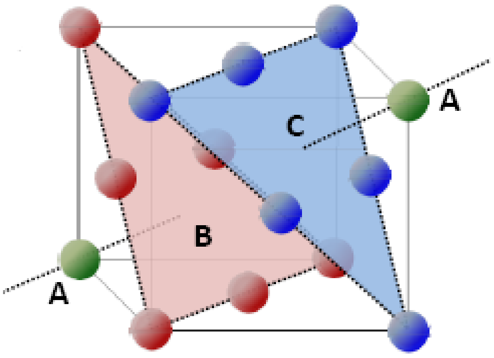




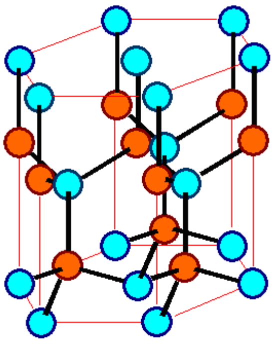
If the *c*/*a* ratio differs considerably from the ideal 1.63 value, the hexagonal structure is no longer close-packed. This is the case of graphite, for example, which is a non-closed-packed hexagonal structure of carbon atoms (see the figure at right), with *a* = 1.42Å and *c* = 3.40 Å, which implies that *c*/*a* = 2.39.

The fact that the hcp structure has the same packing fraction as the fcc structure is easily explained in the figure below. Suppose that we place the first two plane of atoms as in the hcp structure. If the atoms in the third plane are positioned over the centers of the triangles formed by the centers of the atoms in the first plane that have no atoms from the second plane above them, the resulting structure is in fact a fcc. This vertical arrangement is called ABCABC…The hcp and fcc structures differ only by the vertical arrangement (ABAB… or ABCABC…) of hexagonal planes of atoms.

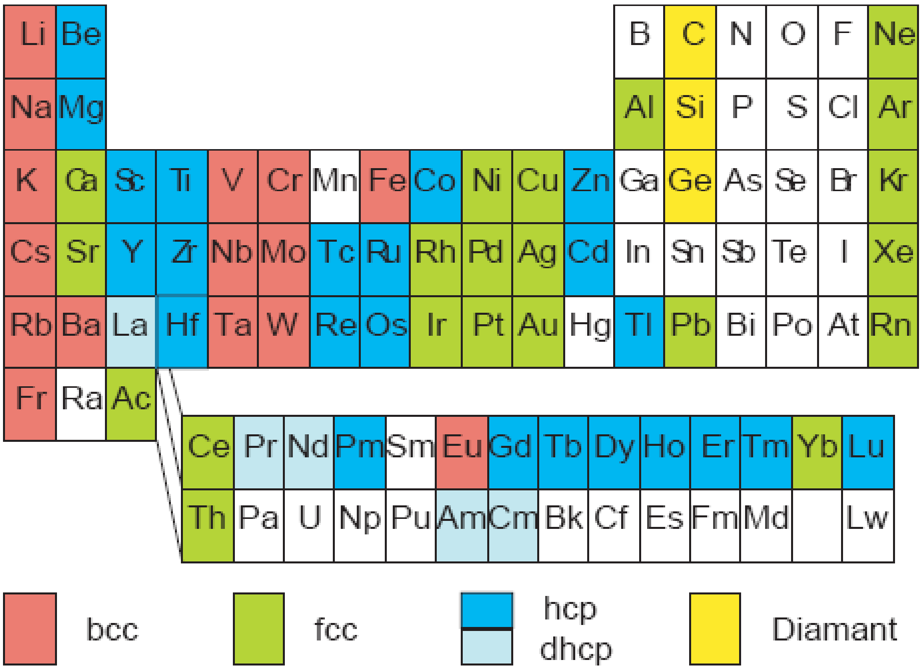




A structure closely related to hcp is wurtzite, generally encountered in binary compound semiconductors such as ZnS (wurtzite), ZnO, BN, CdS, CdSe, GaN, AlN, but sometimes also in ternary compounds such as $Al_{0.25}Ga_{0.5}N$. In binary compounds (see the figure at right), each element has a hcp structure, and the crystal is formed by interpenetrating two such structures, so that an atom in one hcp lattice is equallydistanced from the atoms in the other hcp lattice.



The crystal structure of the elements in the periodic table is indicated in the figure below. Note that several elements can suffer transitions from one crystalline structure to another depending on the external conditions: temperature, pressure, etc. In the table below dhcp stands for double hexagonal closed-packed (the height of the cell along the direction normal to the hexagonal planes is twice that in the hcp structure)



Lattice constants of some elements that crystallize in the fcc crystal structure:

| Crystal | *a* (Å) | Crystal | *a* (Å) | Crystal | *a* (Å) | Crystal | *a* (Å) | Crystal | *a* (Å) |
|---|---|---|---|---|---|---|---|---|---|
| Ar | 5.26 | Au | 4.08 | Cu | 3.61 | Ni | 3.52 | Pt | 3.92 |
| Ag | 4.09 | Ca | 5.58 | Kr | 5.72 | Pb | 4.95 | Sr | 6.08 |
| Al | 4.05 | $\beta$-Co | 3.55 | Ne | 4.43 | Pd | 3.89 | Xe | 6.2 |

Lattice constants of some elements that crystallize in the bcc crystal structure:

| Crystal | *a* (Å) | Crystal | *a* (Å) | Crystal | *a* (Å) | Crystal | *a* (Å) |
|---|---|---|---|---|---|---|---|
| Ba | 5.26 | Fe | 4.08 | Mo | 3.61 | Rb | 3.52 |
| Cr | 4.09 | K | 5.58 | Na | 5.72 | Ta | 4.95 |
| Cs | 4.05 | Li | 3.55 | Nb | 4.43 | V | 3.92 |
| W | 6.08 |  |  |  |  |  |  |

# Reciprocal lattice

The concept of reciprocal lattice is directly connected with the periodicity of crystalline materials and of their physical properties (such as charge density, electric field distribution, etc.). Since the crystal is invariant under any translation with a Bravais lattice vector

$$\boldsymbol{R}_{mnp} = m\boldsymbol{a}_1 + n\boldsymbol{a}_2 + p\boldsymbol{a}_3 \tag{1}$$

for any integers *m*, *n* or *p*, any function $\varphi$ with the same periodicity as the crystalline lattice must satisfy the relation

$$\varphi(\boldsymbol{r}) = \varphi(\boldsymbol{r} + \boldsymbol{R}_{mnp}), \tag{2}$$

where $\boldsymbol{r} = (x_1, x_2, x_3)$ is an arbitrary position vector with coordinates $x_1$, $x_2$, and $x_3$ measured with respect to the (generally non-orthogonal) system of coordinates determined by $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$. This means that

$$\varphi(x_1, x_2, x_3) = \varphi(x_1 + ma_1, x_2 + na_2, x_3 + pa_3) \tag{3}$$

or, for a function that can be expanded in a Fourier series

$$\varphi(x_1, x_2, x_3) = \sum_{G_1, G_2, G_3} \varphi_{\boldsymbol{k}} \exp[i(G_1 x_1 + G_2 x_2 + G_3 x_3)] \tag{4}$$

it follows that, for any *m*, *n*, and *p*,

$$\exp(imG_1 a_1) = 1, \quad \exp(inG_2 a_2) = 1, \quad \exp(ipG_3 a_3) = 1. \tag{5}$$

Thus, $G_i$, with *i* = 1, 2, 3, can only take **discrete** values

$$G_i = 2\pi s_i / a_i, \tag{6}$$

and (4) can be rewritten as

$$\varphi(\boldsymbol{r}) = \sum_{s_1,s_2,s_3} \varphi_{\boldsymbol{k}} \exp(i\boldsymbol{G}\cdot\boldsymbol{r}) \tag{7}$$

where

$$\boldsymbol{G} = s_1\boldsymbol{b}_1 + s_2\boldsymbol{b}_2 + s_3\boldsymbol{b}_3 \tag{8}$$

is a vector in a coordinate system defined by the vectors $\boldsymbol{b}_i$, $i = 1,2,3$, such that

$$\boldsymbol{b}_i \cdot \boldsymbol{a}_j = 2\pi\delta_{ij}\,. \tag{9}$$

Similar to the Bravais lattices that are constructed starting with the primitive vectors $\boldsymbol{a}_i$, one can define a **reciprocal lattice** in terms of the primitive vectors $\boldsymbol{b}_i$, such that $\boldsymbol{G}$ in (8) are points in the reciprocal lattice. A reciprocal lattice can only be defined with respect to a given direct lattice. As demonstrated in the following, the $\boldsymbol{G}$ vectors have dimensions (and meaning of) wavevectors related to plane waves with the periodicity of the direct lattice.

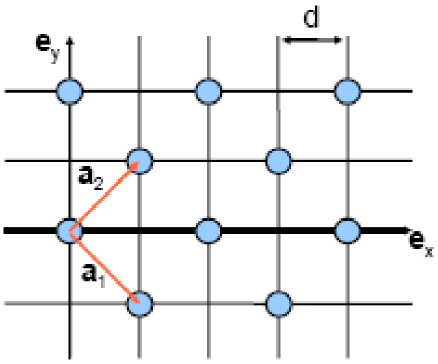
If the vectors $\boldsymbol{a}_i$ are chosen and the volume of the primitive cell in the direct space is $\Omega = |(\boldsymbol{a}_1 \times \boldsymbol{a}_2)\cdot\boldsymbol{a}_3|$, the vectors $\boldsymbol{b}_i$ can be chosen as

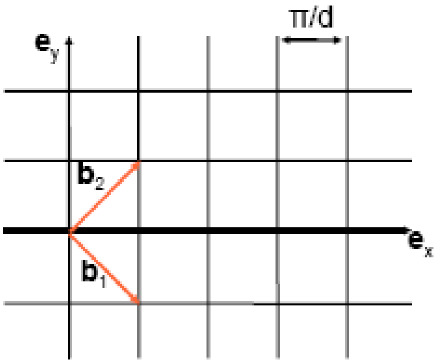
$$\boldsymbol{b}_1 = (2\pi/\Omega)(\boldsymbol{a}_2 \times \boldsymbol{a}_3),\quad \boldsymbol{b}_2 = (2\pi/\Omega)(\boldsymbol{a}_3 \times \boldsymbol{a}_1),\quad \boldsymbol{b}_3 = (2\pi/\Omega)(\boldsymbol{a}_1 \times \boldsymbol{a}_2)\,. \tag{10}$$

It follows then that the volume of the primitive cell of the reciprocal lattice is given by

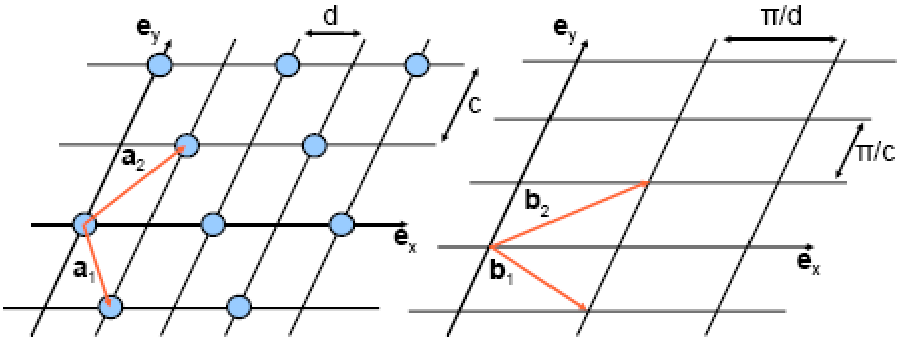
$$\Omega_{rec} = |\boldsymbol{b}_1 \cdot (\boldsymbol{b}_2 \times \boldsymbol{b}_3)| = (2\pi)^3/\Omega\,. \tag{11}$$

Examples of direct and corresponding reciprocal lattices in two dimensions are given in the figures below.





For $\boldsymbol{a}_1 = d(\boldsymbol{x} - \boldsymbol{y})$, $\boldsymbol{a}_2 = d(\boldsymbol{x} + \boldsymbol{y})$, the vectors of the reciprocal lattice are determined from condition (9), and are found to be $\boldsymbol{b}_1 = (\pi / d)(\boldsymbol{x} - \boldsymbol{y})$, $\boldsymbol{b}_2 = (\pi / d)(\boldsymbol{x} + \boldsymbol{y})$.



When $\boldsymbol{x}$ and $\boldsymbol{y}$ are not orthogonal, but $\boldsymbol{x} \cdot \boldsymbol{y} = \varepsilon$ (see the figure above), for $\boldsymbol{a}_1 = d\boldsymbol{x} - c\boldsymbol{y}$ and $\boldsymbol{a}_2 = d\boldsymbol{x} + c\boldsymbol{y}$, we obtain (please check!)

$$\boldsymbol{b}_1 = \pi \frac{c + d\varepsilon}{cd(1-\varepsilon^2)}\boldsymbol{x} - \pi \frac{d + c\varepsilon}{cd(1-\varepsilon^2)}\boldsymbol{y}, \quad \boldsymbol{b}_2 = \pi \frac{c - d\varepsilon}{cd(1-\varepsilon^2)}\boldsymbol{x} + \pi \frac{d - c\varepsilon}{cd(1-\varepsilon^2)}\boldsymbol{y}.$$

In three dimensions, the reciprocal lattices for the Bravais lattices in the cubic system are summarized in the table below

| **Real space** | | **Reciprocal space** | |
|---|---|---|---|
| **Lattice** | **Lattice constant** | **Lattice** | **Lattice constant** |
| SC | $a$ | SC | $2\pi / a$ |
| BCC | $a$ | FCC | $4\pi / a$ |
| FCC | $a$ | BCC | $4\pi / a$ |

The reciprocal lattice of a cubic lattice is also cubic since, in this case, if $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ are orthogonal vectors of unit length, $\boldsymbol{a}_1 = a\boldsymbol{x}$, $\boldsymbol{a}_2 = a\boldsymbol{y}$, $\boldsymbol{a}_3 = a\boldsymbol{z}$ and $\Omega = a^3$, from (10) it follows that $\boldsymbol{b}_1 = (2\pi / a)\boldsymbol{x}$, $\boldsymbol{b}_2 = (2\pi / a)\boldsymbol{y}$, $\boldsymbol{b}_3 = (2\pi / a)\boldsymbol{z}$, i.e. the reciprocal lattice is simple cubic with a lattice constant $2\pi / a$.

Analogously, the reciprocal lattice to the bcc lattice with (see the first course) $\boldsymbol{a}_1 = (a/2)(\boldsymbol{x} + \boldsymbol{y} - \boldsymbol{z})$, $\boldsymbol{a}_2 = (a/2)(-\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z})$, $\boldsymbol{a}_3 = (a/2)(\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{z})$, and $\Omega = a^3/2$ has primitive vectors $\boldsymbol{b}_1 = (2\pi / a)(\boldsymbol{x} + \boldsymbol{y})$, $\boldsymbol{b}_2 = (2\pi / a)(\boldsymbol{y} + \boldsymbol{z})$, $\boldsymbol{b}_3 = (2\pi / a)(\boldsymbol{z} + \boldsymbol{x})$, i.e. is a fcc lattice with a volume (of the primitive unit cell) in reciprocal state of $\Omega_{rec} = 2(2\pi / a)^3$, whereas the reciprocal lattice of the fcc lattice, with $\boldsymbol{a}_1 = (a/2)(\boldsymbol{x} + \boldsymbol{y})$, $\boldsymbol{a}_2 = (a/2)(\boldsymbol{y} + \boldsymbol{z})$,

$\boldsymbol{a}_3 = (a/2)(\boldsymbol{z} + \boldsymbol{x})$, and $\Omega = a^3/4$ is a bcc lattice with $\Omega_{rec} = 4(2\pi/a)^3$ and primitive vectors $\boldsymbol{b}_1 = (2\pi/a)(\boldsymbol{x} + \boldsymbol{y} - \boldsymbol{z})$, $\boldsymbol{b}_2 = (2\pi/a)(-\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z})$, $\boldsymbol{b}_3 = (2\pi/a)(\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{z})$. In both cases the cubic structure of the reciprocal lattice has a lattice constant of $4\pi/a$.

**Observation:** The reciprocal lattice of a reciprocal lattice is the direct lattice.

Because the product of a primitive Bravais lattice vector and of a primitive vector of the reciprocal cell is an integer multiple of $2\pi$, i.e. that

$$\boldsymbol{G}_{mnp} \cdot \boldsymbol{R}_{hkl} = 2\pi(mh + nk + pl), \tag{12}$$

for all integers *m*, *n*, *p* and *h*, *k*, *l*, it follows that $\exp(i\boldsymbol{G} \cdot \boldsymbol{R}) = 1$ for any vector $\boldsymbol{R}$ in the Bravais lattice and any vector $\boldsymbol{G}$ in the reciprocal lattice. This implies that the function $\exp(i\boldsymbol{G} \cdot \boldsymbol{r})$ has the same periodicity as the crystal because $\exp[i\boldsymbol{G} \cdot (\boldsymbol{r} + \boldsymbol{R})] = \exp(i\boldsymbol{G} \cdot \boldsymbol{r})\exp(i\boldsymbol{G} \cdot \boldsymbol{R}) = \exp(i\boldsymbol{G} \cdot \boldsymbol{r})$. As a consequence,

$$\int_{cell} \exp(i\boldsymbol{G} \cdot \boldsymbol{r}) dV \tag{13}$$

is independent of the choice of the cell and a translation with an arbitrary vector $\boldsymbol{d}$ should not change the value of the integral. More precisely, if

$$\int_{cell} \exp[i\boldsymbol{G} \cdot (\boldsymbol{r} + \boldsymbol{d})] dV = \int_{cell} \exp(i\boldsymbol{G} \cdot \boldsymbol{r}) dV \tag{14}$$

then $[\exp(i\boldsymbol{G} \cdot \boldsymbol{d}) - 1]\int_{cell} \exp(i\boldsymbol{G} \cdot \boldsymbol{r}) dV = 0$, from which it follows that

$$\int_{cell} \exp(i\boldsymbol{G} \cdot \boldsymbol{r}) dV = \Omega \delta_{\boldsymbol{G},0} \tag{15}$$

and that the set of functions $\exp(i\boldsymbol{G} \cdot \boldsymbol{r})$ form a complete, orthonormal basis for any periodic function which has the same periodicity as the crystal, i.e. which can be written as

$$\varphi(\boldsymbol{r}) = \sum_{\boldsymbol{G}} \varphi_{\boldsymbol{G}} \exp(i\boldsymbol{G} \cdot \boldsymbol{r}). \tag{16}$$

If the formula above is regarded as a Fourier transformation of the periodic function $\varphi$, the coefficients $\varphi_{\boldsymbol{G}}$ can be retrieved by performing an inverse Fourier transformation. More precisely, since

$$\begin{aligned}\int_{cell}\varphi(\boldsymbol{r})\exp(-i\boldsymbol{G}'\cdot\boldsymbol{r})dV &= \sum_{\boldsymbol{G}}\varphi_{\boldsymbol{G}}\int_{cell}\exp(i\boldsymbol{G}\cdot\boldsymbol{r})\exp(-i\boldsymbol{G}'\cdot\boldsymbol{r})dV \\ &= \sum_{\boldsymbol{G}}\varphi_{\boldsymbol{G}}\int_{cell}\exp[i(\boldsymbol{G}-\boldsymbol{G}')\cdot\boldsymbol{r}]dV = \sum_{\boldsymbol{G}}\varphi_{\boldsymbol{G}}\Omega\delta_{\boldsymbol{G}\boldsymbol{G}'}\end{aligned} \tag{17}$$

it follows that

$$\varphi_{\boldsymbol{G}} = \Omega^{-1}\int_{cell}\varphi(\boldsymbol{r})\exp(-i\boldsymbol{G}\cdot\boldsymbol{r})dV\,. \tag{18}$$

## Relations between the direct and reciprocal lattices

One geometrical property that can be easily shown is that the reciprocal lattice vector

$$\boldsymbol{G}_{mnp} = m\boldsymbol{b}_1 + n\boldsymbol{b}_2 + p\boldsymbol{b}_3 \tag{19}$$

is perpendicular to the plane (actually, to the set of parallel planes) with Miller indices (*mnp*) in the Bravais lattice. The closest plane to the origin from the set of planes (*mnp*) cuts the $\boldsymbol{a}_i$ coordinate axes at $a_1/m$, $a_2/n$, and $a_3/p$, respectively.

To show that (*mnp*) is perpendicular to $\boldsymbol{G}_{mnp}$ it is sufficient to demonstrate that $\boldsymbol{G}_{mnp}$ is perpendicular to two non-collinear vectors in the (*mnp*) plane, which can be chosen as

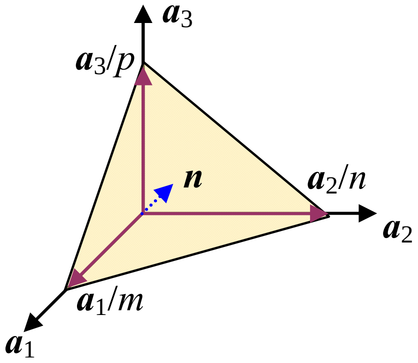
$$\boldsymbol{u} = \boldsymbol{a}_2/n - \boldsymbol{a}_1/m\,, \qquad \boldsymbol{v} = \boldsymbol{a}_3/p - \boldsymbol{a}_1/m\,, \tag{20}$$

and satisfy, indeed, the relations

$$\boldsymbol{u}\cdot\boldsymbol{G}_{mnp} = \boldsymbol{v}\cdot\boldsymbol{G}_{mnp} = 0 \tag{21}$$

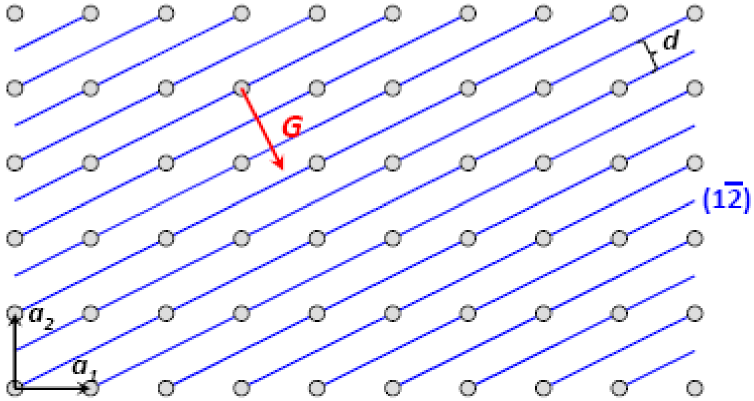
because of (9). Then, it follows that the normal to the (*mnp*) plane that passes through the origin can be expressed as

$$\boldsymbol{n}_{mnp} = \boldsymbol{G}_{mnp}/|\boldsymbol{G}_{mnp}|\,. \tag{22}$$



A consequence of this result is that the distance between two consecutive planes with the same Miller indices (*mnp*) is inversely proportional to the modulus of $\boldsymbol{G}_{mnp}$. Since we can always draw a plane from the (*mnp*) family through the origin, the distance between two successive planes is equal to the distance between the origin and the closest plane to origin from the (*mnp*) family. This distance is obtained by calculating the projection on the normal to the (*mnp*), i.e. on $\boldsymbol{n}_{mnp} = \boldsymbol{G}_{mnp} / |\boldsymbol{G}_{mnp}|$, of any of the vectors $\boldsymbol{a}_1 / m$, $\boldsymbol{a}_2 / n$, or $\boldsymbol{a}_3 / p$. Using (22) it is found that

$$d_{mnp} = \boldsymbol{n} \cdot \frac{\boldsymbol{a}_1}{m} = \boldsymbol{n} \cdot \frac{\boldsymbol{a}_2}{n} = \boldsymbol{n} \cdot \frac{\boldsymbol{a}_3}{p} = \frac{2\pi}{|\boldsymbol{G}_{mnp}|}. \tag{23}$$



So,

$$d_{mnp} = \frac{2\pi}{\sqrt{m^2 b_1^2 + n^2 b_2^2 + p^2 b_3^2 + 2mn(\boldsymbol{b}_1 \cdot \boldsymbol{b}_2) + 2np(\boldsymbol{b}_2 \cdot \boldsymbol{b}_3) + 2pm(\boldsymbol{b}_3 \cdot \boldsymbol{b}_1)}}. \tag{24}$$

As already pointed out in the discussion about Miller indices, the distance between any two planes in the family (*sm,sn,sp*), is $s$ times smaller than between any two planes in the family (*mnp*). The two families/sets of planes are parallel.

In particular, for the simple, body-centered and face-centered cubic Bravais lattices with the primitive translation vectors given in the Crystal Structure section of the course, the distance between two consecutive planes with the same Miller indices is, respectively,

$$d_{mnp}^{sc} = \frac{a}{\sqrt{m^2 + n^2 + p^2}}\,, \tag{25a}$$

$$d_{mnp}^{bcc} = \frac{a}{\sqrt{(n+p)^2 + (p+m)^2 + (m+n)^2}}\,, \tag{25b}$$

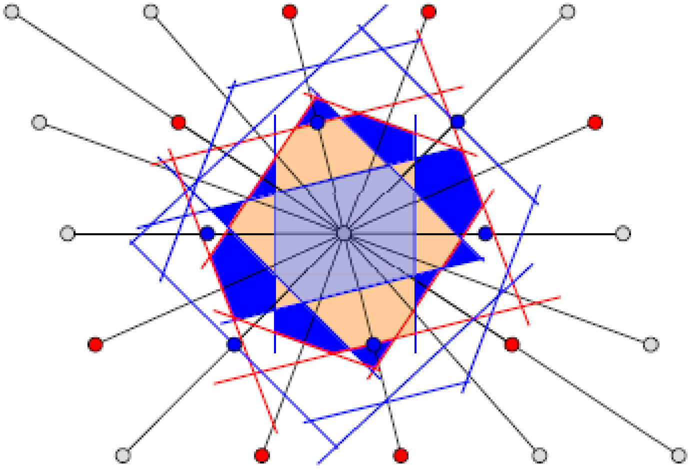
$$d_{mnp}^{fcc} = \frac{a}{\sqrt{(n+p-m)^2 + (p+m-n)^2 + (m+n-p)^2}} \tag{25c}$$

Due to the form of (7), the vectors $\boldsymbol{G}$ of the reciprocal lattice can be understood as wavevectors of plane waves with the periodicity of the lattice and wavelengths $2\pi / |\boldsymbol{G}|$, similar to wave-vectors in optics that are perpendicular to wavefronts and have dimensions related to the wavelength $\lambda$ as $2\pi / \lambda$.

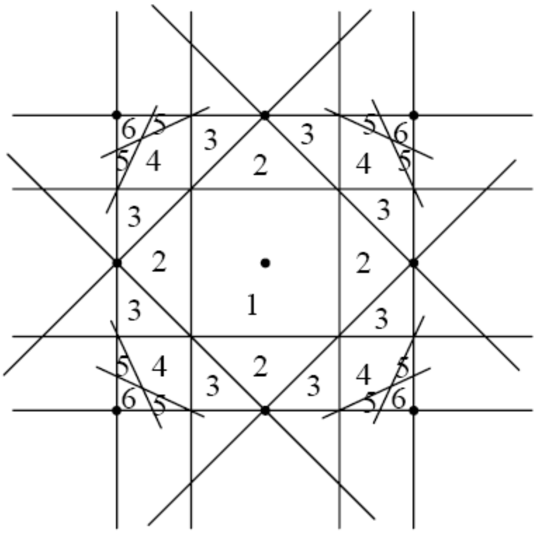
## The first Brillouin zone

Analogous to the Wigner-Seitz cell in direct lattices, one can define a primitive unit cell in the reciprocal lattice that has the same symmetry as this lattice. This primitive unit cell is known as the first Brillouin zone. The construction of the first Brillouin zone is similar to that of the Wigner-Seitz cell, i.e. we draw lines to connect a given lattice point in the reciprocal lattice to all nearby lattice points, and then draw new lines (or planes, in three-dimensional lattices) at the mid point and normal to the first set of lines. These lines (planes) are called Bragg planes since (as we will see later) all $\boldsymbol{k}$ vectors that finish on these surfaces satisfy the Bragg condition. The first Brillouin zone is then the area (volume) in reciprocal space that can be reached from the origin, without crossing any Bragg planes. Higher-order Brillouin zones, say the $n^{\text{th}}$ Brillouin zone, are then defined as the area (volume) in reciprocal space that can be reached from the origin by crossing exactly $n-1$ Bragg planes. The construction of the first (light blue), second (light brown) and third (dark blue) Brillouin zones for a two-dimensional lattice is illustrated in the figure below. The Bragg planes enclosing the $n^{\text{th}}$ Brillouin zone correspond to the $n^{\text{th}}$ order X-ray diffraction.

Although higher order Brillouin zones are fragmented, the fragments, if translated, look like the first Brillouin zone. This process is called reduced zone scheme. All Brillouin zones, irrespective of the order, have the same volume.

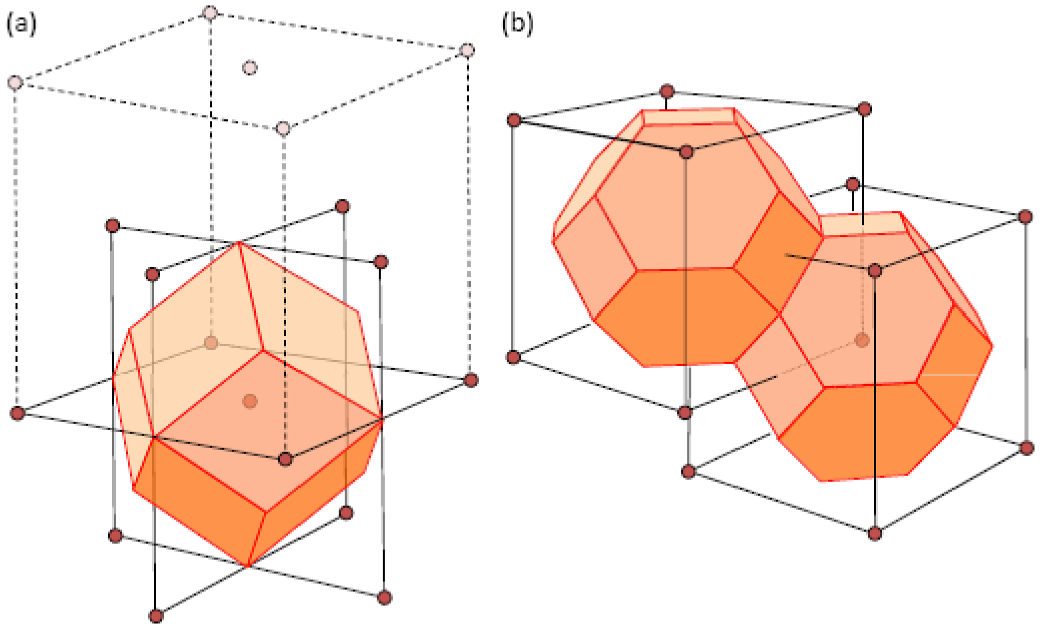


The higher-order Brillouin zones for a two-dimensional square lattice are illustrated in the figure below.



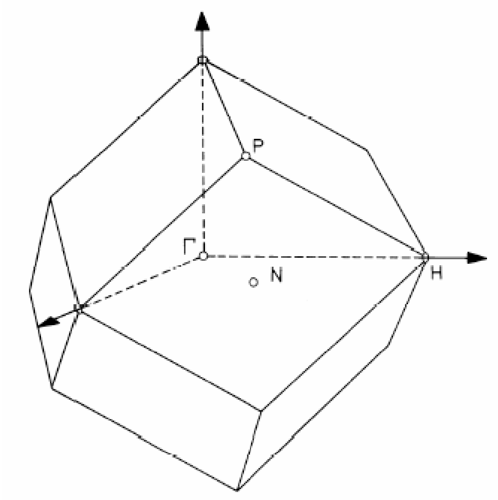
As for Wigner-Seitz cells, the faces of the first Brillouin zone satisfy the relation $\boldsymbol{k}\cdot\boldsymbol{G}=|\boldsymbol{G}|^2/2$, where $|\boldsymbol{G}|$ is the distance to the nearest neighbor in the reciprocal space. This relation can be rewritten as $\boldsymbol{G}^2-2\boldsymbol{k}\cdot\boldsymbol{G}=0$ or, since the equation is equivalent to the replacement of $\boldsymbol{G}$ with $-\boldsymbol{G}$, we obtain $(\boldsymbol{k}+\boldsymbol{G})^2=\boldsymbol{k}^2$, i.e. the first Brillouin zone is the intersection of spheres with the same radius centered at nearest neighbor points in the reciprocal lattice.

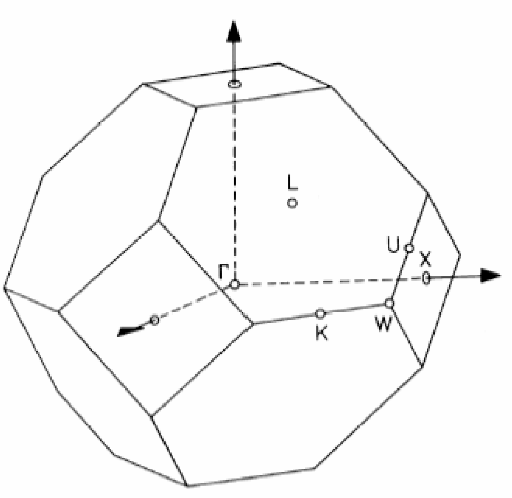
In particular, since the reciprocal lattice of the bcc lattice is a fcc lattice, the first Brillouin zone of the bcc lattice (see the polyhedron in the figure a below) is the Wigner-Seitz cell of the fcc. The reverse is also true: the first Brillouin zone of a fcc lattice (the truncated octahedron/rhombododecahedron in figure b below) is the Wigner-Seitz cell of the bcc lattice.



For certain Bravais lattice, in particular bcc, fcc and hexagonal, the points of highest symmetry in the reciprocal lattice are labeled with certain letters. The center of the Brillouin zone is in all cases denoted by Γ. Other symmetry points are denoted as follows (see also figures):

sc lattice: M – center of an edge
R – corner point
X – center of a face
bcc lattice: H – corner point joining four edges
N – center of a face
P – corner point joining three edges
fcc lattice: K – middle of an edge joining two hexagonal faces
L – center of a hexagonal face
U – middle of an edge joining a hexagonal and a square face
W – corner point
X – center of a square face

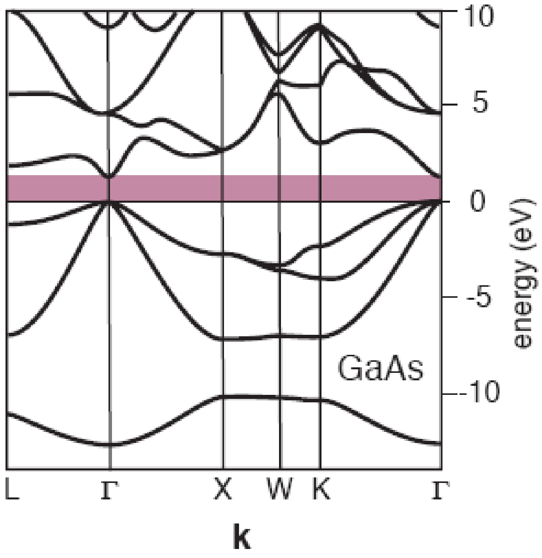


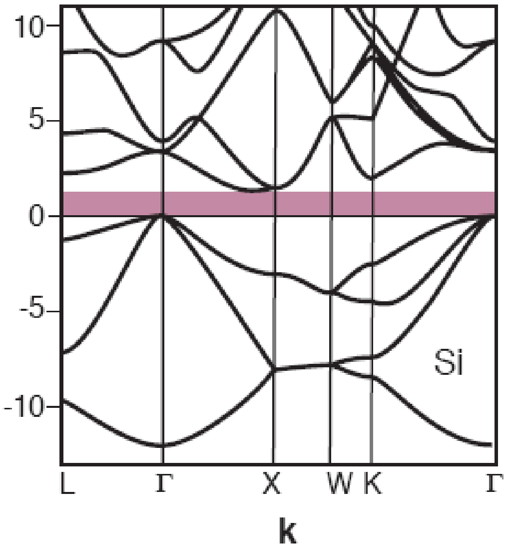


hexagonal lattice:

- A – center of a hexagonal face
- H – corner point
- K – middle of an edge joining two rectangular faces
- L – middle of an edge joining a hexagonal and a rectangular face
- M – center of a rectangular face

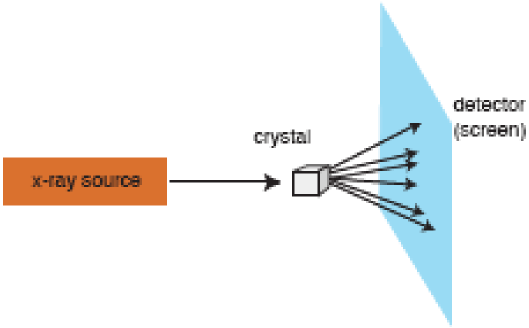
Dispersion relations of electrons and phonons for different crystal directions use this labeling (see the figures below), the labels indicating the direction but also the symmetry of the crystal, since different labels are used for different symmetries.

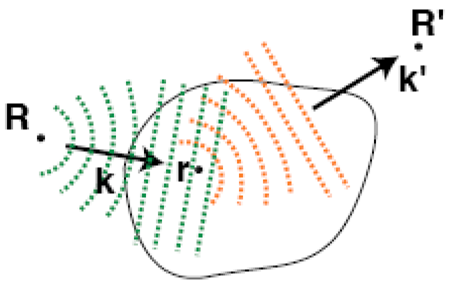


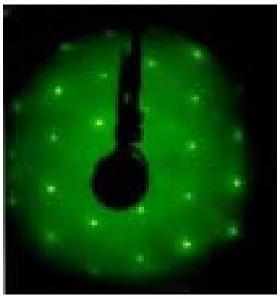


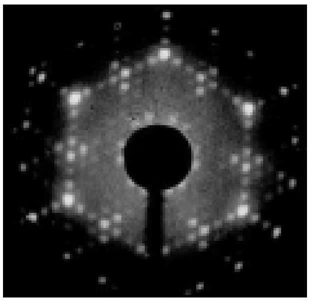
# X-ray diffraction on crystalline structures

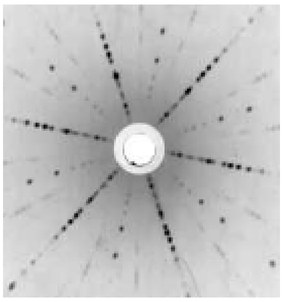
The direct observation of the periodicity of atoms in a crystalline material relies on the X-ray or particle (electron or neutron) diffraction/scattering on these spatially periodic structures, since the wavelength of the incident beam is in these cases comparable to the typical interatomic distance of a few Å. Optical diffraction is not suitable for this purpose since the wavelength of photons is much too long (about 1 μm) in comparison to the lattice constant (a few Angstroms). In a diffraction experiment, both the X-ray or particle source and the detector are placed in vacuum and sufficiently far away from the sample such that, for monochromatic radiation, the incident and outgoing X-ray or particle beams can be approximated by plane waves. The X-rays can be used in either transmission or reflection configurations. The diffraction picture offers information regarding the symmetry of the crystal along a certain axis. In particular, the positions of the spots give information about the lattice and the intensity analysis reveal the composition of the basis.











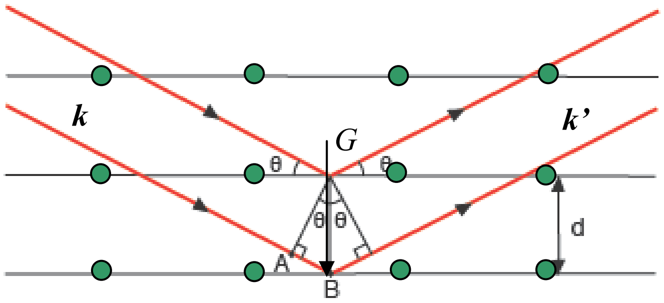
The X-rays penetrate deeply in the material, so that many layers contribute to the reflected intensity and the diffracted peak intensities are very sharp (in angular distribution). To obtain sharp intensity peaks of the scattered radiation, the X-rays should be specularly reflected by the atoms in one plane.

For X-rays, the wavelength is determined from the relation $E = h\nu = hc/\lambda$ or $\lambda = hc/E$, which equals a few Å if $E$ is of the order of few keV. In fact, $\lambda$(Å) = 12.4/$E$(keV). X-rays are scattered mostly by the electronic shells of atoms in a solid, since the nuclei are too heavy to respond.

Electrons can also have de Broglie wavelengths similar to the lattice constants of crystals. In this case $E = (h/\lambda)^2/2m$, and for an electron energy $E$ of 6 eV, the corresponding wavelength $\lambda = h/\sqrt{2mE}$ is about 5 Å. Actually, if the kinetic energy of the electrons is acquired in an acceleration voltage potential $U$, such that $E = eU$, one has $\lambda$(Å) = $12.28/[U(\mathrm{V})]^{1/2}$. For neutron diffraction we have to consider a similar relation, except that the electron mass $m$ has to be replaced by the neutron mass $M$. Then, $\lambda$(Å) = $0.28/[E(\mathrm{eV})]^{1/2}$.

When a wave interacts with the crystal, the plane wave is scattered by the atoms in the crystal, each atom acting like a point source (Huygens' principle). Because a crystal structure consists of a lattice and a basis, the X-ray diffraction is a convolution of diffraction by the lattice points and diffraction by the basis. Generally, the latter term modulates the diffraction by the lattice points. In particular, if each lattice point acts as a coherent point source, each lattice plane acts as a mirror.

The X-rays scattered by all atoms in the crystalline lattice interfere and the problem is to determine the Bravais lattice (including the lattice constants) and the basis from the interference patterns. The wave that is diffracted in a certain direction is a sum of the waves scattered by all atoms. Higher diffraction intensities will be observed along the directions of constructive interference, which are determined by the crystal structure itself.



The diffraction of X-rays by crystals is elastic, the X-rays having the same frequency (and wavelength) before and after the reflection. The path difference between two consecutive planes separated by $d$ is $2\cdot\mathrm{AB} = 2d\sin\theta$. First-order constructive interference occurs if

$$2d\sin\theta = \lambda\,, \tag{1}$$

condition known as **Bragg's law**.

The Bragg law is a consequence of the periodicity of the crystal structure and holds only if $\lambda \le 2d$. This is the reason why the optical radiation is not suitable to detect the crystalline structure, but only X-rays and electron or neutron beams can perform this task.

Higher order diffraction processes are also possible. The Bragg relation determines, through the angle $\theta$, the directions of maximum intensity. These directions are identified as high-intensity points on the detection screen, the position of which reveal the crystal structure. For example, if the sample has a cubic crystal structure oriented such that the direction [111] (the diagonal of the cube) is parallel to the incident beam, the symmetry of the points on the detector screen will reveal a $C_3$ symmetry axis. On the contrary, if the diffraction pattern has a $C_6$ symmetry axis, the crystal is hexagonal, if it has a $C_4$ symmetry axis it is a tetragonal crystal, whereas it is cubic if it shows both a $C_4$ and a $C_3$ symmetry axis.

The Bragg formula says nothing about the intensity and width of the X-ray diffraction peaks, assumes a single atom in every lattice point, and neglects both differences in scattering from different atoms and the distribution of charge around atoms.

A closer look at the interaction between the X-rays and the crystal of volume $V$ reveals that the amplitude of the scattered radiation $F$ (which is proportional to the amplitude of the oscillation of the electric and magnetic fields of the total diffracted ray) is determined by the local electron concentration $n(\boldsymbol{r}) = \sum_{\boldsymbol{G}} n_{\boldsymbol{G}} \exp(i\boldsymbol{G}\cdot\boldsymbol{r})$, which is a measure of the strength of the interaction, and has the same periodicity as the crystalline lattice. The diffraction intensity $I \propto |F|^2$. For elastic X-ray scattering, the phase of the outgoing beam, with wavevector $\boldsymbol{k}'$, differs from that of the incoming beam that propagates with a wavevector $\boldsymbol{k}$ through $\exp[i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}]$, so that

$$F = \int n(\boldsymbol{r})\exp[i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}]dV = \int n(\boldsymbol{r})\exp(-i\Delta\boldsymbol{k}\cdot\boldsymbol{r})dV = \sum_{\boldsymbol{G}} n_{\boldsymbol{G}} \int \exp[i(\boldsymbol{G}-\Delta\boldsymbol{k})\cdot\boldsymbol{r}]dV \tag{2}$$

where $\Delta\boldsymbol{k} = \boldsymbol{k}' - \boldsymbol{k}$ is the scattering vector, which expresses the change in wavevector. The result in the above integral depends on the volume of the crystal. If the crystal has length $L_i$ and $N_i$ primitive cells in the *i* direction ($i$ = 1,2,3) of an orthogonal coordinate system (if the crystal system is not orthogonal, a transformation of coordinates to the $x = x_1$, $y = x_2$, $z = x_3$ axes should be performed), the integral along the *i* direction is given by

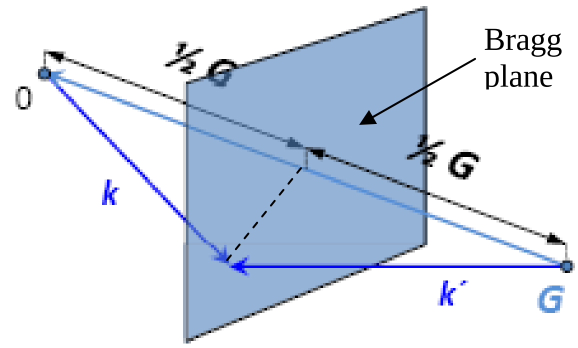
$$\int_{-L_i/2}^{L_i/2} \exp\left[ i\frac{2\pi}{a_i}(s_i - \Delta\xi_i)x_i \right] dx_i = a_i \frac{\sin[\pi(s_i - \Delta\xi_i)N_i]}{\pi(s_i - \Delta\xi_i)} = L_i \mathrm{sinc}[\pi(s_i - \Delta\xi_i)N_i] \qquad (3)$$

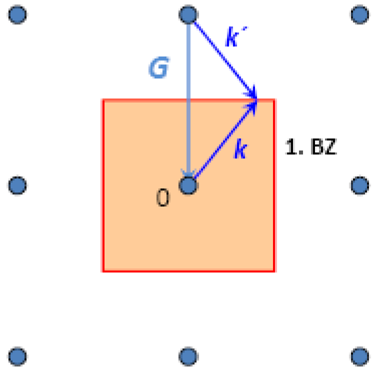
where $s_i$, $\Delta\xi_i$ are the components of $\boldsymbol{G}$ and $\Delta\boldsymbol{k}$ on the *i* axis and $a_i = L_i / N_i$ is the lattice constant on the same direction. The function $\mathrm{sinc}(x) = \sin x / x$ has a maximum value for $x = 0$, and tends to the Dirac delta function for large *x*.

Therefore, in large-volume crystals scattering occurs only if

$$\Delta\boldsymbol{k} = \boldsymbol{G}, \qquad (4)$$

case in which $F = Vn_{\boldsymbol{G}}$. (In finite-volume crystals there is a sort of “uncertainty” in the angular range of $\Delta\boldsymbol{k}$ around $\boldsymbol{G}$ for which the scattering amplitude takes significant values: as the volume decreases, the angular range increases.) The above condition suggests that X-ray diffraction experiments reveal the reciprocal lattice of a crystal, in opposition to microscopy, which exposes the direct lattice (if performed with high-enough resolution).





The diffraction condition $\Delta\boldsymbol{k} = \boldsymbol{k}' - \boldsymbol{k} = \boldsymbol{G}$ can be rewritten as $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{G}$ or $k'^2 = k^2 + G^2 + 2\boldsymbol{k}\cdot\boldsymbol{G}$. In particular, the form $\hbar\boldsymbol{k}' = \hbar\boldsymbol{k} + \hbar\boldsymbol{G}$ of the diffraction condition represents the momentum conservation law of the X-ray photon in the scattering process; the crystal receives the momentum $-\hbar\boldsymbol{G}$. For elastic scattering $|\boldsymbol{k}'| = |\boldsymbol{k}|$ and thus $G^2 + 2\boldsymbol{k}\cdot\boldsymbol{G} = 0$, or $\boldsymbol{k} \cdot \boldsymbol{G} = |\boldsymbol{G}|^2 / 2$, equation that defines the faces of the first Brillouin zone (the Bragg planes). The geometric interpretation of this relation (see the figure above) is that constructive interference/diffraction is the strongest on the faces of the first Brillouin zone. In other words, the first Brillouin zone exhibits all the ***k*** wavevectors that can be Bragg-reflected by the crystal.

The diffraction condition is equivalent to Bragg's law, which can be written for a certain set of planes separated by the distance $d = d_{mnp}$ as $2(2\pi/\lambda)\sin\theta = 2\pi/d_{mnp}$, or $2\boldsymbol{k} \cdot \boldsymbol{G} = G^2$, with $\boldsymbol{G} = m\boldsymbol{b}_1 + n\boldsymbol{b}_2 + p\boldsymbol{b}_3$ (for the direction of ***G*** with respect to the set of planes, see the figure illustrating the Bragg law).

## The Laue condition

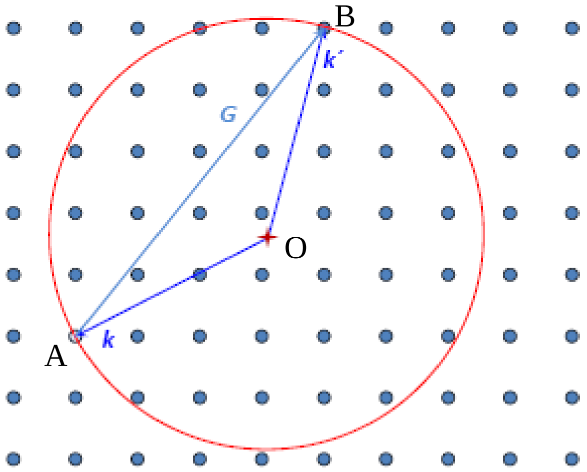
The diffraction condition $\Delta\boldsymbol{k} = \boldsymbol{G}$ can be expressed in still another way: if we multiply both terms of this relation with the primitive translation vectors of the direct lattice, we obtain the Laue conditions

$$\boldsymbol{a}_1 \cdot \Delta\boldsymbol{k} = 2\pi s_1, \qquad \boldsymbol{a}_2 \cdot \Delta\boldsymbol{k} = 2\pi s_2, \qquad \boldsymbol{a}_3 \cdot \Delta\boldsymbol{k} = 2\pi s_3, \tag{5}$$

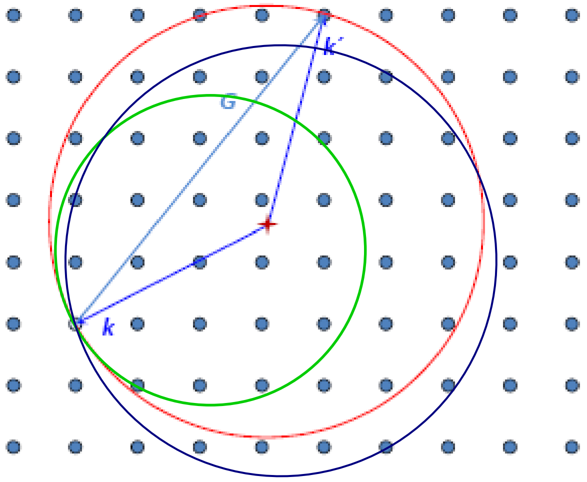
where $s_i$ are integers. The Laue equations have a simple geometrical interpretation: $\Delta\boldsymbol{k}$ lies simultaneously on a cone about $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$, i.e. lies at the common line of intersection of three cones. This condition is quite difficult to satisfy in practice. Moreover, in analogy to optical diffraction experiments, the Laue condition can be viewed as a condition of constructive interference between waves diffracted by two atoms separated by a primitive translation vector or, by extension, between waves diffracted by all atoms in the crystal. At Bragg reflection, the radiation scattered by all atoms arrives in phase at the detector, and intensity peaks are obtained.

## The Ewald sphere

The direction of interference peaks can be easily determined also via a simple geometrical construction suggested by Ewald. Namely, one constructs a sphere (a circle in two dimensions – see the red circle in the figure above) around a point O in the reciprocal lattice chosen such that the incident wavevector with O as origin, ends on an arbitrary lattice point A. The origin of the Ewald sphere (or circle) is not necessary a lattice point.



The radius of the sphere (circle) is the wavenumber of the incident (and outgoing) radiation $k = |\boldsymbol{k}| = |\boldsymbol{k}'|$. A maximum intensity is found around a direction $\boldsymbol{k}'$ if and only if the Ewald sphere (circle) passes through another point B of the reciprocal lattice. The direction $\boldsymbol{k}'$ is determined by the origin O of the Ewald sphere and this lattice point on the surface (circumference), which is separated from the tip of $\boldsymbol{k}$ (from A) by a reciprocal lattice vector. It is possible that for certain incidence angles and wavelengths of the X-rays no such preferential direction $\boldsymbol{k}'$ exists.



Therefore, to obtain peaks in the scattered intensity it is in general necessary to vary either the wavelength or the incidence angle of the incoming X-rays such that a sufficient number of reciprocal lattice points find themselves on the Ewald sphere (circle), in order to determine unambiguously the crystal structure. In the first method, called Laue method, the radius of the Ewald sphere (circle) is varied continuously (see, for example, the green circle in the figure above), while in the second method, called the rotating crystal method or Debye-Scherrer-Hull method, the Ewald sphere (circle) is rotated around the original lattice point with respect to which the Ewald sphere (circle) was constructed. The result is represented with the dark blue circle in the figure above.

In another diffraction method (the Debye-Scherrer method) polycrystalline samples are used, which are either fixed or rotate around an axis. In this case, the incident beam is scattered by only those crystallites (randomly oriented) with planes that satisfy the Bragg condition. Because the sample contains crystallites with all orientations, the diffraction pattern on the screen is no longer formed from discrete points, but from concentric circles.

## The influence of the basis on the scattered amplitude

If the Laue/diffraction condition $\Delta\boldsymbol{k} = \boldsymbol{G}$ is satisfied, an explicit account of the basis influence implies that the assumption of point/spherical sources at the lattice points have to be modified. In this case, we have found that

$$F = Vn_{\boldsymbol{G}} = N\int_{cell} n(\boldsymbol{r})\exp(-i\boldsymbol{G}\cdot\boldsymbol{r})dV = NS_{\boldsymbol{G}}\,, \tag{6}$$

where $n_{\boldsymbol{G}} = \Omega^{-1}\int_{cell} n(\boldsymbol{r})\exp(-i\boldsymbol{G}\cdot\boldsymbol{r})dV$, $N$ is the total number of lattice points, and

$$S_{\boldsymbol{G}} = \int_{cell} n(\boldsymbol{r})\exp(-i\boldsymbol{G}\cdot\boldsymbol{r})dV \tag{7}$$

is called the **structure factor**. It is defined as an integral over a single cell, with $\boldsymbol{r} = 0$ at one corner. If there is only one lattice point in the basis and the electron distribution $n(\boldsymbol{r}) \cong \delta(\boldsymbol{r})$, $S_{\boldsymbol{G}} = 1$.

If there are $s$ atoms in the basis at positions $\boldsymbol{r}_j$, $j = 1,2,..,s$, the total electron density can be expressed as a superposition of electron concentration functions $n_j$ at each atom $j$ in the basis, so that the structure factor is expressed as integrals over the $s$ atoms of a cell:

$$\begin{aligned} S_{\boldsymbol{G}} &= \int\left[\sum_{j=1}^{s} n_j(\boldsymbol{r}-\boldsymbol{r}_j)\right]\exp(-i\boldsymbol{G}\cdot\boldsymbol{r})dV = \int\left[\sum_{j=1}^{s} n_j(\boldsymbol{\rho})\right]\exp(-i\boldsymbol{G}\cdot\boldsymbol{\rho})\exp(-i\boldsymbol{G}\cdot\boldsymbol{r}_j)dV \\ &= \sum_{j=1}^{s}\exp(-i\boldsymbol{G}\cdot\boldsymbol{r}_j)\int n_j(\boldsymbol{\rho})\exp(-i\boldsymbol{G}\cdot\boldsymbol{\rho})dV = \sum_{j=1}^{s} f_j\exp(-i\boldsymbol{G}\cdot\boldsymbol{r}_j) \end{aligned} \tag{8}$$

where $\boldsymbol{\rho} = \boldsymbol{r} - \boldsymbol{r}_j$ and $f_j = \int n_j(\boldsymbol{\rho}) \exp(-i\boldsymbol{G}\cdot\boldsymbol{\rho}) dV$ is the **atomic form factor**, which depends only on the type of element that the atom belongs to. The integral has to be taken over the electron concentration associated with a single atom.

The atomic form factor is a measure of the scattering power of the *j*th atom in the unit cell. If the charge distribution has a spherical symmetry, one can use spherical coordinates chosen such that the polar direction is along ***G***. In this case, $dV = 2\pi\rho^2 \sin\varphi d\rho d\varphi$, $\boldsymbol{G}\cdot\boldsymbol{\rho} = |\boldsymbol{G}|\cdot|\boldsymbol{\rho}|\cdot\cos\varphi = G\rho\cos\varphi$, where $\varphi$ is the angle between $\boldsymbol{\rho}$ and ***G***, and the atomic form factor becomes

$$f_j = 2\pi\int_0^\infty n_j(\rho)\rho^2 d\rho \int_0^\pi \exp(-iG\rho\cos\varphi)\sin\varphi d\varphi = 4\pi\int_0^\infty n_j(\rho)\rho^2(\sin G\rho / G\rho) d\rho\,. \tag{9}$$

The atomic form factor decreases rapidly with the distance and, in the limit $\rho \to 0$, when $\sin G\rho / G\rho \to 0$,

$$f_j \to 4\pi\int_0^\infty n_j(\rho)\rho^2 d\rho = Z\,, \tag{10}$$

where $Z$ is the number of electrons in an atom. Also, when $\boldsymbol{G} = \Delta\boldsymbol{k} = 0$ (for a diffracted ray collinear with the incident ray), the phase difference vanishes and again $f_j(G = 0) = Z$.

*f* can be viewed as the ratio of the radiation amplitude scattered by the electron distribution in an atom to that scattered by one electron localized at the same point as the atom. The overall electron distribution in a solid, as obtained from X-ray diffraction experiments, is almost the same as for free atoms, i.e. atoms in which the outermost (valence) electrons are not redistributed in forming the solid. X-ray diffraction experiments are thus not very sensitive to small redistributions of electrons.

**Example**: consider a bcc lattice as a sc lattice with a basis consisting of two atoms at [[000]] and [[1/2,1/2,1/2]]. The primitive lattice vectors for the Bravais and the reciprocal lattices are in this case $\boldsymbol{a}_1 = a\boldsymbol{x}$, $\boldsymbol{a}_2 = a\boldsymbol{y}$, $\boldsymbol{a}_3 = a\boldsymbol{z}$, and $\boldsymbol{b}_1 = (2\pi/a)\boldsymbol{x}$, $\boldsymbol{b}_2 = (2\pi/a)\boldsymbol{y}$, $\boldsymbol{b}_3 = (2\pi/a)\boldsymbol{z}$, respectively. The diffraction peak of the sc lattice that is labeled by (*mnp*) corresponds to $\boldsymbol{G} = m\boldsymbol{b}_1 + n\boldsymbol{b}_2 + p\boldsymbol{b}_3 = (2\pi/a)(m\boldsymbol{x} + n\boldsymbol{y} + p\boldsymbol{z})$ and for this diffraction peak

$$\begin{aligned} S_{mnp} &= \sum_{j=1}^{2} f_j \exp(i\boldsymbol{G}\cdot\boldsymbol{r}_j) = f_1 \exp[i(2\pi/a)(m\boldsymbol{x}+n\boldsymbol{y}+p\boldsymbol{z})\cdot\boldsymbol{0}] \\ &\quad + f_2 \exp[i(2\pi/a)(m\boldsymbol{x}+n\boldsymbol{y}+p\boldsymbol{z})\cdot(a/2)(\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z})] \\ &= f_1 + f_2 \exp[i\pi(m+n+p)] \end{aligned} \tag{11}$$

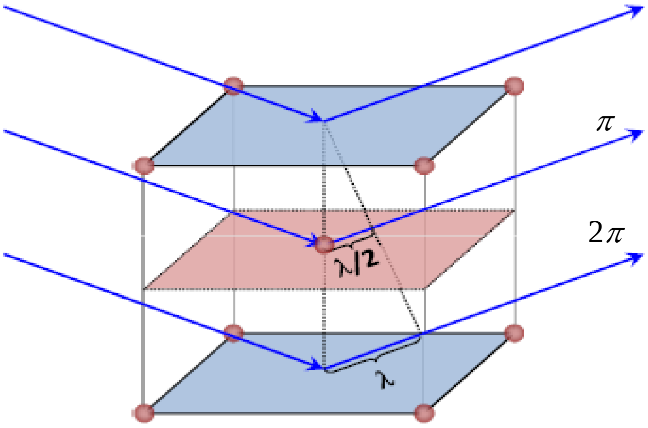
The bcc diffraction intensity is given by

$$I_{mnp} \propto |S_{mnp}|^2 = f_1^2 + f_2^2 + 2\,\mathrm{Re}[f_1 f_2 \exp[i\pi(m+n+p)]]\,. \tag{12}$$

If $f_1 = f_2 = f$,

$$I_{mnp} \propto 2f^2[1+\exp[i\pi(m+n+p)]] = \begin{cases} 4f^2, \text{ if } m+n+p = \text{even} \\ 0, \quad \text{if } m+n+p = \text{odd} \end{cases} \tag{13}$$

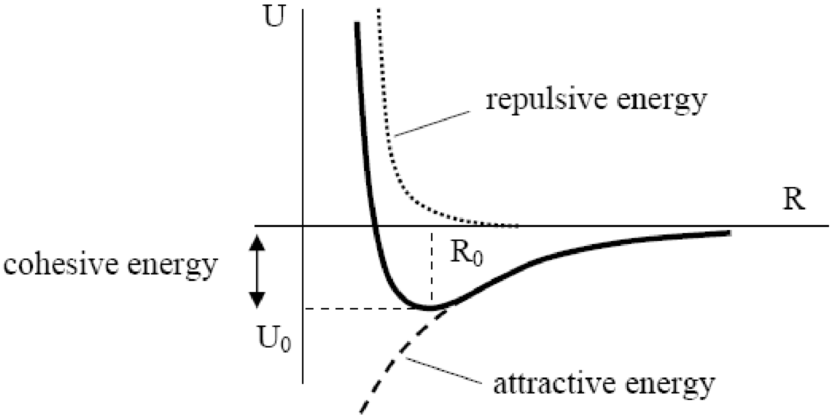
So, for the bcc structure with the same type of atoms, the (*mnp*) diffraction peaks of the sc lattice disappear whenever $m+n+p$ is an odd integer. In particular, it disappears for a (100) reflection (see the figure below) since the phase difference between successive planes is $\pi$, and the reflected amplitudes from two adjacent planes are out-of-phase/destructive interference occurs.



**Observation:** for a sc lattice with one atom in the basis, the diffraction intensity would have been the same, irrespective of the parity (even or odd) of $m+n+p$. This example illustrates the effect of the basis on the diffraction intensity.

# Crystal binding

The stability of solid state materials is assured by the existing interactions (attractive and repulsive) between the atoms in the crystal. The crystal itself is definitely more stable than the collection of the constituent atoms. This means that there exist attractive interatomic forces and that the energy of the crystal is lower than the energy of the free atoms. On the other hand, repulsive forces must exist at small distance in order to prevent the collapse of the material. One measure of the strength of the interatomic forces is the so-called **cohesive energy** of the crystal, defined as the difference between the energy of free atoms and the crystal energy. Similarly, the cohesive energy per atom $U_0$ is defined as the ratio between the cohesive energy of the crystal and the number of atoms. Typical values of the cohesive energy per atom range from 1 to 10 eV/atom, with the exception of inert gases, where the cohesive energy is about 0.1 eV/atom. In particular, the cohesive energy determines the melting temperature of solid state materials. Crystals with $|U_0| < 0.5$ eV have weak crystal bindings, while the others are characterized by strong crystal bindings.



As shown in the figure above, the potential/binding energy *U*, which describes the interaction between two atoms, approach 0 (or infinity) for an interatomic distance $R \to \infty$ (or to 0), and has a minimum at a certain distance $R = R_0$. It is composed of an attractive energy part, dominant at $R > R_0$, and a repulsive energy part that prevails at $R < R_0$. Then, the most stable state of the system, which occurs at the lowest possible energy, is characterized by the cohesive energy $U_0$, the corresponding interatomic distance, $R_0$, being known as the

**equilibrium interatomic distance**. The last parameter has typical values of 2–3 Å, which implies that the stability of the crystal is determined by short-range forces.

The interatomic force, defined as

$$F(R) = -\partial U / \partial R, \tag{1}$$

is negative (attractive) for $R > R_0$, and positive (repulsive) for $R < R_0$. The attractive and repulsive forces, which have different origins, cancel each other at the equilibrium interatomic distance.

The general form of the potential energy is

$$U(r) = \frac{A}{r^n} - \frac{B}{r^m}, \quad \text{with } n > m. \tag{2}$$

The repulsive force between atoms in the solid has the same origin in all crystals: Pauli exclusion principle, which forbids two electrons to occupy the same orbital (the same quantum state). The repulsive force is characterized (see the formula above) by the power-law expression $U = A / r^n$, with $n > 6$ or, sometimes, by the exponential expression $U = \lambda \exp(-r / \rho)$, where $\lambda$ and $\rho$ are empirical constants that can be determined from the lattice parameters and the compressibility of the material. Which expression is better suited to describe the repulsive force depends on which one better fits with experimental values. The repulsive potential is short-ranged and thus it is effective only for nearest neighbors.

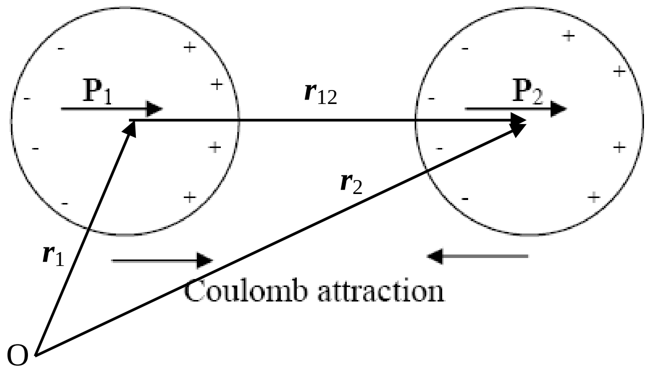
The attractive forces create **bonds** between atoms/molecules in the solid, which guarantee the crystal stability and are of different types depending on the crystal. Only the outer (valence) electrons participate in the bonding. There are several types of bonding, depending on the mechanism responsible for crystal cohesion: ionic, covalent and metallic, which give rise to strong crystal bindings, and hydrogen bonding and van der Waals interaction, which determine weak crystal bindings.

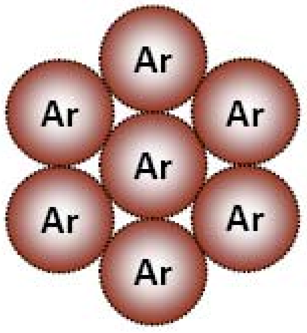
## Crystal binding in inert/noble gases. Van der Waals-London interaction

The crystals of inert gases have low cohesion energy and melting temperature, and high ionization energies. They are the simplest crystals, with an electron distribution close to that of free atoms. From an electrical point of view they are isolators, and from an optical point of view, are transparent in the visible domain. The weak binding between the constituent atoms favors compact crystalline structures, in particular fcc Bravais lattices with one atom in the basis (the only exceptions are $He^3$ and $He^4$, which crystallize in the hcp crystal structure).

Individual atoms of Ne, Ar, Kr, or Xe have completely occupied external shells, with a spherically symmetric electronic charge distribution. In crystals, the presence of other atoms induces a redistribution of the electric charge and a perturbation of the spherical charge symmetry that can be described within the model of fluctuating dipoles. Coulomb attraction can occur between two neutral spheres, as long as their internal charges polarize the spheres.

In a classical formalism (valid since electrostatic forces have a long range), this model assumes that the movement of the electron in atom 1 induces an instantaneous dipole moment $\boldsymbol{p}_1$ which generates an electric field





$$\boldsymbol{E}(\boldsymbol{r}_{12}) = -\frac{1}{4\pi\varepsilon_0}\left(\frac{\boldsymbol{p}_1}{r_{12}^3} - \frac{3(\boldsymbol{p}_1 \cdot \boldsymbol{r}_{12})}{r_{12}^5}\boldsymbol{r}_{12}\right) \tag{3}$$

at the position of atom 2 separated from atom 1 through a distance $r_{12} = |\boldsymbol{r}_{12}|$. This electric field induces a fluctuating dipole in atom 2 (the distance between the atoms as well as the magnitude and direction of $\boldsymbol{p}_1$ fluctuate in time), with a moment

$$\boldsymbol{p}_2 = \alpha \boldsymbol{E}(\boldsymbol{r}_{12}), \tag{4}$$

where $\alpha$ is the atomic polarizability. The energy of the dipole-dipole interaction between the two fluctuating dipoles is

$$U_{attr}(\boldsymbol{r}_{12}) = -\boldsymbol{p}_2 \cdot \boldsymbol{E}(\boldsymbol{r}_{12}) = \frac{1}{4\pi\varepsilon_0}\left(\frac{\boldsymbol{p}_1 \cdot \boldsymbol{p}_2}{r_{12}^3} - 3\frac{(\boldsymbol{p}_1 \cdot \boldsymbol{r}_{12})(\boldsymbol{p}_2 \cdot \boldsymbol{r}_{12})}{r_{12}^5}\right), \tag{5}$$

and its minimum value is attained when $\boldsymbol{p}_1 \parallel \boldsymbol{p}_2 \parallel \boldsymbol{r}_{12}$, case in which, replacing the value of $\boldsymbol{p}_2$ in (5) with its expression in (4), we get

$$U_{attr,\min}(r_{12}) = -\left(\frac{1}{4\pi\varepsilon_0}\right)^2 \frac{4\alpha p_1^2}{r_{12}^6} = -\frac{C}{r_{12}^6}\,. \tag{6}$$

This van der Waals (or London) interaction is the dominant attractive interaction in noble gases. The higher-order contributions of the dipole-quadrupole and quadrupole-quadrupole interactions are characterized by the respective potentials $-C_1/r_{12}^8$ and $C_2/r_{12}^{10}$, and do not contribute significantly to the cohesion energy of the noble gases crystals. The same $-C/r_{12}^6$ dependence of the energy is recovered in a quantum treatment, in the second-order perturbation theory.

Assuming a power-law expression for the repulsive forces with $n = 12$, the interaction potential is given by the Lenard-Jones formula

$$U(r_{12}) = 4\gamma\left[\left(\frac{\sigma}{r_{12}}\right)^{12} - \left(\frac{\sigma}{r_{12}}\right)^{6}\right], \tag{7}$$

where the parameters $\gamma$ and $\sigma$ are determined from X-ray and cohesion energy experiments.

The interaction energy of atom 1 (atom *i*, in general) with all other atoms in the crystal is then

$$U_i = \sum_{j\neq i} U(r_{ij}) = \sum_{j\neq i} 4\gamma\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{8}$$

and the energy of the crystal composed of *N* atoms is $U_{cryst} = (N/2)U_i$. For a periodic arrangement of atoms in the lattice, with nearest-neighbors at a distance *R*, $r_{ij} = p_{ij}R$ and

$$U_{cryst} = 2N\gamma\left[S_{12}\left(\frac{\sigma}{R}\right)^{12} - S_6\left(\frac{\sigma}{R}\right)^{6}\right] \tag{9}$$

where

$$S_6 = \sum_{j \neq i} \left( \frac{1}{p_{ij}} \right)^6 , \quad S_{12} = \sum_{j \neq i} \left( \frac{1}{p_{ij}} \right)^{12} \tag{10}$$

are rapid convergent series, that can be calculated after the crystalline structure is determined by X-ray measurements. Their values are, respectively, 12.132 and 14.454 for the fcc structure, with almost the same values for hcp structures.

The crystal energy is minimum value for the $R$ value which is the solution of $\partial U_{cryst} / \partial R = 2N\gamma[12S_{12}(\sigma / R)^{11} - 6S_6(\sigma / R)^5] = 0$, i.e. for

$$R_0 = \sigma(2S_{12} / S_6)^{1/6} . \tag{11}$$

The ratio $R_0 / \sigma = 1.09$ for a fcc Bravais lattice, the corresponding cohesion energy per atom (at zero temperature and pressure) being

$$U_0 = U_{cryst}(R_0) / N = 2\gamma \left[ S_{12} \left( \frac{S_6}{2S_{12}} \right)^2 - S_6 \left( \frac{S_6}{2S_{12}} \right) \right] = -\gamma \frac{S_6^2}{2S_{12}} = -8.6\gamma . \tag{12}$$

Quantum corrections reduce the binding energy above by 28%, 10%, 6%, and 4% for Ne, Ar, Kr, and Xe, respectively. The quantum corrections are more important for inert gas crystals with smaller equilibrium interatomic distance (smaller lattice constants).

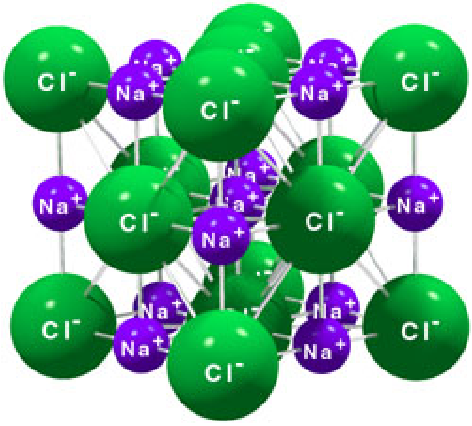
The above model determines also the **compressibility modulus** of noble gases with volume $V$ (and volume per atom $v = V / N = R^3 / \sqrt{2}$), defined at low temperatures as

$$B_0 = -V \left( \frac{\partial p}{\partial V} \right)_{T=const} = V \left( \frac{\partial^2 U_{cryst}}{\partial V^2} \right)_{R=R_0} = v \left( \frac{\partial^2 U}{\partial v^2} \right)_{R=R_0} = 4 \frac{\gamma}{\sigma^3} S_{12} \left( \frac{S_6}{S_{12}} \right)^{5/2} = 75 \frac{\gamma}{\sigma^3} . \tag{13}$$

| | **Ne** | **Ar** | **Kr** | **Xe** |
|---|---|---|---|---|
| $R_0$ (Å) | 3.05 | 3.74 | 4 | 4.34 |
| $U_0$ (eV) | -0.024 | -0.085 | -0.118 | -0.171 |
| $T_{\mathrm{melt}}$ (K) | 24 | 84 | 117 | 161 |
| $\gamma$ (eV) | 0.031 | 0.01 | 0.014 | 0.02 |
| $\sigma$ (Å) | 2.74 | 3.4 | 3.05 | 3.98 |
| $B_0$ ($10^9$ Pa) | 1.45 | 2.95 | 3.48 | 3.7 |

## Ionic binding

The ionic binding is found in ionic crystals formed from positive and negative ions, for example $Na^+$ and $Cl^-$ in NaCl. In this bonding type, electrons are transferred from the low electronegative atom, which becomes a positive ion, to the high electronegative atom, which is transformed into a negative ion (see the figure below).

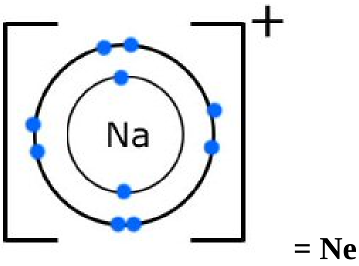


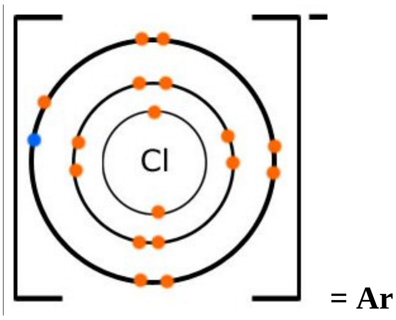
The **electronegativity** is the average of the first ionization energy and the electron affinity. It measures the ability of an atom or molecule to attract electrons in the context of a chemical bond. In NaCl the **ionization energy** (actually the first ionization energy $E_i$, which is the energy required to move an electron from a neutral isolated atom to form an ion with one positive charge: $Na + E_i \rightarrow Na^+ + e^-$) of Na is 5.14 eV and the **electron affinity** (the energy $E_a$ absorbed when an electron is added to a neutral isolated atom to form an ion with one negative charge: $Cl + e^- \rightarrow Cl^- + E_a$) of Cl is 3.56 eV. The electron affinity is negative if energy is released in the process. For most elements the electron affinity is negative, but it takes positive values in atoms with a complete shell. The net energy cost of the ionic bonding (i.e. the difference between the energy of the ions and that of the two atoms) is then $E_i - E_a = 5.14\text{ eV} - 3.56\text{ eV} = 1.58\text{ eV}$ per pair of ions, without taking into account the Coulomb energy between the ions.

In general, the electronegativity increases with the group number in the periodic element table, from the first to the seventh group (elements in the eight group have complete shells). Depending on the difference in electronegativity between two atoms, the bonding between them is

- Ionic (for large difference). Example: Na-Cl.
- Polar covalent bonding (for moderate difference). Example: H-O.
- Covalent bonding (for small difference). Examples: C-O, O-O

In ionic crystals the bonding is achieved by the long-range electrostatic force and so, a classical treatment is meaningful. The electronic configuration of the ions is similar to that of inert/noble gases, i.e. the electronic charge has a spherical symmetry, which is only slightly perturbed in crystal. The perturbations are localized in the regions in which the ions are closer. In particular, in NaCl the electronic configurations of the $Na^+$ and $Cl^-$ ions are similar to that of noble gases $Ne^{10}$ ($1s^2 2s^2 2p^6$) and $Ar^{18}$ ($1s^2 2s^2 2p^6 3s^2 3p^6$), respectively (see below).





In ionic crystals, the cohesion energy $U_0$ is no longer equal to the difference between the attractive and the repulsive potentials that act upon an ion at the equilibrium position, denoted in this case by $U_{\min}$ (and which still determines the echilibrium interatomic distance), but has a correction term equal to $E_a - E_i$, such that the difference between the energy of free atoms and of the ions in the crystal (which defines the cohesion energy) is $U_0 = U_{\min} + E_a - E_i$. In other words,

$$\mathrm{Na} + \mathrm{Cl} \rightarrow \underbrace{\mathrm{Na}^{+}\mathrm{Cl}^{-}}_{\text{crystal}} + U_0 + E_a - E_i$$

and $U_{\min} + E_a - E_i$ is the energy released per molecule when the neutral constituents form a ionic crystal..

The Coulomb force between one positive Na ion and one negative Cl ion, separated by a distance $R$ is given by

$$F_{Coulomb} = -\frac{e^2}{4\pi\varepsilon_0 R^2} \tag{14}$$

with $R$ = 2.81 Å the nearest-neighbor distance in NaCl, so that the respective attractive potential energy,

$$U_{Coul} = -\frac{e^2}{4\pi\varepsilon_0 R}, \tag{15}$$

equals −5.12 eV per pair. It follows then that the net energy gain in the ionic bonding, is 5.12 eV – 1.58 eV = 3.54 eV per pair of ions.

The electrostatic energy gain per NaCl molecule in a fcc crystal is obtained by adding different contributions:

- that of the (opposite type) 6 nearest-neighbors of a certain ion, $U_1 = -6\frac{e^2}{4\pi\varepsilon_0 R}$,
- that of the 12 second nearest-neighbors (of the same ion type), $U_2 = 12\frac{e^2}{4\pi\varepsilon_0 R\sqrt{2}}$,
- that of the 8 third nearest neighbors of opposite type, $U_3 = -8\frac{e^2}{4\pi\varepsilon_0 R\sqrt{3}}$, and so on.

The result is

$$U_{ion} = -\frac{e^2}{4\pi\varepsilon_0 R}\left(6 - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - .....\right) = -\frac{e^2}{4\pi\varepsilon_0 R}\sum_{j\neq i}\frac{(\pm)}{p_{ij}} \tag{16}$$

The series above converge eventually to $U_{ion} = -1.748\frac{e^2}{4\pi\varepsilon_0 R} = -M\frac{e^2}{4\pi\varepsilon_0 R}$, where $M$ is the Madelung constant, which takes specific values for each crystal structure. For other crystal structures: CsCl, zinc blende, and wurtzite, we have, respectively, $M$ = 1.763, 1.638 and 1.641. (If the series is slowly convergent or even divergent, the terms in the sum are rearranged such that the terms corresponding to each cell cancel each other – the cell remains neutral in charge.) The total attractive energy in a NaCl crystal with $N$ ion pairs is given by $U_{attr} = 2U_{ion} \times N/2$, where the factor 2 in the numerator accounts for the fact that there are two types of ions: Na and Cl, and the factor 2 in the denominator is introduced in order to count every ion pair only once. For NaCl, $U_{attr}$ = 861 kJ/mol (experiments give 776 kJ/mol). The discrepancy (of about 10%) between the experimental and theoretical values is explained by the existence of the (non-classical) repulsive forces.

Similarly, if we add up the repulsive potential felt by an atom from all others (the exponential form is used now), we obtain

$$U_{rep} = \sum_{j\neq i} \lambda \exp(-r_{ij}/\rho) \cong z\lambda \exp(-R/\rho) \tag{17}$$

where we consider that $\rho << R$ and $z$ is the number of nearest neighbors. The interaction energy of the whole crystal, which consists of $N$ ion pairs/molecules is

$$U_{cryst}(R) = N\left( z\lambda \exp(-R/\rho) - \frac{Me^2}{4\pi\varepsilon_0 R}\right), \tag{18}$$

and its minimum value per molecule,

$$U_{\min} = \frac{U_{cryst,0}}{N} = -\frac{e^2 M}{4\pi\varepsilon_0 R_0}\left(1 - \frac{\rho}{R_0}\right), \tag{19}$$

occurs for the equilibrium interatomic distance $R_0$ found from the condition $dU_{cryst}/dR = -(z\lambda/\rho)\exp(-R/\rho) + Me^2/(4\pi\varepsilon_0 R^2) = 0$. The first term (the Madelung term) in (19), which expresses the electrostatic contribution of the interactions, is dominant since $\rho << R_0$.

With the same definition as above, the compressibility modulus takes now the form

$$B_0 = \frac{1}{9fR_0}\frac{d^2U}{dR^2}\bigg|_{R=R_0} = \frac{e^2 M}{36\pi\varepsilon_0 fR_0^4}\left(\frac{R_0}{\rho} - 2\right) \tag{20}$$

where $f = v/R^3$ is the ratio between the volume per particle $v$ and the third power of the nearest-neighbor distance. $f$ = $8/3\sqrt{3}$, 2 and $16/3\sqrt{3}$ for CsCl, NaCl, and zinc blende, respectively.

| | LiF | LiCl | NaF | NaCl | KF | KCl | RbF | RbCl |
|---|---|---|---|---|---|---|---|---|
| $R_0$ (Å) | 2.01 | 2.56 | 2.31 | 2.82 | 2.67 | 3.15 | 2.82 | 3.29 |
| $U_0$ (eV) | -10.5 | -8.45 | -9.31 | -7.86 | -8.23 | -7.1 | -7.85 | -6.84 |
| $B_0$ ($10^{10}$ Pa) | 6.71 | 2.98 | 4.65 | 2.4 | 3.05 | 1.75 | 2.62 | 1.56 |
| $\lambda$ ($10^2$ eV) | 3.08 | 5.1 | 6.24 | 10.93 | 13.63 | 21.35 | 18.54 | 33.24 |
| $\rho$ (Å) | 0.291 | 0.33 | 0.29 | 0.321 | 0.298 | 0.326 | 0.301 | 0.323 |

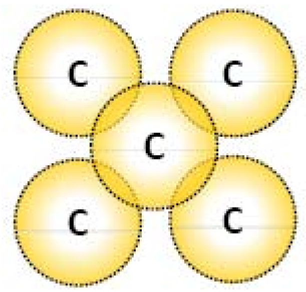
## Covalent bonding

The covalent bonding forms in molecules composed of identical particles, for example hydrogen. In this case two atoms form a (homopolar) bond by sharing a pair of electrons (one from each atom, with opposite spins). Most atoms can form more than one covalent bond. For example, C has four outer electrons (of $2sp^3$ type) and thus can form 4 covalent bonds. The covalent bond is highly directional and different bonds repel each other. Therefore, the corresponding crystal has generally a low packing ratio. For example, C and Si can have diamond structure, with atoms joined to 4 nearest neighbors at tetrahedral angles; this structure has a packing ratio of only 0.34 compared to 0.74 for close-packed structures. The electrons in covalent bonds are strongly localized along the bond, so that the crystals are semiconductors or isolators, with not very good electrical conductivity.

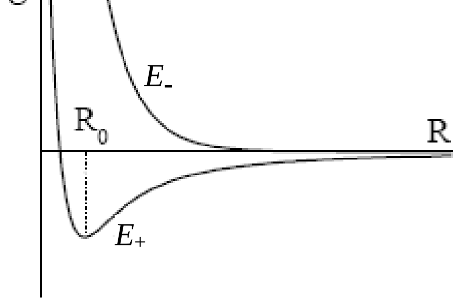
To describe the covalent bonding in hydrogen, we introduce the normalized atomic orbitals 1*s* for the *j* (*j* = 1,2) electron that can belong to either atom A or B as $\psi_{1s}^{A,B}(j)$, so that the normalized wavefunction of the total system can be either symmetric (labeled with +) or antisymmetric (labeled with −)

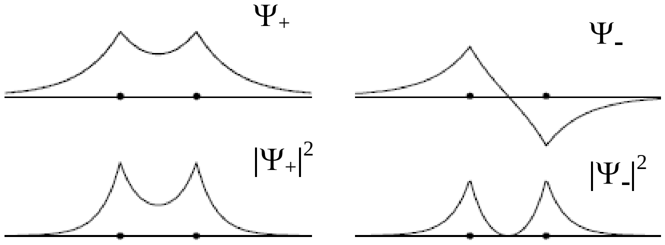
$$\Psi_+ = [1/\sqrt{2(1+S_{AB}^2)}][\psi_{1s}^A(1)\psi_{1s}^B(2)+\psi_{1s}^B(1)\psi_{1s}^A(2)] \qquad (21a)$$

$$\Psi_- = [1/\sqrt{2(1-S_{AB}^2)}][\psi_{1s}^A(1)\psi_{1s}^B(2)-\psi_{1s}^B(1)\psi_{1s}^A(2)] \qquad (21b)$$



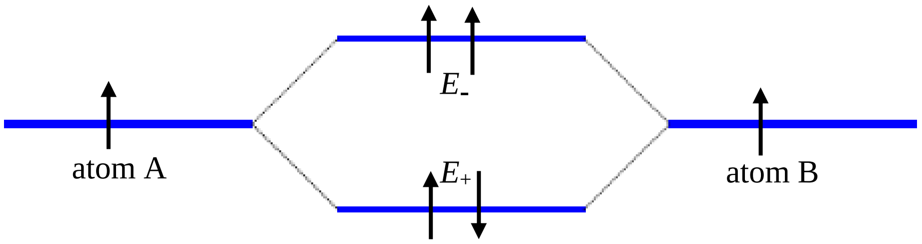
where $S_{AB} = \int (\psi_{1s}^A(1))^* \psi_{1s}^B(2) d\mathbf{r}$ is the overlap integral. Note that the symmetric wavefunction for ionic crystals can be expressed as $\Psi_+ \propto [\psi^A(1)\psi^A(2)+\psi^B(1)\psi^B(2)]$.



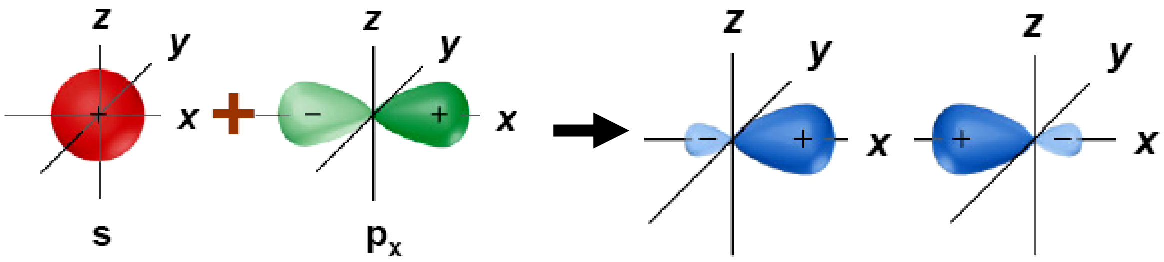


The symmetric wavefunction (also called singlet) corresponds to two antiparallel spin, with quantum number $S = 0$ of the operator $\boldsymbol{S}^2$ (with $\boldsymbol{S}$ the total spin), while the antisymmetric wavefunction (also called triplet) corresponds to parallel spins, i.e. $S = 1$ (with the spin projection quantum number $m_s = -1$, 0, and 1; there are three antisymmetric wavefunctions!). The form of the wavefunctions above is determined from the condition that the total wave function for fermions (including spin) must be antisymmetric upon particle exchange.

The energy eigenvalues are represented above as a function of the distance between the atoms. A bound state can exist in the singlet state, with $E_+ = -3.14$ eV if the covalent bonding forms between H atoms, i.e. the strongest binding occurs if the spins of the two electrons are antiparallel.

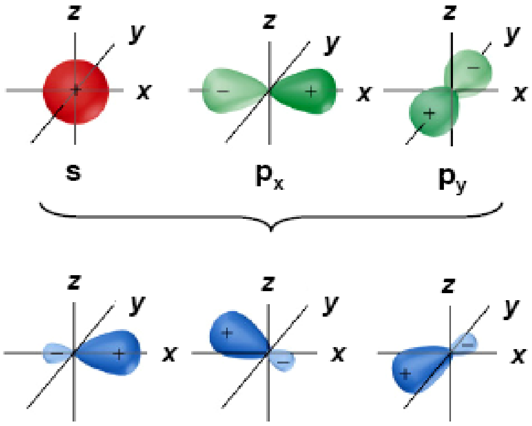


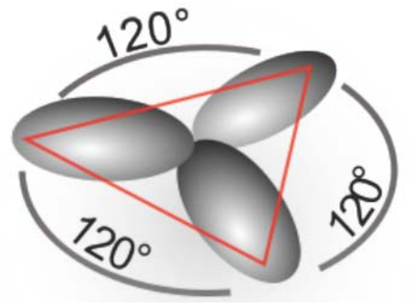
To characterize the crystalline structure of diamond one must generalize the previous formula in order to incorporate the *p* atomic orbitals. Indeed, the last occupied orbitals of these materials are: C($2s^22p^2$), Si($3s^23p^2$), and Ge($4s^24p^2$). When both *s* and *p*-type orbitals are involved, they hybridize (see figures below). [The *s* atomic orbitals have quantum numbers $n = 1,2,3\ldots$ (principal quantum number), $l = 0$ (orbital quantum number), and $m = 0$ (magnetic quantum number; the projection of $l$). The *p* orbitals have $n = 1,2,3\ldots$, $l = 1$, and $m = -1,0,1$.] The *s* and *p* atomic orbitals hybridize when the energy difference between them is much smaller than the binding energy.



In particular, when one *s* orbital with wavefunction $\Psi_s$ and one *p* orbital, say the $p_x$ orbital, with wavefunction $\Psi_{p_x}$, hybridize, the result is two linear *sp* orbitals (see figure above) with wavefunctions

$$\Psi_1 = 2^{-1/2}(\Psi_s + \Psi_{p_x}), \qquad \Psi_2 = 2^{-1/2}(\Psi_s - \Psi_{p_x}). \tag{22}$$





On the contrary, $sp^2$ hybrid orbitals form between one *s* orbital and two *p* orbitals, the resulting planar structure (see figure above) having orbitals arranged in plane with an angle of 120° between them. The electrons in the hybrid orbitals are strongly localized and form $\sigma$ bonds; they do not participate in electrical conduction. One *p* orbital remains perpendicular to the plane, where it forms a $\pi$ bond with other out-of-plane *p* orbitals from neighboring atoms; this is the case of graphite or graphene (bidimensional crystal). The electrons in the $\pi$ orbitals are delocalized and participate in electrical conduction. The three hybrid orbitals are given by

$$\Psi_1 = 3^{-1/2}(\Psi_s + \sqrt{2}\Psi_{p_x}), \tag{23}$$
$$\Psi_2 = 3^{-1/2}[\Psi_s - (1/\sqrt{2})\Psi_{p_x} + \sqrt{3/2}\Psi_{p_y}], \quad \Psi_3 = 3^{-1/2}[\Psi_s - (1/\sqrt{2})\Psi_{p_x} - \sqrt{3/2}\Psi_{p_y}].$$

Similarly, the electronic configurations that forms from one *s* orbital and three *p* orbitals is called $sp^3$. This electronic configuration is characteristic for diamond. The angular part of the *s* and *p* orbitals are (in polar coordinates)

$$\Psi_s = (4\pi)^{-1/2}, \tag{24}$$
$$\Psi_{p_x} = (3/4\pi)^{1/2}\sin\theta\cos\varphi, \quad \Psi_{p_y} = (3/4\pi)^{1/2}\sin\theta\sin\varphi, \quad \Psi_{p_z} = (3/4\pi)^{1/2}\cos\theta$$

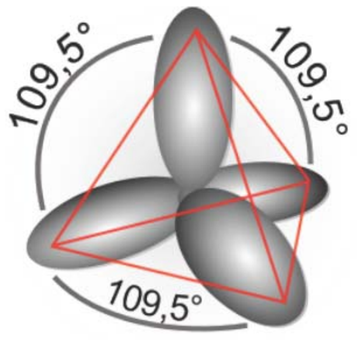
so that the four hybrid atomic orbitals that are linear combinations of atomic orbitals form a tetrahedron (see figure below) and are given by

$$\Psi_1 = (1/2)(\Psi_s + \Psi_{p_x} + \Psi_{p_y} + \Psi_{p_z}), \qquad (25a)$$

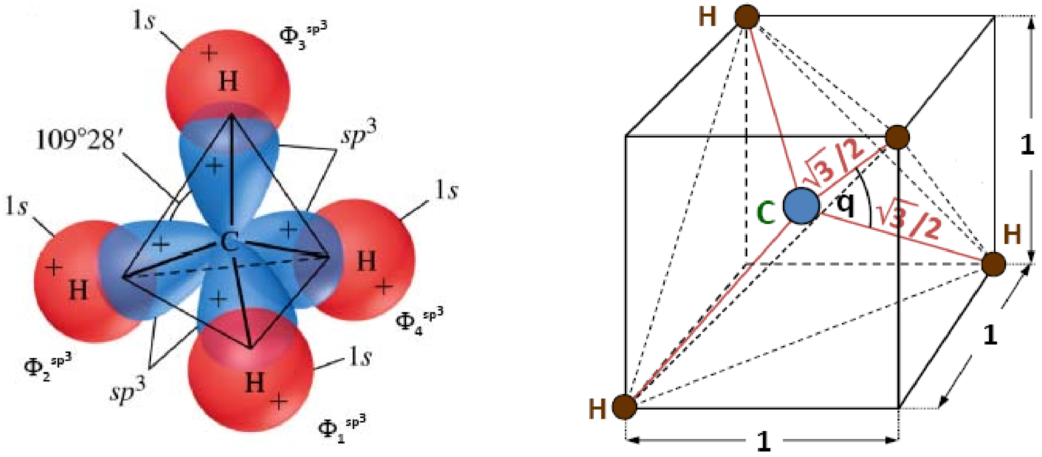
$$\Psi_2 = (1/2)(\Psi_s + \Psi_{p_x} - \Psi_{p_y} - \Psi_{p_z}), \qquad (25b)$$

$$\Psi_3 = (1/2)(\Psi_s - \Psi_{p_x} + \Psi_{p_y} - \Psi_{p_z}), \qquad (25c)$$

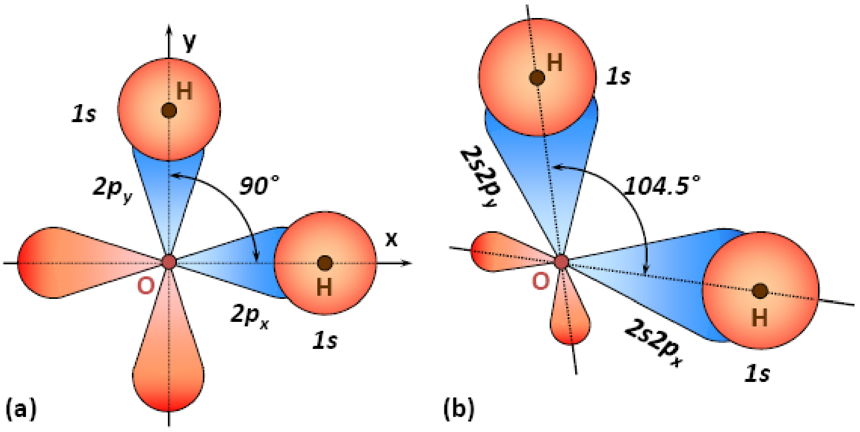
$$\Psi_4 = (1/2)(\Psi_s - \Psi_{p_x} - \Psi_{p_y} + \Psi_{p_z}). \qquad (25d)$$



C, Si and Ge form crystals in which the covalent binding is dominant, the van der Waals contribution to the cohesion energy, also encountered in crystals from a single element, being negligible. However, in crystals with a basis composed of two atoms A and B, with *n* and, respectively, $8-n$ valence electrons, the covalent binding is accompanied by a ionic contribution. The resulting bond is called polar covalent bond. The ionic contribution (in fraction) is 0.18 in SiC, 0.26 in GaSb, 0.32 in GaAs, and 0.44 in InP. Similarly, in ionic crystals the covalent binding can also contribute to the cohesion energy, the fraction of the ionic contribution being only 0.86 in AgCl, 0.94 in NaCl, and 0.96 in RbF. When covalent bonding forms between different atoms, the hybrid orbitals considered above are modified, as can be seen from the figures below.



Electronic configuration in the $CH_4$ molecule.



Bonding between the 1*s* orbitals of the H atom and the 2$p_x$ and 2$p_y$ orbitals of O atom in $H_2O$ (a) without, and (b) with hybridization

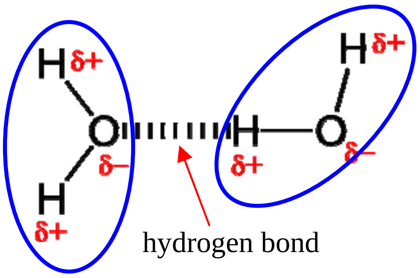
Covalent crystals are characterized by:

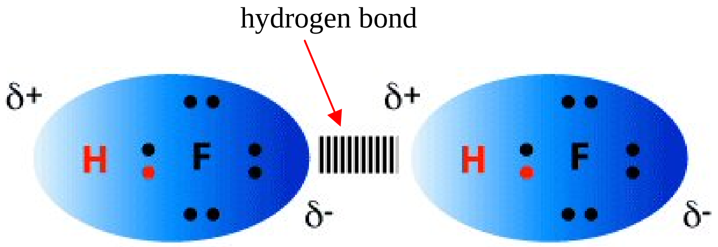
- high melting temperatures (the cohesion energy per atom is about –10 eV)
- hardness (but also friable)
- their conductivity depends strongly on temperature and impurity atoms
- high value of the dielectric constant
- generally transparent in IR, but strongly absorbent in visible and near-IR.

**Note:** Crystal bonds form between valence electrons, i.e. the electrons on the outer shells, which participate in chemical reactions/determine the physical properties of the material. In contrast, core electrons are those on inner shells.

## Hydrogen binding of crystals

Because neutral hydrogen has only one electron, it should form a covalent bond with only one other atom. However, just as oppositely charged ions are attracted to one another and can form ionic bonds, the partial charges that exist at different atoms in polar covalent bonds can interact with other partially charged atoms/molecules. Particularly strong polar covalent bonds are found, for example, when a hydrogen atom bonds to extremely electronegative ions such as O in water/ice (see the figure below, left), F (see the figure below, right), N or Cl. The partial charges in the figure below are denoted by δ. The hydrogen bond forms between the hydrogen atom with a strong partial positive charge and electronegative ions with strong partial negative charges in neighboring molecules. The binding energy is of the order of 0.1 eV. For example, the cohesion energy per molecule in ice is –0.3 eV.





The hydrogen bond is weaker than, although similar to, ionic bond since it forms between partial charges rather than full (complete) charges. In hydrogen bonds the hydrogen atom is the hydrogen bond donor and the electronegative ion is the hydrogen bond acceptor. As the polar covalent binding, the hydrogen bond can be viewed as a mixture of ionic and covalent bonding, the ionic bonding being dominant. For example, in the typical hydrogen bond that links two $H_2O$ molecules in ice, the binding can be considered as a superposition of three binding types:

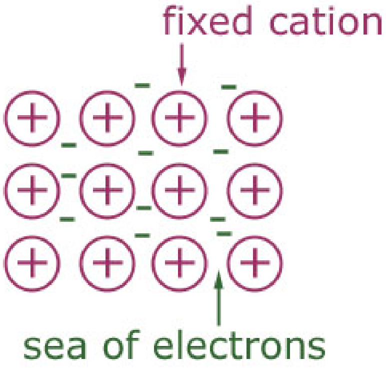
O (covalent) H (ionic) O
O (ionic) H (ionic) O
O (ionic) H (covalent) O

In hydrogen, the proton radius is with five orders of magnitude smaller than the radius of any other ion, and so it allows the existence of only two nearest neighbors of the proton (more than two atoms would get in each other’s way), i.e. the hydrogen bond is directional.

Despite it is weak, the hydrogen bond is extremely important in living organisms, which are mainly composed of water, since water as well as proteins and nucleic acids posses a great capacity to form hydrogen bonds. In particular, the hydrogen binding occurs as intra-molecular binding between the organic complementary bases thymine and adenine, and cytosine and guanine in DNA. It can also be encountered between constituents of crystals such as $KH_2PO_4$, $KD_2PO_4$ (KDP), $Ca(OH)_2$, or $Mg(OH)_2$.

## Metallic bonding

The metallic bonding can be understood as the bonding between positively charged metallic nuclei/ions and delocalized conduction electrons, seen as a “sea of free electrons”. It prevails in elements in which the valence electrons are not tightly bound with the nucleus (in metals, for example). However, in the metallic bond we cannot speak about ions, since there is no particular electron that is “lost” to another ion. Unlike other bonding types, the metallic bonding is collective in nature, so that no single “metallic bond” exists. It is neither intra- nor intermolecular since no molecule can be distinguished in metals. Metallic bonding can be understood as a nonmolecular, extremely delocalized communal form of covalent bonding. The delocalization is most pronounced for *s* and *p* electrons, with $l = 0$ and $l = 1$, respectively, being much weaker for *d* and *f* electrons, which have quantum numbers $l = 2$ and $l = 3$, respectively.



In metals, an atom achieves a more stable configuration by sharing all its valence electrons with all other atoms in the crystal. However, besides delocalization, metallic bonding also requires the availability of a far larger number of delocalized energy states than of delocalized electrons. These states are referred to as electron deficiency; they assure the kinetic energy for delocalization.

The metallic bonding is encountered, for example, in alkaline metals such as Li, K, Na, with electronic configurations that resemble those of noble gases with an additional *s* electron on the outer shell. Having few electrons on their outer shells, alkali metals have only partly filled energy levels, and therefore are electron deficient. In forming the crystal, the wavefunctions of the outer *s* electrons overlap with those of their nearest neighbors, and the electrons become delocalized. Their dynamics resembles that of free electrons, so that in alkaline metals the lattice is occupied by the positively charged ions with the noble gas structure (they occupy in fact only as much as 20% from the volume of the crystal), while the valence electrons occupy the remaining volume. Unlike in covalent crystals, where the electronic charge is distributed in a strongly nonuniform manner (the bonds are spatially oriented), the electronic density in metallic crystals is highly uniform. This explains the high elasticity and malleability of these materials.

The total Coulomb potential, which includes electron-electron, electron-ion and ion-ion interactions, is $U_{Coul} = U_{e-e} + U_{e-i} + U_{i-i} < 0$ (the first and third terms on the rhs are positive, the middle one is negative). Therefore, the attractive potential is of electrostatic nature, being balanced by the repulsive interaction due to the Pauli exclusion principle. It should be mentioned that it is not necessary for metals to have metallic bonding. For example, many transition metals show covalent properties (not all electrons participate in covalent bonds, and are good electrical conductors).

**Note:** In transition metals (Fe, Co, Ni, Cu, Zn, Ag, Au, Mn, etc.) the *d* orbitals are only partially occupied and the outermost *s* orbitals are fully occupied.
Example: $4s^2$–full, 3d–incomplete.

Crystals with metallic bonding are usually characterized by

- high electrical and thermical conductivity, with weak temperature dependence
- high elasticity
- high optical reflectivity in a large frequency bandwidth
- broad range of melting temperatures: low melting temperatures for alkaline metals (Li, Na, K, Rb, Cs), intermediate for noble metals (Cu, Ag, Au), and high values for metals such as Ti, Zr, Mo, W. The corresponding cohesion energies vary between −1 eV and −5 eV.

# Lattice oscillations. Phonons

Let us consider a crystalline material consisting of a large number $N_{ion}$ of heavy positively-charged ions (composed of the nucleus and the valence electrons on the inner atomic orbitals) with masses $M_\alpha$ and situated at positions $\boldsymbol{R}_\alpha$, $\alpha = 1,.., N_{ion}$, surrounded by and in interaction with $N_{el}$ electrons on the outer atomic orbitals with masses $m$ and at positions denoted by $\boldsymbol{r}_i$, $i = 1,.., N_{el}$. The total Hamiltonian of the system is then

$$\begin{aligned} H &= T_{el} + T_{ion} + V_{el} + V_{ion} + V_{el-ion} \\ &= \sum_i -\frac{\hbar^2}{2m}\nabla_i^2 + \sum_\alpha -\frac{\hbar^2}{2M_\alpha}\nabla_\alpha^2 + \frac{1}{2}\sum_{\substack{i,j \\ i\neq j}} U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) + \frac{1}{2}\sum_{\substack{\alpha,\beta \\ \alpha\neq\beta}} U_{ion}(\boldsymbol{R}_\alpha - \boldsymbol{R}_\beta) + \sum_{i,\alpha} V(\boldsymbol{r}_i - \boldsymbol{R}_\alpha) \end{aligned} \qquad (1)$$

The terms of the right-hand-side denote, in order, the kinetic energy of the electrons, the kinetic energy of the ions, the (Coulomb) interaction energy of electron pairs, the interaction energy of ion pairs, and the interaction energy between electrons and ions.

Since $m << M_\alpha$, the electron velocities are much higher than the ion velocities, so that the electrons "see" a "frozen" distribution of ions, while the ions can only sense the average (not instantaneous!) spatial distribution of electrons. In other words, for a given ion configuration the electrons are in a quasi-equilibrium state that is slowly varying in time due to ion's motion, whereas the ions evolve slowly in a potential distribution generated by the average configuration of the electrons. This adiabatic approximation, known also as the Born-Oppenheimer approximation, allows a factorization of the total wavefunction of the system, $\Psi(\boldsymbol{r},\boldsymbol{R})$ with $\boldsymbol{r} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_{N_{el}}\}$, $\boldsymbol{R} = \{\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_{N_{ion}}\}$ into an electronic part, $\psi(\boldsymbol{r};\boldsymbol{R})$, in which the ion's positions are considered as parameters, and into an ionic part, $\phi(\boldsymbol{R})$:

$$\Psi(\boldsymbol{r},\boldsymbol{R}) = \psi(\boldsymbol{r};\boldsymbol{R})\phi(\boldsymbol{R}). \qquad (2)$$

These electronic and ionic parts satisfy the following equations:

$$\left(\sum_i -\frac{\hbar^2}{2m}\nabla_i^2 + \frac{1}{2}\sum_{\substack{i,j \\ i\neq j}} U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) + \sum_{i,\alpha} V(\boldsymbol{r}_i - \boldsymbol{R}_\alpha)\right)\psi(\boldsymbol{r};\boldsymbol{R}) = E_{el}(\boldsymbol{R})\psi(\boldsymbol{r};\boldsymbol{R}) \qquad (3)$$

$$\left(\sum_{\alpha}-\frac{\hbar^2}{2M_\alpha}\nabla_\alpha^2+\frac{1}{2}\sum_{\substack{\alpha,\beta\\ \alpha\neq\beta}}U_{ion}(\boldsymbol{R}_\alpha-\boldsymbol{R}_\beta)+E_{el}(\boldsymbol{R})\right)\phi(\boldsymbol{R})=E\phi(\boldsymbol{R}), \tag{4}$$

where $E$ is the energy of the whole system and $E_{el}$ is the energy of the (sub)system of electrons.

Let us assume further that in a crystalline lattice with $s$ atoms in the basis, the ions move around their equilibrium positions $\boldsymbol{R}_\alpha^0$, so that $|\,\boldsymbol{R}_\alpha-\boldsymbol{R}_\alpha^0=\boldsymbol{u}_\alpha\,|<<\boldsymbol{R}_\alpha^0$. Then, the interaction energy between pairs of ions can be expanded in a Taylor series around the equilibrium positions. Taking into account that

$$\nabla U_{ion}(\boldsymbol{R}_\alpha-\boldsymbol{R}_\beta)=\left(\frac{\partial U_{ion}(\boldsymbol{R}_\alpha-\boldsymbol{R}_\beta)}{\partial\boldsymbol{R}_\alpha}\boldsymbol{u}_\alpha+\frac{\partial U_{ion}(\boldsymbol{R}_\alpha-\boldsymbol{R}_\beta)}{\partial\boldsymbol{R}_\beta}\boldsymbol{u}_\beta\right)=0 \tag{5}$$

since the force that acts upon an ion at equilibrium (which is proportional to this derivative) vanishes, we find that

$$\sum_{\substack{\alpha,\beta\\ \alpha\neq\beta}}U_{ion}(\boldsymbol{R}_\alpha-\boldsymbol{R}_\beta)=\sum_{\substack{\alpha,\beta\\ \alpha\neq\beta}}U_{ion}(\boldsymbol{R}_\alpha^0-\boldsymbol{R}_\beta^0)+\sum_{\alpha,\beta,\mu,\nu}A_{\alpha\beta}^{\mu\nu}u_\alpha^\mu u_\beta^\nu, \tag{6}$$

where

$$A_{\alpha\beta}^{\mu\nu}=\left(\frac{\partial^2U_{ion}(\boldsymbol{R}_a-\boldsymbol{R}_\beta)}{\partial R_\alpha^\mu\partial R_\beta^\nu}\right)_{R_{\alpha,\beta}=R_{\alpha,\beta}^0} \tag{7}$$

and the indices $\mu = x,y,z$ (and $\nu$) denote the projections of the position vectors on a Cartesian coordinate system, the first spatial derivative of $U_{ion}$ vanishes due to the requirement that the force (which is proportional to this derivative) that acts upon an ion at equilibrium vanishes, and higher-order terms in the Taylor expansion are neglected. The last approximation is called harmonic. Because the first term in the Taylor expansion of $U_{ion}$ is constant, it can (together with $E_{el}$) be included in the reference energy of the system, so that in the harmonic approximation the lattice dynamics is described by the Hamiltonian

$$H_{ion} = \sum_{\alpha,\mu} \frac{(p_\alpha^\mu)^2}{2M_\alpha} + \frac{1}{2} \sum_{\alpha,\beta,\mu,\nu} A_{\alpha\beta}^{\mu\nu} u_\alpha^\mu u_\beta^\nu . \tag{8}$$

The dynamics of the lattice oscillations can then be expressed in the canonical form as

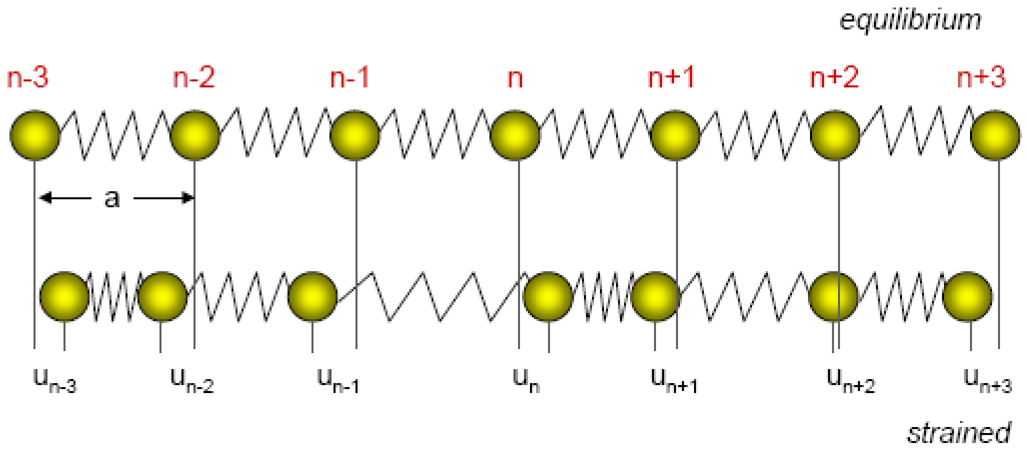
$$\dot{p}_\alpha^\mu = -\frac{\partial H_{ion}}{\partial u_\alpha^\mu} = -\sum_{\beta,\nu} A_{\alpha\beta}^{\mu\nu} u_\beta^\nu , \qquad \dot{u}_\alpha^\mu = \frac{\partial H_{ion}}{\partial p_\alpha^\mu} = \frac{p_\alpha^\mu}{M_\alpha} , \tag{9}$$

and the equation of motion for the displacement of ion $\alpha$ in the $\mu$ direction,

$$M_\alpha \ddot{u}_\alpha^\mu + \sum_{\beta,\nu} A_{\alpha\beta}^{\mu\nu} u_\beta^\nu = 0 , \tag{10}$$

describes in fact a set of coupled harmonic oscillators. The coupling strength with neighboring ions is characterized by the coefficients $A_{\alpha\beta}^{\mu\nu}$. A harmonic potential energy corresponds to forces that are linear in the displacements.

In the harmonic approximation one can view the lattice vibration as an interaction of connected elastic springs (classical harmonic oscillators), as in the figure below. The lattice oscillations are thus similar to elastic waves that propagate through such a chain of connected springs. If an atom is displaced from its equilibrium site by a small amount, it will tend to return to its equilibrium position due to the force acting on it. This results in lattice vibrations. Due to interactions between atoms, various atoms move simultaneously, so we have to consider the motion of the entire lattice.



Note that in coupled harmonic oscillators, the force that acts on an ion $\alpha$ from other ions $\beta$, given by $F_\alpha^\mu = M_\alpha \ddot{u}_\alpha^\mu = -\sum_{\beta,\nu} A_{\alpha\beta}^{\mu\nu} u_\beta^\nu$, is proportional to the relative displacement, $\boldsymbol{u}_\alpha - \boldsymbol{u}_\beta$. This means that the expression of the force should be

$$F_\alpha^\mu = -\sum_{\beta,\nu} A_{\alpha\beta}^{\mu\nu} u_\beta^\nu = -\sum_{\beta,\nu} A_{\alpha\beta}^{\mu\nu} (u_\beta^\nu - u_\alpha^\nu)\,. \tag{11}$$

This is possible only if the following equation is satisfied:

$$\sum_\beta A_{\alpha\beta}^{\mu\nu} = A_{\alpha\alpha}^{\mu\nu} + \sum_{\beta\neq\alpha} A_{\alpha\beta}^{\mu\nu} = 0\,. \tag{12}$$

The physical relevant solutions for the system of harmonic oscillators are of plane-wave type, i.e. are oscillatory in time, with the same frequency for all ions. These are the normal oscillations. Because of the periodicity of the crystalline lattice, the amplitude of ions' displacements in different unit cells are the same, so that only the phase of the oscillations vary from one unit cell to the other. So, for a crystal with *s* atoms in the basis and *N* elementary cells, $\alpha = \{\chi, n\}$, $\chi$ = 1,…,*s*, *n* = 1,…,*N*, we look for solutions of the form ($N_{ion} = sN$)

$$u_{n\chi}^\mu(t) = u_{0,\lambda}(\boldsymbol{k}) e_{\chi\lambda}^\mu(\boldsymbol{k}) \exp[i(\boldsymbol{k}\cdot\boldsymbol{R}_{n\chi}^0 - \omega_\lambda(\boldsymbol{k})t)] \tag{13}$$

where $u_{0,\lambda}(\boldsymbol{k})$ is the amplitude of the normal oscillations of type $\lambda$ (several longitudinal and transverse oscillations) that propagate along the direction $\boldsymbol{k}/|\boldsymbol{k}|$, and $\boldsymbol{e}_{\chi\lambda}(\boldsymbol{k})$ is the polarization vector of $\chi$-th atom in the unit cell (not normalized to unity!). Introducing this solution in (10), we find that the polarization vectors of the atoms satisfy the following system of 3*s* equations ($\nu$ = 1,2,3, and $\gamma$ = 1,…,*s*)

$$M_\chi \omega_\lambda^2(\boldsymbol{k}) e_{\chi\lambda}^\mu = \sum_{m,\gamma,\nu} A_{n\chi,m\gamma}^{\mu\nu} \exp[i\boldsymbol{k}\cdot(\boldsymbol{R}_{m\gamma}^0 - \boldsymbol{R}_{n\chi}^0)] e_{\gamma\lambda}^\nu = \sum_{\gamma,\nu} B_{\chi\gamma}^{\mu\nu}(\boldsymbol{k}) e_{\gamma\lambda}^\nu\,, \tag{14}$$

where $B^{\mu\nu}_{\chi\gamma}(\boldsymbol{k})=\sum_m A^{\mu\nu}_{n\chi,m\gamma}\exp[i\boldsymbol{k}\cdot(\boldsymbol{R}^0_{m\gamma}-\boldsymbol{R}^0_{n\chi})]$. In general, we have a system of *ds* equations, where *d* is the number of dimensions.

It can easily be demonstrated that the matrix of the coefficients $B^{\mu\nu}_{\chi\gamma}(\boldsymbol{k})$ is hermitic, which implies the orthogonality of the polarization vectors

$$\sum_{\chi,\mu}[e^{\mu}_{\chi\lambda}(\boldsymbol{k})]^{*}e^{\mu}_{\chi\lambda'}(\boldsymbol{k})=s\delta_{\lambda\lambda'}\,, \tag{15}$$

from which it follows that the system of 3*s* equations has non-trivial solutions only when

$$\mathrm{Det}\left|B^{\mu\nu}_{\chi\gamma}(\boldsymbol{k})-M_{\chi}\omega^2_{\lambda}(\boldsymbol{k})\delta_{\chi\gamma}\delta_{\mu\nu}\right|=0\,. \tag{16}$$

This condition represents a characteristic equation for $\omega^2_\lambda$, which has 3*s* solutions/branches for a given ***k***, called the normal oscillation frequencies of the lattice, with corresponding polarization vectors $e^{\mu}_{\chi\lambda}(\boldsymbol{k})$. The dependence of the oscillation frequencies $\omega_\lambda$ on ***k*** is called the dispersion relation of the normal oscillation of the $\lambda$-th branch.

From the definition of $B^{\mu\nu}_{\chi\gamma}(\boldsymbol{k})$, if $A^{\mu\nu}_{n\chi,m\gamma}$ are real, it follows that $B^{\mu\nu}_{\chi\gamma}(\boldsymbol{k})=[B^{\mu\nu}_{\chi\gamma}(-\boldsymbol{k})]^{*}$ and thus

$$e^{\mu}_{\chi\lambda}(\boldsymbol{k})=[e^{\mu}_{\chi\lambda}(-\boldsymbol{k})]^{*}\,,\qquad \omega^2_{\lambda}(\boldsymbol{k})=[\omega^2_{\lambda}(-\boldsymbol{k})]^{*} \tag{17}$$

or $\omega_\lambda(\boldsymbol{k})=\omega_\lambda(-\boldsymbol{k})$ since the oscillation frequencies are real and positive.

If, in three dimensions, we consider a lattice with one atom in the basis, i.e. with $s$ = 1, then we have three oscillations branches (there are three degrees of freedom for each atom), $\lambda$ = 1,2,3, and in the limit ***k*** = 0 of long wavelengths,

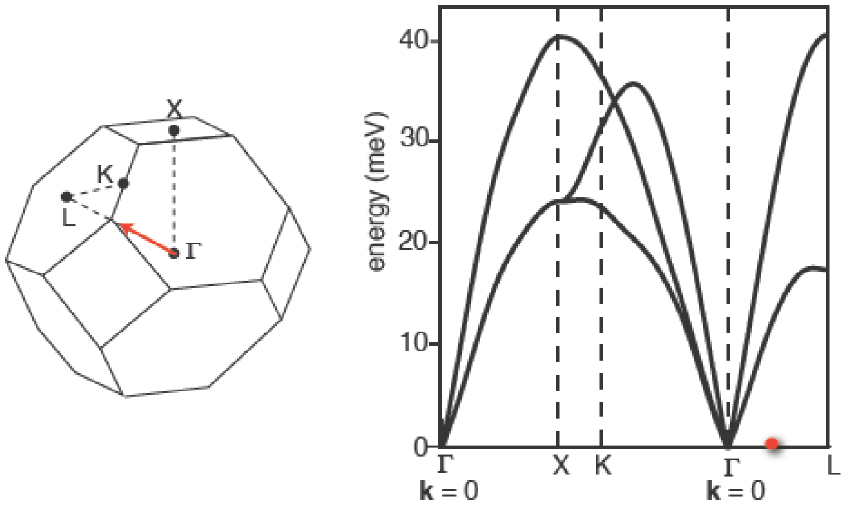
$$M_1\omega^2_{\lambda}(0)e^{\mu}_{1\lambda}(0)=\sum_{\nu}B^{\mu\nu}_{11}(0)e^{\nu}_{1\lambda}(0) \tag{18}$$

$$B^{\mu\nu}_{11}(0)=\sum_{m}A^{\mu\nu}_{n1,m1}=A^{\mu\nu}_{n1,n1}+\sum_{m\neq n}A^{\mu\nu}_{n1,m1}=0\,, \tag{19}$$

the last equality following from (12). Because the left-hand-side of (19) vanishes and the polarization vectors are finite, it follows that $\omega_\lambda(0) = 0$ and the dispersion law for the three branches in the long-wavelength limit can be written as

$$\omega_\lambda(\boldsymbol{k}) = v_{ac,\lambda}(\boldsymbol{k})k \,. \tag{20}$$

The parameters $v_{ac,\lambda}$ are called acoustic velocities since a similar relation as that above holds for acoustic waves propagating in a continuum, elastic and isotropic medium. Moreover, the oscillations that take place in the direction of the propagation vector $\boldsymbol{k}$ are called longitudinal and those normal to $\boldsymbol{k}$ are transverse: we have one longitudinal and two transverse acoustical oscillations. The figure below illustrates the dispersion relation of a crystal, in which we can identify only acoustical branches, which means that the there is only one atom in the basis of this crystal.



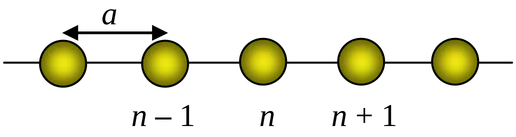
For complex lattices, with $s > 1$, there are again three oscillation branches with an acoustic-like dispersion relation in the limit $\boldsymbol{k} \to 0$, as above, corresponding to the situation in which all atoms in the lattice have the same displacements (oscillate in phase), and thus the complex structure of the lattice is not manifest. These are the acoustic oscillation branches. However, in this case we have also $3s-3$ oscillation branches with no analog in the dynamic of continuum media and for which the dispersion relation in the long-wavelength limit has the form

$$\omega_\lambda(\boldsymbol{k}) = \omega_{0\lambda} - \beta_\lambda k^2 . \tag{21}$$

These are the optical oscillation branches, characterized by the cut-off oscillation frequencies $\omega_{0\lambda}$; the parameters $\beta_\lambda$ are generally positive. This type of oscillation branches is called optical because, when the unit cell consists of ions with different type of charges (positive and negative), these oscillations form an instantaneous dipolar moment that interacts strongly with the electromagnetic radiation. In this case ions with different sign oscillate in anti-phase, i.e. their displacements are in opposite directions. As for the acoustic oscillations, in three dimensions we have one longitudinal and two transverse optical oscillation branches for each *s* value.

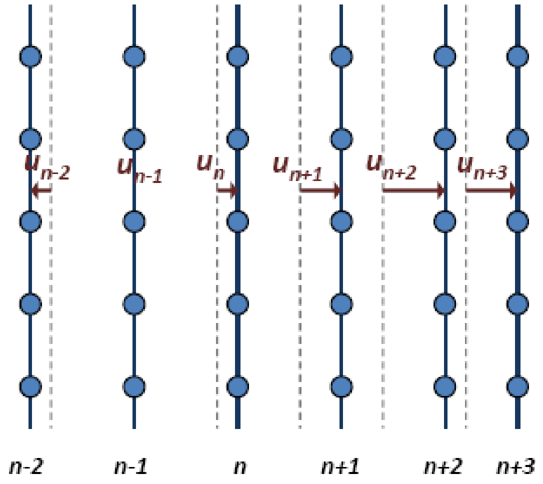
## Oscillations of an infinite atomic chain with one atom in the basis

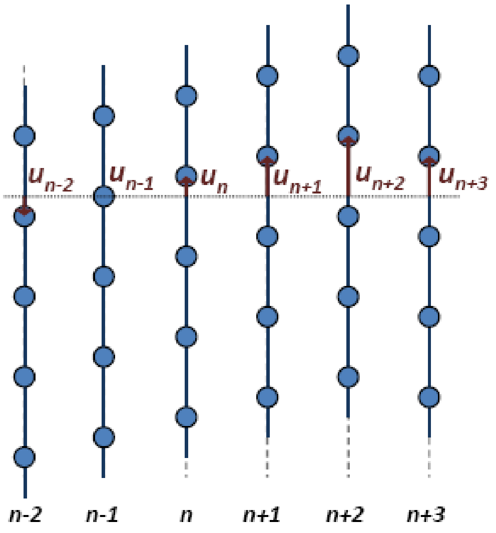
For exemplification, let us consider first a simple one-dimensional infinite lattice (an "atomic chain") consisting of identical atoms (more precisely, ions) with mass *M*, separated by the lattice constant *a*, as in the figure below. We expect $ds = 1$ (for $d = 1$ and $s = 1$), i.e. a single, acoustical oscillation branch.



For (thermal) vibrations of the crystalline lattice, in which the ions move slightly around their equilibrium positions $R_n^0 = na$, their actual positions $R_n$ satisfy the relations $| R_n - R_n^0 = u_n | << R_n^0$, where the displacements can occur either along the chain or transverse to the chain of atoms.

It should be noted that one-dimensional lattice vibrations are not encountered only in atomic chains. For example, in a simple cubic crystal with one atom in the primitive cell, when a wave propagates along the directions of the cube edge, face diagonal, or body diagonal, entire planes of atoms move in phase with displacements either parallel or perpendicular to the direction of the wavevector (see the figures below).





We can describe the displacements of the plane *n* from its equilibrium position with a single coordinate, $u_n$. The problem becomes in this way one-dimensional. For each wavevector $u_n$ there is one solution with longitudinal polarization and two with transverse polarization. The parameter *A* is different for longitudinal and transverse waves.

We have already seen that, in the harmonic approximation, the dynamics of the system is equivalent to that of coupled harmonic oscillators, the harmonic potential acting on an ion describing a force that is linear in the displacement. For simplicity, we assume further that only the interaction between nearest neighbors is significant, case in which the force exerted on th *n*-th atom in the lattice is linear in the ion's (relative) displacements and hence given by

$$F_n = A(u_{n+1} - u_n) + A(u_{n-1} - u_n) \tag{22}$$

where *A* is the interatomic force (or, equivalently, the elastic constant between nearest-neighbor ions).

Applying Newton's second law to the motion of the *n*-th atom with mass *M*, $M(d^2u_n/dt^2) = F_n$, we obtain

$$M\frac{d^2u_n}{dt^2} = A(u_{n+1} - u_n) + A(u_{n-1} - u_n) = -A(2u_n - u_{n+1} - u_{n-1})\,. \tag{23}$$

A similar equation should be written for each atom in the lattice. The solutions of the equation above have the form

$$u_n(t) = u_0 \exp[i(kR_n^0 - \omega t)] = u_0 \exp[i(kan - \omega t)]\,. \qquad (24)$$

Such solutions represent traveling waves, in which all the atoms oscillate with the same frequency $\omega$ and the same amplitude $u_0$ and have wavevector $k$. Solutions of this form are only possible because of the translational symmetry of the lattice. Note that in (24) there is no need for the indices $\chi$, $\lambda$ or $\mu$ because here is only one atom in the basis ($\chi = 1$), we have one oscillation ($\lambda = 1$) and one dimension ($\mu = 1$). In addition, all atoms oscillate with the same amplitude, and no polarization vectors need be introduced (it can be included in $u_0$).

Inserting (24) into the equation of motion (23) we obtain

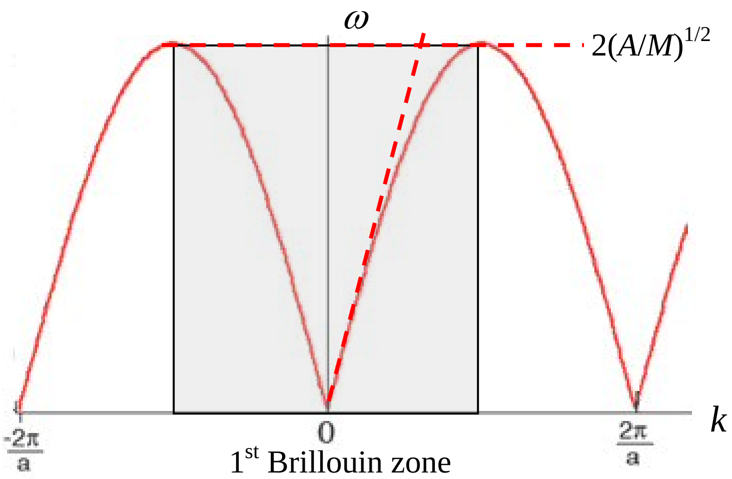
$$-M\omega^2 = -A[2 - \exp(-ika) - \exp(ika)] \qquad (25)$$

or

$$M\omega^2 = 2A\left(1 - \frac{\exp(ika) + \exp(-ika)}{2}\right) = 2A[1 - \cos(ka)] = 4A\sin^2(ka/2)\,. \qquad (26)$$

and the dispersion relation, represented in the figure below, is

$$\omega(k) = 2\sqrt{A/M}\,|\sin(ka/2)|\,. \qquad (27)$$



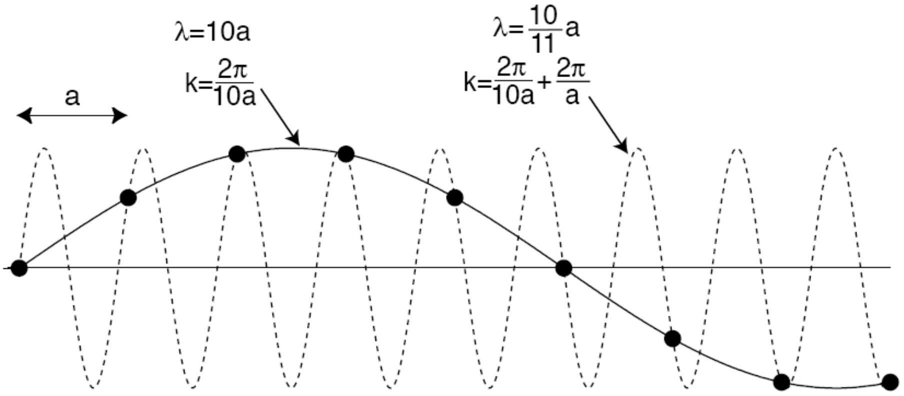
Please observe that equation (23) follows also directly, by particularizing (10) for one-dimensional motion and assuming that only the nearest-neighbor atoms interact with each another, so that $A_{n,m} = 0$ for $m \neq n, n \pm 1$. In this case, $A_{n,n-1} = A_{n,n+1} = -A$, $A_{n,n} = 2A$, such that the sum rule (12) is satisfied. Similarly, equation (25) could have been directly written from $M\omega^2 = [A_{n,n-1}\exp(-ika) + A_{n,n} + A_{n,n+1}\exp(ika)]$.

Because the dispersion relation is periodic: $\omega(k) = \omega(k + 2\pi / a)$, with the periodicity given by the reciprocal lattice vector, all distinct frequency values can be found in the $k$ interval

$$-\pi / a \leq k < \pi / a\,, \tag{28}$$

which corresponds to the first Brillouin zone. The maximum (cut-off) frequency $\omega_{\max} = 2\sqrt{A/M}$ is obtained for the minimum wavelength of $\lambda_{\min} = 2\pi /(\pi / a) = 2a$. The existence of a minimum wavelength can be understood as resulting from the condition that waves with wavelengths smaller than $2a$ cannot propagate in the lattice, being reflected at the boundaries of the first Brillouin zone.

The significance of the periodicity of the dispersion relation is evident from the figure below: changing $k$ by one reciprocal lattice vector gives exactly the same movement of atoms.



In the long-wavelength limit $ka/2 << 1$, we have $\sin(ka/2) \cong ka/2$, and

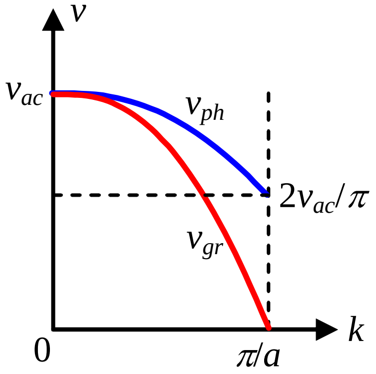
$$\omega(k) = \sqrt{A/M}\,ak = v_{ac}k \tag{29}$$

so that the oscillations are acoustic and characterized by the acoustic velocity $v_{ac} = a\sqrt{A/M}$.

Moreover, since the oscillation frequency does not depend linearly with *k*, we can define separately the phase velocity, i.e. the velocity of the phase of the wavefront, and the group velocity, i.e. the propagation velocity of the wavepacket and of the wave energy. Their modulus are given, respectively, by

$$v_{ph} = \left|\frac{\omega}{k}\right| = \sqrt{\frac{A}{M}}a\left|\frac{\sin(ka/2)}{ka/2}\right| = v_{ac}\left|\frac{\sin(ka/2)}{ka/2}\right|, \tag{30}$$

$$v_{gr} = \left|\frac{d\omega}{dk}\right| = \sqrt{\frac{A}{M}}a\left|\cos(ka/2)\right| = v_{ac}\left|\cos(ka/2)\right|. \tag{31}$$

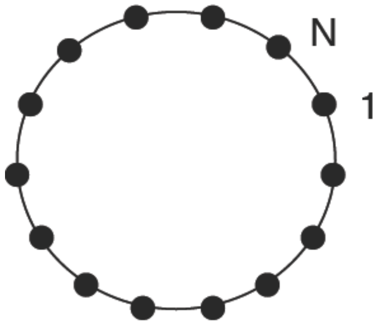


In the long-wavelength range, $k \to 0$, $v_{ph} = v_{gr} = v_{ac}$, while at the edges of the first Brilouin zone, for $k = \pm\pi/a$, $v_{ph} = 2v_{ac}/\pi$ and $v_{gr} = 0$.

## Finite lattices

For finite one-dimensional lattices consisting of *N* identical atoms, the requirement of symmetry (of equivalence of physical properties) when the equation of motion refers to different atoms imposes the cyclic boundary condition $u_n = u_{n+N}$. This so-called Born-Karman condition expresses the independence of the properties on the surface, i.e. we have a finite solid, with no surfaces; a finite chain with no end. From the cyclic boundary condition it follows that $\exp(ikNa) = 1$, or

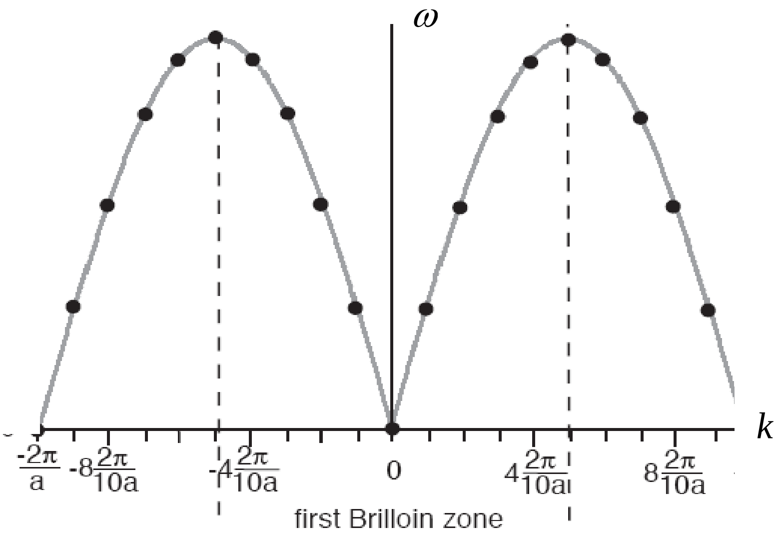
$$k = \frac{2\pi}{Na}m, \tag{33}$$



with *m* an integer. There are *N* allowed *m* values for *k* in the first Brillouin zone:

$$-N/2 \le m < N/2, \tag{34}$$

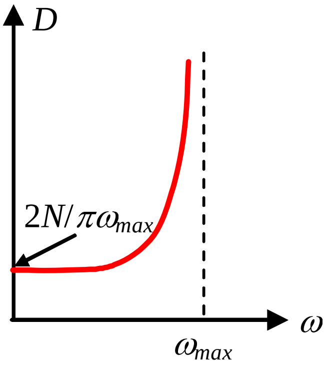
which correspond to the *N* degrees of freedom of the system. Because *N* is usually a large number, the discrete nature of the wavenumber is disregarded and it is considered as a continuous variable. Below: example for *N* = 10.



## Density of oscillations in a simple finite one-dimensional lattice

How many oscillations (with different $k$ values) exist then in the frequency interval $(\omega, \omega + d\omega)$? This number is referred to as the density of oscillations per unit frequency and is denoted by $dN_{osc}/d\omega = D(\omega)$. Because for the finite one-dimensional lattice the wavenumber varies only in discrete steps of $\Delta k = 2\pi/Na$, there is only one oscillation possible in this wavenumber range, so that

$$\frac{dN_{osc}}{dk} = \frac{1}{\Delta k} = \frac{Na}{2\pi} \tag{35}$$

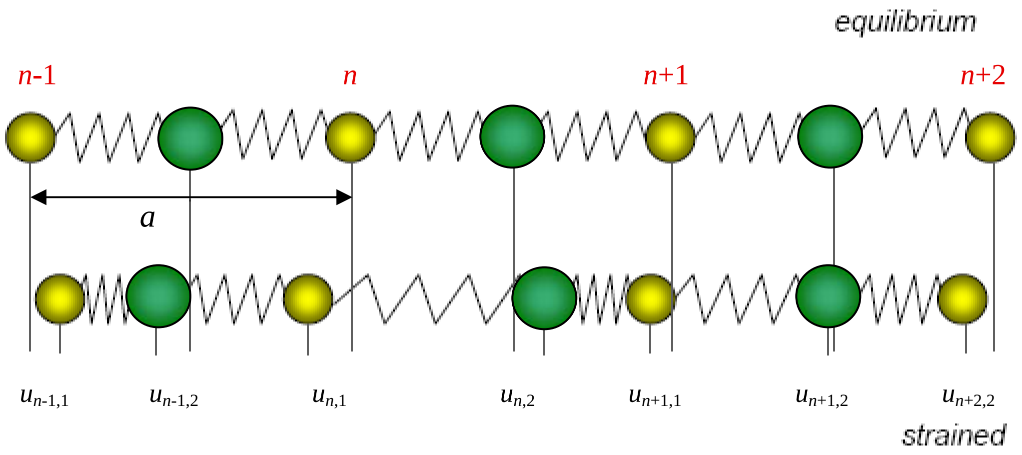


and, taking into account the double degeneracy due to the symmetry of $\omega(k)$ (two $k$ values correspond to the same $\omega$), we obtain

$$\begin{aligned} D(\omega) &= \frac{dN_{osc}}{d\omega} = 2\frac{dN_{osc}}{dk}\frac{dk}{d\omega} = \frac{N}{\pi}\frac{1}{\sqrt{A/M}\,|\cos(ka/2)|} \\ &= \frac{N}{\pi}\frac{1}{\sqrt{A/M}\sqrt{1-\sin^2(ka/2)}} = \frac{N}{\pi}\frac{1}{\sqrt{A/M}\sqrt{1-\omega^2/\omega_{\max}^2}} = \frac{2N}{\pi}\frac{1}{\sqrt{\omega_{\max}^2-\omega^2}} \end{aligned} \tag{36}$$

## Oscillations of an infinite atomic chain with two atoms in the basis

Let us assume that we have an infinite one-dimensional lattice with lattice constant *a*, consisting of equally spaced ions with different masses $M_1$ and $M_2 > M_1$ (see the figure below). The basis has therefore two atoms, placed at equilibrium positions $R^0_{n,1} = na$, and $R^0_{n,2} = (n+1/2)a$. In this case *d* = 1, *s* = 2, so there should be two oscillation branches, one acoustical and one optical.



Similarly to the atomic chain with one atom in the basis, when there are two atoms per unit cell we have two equations of motion of the general form $F = M(d^2u/dt^2)$, one for every type of atom. To distinguish between the displacements of the two atoms, we denote with $u_{n,1}(t) = u_0 e_1 \exp[i(kan - \omega t)]$ the displacement of atoms with mass $M_1$ (the yellow ones in the figure above) and with $u_{n,2}(t) = u_0 e_2 \exp[i(ka(n+1/2) - \omega t)]$ that of atoms with mass $M_2$ (the green ones). (Note that, for simplicity, we omit for now the index $\lambda$, which has two values: *ac* and *opt*, corresponding to the acoustical and optical oscillations.) So, we have

$$M_1 \frac{d^2 u_{n,1}}{dt^2} = A(u_{n-1,2} - u_{n,1}) + A(u_{n,2} - u_{n,1}) = -A(2u_{n,1} - u_{n,2} - u_{n-1,2}), \tag{37a}$$

$$M_2 \frac{d^2 u_{n,2}}{dt^2} = A(u_{n,1} - u_{n,2}) + A(u_{n+1,1} - u_{n,2}) = -A(2u_{n,2} - u_{n,1} - u_{n+1,1}), \tag{37b}$$

or

$$-M_1\omega^2 e_1 = -2Ae_1 + Ae_2[\exp(ika/2) + \exp(-ika/2)] = -2Ae_1 + 2Ae_2\cos(ka/2), \tag{38a}$$

$$-M_2\omega^2 e_2 = -2Ae_2 + Ae_1[\exp(-ika/2) + \exp(ika/2)] = -2Ae_2 + 2Ae_1\cos(ka/2). \tag{38b}$$

This system of coupled equations has solution only when its determinant vanishes, i.e. when

$$\begin{vmatrix} 2A - M_1\omega^2 & -2A\exp\cos(ka/2) \\ -2A\cos(ka/2) & 2A - M_2\omega^2 \end{vmatrix} = 0\,, \tag{39}$$

or

$$\omega^4 - \omega^2\omega_0^2 + \frac{\omega_0^4\gamma^2}{4}\sin^2\left(\frac{ka}{2}\right) = 0 \tag{40}$$

with $\omega_0^2 = 2A\dfrac{M_1 + M_2}{M_1 M_2}$, $\gamma^2 = 4\dfrac{M_1 M_2}{(M_1 + M_2)^2}$.

From (40) it follows that there are two solutions (two types of lattice oscillations) for every value of *k*; these are the optical and acoustic branches. The two solutions are

$$\omega_1^2(k) = \frac{\omega_0^2}{2}\left[1 - \sqrt{1 - \gamma^2\sin^2\left(\frac{ka}{2}\right)}\right], \qquad \omega_2^2(k) = \frac{\omega_0^2}{2}\left[1 + \sqrt{1 - \gamma^2\sin^2\left(\frac{ka}{2}\right)}\right]. \tag{41}$$

As for the one-dimensional lattice with one atom in the basis, $\omega_{1,2}(k + 2\pi/a) = \omega_{1,2}(k)$, so that all relevant values are found in the first Brillouin zone.

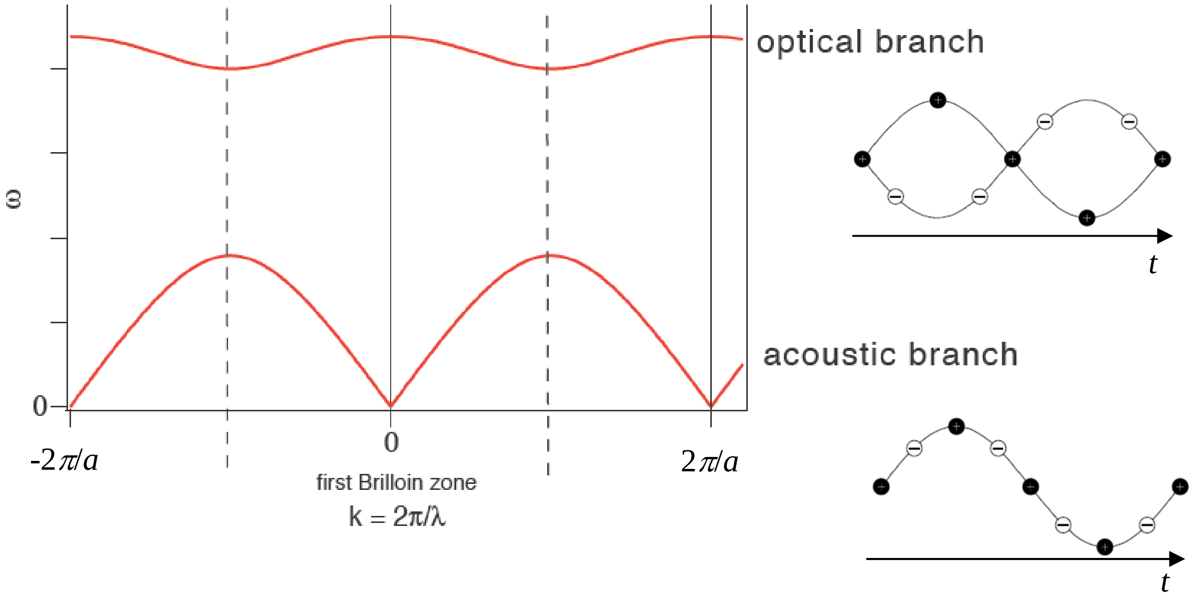
Again, the dispersion relation could have been obtained directly from (14). More precisely, when only the nearest neighbors interact we would have obtained $M_1\omega^2 e_1(k) = A_{n,1;n-1,2}\exp(-ika/2)e_2(k) + A_{n,1;n,1}e_1(k) + A_{n,1;n+1,2}\exp(ika/2)e_2(k)$, $M_2\omega^2 e_2(k) = A_{n,2;n-1,1}\exp(-ika/2)e_1(k) + A_{n,2;n,2}e_2(k) + A_{n,2;n+1,1}\exp(ika/2)e_1(k)$, so that the dispersion relation is recovered if $A_{n,1;n-1,2} = A_{n,1;n+1,2} = A_{n,2;n-1,1} = A_{n,2;n+1,1} = -A$ and $A_{n,1;n,1} = A_{n,2;n,2} = 2A$ (the sum rule (12) is again satisfied!).

From the solutions (41) one can identify the oscillation branches: the acoustic one corresponds to the first solution, for which

$$\omega_{ac}(0) = \omega_1(0) = 0\,, \qquad \omega_{ac}\left(\pm\frac{\pi}{a}\right) = \frac{\omega_0}{\sqrt{2}}\sqrt{1 - \sqrt{1 - \gamma^2}} > \omega_{ac}(0)\,, \tag{42}$$

and the optical branch is consistent with the second solution, for which

$$\omega_{opt}(0) = \omega_2(0) = \omega_0\,, \qquad \omega_{opt}\left(\pm\frac{\pi}{a}\right) = \frac{\omega_0}{\sqrt{2}}\sqrt{1+\sqrt{1-\gamma^2}} < \omega_{opt}(0)\,. \tag{43}$$



Because for the acoustic branch $\omega_{ac}(0) = 0$ for $k$ = 0, from the system of coupled equations (38) it follows that $(e_1/e_2)_{ac} = 1$, which implies that the displacement of the two types of ions is the same/occurs in the same direction and the unit cell moves as a whole; it oscillates in phase (see the figure above, bottom, right). On the other hand, in the long-wavelength limit of the optical branch, $(e_1/e_2)_{opt} = -M_2/M_1$, i.e. the ions are displaced in opposite direction and we have out-of-phase oscillations. The oscillations occur such that the center of mass of each ion pair is fixed, i.e. $M_1e_1 + M_2e_2 = 0$ (see the figure above, top, right).

In the long-wavelength limit, when $\sin(ka/2) \cong ka/2$, the dispersion relation of the acoustic branch can be approximated as

$$\omega_{ac} \cong \frac{\omega_0}{\sqrt{2}}\sqrt{1-\left[1-\frac{\gamma^2}{2}\left(\frac{ka}{2}\right)^2\right]} \cong \frac{\omega_0}{4}\gamma ka\,, \tag{44}$$

i.e.

$$\omega_{ac}(k) = v_{ac} k \,, \qquad v_{ac} = \frac{1}{4}\omega_0 \gamma a = \sqrt{\frac{A}{2(M_1 + M_2)}}a \,, \tag{45}$$

while the dispersion relation for the optical branch becomes

$$\omega_{opt} \cong \frac{\omega_0}{\sqrt{2}}\sqrt{1 + \left[1 - \frac{\gamma^2}{2}\left(\frac{ka}{2}\right)^2\right]} \cong \omega_0\left[1 - \frac{\gamma^2}{8}\left(\frac{ka}{2}\right)^2\right] \tag{46}$$

or

$$\omega_{opt}(k) = \omega_0 - \beta k^2 \,, \qquad \beta = \omega_0 \gamma^2 a^2 / 32 \,. \tag{47}$$

Note that, to calculate the density of states for the **finite** one-dimensional lattice with two atoms in the basis, we can follow the same treatment as for the atomic chain with one atom in the basis, taking into account that we must calculate separately the density of states for the two oscillation types, which have different dispersion relations.

## Density of states in a finite three-dimensional crystal

The density of states/oscillations in a three-dimensional crystal is obtained, as in the one-dimensional crystal, by imposing the appropriate boundary conditions for $\boldsymbol{u}_{n\chi}$. Because there is a large number of atoms in a crystal, which interact strongly with their neighbors, the contribution of the atoms at the surface of the crystal to any physical phenomena is negligible. Therefore, we can employ again the Born-Karman cyclic condition

$$u_{n\chi}^{\mu} = u_{(n+N_\mu)\chi}^{\mu} \tag{48}$$

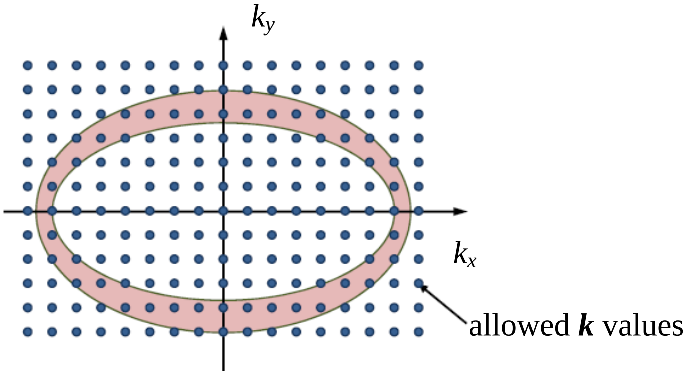
with $u_{n\chi}^{\mu} = u_0 e_{\lambda}^{\mu} \exp(i\boldsymbol{k} \cdot \boldsymbol{R}_{n\chi}^{0} - i\omega t)$, where $N_\mu$ is the number of atoms in the $x_\mu$ direction, $\mu$ = 1,2,3, with $x_1 = x$, $x_2 = y$, $x_3 = z$. As in the one-dimensional case, the wavevector component along the $\mu$ direction in the first Brillouin zone has a discrete spectrum,

$$k_\mu = \frac{2\pi}{N_\mu a_\mu} m_\mu \,, \tag{49}$$

with

$$-\frac{N_\mu}{2} \le m_\mu < \frac{N_\mu}{2} \tag{50}$$

where $a_\mu$ is the lattice constant along $x_\mu$, and $m_\mu$ are integers, i.e. it varies in steps of $\Delta k_\mu = 2\pi / L_\mu$, with $L_\mu = N_\mu a_\mu$ the length of the crystal along the $x_\mu$ direction. The discrete ***k*** values in a two-dimensional lattice are represented by points in the figure below.



It follows then that a state/oscillation occupies a volume in the ***k*** space given by

$$\Delta \mathbf{k} = \Delta k_1 \Delta k_2 \Delta k_3 = \frac{(2\pi)^3}{V}, \tag{51}$$

where $V = L_1 L_2 L_3 = N\Omega$ is the volume of the crystal with $sN = sN_1 N_2 N_3$ atoms that form a lattice with a primitive cell of volume $\Omega = a_1 a_2 a_3$ and *s* atoms in the basis.

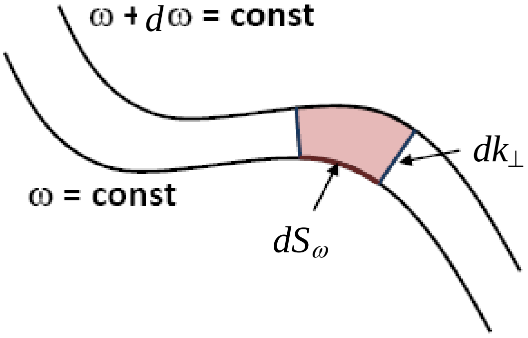
The density of states/oscillations in the ***k*** space is then defined as

$$\frac{dN_{osc}}{d\mathbf{k}} = \frac{1}{\Delta \mathbf{k}} = \frac{V}{(2\pi)^3}. \tag{52}$$

The density of oscillations in the frequency space, defined as

$$D(\omega) = \frac{dN_{osc}}{d\omega} = \frac{dN_{osc}}{d\mathbf{k}} \frac{d\mathbf{k}}{d\omega}, \tag{53}$$

represents the number of oscillations (with different $\boldsymbol{k}$ values) that exist in the frequency interval $(\omega, \omega + d\omega)$. In this formula $d\boldsymbol{k}$ is the volume in $\boldsymbol{k}$ space between the surfaces $\omega(\boldsymbol{k})$ and $\omega(\boldsymbol{k}) + d\omega(\boldsymbol{k})$ (see the figure below). The density of oscillations is discrete (as for one-dimensional crystals) but, for sufficiently large crystals, the sum over the discrete states can be replaced by an integral. We can calculate it observing that



$$d\boldsymbol{k} = \int_{\omega(\boldsymbol{k})=const} dS_\omega dk_\perp = \int_{\omega(\boldsymbol{k})=const} dS_\omega \frac{d\omega}{|\nabla_{\boldsymbol{k}}\omega|}, \tag{54}$$

from which it follows that

$$D(\omega) = \frac{V}{(2\pi)^3} \int_{\omega(\boldsymbol{k})=const} \frac{dS_\omega}{|\nabla_{\boldsymbol{k}}\omega|}. \tag{55}$$

We can express also the density of states as $D(\omega)d\omega = [V/(2\pi)^3]d\boldsymbol{k}$, or

$$\int D(\omega)d\omega = \frac{V}{(2\pi)^3} \int_{1^{st}BZ} d\boldsymbol{k}, \tag{56}$$

which represents a particular case of approximating a sum over $\boldsymbol{k}$ in the first Brillouin zone by an integral, approximation that for an arbitrary function $f(\boldsymbol{k})$ is

$$\sum_{\boldsymbol{k}} f(\boldsymbol{k}) = \frac{V}{(2\pi)^3} \sum_{\boldsymbol{k}} f(\boldsymbol{k})\Delta\boldsymbol{k} \to \frac{V}{(2\pi)^3} \int_{1^{st}BZ} f(\boldsymbol{k})d\boldsymbol{k}. \tag{57}$$

For $f(\boldsymbol{k}) = F(\omega(\boldsymbol{k}))$, we have

$$\sum_{\boldsymbol{k}} F(\omega(\boldsymbol{k})) = \int F(\omega) D(\omega) d\omega \tag{58}$$

if the density of states is normalized in each branch $\chi$, $\chi = 1,\ldots,s$ such that

$$\int_0^{\omega_{\max}} D(\omega) d\omega = 3N \tag{59}$$

with $\omega_{\max}$ the maximum value of the oscillation frequency in the branch.

Note that for the three-dimensional crystals, we have not defined the density of states normalized at unit volume (as for the atomic chain with one atom in the basis), but have kept the crystal volume throughout the calculations!

## Quantized oscillations/phonons in a one-dimensional finite lattice with one atom in the basis

We have seen that in the one-dimensional lattice with one atom in the basis, the ions act as coupled harmonic oscillators. Here we show that this system of coupled oscillators can be reduced to an equivalent system of independent harmonic oscillators by the introduction of normal coordinates. Then, we associate a normal oscillation to each normal coordinate. The Hamiltonian of the finite one-dimensional lattice with one atom in the basis can be written as

$$H_{ion} = \sum_{n=1}^{N} \frac{p_n^2}{2M} + \frac{1}{2} \sum_{n,n'=1}^{N} A_{nn'} u_n u_{n'} \,. \tag{60}$$

In the quantum treatment of the system of coupled harmonic oscillators, the position and momentum coordinates become (conjugate) operators, such that

$$[\hat{u}_n, \hat{p}_{n'}] = \hat{u}_n \hat{p}_{n'} - \hat{p}_{n'} \hat{u}_n = i\hbar \delta_{nn'} \,, \qquad [\hat{u}_n, \hat{u}_{n'}] = 0 \,, \qquad [\hat{p}_n, \hat{p}_{n'}] = 0 \,. \tag{61}$$

We can introduce **normal** position and momentum operators,

$$\hat{Q}_k = N^{-1/2} \sum_n \hat{u}_n \exp(-ikna) \,, \qquad \hat{P}_k = N^{-1/2} \sum_n \hat{p}_n \exp(ikna) \,, \tag{62}$$

in terms of which the original operators can be expressed as

$$\hat{u}_n = N^{-1/2}\sum_k \hat{Q}_k \exp(ikna)\,, \qquad \hat{p}_n = N^{-1/2}\sum_k \hat{P}_k \exp(-ikna)\,. \tag{63}$$

The normal operators satisfy the commutation relations

$$\begin{aligned}&[\hat{Q}_k, \hat{P}_{k'}] = N^{-1}\sum_{n,n'}[\hat{u}_n, \hat{p}_{n'}]\exp(ik'n'a - ikna) = N^{-1}i\hbar\sum_n \exp[i(k'-k)na] = i\hbar\delta_{kk'}\,,\\ &[\hat{Q}_k, \hat{Q}_{k'}] = 0\,, \qquad [\hat{P}_k, \hat{P}_{k'}] = 0\,,\end{aligned} \tag{64}$$

since

$$\sum_{n=1}^{N}\exp[i(k'-k)na] = \sum_{n=1}^{N}\exp[i2\pi nm/N] = \frac{\exp(i2\pi m/N)[\exp(i2\pi m)-1]}{\exp(i2\pi m/N)-1} = N\delta_q = N\delta_{kk'} \tag{65}$$

for $k$, $k'$ in the first Brillouin zone, for which $k'-k = 2\pi m/Na$ with $m$ an integer. From the commutation relations it follows that the normal position and momentum operators for the same wavenumber $k$ are also conjugate.

Because there are $N$ normal oscillations and $\hat{u}_n$, $\hat{p}_n$ are hermitic operators, $\hat{Q}_k^+ = \hat{Q}_{-k}$, $\hat{P}_k^+ = \hat{P}_{-k}$, and, for $k$ in the first Brillouin zone (where it takes $N$ discrete values) we have $2N$ normal operators: $N$ normal position operators $\hat{Q}_k$, and $N$ normal momentum operators $\hat{P}_k$.

In the normal position and momentum operators, the kinetic energy term in the Hamiltonian becomes

$$\sum_{n=1}^{N}\frac{\hat{p}_n^2}{2M} = \frac{1}{2MN}\sum_{k,k'}\hat{P}_k\hat{P}_{k'}\sum_{n=1}^{N}\exp[-i(k+k')na] = \frac{1}{2M}\sum_k \hat{P}_k\hat{P}_{-k} \tag{66}$$

with $k$ in the first Brillouin zone. In a similar manner,

$$\begin{aligned}&\frac{1}{2}\sum_{n,n'}A_{nn'}\hat{u}_n\hat{u}_{n'} = \frac{1}{2N}\sum_{k,k'}\hat{Q}_k\hat{Q}_{k'}\sum_{n,n'}A_{nn'}\exp(ikna)\exp(ik'n'a)\\ &\qquad = \frac{1}{2N}\sum_{k,k'}\hat{Q}_k\hat{Q}_{k'}\sum_n \exp[i(k'+k)na]\sum_{n'}A_{nn'}\exp[ik'(n'-n)a] = \frac{M}{2}\sum_k \omega^2(k)\hat{Q}_k\hat{Q}_{-k}\end{aligned} \tag{67}$$

if we take into account that $\sum_{n'}A_{nn'}\exp[ik'(n'-n)a] = M\omega^2(k')$ and $\omega(-k) = \omega(k)$.

The Hamiltonian of the system of ions,

$$\hat{H}_{ion} = \sum_k \left( \frac{\hat{P}_k \hat{P}_{-k}}{2M} + \frac{M\omega^2(k)\hat{Q}_k \hat{Q}_{-k}}{2} \right), \tag{68}$$

describes a system of harmonic oscillators that are not, however, totally independent since the terms corresponding to $k$ and $-k$ are still coupled. To avoid this situation, we introduce annihilation and creation operators for each $k$, $\hat{a}_k$ and $\hat{a}_k^+$, and express the normal position and momentum operators as

$$\hat{Q}_k = \sqrt{\frac{\hbar}{2M\omega(k)}}(\hat{a}_k + \hat{a}_{-k}^+)\,, \qquad \hat{P}_k = -\frac{i\hbar}{2}\sqrt{\frac{2M\omega(k)}{\hbar}}(\hat{a}_k - \hat{a}_{-k}^+)\,. \tag{69}$$

The creation and annihilation operators satisfy the commutation relations

$$[\hat{a}_k, \hat{a}_{k'}^+] = \hat{a}_k \hat{a}_{k'}^+ - \hat{a}_{k'}^+ \hat{a}_k = \delta_{kk'}\,, \qquad [\hat{a}_k, \hat{a}_{k'}] = 0\,, \qquad [\hat{a}_k^+, \hat{a}_{k'}^+] = 0\,. \tag{70}$$

For $k$ in the first Brillouin zone, the Hamiltonian operator of the lattice is now a sum of $N$ independent harmonic oscillators:

$$\hat{H}_{ion} = \sum_k \hat{H}_k = \sum_k \frac{\hbar\omega(k)}{2}(\hat{a}_{-k}^+ \hat{a}_{-k} + \hat{a}_k \hat{a}_k^+) = \sum_k \hbar\omega(k)\left( \hat{a}_k^+ \hat{a}_k + \frac{1}{2} \right) = \sum_k \hbar\omega(k)\left( \hat{N}_k + \frac{1}{2} \right), \tag{71}$$

with $[\hat{H}_k, \hat{H}_{k'}] = 0$, $\hat{N}_k = \hat{a}_k^+ \hat{a}_k$ the hermitic number operator, for which $[\hat{N}_k, \hat{N}_{k'}] = 0$ and $[\hat{N}_k, \hat{H}_k] = 0$, (in the third term of the equation above, we substitute first $k$ for $-k$).

For each $k$ mode, the eigenstates (Fock states) and eigenvalues of the Hamiltonian and the number operators are

$$|\, n_k \rangle = \frac{(\hat{a}_k^+)^n}{\sqrt{n_k!}}\,|\, 0_k \rangle\,, \tag{72a}$$

$$E_k = \hbar\omega(k)\left( n_k + \frac{1}{2} \right), \qquad n_k = 0,\ 1,\ 2, \ldots\ldots \tag{72b}$$

where $|\, 0_k \rangle$ is the fundamental state of the oscillator, and

$$\hat{a}_k \mid n_k\rangle = \sqrt{n_k} \mid n_k - 1\rangle, \qquad \hat{a}_k^+ \mid n_k\rangle = \sqrt{n_k + 1} \mid n_k + 1\rangle, \qquad \hat{N}_k \mid n_k\rangle = n_k \mid n_k\rangle \tag{73}$$

with $\hat{a}_k \mid 0_k\rangle = 0$. The eigenstate of the lattice Hamiltonian is then (there are *N* discrete values for *k*)

$$\mid n_{k_1}, n_{k_2}, ..., n_{k_N}\rangle = \mid n_{k_1}\rangle \otimes \mid n_{k_2}\rangle \otimes ... \otimes \mid n_{k_N}\rangle \tag{74}$$

and the energy of the collection of harmonic oscillators is

$$E = \sum_k E_k = \sum_k \hbar\omega(k)\left(n_k + \frac{1}{2}\right) = E_0 + E_{ph}\,. \tag{75}$$

where $E_0 = \sum_k \hbar\omega(k)/2$ is the zero (fundamental) energy, and $E_{ph} = \sum_k \hbar\omega(k) n_k$ is the energy of the quantized oscillations of the lattice in an excited state.

The state of each *k* quantum oscillator can be seen as that of $n_k$ excitation quanta, each with an associated energy $\hbar\omega(k)$. This excitation quanta is associated to a quasi-particle named phonon, in analogy with the photon, which is the quanta of the electromagnetic field. In the one-dimensional lattice considered here, the phonon is called acoustic phonon since there are only acoustic oscillation branches in a one-dimensional lattice with one atom in the basis, and the lattice has $n_{k_1}$ phonons at the wavevector $k_1$, $n_{k_2}$ phonons at the wavevector $k_2$, and so on.

Similarly, in a one-dimensional finite lattice with two atoms in the basis and *N* values for *k*,

$$\hat{H}_{ion} = \sum_{k,\lambda=1,2} \hbar\omega_\lambda(k)\left(\hat{a}_{k,\lambda}^+ \hat{a}_{k,\lambda} + \frac{1}{2}\right), \tag{76a}$$

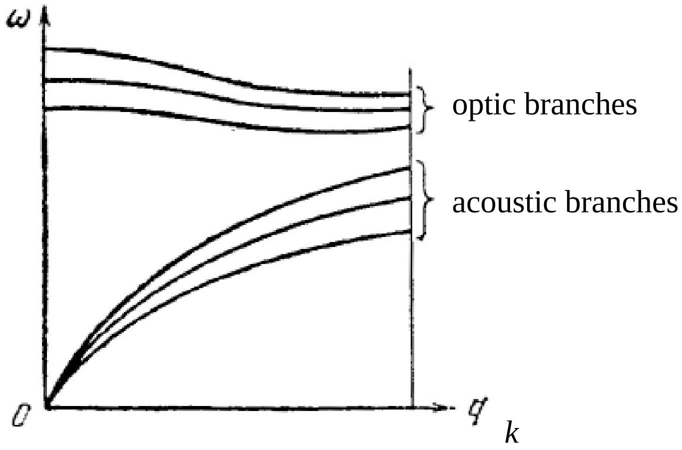
$$\mid n_{k_1,\lambda}, n_{k_2,\lambda}, ..., n_{k_N,\lambda}\rangle = \frac{(\hat{a}_{k_1,1}^+)^{n_{k1,1}} (\hat{a}_{k_1,2}^+)^{n_{k1,2}} ... (\hat{a}_{k_N,1}^+)^{n_{kN,1}} (\hat{a}_{k_N,2}^+)^{n_{kN,2}}}{\sqrt{(n_{k_1,\lambda})!(n_{k_2,\lambda})!...(n_{k_N,\lambda})!}} \mid 0\rangle \tag{76b}$$

with $\mid 0\rangle = \mid 0,0,...,0\rangle$, and

$$E = \sum_{k,\lambda=1,2} \hbar\omega_\lambda(k)\left(n_{k,\lambda} + \frac{1}{2}\right), \tag{77}$$

where the phonons associated with the $\lambda = 1$ branch are the acoustic phonons, and those associated with $\lambda = 2$ are the optical phonons.

In a three-dimensional lattice with $s$ atoms in the basis, there are three degrees of freedom for every atom, and hence there are 3 acoustic phonons and $3s - 3$ optical phonons. A typical phonon dispersion spectrum for $s$ = 2 is illustrated below.



Phonons are quanta of the collective/thermal lattice oscillations. The crystalline lattice can be viewed either as a collection of coupled harmonic oscillators or as a gas of free/non-interacting phonons, which obey the laws of quantum statistics. In particular, phonons are bosons and obey the Bose-Einstein statistics. However, since they are not real particles, their number is not independent of temperature and volume, so that the electro-chemical potential of the phonon gas must be zero. Then, the thermal equilibrium number of phonons with frequency $\omega_\lambda(\boldsymbol{k})$ is given by the Planck distribution

$$n_{ph} = n_{\boldsymbol{k},\lambda} = \frac{1}{\exp[\hbar\omega_\lambda(\boldsymbol{k}) / k_B T] - 1}. \tag{78}$$

So, the number of phonons is small at low temperatures $\hbar\omega_\lambda >> k_B T$, for which $n_{ph} \cong \exp[-\hbar\omega_\lambda / k_B T]$, but becomes high at large temperatures $\hbar\omega_\lambda << k_B T$, for which $n_{ph} \cong 1/[1 + (\hbar\omega_\lambda / k_B T) - 1] = k_B T / \hbar\omega_\lambda$.

# Phononic heat capacity

The thermal properties of solids, and in particular the heat capacity, are determined by both phonons and electrons. We refer now to the phonon, or lattice, contribution to the heat capacity. The heat capacity is defined as the heat $\Delta Q$ required to raise the temperature by $\Delta T$, i.e. $C = \Delta Q / \Delta T$. If the process is carried out at constant volume, $\Delta Q$ must be replaced by $\Delta U$, which represents the increase in the internal energy $U$ of the system. Then, the heat capacity at constant volume is

$$C_V = \left( \frac{\partial U}{\partial T} \right)_V . \tag{1}$$

The phonon contribution at the heat capacity is obtained from the lattice energy term

$$E = \sum_{\boldsymbol{k},\lambda} \hbar\omega_{\boldsymbol{k},\lambda} \left( n_{\boldsymbol{k},\lambda} + \frac{1}{2} \right) = \sum_{\boldsymbol{k},\lambda} \hbar\omega_{\boldsymbol{k},\lambda} \left( \frac{1}{\exp(\hbar\omega_{\boldsymbol{k},\lambda} / k_B T) - 1} + \frac{1}{2} \right) \tag{2}$$

of the internal energy $U = E_{eq} + E$, with $E_{eq}$ the energy in the equilibrium configuration of the system. Actually, expressing the internal energy of the system in terms of the free energy $F$ and the entropy $S$ as

$$U = F + TS , \tag{3}$$

apart from the heat capacity at constant volume $C_V = (\partial U / \partial T)_V = T(\partial S / \partial T)_V$ we can define also a heat capacity at constant pressure, $C_P = T(\partial S / \partial T)_P$. These two parameters are related through $C_P - C_V = -T(\partial V / \partial T)_P^2 / (\partial V / \partial P)_T$, and one can be determined from the other. These parameters are the same only in the harmonic approximation of lattice oscillations.

In a classical statistical theory, based on the classical partition function, the mean energy of a one-dimensional oscillator (resulting equally from its kinetic and potential energy parts) is $k_B T$, value that becomes $3k_B T$ for a three-dimensional oscillator. Then, for $sN$ three-dimensional oscillators $\langle E \rangle = 3sNk_B T$, and the phononic heat capacity at constant volume is

$$C_{ph} = \left(\frac{\partial E}{\partial T}\right)_V = 3sNk_B\,, \tag{4}$$

i.e. $C_{ph} = 3k_B$ per atom, or $C_{ph} = 3N_A k_B = 3R = 6\,\text{cal/molK}$ per mole, with $N_A$ the Avogadro number. This is the **Dulong-Petit law**, and it predicts a temperature-independent heat capacity. This prediction agrees with experimental data at high temperatures, but not at low temperatures, where experiments indicate that $C_{ph} \propto T^3 \to 0$ as $T \to 0$.

To explain the low-temperature behavior of the heat capacity, one should disregard the classical statistical theory, which is no longer valid when the separation between the energy levels of the oscillator is comparable to or higher than $k_B T$, and use instead the quantum statistical mechanics. The specific heat of the lattice is then defined as

$$C_{ph} = \frac{dE}{dT} = k_B \sum_{\boldsymbol{k},\lambda} \left(\frac{\hbar\omega_{\boldsymbol{k},\lambda}}{k_B T}\right)^2 \frac{\exp(\hbar\omega_{\boldsymbol{k},\lambda}/k_B T)}{[\exp(\hbar\omega_{\boldsymbol{k},\lambda}/k_B T)-1]^2}\,. \tag{5}$$

This expression does not involve the zero energy of the lattice and is called for this reason the phononic heat capacity. To specifically calculate the phononic heat capacity we need to know the phonon dispersion relation. This relation is quite complicated for three-dimensional crystals and therefore approximations are generally made.

## The Einstein model

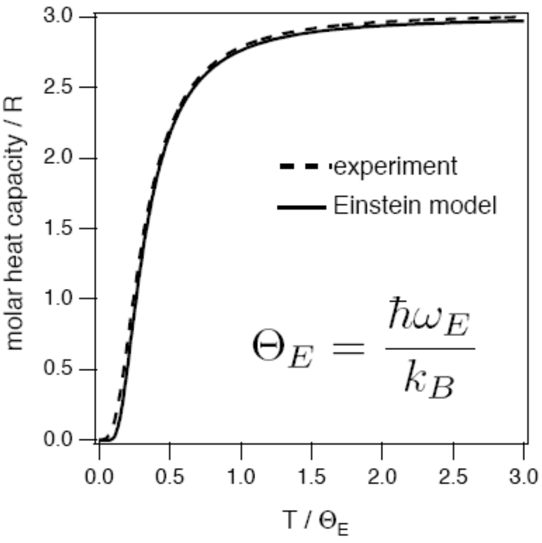
In this model each atom or molecule is considered as a particle that oscillates in the average field of its neighbors. Therefore in the system with $3sN$ degrees of freedom all particles have the same oscillation frequency $\omega_E$. The phonon heat capacity is in this case given by

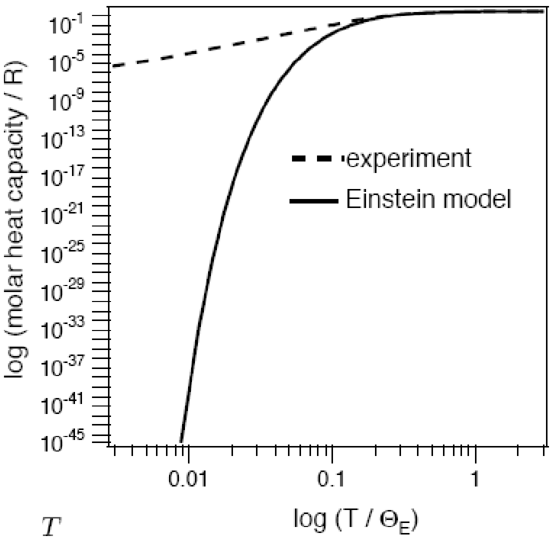
$$C_{ph}(T) = 3sNk_B B_2(\Theta_E/T)\,, \tag{6}$$

where $\Theta_E$ is the Einstein temperature, defined by $\hbar\omega_E = k_B\Theta_E$ and

$$B_2(\xi) = \frac{\xi^2 \exp(\xi)}{[\exp(\xi)-1]^2} \tag{7}$$

is the Einstein function.





At high temperatures, for $\xi = \Theta_E / T << 1$, $B_2(\xi) \cong \dfrac{\xi^2 \exp(\xi)}{(1+\xi-1)^2} = \exp(\xi) \cong 1$, so that

$$C_{ph}(T) \cong 3sNk_B\,, \tag{8}$$

as in the Dulong-Petit law, but at low temperatures, for $\xi >> 1$, where $B_2(\xi) \cong \xi^2 \exp(-\xi)$, the heat capacity has an exponential temperature dependence of the form

$$C_{ph}(T) \cong 3sNk_B \left(\frac{\Theta_E}{T}\right)^2 \exp(-\Theta_E / T)\,. \tag{9}$$

Although $\lim_{T\to 0} C_{ph}(T) \to 0$, the low-temperature dependence of the heat capacity is not proportional to $T^3$ (see the figure above, right). The discrepancy is due to the inappropriate treatment of the acoustic phonon contribution to the heat capacity. Unlike for optical phonons, for which the frequency is almost constant as a function of ***k***, the frequency of acoustic phonons has a much wider interval of variation and the oscillations in different lattice cells must be considered as correlated (the atoms oscillate in phase!). Therefore, since the Einstein model describes in a satisfactory manner the optical phonon contribution to $C_{ph}$, the heat capacity is expressed as

$$C_{ph}(T) = C_{ph}^{opt}(T) + C_{ph}^{ac}(T) \tag{10}$$

where

$$C_{ph}^{opt}(T) = 3(s-1)Nk_B B_2(\Theta_E / T) \tag{11}$$

and the contribution of the acoustic phonons is estimated from the Debye model.

## The Debye model

In the Debye model the frequency of acoustic phonons in a general, anisotropic crystal is written as

$$\omega_{\boldsymbol{k},\lambda} = \omega_\lambda(\boldsymbol{k}) = v_{ac,\lambda}(\theta,\varphi)k\,, \quad \lambda = 1,2,3 \tag{12}$$

with $\theta$, $\varphi$ the polar angles, and their contribution to the heat capacity is given by

$$\begin{aligned} C_{ph}^{ac} &= k_B \sum_{\lambda=1}^{3}\sum_{\boldsymbol{k}} \left(\frac{\hbar\omega_{\boldsymbol{k},\lambda}}{k_B T}\right)^2 \frac{\exp(\hbar\omega_{\boldsymbol{k},\lambda}/k_B T)}{[\exp(\hbar\omega_{\boldsymbol{k},\lambda}/k_B T)-1]^2} \\ &= k_B \sum_{\lambda=1}^{3}\int \left(\frac{\hbar\omega_{\boldsymbol{k},\lambda}}{k_B T}\right)^2 \frac{\exp(\hbar\omega_{\boldsymbol{k},\lambda}/k_B T)}{[\exp(\hbar\omega_{\boldsymbol{k},\lambda}/k_B T)-1]^2} D(\omega_{\boldsymbol{k},\lambda})d\omega_{\boldsymbol{k},\lambda} \end{aligned} \tag{13}$$

where

$$\begin{aligned} D(\omega_{\boldsymbol{k},\lambda}) &= \frac{V}{(2\pi)^3}\int_{\omega(\boldsymbol{k})=const} \frac{dS_\omega}{|\nabla_{\boldsymbol{k}}\omega|} = \frac{V}{(2\pi)^3}\int_{\omega(\boldsymbol{k})=const} \frac{k^2 d\Omega_\omega}{|d\omega_{\boldsymbol{k},\lambda}/dk|} \\ &= \frac{V}{(2\pi)^3}\int_{\omega(\boldsymbol{k})=const} \frac{\omega_{\boldsymbol{k},\lambda}{}^2 d\Omega_\omega}{v_{ac,\lambda}^3(\theta,\varphi)} \end{aligned} \tag{14}$$

is the density of states. The frequency integral in the expression of $C_{ph}^{ac}$ is performed between 0 and $\omega_{\max,\lambda}(\theta,\varphi)$. If

$$\frac{1}{4\pi}\sum_{\lambda=1}^{3}\int_{\omega=const} \frac{d\Omega_\omega}{v_{ac,\lambda}^3(\theta,\varphi)} = \frac{3}{v_{ac}^3} \tag{15}$$

is an angular average of the acoustic velocity (the equality holds as identity in the isotropic crystal), then we can introduce also an angle-independent maximum oscillation frequency (the Debye frequency) $\omega_{\max} = \omega_D$ which, in the Debye model, is also independent on the polarization $\lambda$. This maximum oscillation frequency follows from the normalization condition of the $3N$ acoustic oscillation branches:

$$\sum_{\lambda=1}^{3}\int_0^{\omega_D} D(\omega_{\boldsymbol{k},\lambda})d\omega_{\boldsymbol{k},\lambda} = \left(\frac{1}{4\pi}\sum_{\lambda=1}^{3}\int_{\omega=const}\frac{d\Omega_\omega}{v_{ac,\lambda}^3(\theta,\varphi)}\right)\frac{V}{2\pi^2}\int_0^{\omega_D}\omega^2 d\omega = \frac{V\omega_D^3}{2\pi^2 v_{ac}^3} = 3N\ , \tag{16}$$

and so

$$\omega_D = v_{ac}\sqrt[3]{6\pi^2 N/V}\ . \tag{17}$$

In this case

$$\begin{aligned} C_{ph}^{ac} &= \frac{3}{2}\frac{V}{\pi^2}\frac{k_B}{v_{ac}^3}\left(\frac{k_B T}{\hbar}\right)^2\int_0^{\omega_D}\left(\frac{\hbar\omega}{k_B T}\right)^4\frac{\exp(\hbar\omega/k_B T)}{[\exp(\hbar\omega/k_B T)-1]^2}d\omega \\ &= 9k_B N\left(\frac{T}{\Theta_D}\right)^3 J_4(\Theta_D/T) \end{aligned} \tag{18}$$

where $\Theta_D$ defined through $\hbar\omega_D = k_B\Theta_D$ is the Debye temperature, and

$$J_n(\xi) = \int_0^{\xi}\frac{x^n\exp(x)}{[\exp(x)-1]^2}dx \tag{19}$$

is the Debye-Grüneisen integral, which has no analytical solution. The Debye temperature is proportional to the acoustic velocity, and so is higher for high Young modulus values and for lower crystal densities. It is usually determined by measuring the temperature dependence of the resistance around the Debye temperature.

At high temperatures, for $T >> \Theta_D$, the argument in the $J_4$ integral is very small, since $x << 1$, and after expanding it in series one obtains

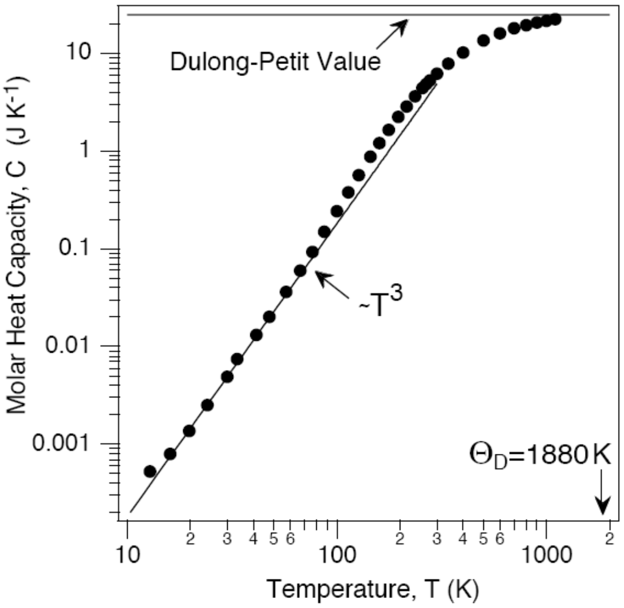
$$
\begin{aligned}
C_{ph}^{ac} &\cong 9k_B N\left(\frac{T}{\Theta_D}\right)^3 \int_0^{\Theta_D/T} \frac{x^4\exp(x)}{(1+x-1)^2}dx = 9k_B N\left(\frac{T}{\Theta_D}\right)^3 \int_0^{\Theta_D/T} x^2\exp(x)dx \\
&\cong 9k_B N\left(\frac{T}{\Theta_D}\right)^3 \int_0^{\Theta_D/T} x^2 dx = 3k_B N
\end{aligned}
\tag{20}
$$

while at low temperatures $T << \Theta_D$ the upper limit of the integral can be extended to $\infty$, so that

$$
C_{ph}^{ac} \cong 9k_B N\left(\frac{T}{\Theta_D}\right)^3 \int_0^{\infty} \frac{x^4\exp(x)}{[\exp(x)-1]^2}dx = \frac{12\pi^4}{5}k_B N\left(\frac{T}{\Theta_D}\right)^3 \propto T^3 \tag{21}
$$

This temperature dependence can be understood from a qualitative argument: at low temperatures only the phonon modes with energy $\hbar\omega < k_B T$ are excited. These modes are, in the ***k*** space, inside a sphere (the thermal sphere), so that the number of modes is proportional to $k^3 \propto \omega^3 \propto T^3$. If each mode has an average excitation energy of $k_B T$, the total energy of excitation is proportional to $T^4$ and hence the heat capacity is proportional to $T^3$.

The total phononic heat capacity, $C_{ph} = C_{ph}^{ac} + C_{ph}^{opt}$, is now in agreement with experiments for both high and low temperatures (see the figure below).



The Debye temperatures for some elemental crystals are given in the table below

| Element | $\Theta_D$ (K) | Element | $\Theta_D$ (K) | Element | $\Theta_D$ (K) |
|---|---|---|---|---|---|
| Li | 344 | Be | 1440 | Cu | 343 |
| Na | 158 | Mg | 400 | Ag | 225 |
| K | 91 | Ca | 230 | C | 2230 |
| Rb | 56 | Sr | 177 | Si | 645 |
| Cs | 38 | Ba | 110 | Ge | 374 |

# Electron dynamics in the crystalline lattice

We have seen that, in the adiabatic approximation, the equation satisfied by the (sub)system of valence electrons with mass *m* and positions $\boldsymbol{r}_i$, $i = 1,\dots, N_{el}$ in the field of ions at positions $\boldsymbol{R}_\alpha$, $\alpha = 1,\dots, N_{ion}$ is

$$\left(\sum_i -\frac{\hbar^2}{2m}\nabla_i^2 + \frac{1}{2}\sum_{\substack{i,j\\ i\neq j}} U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) + \sum_{i,\alpha} V(\boldsymbol{r}_i - \boldsymbol{R}_\alpha)\right)\psi(\boldsymbol{r};\boldsymbol{R}) = E_{el}(\boldsymbol{R})\psi(\boldsymbol{r};\boldsymbol{R}), \tag{1}$$

where $\boldsymbol{r} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_{N_{el}}\}$, $\boldsymbol{R} = \{\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_{N_{ion}}\}$, and the interaction energy between electrons, $U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) = e^2/(4\pi\varepsilon_0 |\boldsymbol{r}_i - \boldsymbol{r}_j|)$, is a bi-particle term of Coulomb type. This equation is quite difficult to solve and, therefore approximate methods are employed. In one of them, called the self-consistent field method (or Hartree-Fock method), the interaction energy between electrons $\frac{1}{2}\sum_{\substack{i,j\\ i\neq j}} U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j)$ is replaced by an effective field of the remaining $N_{el} - 1$ electrons, which can be expressed as a sum of one-particle terms, $\sum_i U_i^{ef}$. More exactly, if $\psi_i(\boldsymbol{r}_i, s_i)$ is the one-particle wavefunction for the *i*th electron, with $s_i$ the spin value, the antisymmetric (with respect to the permutation operator of two electrons) wavefunction of the system of electrons can be expressed as a Slater determinant,

$$\psi(\boldsymbol{r}) = \frac{1}{\sqrt{N_{el}!}}\begin{vmatrix} \psi_1(\boldsymbol{r}_1, s_1) & \psi_1(\boldsymbol{r}_2, s_2) & \cdots & \psi_1(\boldsymbol{r}_{N_{el}}, s_{N_{el}}) \\ \psi_2(\boldsymbol{r}_1, s_1) & \psi_2(\boldsymbol{r}_2, s_2) & \cdots & \psi_2(\boldsymbol{r}_{N_{el}}, s_{N_{el}}) \\ \vdots & & & \vdots \\ \psi_{N_{el}}(\boldsymbol{r}_1, s_1) & \psi_{N_{el}}(\boldsymbol{r}_2, s_2) & \cdots & \psi_{N_{el}}(\boldsymbol{r}_{N_{el}}, s_{N_{el}}) \end{vmatrix}, \tag{2}$$

and the energy of the system is given by

$$E_{el} = \sum_i \langle\psi_i | H_i | \psi_i\rangle + \frac{1}{2}\sum_{\substack{i,j\\ j\neq i}} \langle\psi_i\psi_j | U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) | \psi_i\psi_j\rangle - \frac{1}{2}\sum_{\substack{i,j\\ j\neq i}} \langle\psi_i\psi_j | U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) | \psi_j\psi_i\rangle, \tag{3}$$

with

$$H_i = -\frac{\hbar^2}{2m}\nabla_i^2 + \sum_\alpha V(\boldsymbol{r}_i - \boldsymbol{R}_\alpha) = -\frac{\hbar^2}{2m}\nabla_i^2 + V_i^{ion} \tag{4}$$

the one-particle Hamiltonian. The second term in the right-hand-side of $E_{el}$ is the Hartree term, while the third term is the Hartree-Fock or exchange term.

If the total energy of the electron system could be written as a sum of one-electron energies:

$$E_{el} = \sum_i E_i \langle \psi_i | \psi_i \rangle\,, \tag{5}$$

then the system of interacting electrons would be equivalent to a system of independent electrons that satisfy the Schrödinger equation of motion

$$\left(-\frac{\hbar^2}{2m}\nabla_i^2 + V_i^{ion} + U_i^{ef}\right)\psi_i = (H_i + U_i^{ef})\psi_i = E_i\psi_i\,, \tag{6}$$

with

$$\sum_i \langle \psi_i | U_i^{ef} | \psi_i \rangle = \frac{1}{2}\sum_{i,j\neq i}\langle \psi_i\psi_j | U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) | \psi_i\psi_j \rangle - \frac{1}{2}\sum_{i,j\neq i}\langle \psi_i\psi_j | U_{el}(\boldsymbol{r}_i - \boldsymbol{r}_j) | \psi_j\psi_i \rangle\,. \tag{7}$$

The effective field $U_i^{ef}$ can be determined if the one-electron wavefunctions $\psi_i$ are known, but the latter can only be found if the effective field is given. It is a self-consistent problem, which can only be solved iteratively. More precisely, a set of trial functions is first chosen, such that $H_i\psi_i^{(0)} = E_i\psi_i^{(0)}$, with the help of which the effective potential $U_i^{ef(0)}$ is calculated. If this effective potential is introduced in the one-electron equation, one obtains $\psi_i^{(1)}$ as a solution of $(H_i + U_i^{ef(0)})\psi_i^{(1)} = E_i\psi_i^{(1)}$, and so on. The iterative process stops after, let's say *p* steps, when $\psi_i^{(p-1)} \cong \psi_i^{(p)}$. In conclusion, the motion of one electron in a crystal can be described with the one-electron approximation. We use further this approximation and drop the subscript *i* for simplicity.

## Electrons in a periodic lattice. Bloch functions

We have seen that an electron in a lattice is described by the Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r})\right)\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}), \tag{8}$$

where the energy potential $V = V^{ion} + U^{ef}$ describes the interaction with all positive ions and the remaining electrons, and

$$V(\boldsymbol{r}) = V(\boldsymbol{r} + \boldsymbol{R}_n) \tag{9}$$

for any translation vector $\boldsymbol{R}_n$ of the infinite lattice.

The influence of lattice periodicity on the electron wavefunction can be deduced observing that

$$\left(-\frac{\hbar^2}{2m}\nabla^2_{\boldsymbol{r}+\boldsymbol{R}_n} + V(\boldsymbol{r}+\boldsymbol{R}_n)\right)\psi(\boldsymbol{r}+\boldsymbol{R}_n) = E\psi(\boldsymbol{r}+\boldsymbol{R}_n) = \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r})\right)\psi(\boldsymbol{r}+\boldsymbol{R}_n), \tag{10}$$

since $\nabla_{\boldsymbol{r}+\boldsymbol{R}_n} = \nabla_{\boldsymbol{r}} = \nabla$. This implies that the electron wavefunction after the translation with $\boldsymbol{R}_n$ satisfies the same equation as before, and hence

$$\psi(\boldsymbol{r}+\boldsymbol{R}_n) = C_n\psi(\boldsymbol{r}), \tag{11}$$

with $|C_n| = 1$, because the normalization condition on the volume $V$ imposes that $\int_V |\psi(\boldsymbol{r}+\boldsymbol{R}_n)|^2\, d\boldsymbol{r} = |C_n|^2 \int_V |\psi(\boldsymbol{r})|^2\, d\boldsymbol{r} = 1$. Moreover, at two successive translations we have $C_nC_m = C_{n+m}$, so that

$$C_n = \exp(i\boldsymbol{k}\cdot\boldsymbol{R}_n), \tag{12}$$

or $\psi(\boldsymbol{r}+\boldsymbol{R}_n) = \exp(i\boldsymbol{k}\cdot\boldsymbol{R}_n)\psi(\boldsymbol{r})$, where $\boldsymbol{k}$ is, for now, a wavevector (more precisely, a quasi-wavevector).

The symmetry properties of the electron wavefunction at translations can be used to find the general form of the eigenfunctions of the Schrödinger equation for electrons in a crystalline lattice. More precisely, since $\psi(\boldsymbol{r}) = \exp(-i\boldsymbol{k}\cdot\boldsymbol{R}_n)\psi(\boldsymbol{r}+\boldsymbol{R}_n) = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})\times$ $\{\exp[-i\boldsymbol{k}\cdot(\boldsymbol{r}+\boldsymbol{R}_n)]\psi(\boldsymbol{r}+\boldsymbol{R}_n)\}$, the wavefunctions can be written as

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r})\exp(i\boldsymbol{k}\cdot\boldsymbol{r}), \tag{13}$$

where $u_{\boldsymbol{k}}(\boldsymbol{r}) = \exp[-i\boldsymbol{k}\cdot(\boldsymbol{r}+\boldsymbol{R}_n)]\psi(\boldsymbol{r}+\boldsymbol{R}_n)$ is a periodic function with the same periodicity as the lattice, $u_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}_n)$. Electron wavefunctions of the form given above are called Bloch functions.

Note that for free electrons, which satisfy the Schrödinger equation with vanishing potential energy, and which have wavefunctions $\psi_{\boldsymbol{k}}(\boldsymbol{r}) = V^{-1/2}\exp(i\boldsymbol{k}\cdot\boldsymbol{r})$, with $V$ the normalization volume, all points in space are equivalent (the probability to find the electron at a point $\boldsymbol{r}$, $|\psi_{\boldsymbol{k}}(\boldsymbol{r})|^2 = 1/V$, is independent of $\boldsymbol{r}$), and the wavefunction is at the same time an eigenfunction of the momentum $\hat{\boldsymbol{p}} = -i\hbar\nabla$ with eigenvalues $\boldsymbol{p} = \hbar\boldsymbol{k}$ (besides being an eigenfunction of the Hamiltonian, with eigenvalues $E_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m = \boldsymbol{p}^2/2m$). In contrast, for Bloch waves $|\psi_{\boldsymbol{k}}(\boldsymbol{r})|^2 = |u_{\boldsymbol{k}}(\boldsymbol{r})|^2 \neq \text{const.}$, and only points that differ through a translation vector $\boldsymbol{R}_n$ are equivalent. Therefore $\hbar\boldsymbol{k}$ is the quasi-momentum of the electron in the crystal and $\boldsymbol{k}$ is the quasi-wavevector; in the following, we will still refer to $\boldsymbol{k}$ as wavevector, for simplicity.

The influence of lattice periodicity on the electron energy can be inferred observing first that the electron energy in a crystal is $\boldsymbol{k}$-dependent and real (the Hamiltonian is a hermitic operator): $E = E(\boldsymbol{k})$. The Schrödinger equation can be expressed also as

$$\left(-\frac{\hbar^2}{2m}(\nabla + i\boldsymbol{k})^2 + V(\boldsymbol{r})\right)u_{\boldsymbol{k}}(\boldsymbol{r}) = E(\boldsymbol{k})u_{\boldsymbol{k}}(\boldsymbol{r}) \tag{14}$$

since

$$\nabla\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \nabla[u_{\boldsymbol{k}}(\boldsymbol{r})\exp(i\boldsymbol{k}\cdot\boldsymbol{r})] = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})[\nabla u_{\boldsymbol{k}} + i\boldsymbol{k}u_{\boldsymbol{k}}] = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})(\nabla + i\boldsymbol{k})u_{\boldsymbol{k}} \tag{15}$$

and hence $\nabla^2\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})(\nabla + i\boldsymbol{k})^2 u_{\boldsymbol{k}}$. Because the energy is real, from the (identical) Schrödinger equations satisfied by the wavefunction for $-\boldsymbol{k}$ and its complex conjugate:

$$\left(-\frac{\hbar^2}{2m}(\nabla-i\boldsymbol{k})^2+V(\boldsymbol{r})\right)u_{-\boldsymbol{k}}(\boldsymbol{r})=E(-\boldsymbol{k})u_{-\boldsymbol{k}}(\boldsymbol{r})\,, \tag{16}$$

$$\left(-\frac{\hbar^2}{2m}(\nabla-i\boldsymbol{k})^2+V(\boldsymbol{r})\right)u_{\boldsymbol{k}}^*(\boldsymbol{r})=E(\boldsymbol{k})u_{\boldsymbol{k}}^*(\boldsymbol{r})\,, \tag{17}$$

it follows that the energy is an even function of ***k***:

$$E(\boldsymbol{k})=E(-\boldsymbol{k})\,. \tag{18}$$

In particular, the iso-energetic surfaces of an electron in a lattice, i.e. the surfaces in ***k*** space for which $E(\boldsymbol{k})=$const., have an inversion symmetry; they are not, however, spheres as for free electrons.

If in the crystalline lattice we perform a translation of ***k*** with a vector in the reciprocal space, ***G***, we obtain

$$E(\boldsymbol{k})=E(\boldsymbol{k}+\boldsymbol{G})\ , \tag{19}$$

because, by replacing ***k*** with $\boldsymbol{k}+\boldsymbol{G}$ in (11) we obtain $\psi(\boldsymbol{r}+\boldsymbol{R}_n)=\exp(i\boldsymbol{k}\cdot\boldsymbol{R}_n)\psi(\boldsymbol{r})$ $=\exp[i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{R}_n]\psi(\boldsymbol{r})$ since $\exp(i\boldsymbol{G}\cdot\boldsymbol{R}_n)=1$ (the Schrödinger equation and the wave-function are identical after a translation in the ***k*** space). Therefore, the energy is a periodic function in the reciprocal lattice space (in ***k*** space), and takes distinct values only inside the first Brillouin zone (remember that the whole ***k*** space can be divided in Brillouin zones with all possible orders *n*, which have the same volume and can be reduced to the first Brillouin zone by applying symmetry operations). In the Schrödinger equation for the electron ***k*** is a parameter. As for phonons, where we have different solutions $\omega(\boldsymbol{k})$ for a given ***k***, associated to different oscillation branches, we can have also different eigenfunctions and eigenvalues of the electronic Hamiltonian (different energy values $E_n(\boldsymbol{k})$) for a given wavevector ***k***.

## Electron velocity in a crystal

We have seen that the electrons in a crystalline lattice are Bloch waves. The so-called Bloch electrons are quasi-particles since their properties are dependent on the crystalline lattice. If an electron with wavevector ***k*** is inside a lattice, one can introduce the concept of average velocity through the quantum mechanical definition

$$\langle \boldsymbol{v}\rangle = \boldsymbol{v} = \frac{1}{m}\langle \psi_{\boldsymbol{k}} \mid \hat{\boldsymbol{p}} \mid \psi_{\boldsymbol{k}}\rangle = \frac{1}{m}\int_V \psi^*_{\boldsymbol{k}}(\boldsymbol{r})\frac{\hbar}{i}\nabla \psi_{\boldsymbol{k}}(\boldsymbol{r})d\boldsymbol{r}\ , \tag{20}$$

with $\hat{\boldsymbol{p}} = (\hbar / i)\nabla$ the momentum operator. For Bloch electrons the average velocity can be expressed as

$$\boldsymbol{v} = \frac{1}{m}\int_V u^*_{\boldsymbol{k}}(\boldsymbol{r})\frac{\hbar}{i}(\nabla + i\boldsymbol{k})u_{\boldsymbol{k}}(\boldsymbol{r})d\boldsymbol{r} = \frac{\hbar \boldsymbol{k}}{m} + \frac{\hbar}{im}\int_V u^*_{\boldsymbol{k}}(\boldsymbol{r})\nabla u_{\boldsymbol{k}}(\boldsymbol{r})d\boldsymbol{r}\ , \tag{21}$$

which differs from the expression for free electrons, $\boldsymbol{v} = \hbar\boldsymbol{k}/m = \boldsymbol{p}/m$, through the second term.

To calculate the average electron velocity, we differentiate (14) with respect to ***k*** (we apply the gradient in ***k*** space), and obtain

$$\left( i\frac{\hbar^2}{m}(\nabla + i\boldsymbol{k}) + \nabla_{\boldsymbol{k}}E \right)u_{\boldsymbol{k}}(\boldsymbol{r}) = 0\ . \tag{22}$$

if the potential energy does not depend explicitly on ***k***. From this result and the normalization condition $\int_V u^*_{\boldsymbol{k}}(\boldsymbol{r})u_{\boldsymbol{k}}(\boldsymbol{r})d\boldsymbol{r} = 1$ we can express (21) as

$$\frac{\hbar}{im}\int_V u^*_{\boldsymbol{k}}(\boldsymbol{r})(\nabla + i\boldsymbol{k})u_{\boldsymbol{k}}(\boldsymbol{r})d\boldsymbol{r} = \frac{1}{\hbar}\nabla_{\boldsymbol{k}}E\ , \tag{23}$$

so that, finally,

$$\boldsymbol{v} = \hbar^{-1}\nabla_{\boldsymbol{k}}E\ . \tag{24}$$

This relation is identical to that obtained for free electrons with a parabolic dispersion relation, in which $E = \hbar^2\boldsymbol{k}^2/2m$ and, thus, $\nabla_{\boldsymbol{k}}E = \hbar^2\boldsymbol{k}/m$ and $\boldsymbol{v} = \hbar\boldsymbol{k}/m = \hbar^{-1}\nabla_{\boldsymbol{k}}E$. Therefore, (24) is a more general expression of average velocity than (21), the difference being that in (24) the mass of electrons does not appear explicitly; in a crystalline lattice the free electron mass *m* can be replaced by an effective mass, so that (24) holds, but not (21).

## Electron acceleration in a crystalline lattice. Effective mass

Let us now consider that the electrons in the crystal are accelerated by an external force $\boldsymbol{F}$. We separate the effects of the external forces on the electron and the Coulomb interactions inside the crystal, by modeling the Bloch electron as a quasi-particle that reacts only to the external forces. The influence of the Coulomb interactions on the electron dynamics is characterized by an effective electron mass.

Because, according to the Ehrenfest's theorem, the classical equations of motion are valid for the average values of the quantum observables, we have

$$\frac{dE(\boldsymbol{k})}{dt} = \nabla_{\boldsymbol{k}} E \frac{d\boldsymbol{k}}{dt} = \boldsymbol{v} \cdot \frac{d(\hbar \boldsymbol{k})}{dt} = \boldsymbol{v} \cdot \frac{d\boldsymbol{p}}{dt} = \boldsymbol{v} \cdot \boldsymbol{F} \ , \tag{25}$$

i.e. the quasi-momentum $\boldsymbol{p} = \hbar \boldsymbol{k}$ of an electron in the crystal plays the same role as the momentum of a free electron. Here $\boldsymbol{F}$ is only the external force; the internal forces in the crystal are taken into account when calculating the dispersion relation $E(\boldsymbol{k})$, and thus are included in the expression of the average velocity.

The acceleration of the electron in the crystal can be defined as

$$\boldsymbol{a} = \frac{d\boldsymbol{v}}{dt} = \frac{1}{\hbar} \nabla_{\boldsymbol{k}} \left( \frac{dE}{dt} \right) = \frac{1}{\hbar} \nabla_{\boldsymbol{k}} \left( \nabla_{\boldsymbol{k}} E \frac{d\boldsymbol{k}}{dt} \right) = \frac{1}{\hbar^2} \nabla_{\boldsymbol{k}} (\nabla_{\boldsymbol{k}} E \cdot \boldsymbol{F}) \tag{26}$$

or, written on Cartesian components $\mu, \nu = x,y,z$,

$$a_\mu = \frac{dv_\mu}{dt} = \sum_\nu \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_\mu \partial k_\nu} F_\nu = \sum_\nu \left( \frac{1}{m_{eff}} \right)_{\mu\nu} F_\nu \ . \tag{27}$$

From the relation above it follows that, in general, the acceleration of an electron in the crystal has a different direction than that of the applied force. The relation between the acceleration and the applied force can be used to introduce, as for the case of free electrons, a second-order tensor parameter, called effective mass. Its inverse, defined as

$$\left( \frac{1}{m_{eff}} \right)_{\mu\nu} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_\mu \partial k_\nu} , \tag{28}$$

is a symmetric tensor, since $(1/m_{eff})_{\mu\nu} = (1/m_{eff})_{\nu\mu}$. The inverse of the effective mass can take both positive and negative values, unlike for a free electron, where it is a scalar parameter and always positive. The effective mass incorporates the effect of the crystal lattice on the motion of electrons; it is not a characteristic of the electron (as is its free mass) but of the electron-lattice interaction. The effective mass is inversely proportional to the curvature of the dispersion relation and is infinite at inflexion points, where the curvature vanishes.

If $\boldsymbol{k}_0$ is an extremum (maximum or minimum) of the energy dispersion relation in the $\boldsymbol{k}$ space, for which $(\nabla_{\boldsymbol{k}} E)_{\boldsymbol{k}=\boldsymbol{k}_0} = 0$, a Taylor series expansion of $E$ around $\boldsymbol{k}_0$ in the harmonic approximation (in which only terms up to the second order are considered) can be written as

$$E(\boldsymbol{k}) = E(\boldsymbol{k}_0) + \frac{1}{2}\sum_{\mu,\nu}\left(\frac{\partial^2 E}{\partial k_\mu \partial k_\nu}\right)_{k=k_0}(k_\mu - k_\mu^0)(k_\nu - k_\nu^0) = E_0 + \frac{\hbar^2}{2}\sum_{\mu,\nu}\left(\frac{1}{m_{eff}}\right)_{\mu\nu}(k_\mu - k_\mu^0)(k_\nu - k_\nu^0) \tag{29}$$

We have denoted $E(\boldsymbol{k}_0) = E_0$. The above form is positively defined if the effective mass takes only positive values, and, reducing it to the principal axes we obtain an ellipsoidal iso-energetic surface that satisfies the equation

$$E(\boldsymbol{k}) = E_0 + \frac{\hbar^2}{2}\left(\frac{k_1^2}{m_{eff,1}} + \frac{k_2^2}{m_{eff,2}} + \frac{k_3^2}{m_{eff,3}}\right). \tag{30}$$

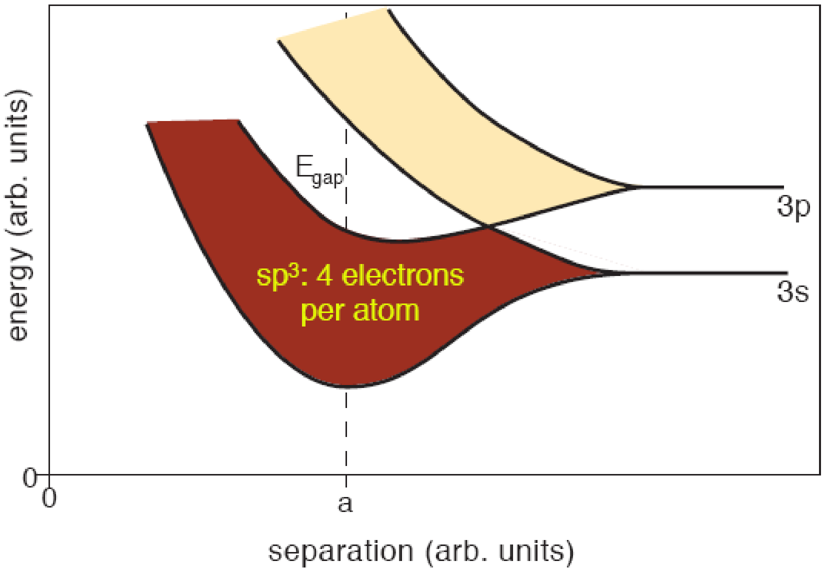
If two of the semiaxes of the ellipsoid are equal, for example if $m_{eff,1} = m_{eff,2}$ the common value of the effective mass is called transverse effective mass $m_t$, and $m_{eff,3} = m_l$ is the longitudinal effective mass; the ellipsoid is in this case an ellipsoid of rotation. If all diagonal components of the effective mass tensor are equal, the iso-energetic surfaces become spheres, $m_{eff,1} = m_{eff,2} = m_{eff,3} = m_{eff}$, and the dispersion relation of the electron in the crystal is similar to that of the free electron with mass *m*, except that *m* is replaced with the effective mass:

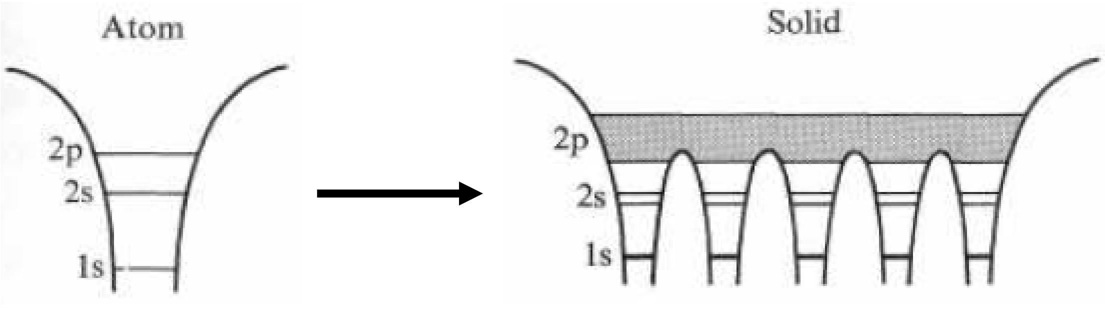
$$E(\boldsymbol{k}) = \frac{\hbar^2 \boldsymbol{k}^2}{2m_{eff}}, \tag{31}$$

if the reference energy is chosen such that $E_0 = 0$. The acceleration has the same direction as the external force only along the principal axes of the ellipsoid, or for spherical iso-energetic surfaces.

## Energy bands

We have seen that the electron energy can take multiple values for a given ***k*** in the first Brillouin zone. The energy spectrum for the electrons in a crystal can be quantitatively determined starting from two models: the approximation of quasi-free electrons (the weak-binding approximation) or the approximation of quasi-bound electrons (the tight-binding approximation). These approximations correspond to two extreme cases. In the first one it is assumed that the state of the electron in the crystal can be modeled as a perturbed state of a free electron, while in the second approximation, the state of the electron is a perturbed state of a bound electron in an isolated atom, the perturbation in both cases being due to the periodic lattice potential. The weak-binding approximation is particularly suitable for treating the energetic spectrum of valence electrons in metals, while the tight-binding approximation is more suitable for semiconductor and isolating materials. In both cases the periodicity of the crystalline lattice leads to the formation of allowed and forbidden energy bands. In the figure below it is shown how energy bands, separated by gaps, form from atomic *s* and *p* orbitals as the separation between atoms decreases in Si; *a* is the lattice constant.





## The tight-binding approximation

Let us assume that the wavefunction of the Bloch electron in the crystal, $\psi_{\boldsymbol{k}}(\boldsymbol{r})$, can be expressed as a linear combination of atomic wavefunctions:

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_n C_n \varphi_n(\boldsymbol{r} - \boldsymbol{R}_n) \tag{32}$$

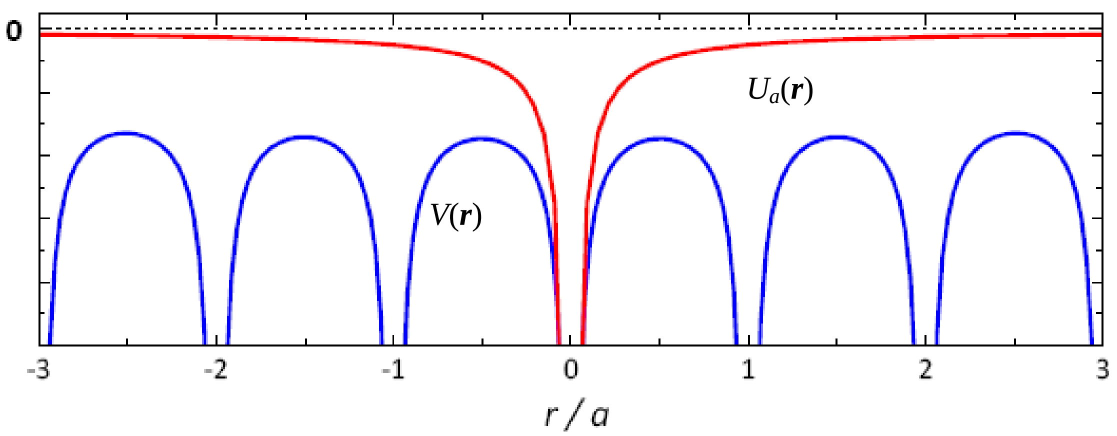
where $\varphi_n(\boldsymbol{r} - \boldsymbol{R}_n)$ is the wavefunction of the atom at position $\boldsymbol{R}_n$ and $C_n = \exp(i\boldsymbol{k} \cdot \boldsymbol{R}_n)$ since the electron wavefunction must be invariant (up to a phase factor) at translation. The atomic wavefunctions satisfy the Schrödinger equation for the isolated atom,

$$H_a \varphi_n = \left( -\frac{\hbar^2}{2m} \nabla^2 + U_a(\boldsymbol{r} - \boldsymbol{R}_n) \right) \varphi_n = E_a \varphi_n , \tag{33}$$

where $H_a$ is the Hamiltonian and $U_a$ is the potential energy of an isolated atom and $E_a$ is the corresponding energy, whereas the Schrödinger equation for the electron in the crystal is

$$H\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \left( -\frac{\hbar^2}{2m} \nabla^2 + V(\boldsymbol{r}) \right) \psi_{\boldsymbol{k}}(\boldsymbol{r}) = E_{\boldsymbol{k}} \psi_{\boldsymbol{k}}(\boldsymbol{r}) = [H_a + W(\boldsymbol{r})] \psi_{\boldsymbol{k}}(\boldsymbol{r}) , \tag{34}$$

with $W(\boldsymbol{r}) = V(\boldsymbol{r}) - U_a(\boldsymbol{r} - \boldsymbol{R}_n) < 0$ the perturbation energy due to the crystal. The fact that $V(\boldsymbol{r}) < U_a(\boldsymbol{r} - \boldsymbol{R}_n)$, and hence $W(\boldsymbol{r}) < 0$, expresses the stability of the crystal, the potential energy in the lattice being lower than in an isolated atom. The electron energies in the crystal and in the isolated atom are illustrated in the figure below.



The last equation can be re-written as

$$\sum_n \exp(i\boldsymbol{k}\cdot\boldsymbol{R}_n)[H_a + W(\boldsymbol{r}) - E_{\boldsymbol{k}}]\varphi_n = 0\,, \tag{35}$$

or

$$\sum_n \exp(i\boldsymbol{k}\cdot\boldsymbol{R}_n)[W(\boldsymbol{r})\varphi_n + (E_a - E_{\boldsymbol{k}})\varphi_n] = 0\,, \tag{36}$$

which becomes, after left-side multiplication with $\exp(-i\boldsymbol{k}\cdot\boldsymbol{R}_m)\varphi_m^*(\boldsymbol{r}-\boldsymbol{R}_m)$ and integration over the crystal volume *V*,

$$\sum_n \exp[i\boldsymbol{k}\cdot(\boldsymbol{R}_n - \boldsymbol{R}_m)][A(\boldsymbol{R}_n - \boldsymbol{R}_m) + (E_a - E_{\boldsymbol{k}})S(\boldsymbol{R}_n - \boldsymbol{R}_m)] = 0\,. \tag{37}$$

Here

$$A(\boldsymbol{R}_n - \boldsymbol{R}_m) = \int_V \varphi_m^*(\boldsymbol{r}-\boldsymbol{R}_m)W(\boldsymbol{r})\varphi_n(\boldsymbol{r}-\boldsymbol{R}_n)d\boldsymbol{r} \tag{38}$$

is the exchange integral, which depends on the overlap of the atomic wavefunctions and the crystal perturbation and defines the exchange interaction energy, and

$$S(\boldsymbol{R}_n - \boldsymbol{R}_m) = \int_V \varphi_m^*(\boldsymbol{r}-\boldsymbol{R}_m)\varphi_n(\boldsymbol{r}-\boldsymbol{R}_n)d\boldsymbol{r} \tag{39}$$

is the overlap integral, which depends only on the overlapping degree of the atomic wavefunctions. If $\boldsymbol{R}_n - \boldsymbol{R}_m = \boldsymbol{R}_l$ is the translation vector between the *n*th and *m*th atoms, it follows finally that

$$E_{\boldsymbol{k}} = E_a + \frac{\sum_l A(\boldsymbol{R}_l)\exp(i\boldsymbol{k}\cdot\boldsymbol{R}_l)}{\sum_l S(\boldsymbol{R}_l)\exp(i\boldsymbol{k}\cdot\boldsymbol{R}_l)}\,. \tag{40}$$

Thus, the energy of the electron in the crystal differs from the energy in an isolated atom through a periodic function of ***k***. Instead of a discrete energy level (as in atoms) we have now an energy band with a width determined by the maximum and minimum values of the perturbation term.

In the simplifying assumption that the atomic wavefunctions are rapidly decreasing with the distance, such that the overlap integral is negligible even for neighboring atoms, i.e. such that

$$S(\mathbf{R}_l)=\begin{cases}1, & \mathbf{R}_l=0\ (n=m)\\ 0, & \mathbf{R}_l\neq 0\ (n=m)\end{cases},\qquad \text{and hence}\qquad \sum_l S(\mathbf{R}_l)\exp(i\mathbf{k}\cdot\mathbf{R}_l)=1, \tag{41}$$

we obtain that

$$E_{\mathbf{k}}=E_a-C-A\sum_{l=nn}\exp(i\mathbf{k}\cdot\mathbf{R}_l), \tag{42}$$

where the sum extends now only over the nearest neighbors, for which $A(\mathbf{R}_l)=-A<0$ for *s* atomic orbitals, for which $A > 0$, and

$$-C=A(0)=\int_V \varphi_n^*(\mathbf{r}-\mathbf{R}_n)W(\mathbf{r})\varphi_n(\mathbf{r}-\mathbf{R}_n)d\mathbf{r}<0 \tag{43}$$

is a constant (independent of $\mathbf{k}$) parameter. For a simple cubic lattice with period $a$, in which an atom at the origin of the coordinate system has 6 nearest neighbors at positions [[100]], $[[\bar{1}00]]$, [[010]], $[[0\bar{1}0]]$, [[001]], and $[[00\bar{1}]]$,

$$\begin{aligned}\sum_{l=nn}\exp(i\mathbf{k}\cdot\mathbf{R}_l)&=\exp(ik_1a)+\exp(-ik_1a)+\exp(ik_2a)+\exp(-ik_2a)+\exp(ik_3a)+\exp(-ik_3a)\\ &=2[\cos(k_1a)+\cos(k_2a)+\cos(k_3a)]\end{aligned} \tag{44}$$

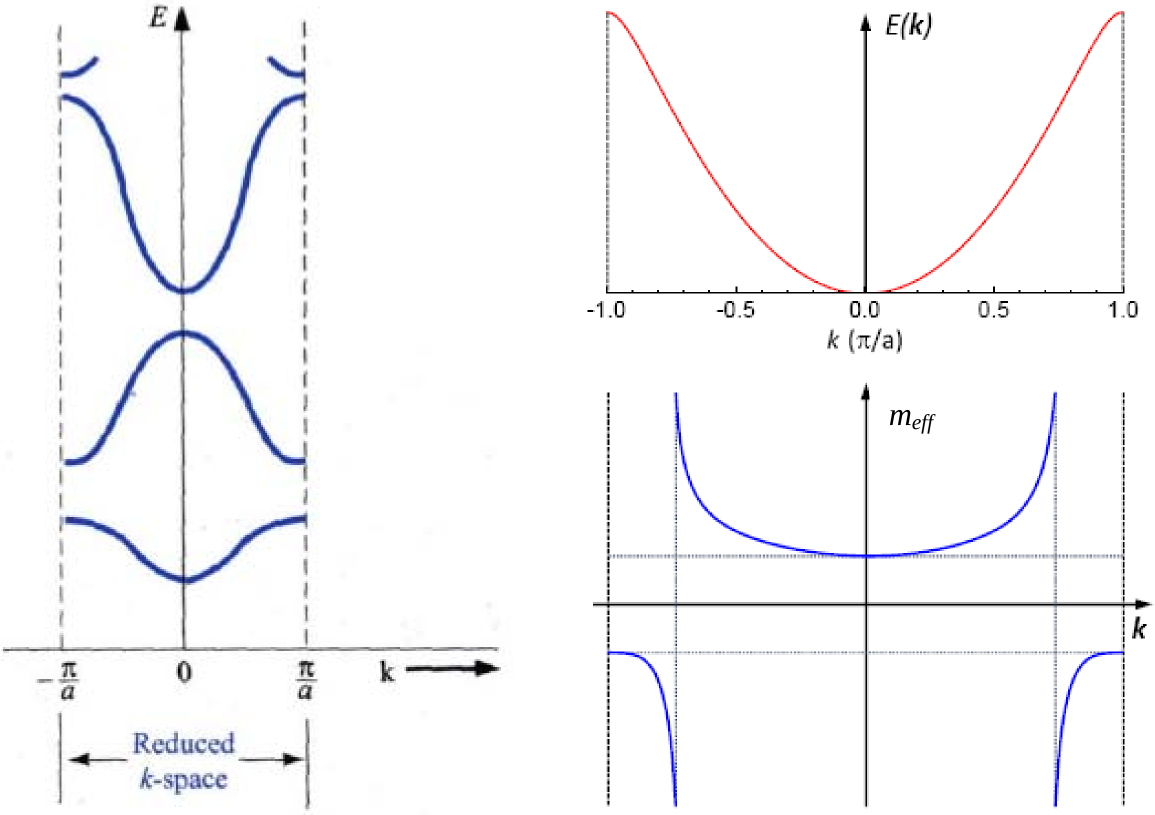
$$E_{\mathbf{k}}=E_a-C-2A[\cos(k_1a)+\cos(k_2a)+\cos(k_3a)], \tag{45}$$

which is an even function of the wavevector components; all possible energy values are obtained for wavevector components in the first Brillouin zone, i.e. for $-\pi/a\le k_i\le\pi/a$. The expression above indicates that, in the crystal, the energy level of the isolated atom shifts downwards with $C$ due to the interaction between atoms, which renders the crystal more stable than an isolated atom, and transforms into an energy band, which is periodic in the wavevector components and extends between $E_{\min}$ and $E_{\max}$, with

$$E_{\min} = E_a - C - 6A, \quad \text{for} \quad k_i = 0\,, \tag{46a}$$

$$E_{\max} = E_a - C + 6A, \quad \text{for} \quad k_i = \pm\pi / a\,. \tag{46b}$$

The widths of the energy bands, $\Delta E = 12A$, increase with the exchange integral (with the overlap of the atomic orbitals) and are wider for the higher discrete atomic energy levels since the wavefunctions of the outer atomic levels are more extended in space. Different energy bands form starting from different atomic orbitals, and an additional subscript *i* label these energy bands. In the figure below (left) different energy bands are represented in the first Brillouin zone. These bands are formed (in the increasing energy order) from atomic orbitals for which *A* is positive, negative, positive, etc.



These (allowed) energy bands are separated by energy gaps (forbidden energy bands) with widths $E_{gi}$. The width of the first energy band, for example, is given by

$$E_{g1} = E_{\min,2} - E_{\max,1} = (E_{a,2} - E_{a,1}) - (C_2 - C_1) - 6(A_1 + A_2)\,. \tag{47}$$

If both $A_1$ and $A_2$ correspond to *s* states, the energy gap is determined from the extremities of the two bands at different ***k*** values (center and edges of the first Brillouin zone), whereas the energy gap between bands that form from *s* and *p* atomic orbitals is determined by states with the same ***k***: at the edges of the first Brillouin zone, for $A_2 = A_p < 0$ (the exchange integral is positive) and $A_1 = A_s > 0$, and at the center of the first Brillouin zone if $A_2 > 0$ and $A_1 < 0$.

## Effective mass in electronic energy bands

Let us calculate the effective electron mass at the extreme points of the energy band with the dispersion relation $E_{\boldsymbol{k}} = E_a - C - 2A[\cos(k_1 a) + \cos(k_2 a) + \cos(k_3 a)]$. Near the center of the first Brillouin zone, when $k_i a << 1$ and $\cos(k_i a) \cong 1 - (k_i a)^2 / 2$,

$$E_{\boldsymbol{k}} = E_a - C - 6A + Aa^2 (k_1^2 + k_2^2 + k_3^2) = E_{\min} + Aa^2 k^2, \tag{48}$$

and hence the energy depends quadratically on the wavevector, $E_{\boldsymbol{k}} = E_{\min} + \hbar^2 k^2 / 2m_{eff}$, as for free electrons, with an effective mass

$$m_{eff} = \hbar^2 (\nabla_{\boldsymbol{k}}^2 E_{\boldsymbol{k}})^{-1} = \hbar^2 / 2Aa^2 . \tag{49}$$

The effective mass is positive for energy bands that form from *s* atomic orbitals, for which $A > 0$. For the particular case considered here, that of a simple cubic lattice, the iso-energetic surfaces $E_{\boldsymbol{k}}$ = const. in the neighborhood of the center of the first Brillouin zone are spheres. Note that $m_{eff}$ depends on the dispersion relation, and hence on the crystal structure.

On the contrary, at the edges of the first Brillouin zone, introducing the new variables $k_i' = \pm(\pi / a - k_i)$, such that $\cos(k_i a) = \cos(\mp k_i' a + \pi) = -\cos(k_i' a)$, the dispersion relation can be expressed as

$$E_{\boldsymbol{k}'} = E_a - C + 2A[\cos(k_1' a) + \cos(k_2' a) + \cos(k_3' a)] \tag{50}$$

and, for $k_i' a << 1$ and $\cos(k_i' a) \cong 1 - (k_i' a)^2 / 2$,

$$E_{\boldsymbol{k}'} = E_a - C + 6A - Aa^2 (k_1'^2 + k_2'^2 + k_3'^2) = E_{\max} - Aa^2 k'^2 . \tag{51}$$

The dispersion relation is again similar to that of a free electron, i.e. has the form $E_{\boldsymbol{k}'} = E_{\max} + \hbar^2 k'^2 / 2m_{eff}$, with an effective mass

$$m_{eff} = \hbar^2 (\nabla_{\boldsymbol{k}}^2 E_{\boldsymbol{k}})^{-1} = -\hbar^2 / 2Aa^2\,, \tag{52}$$

which is negative if $A > 0$. An example of such a situation is represented in the figure above (right). Note that the effective mass is negative at the center of the first Brillouin zone and positive at its edges for energy bands that form from *p* atomic orbitals, for which $A < 0$. In general, $m_{eff} > 0$ in the neighborhood of the minimum energy value in the band and $m_{eff} < 0$ near the maximum energy value in the band.

Although (51) shows that the iso-energetic surfaces are also spherical at the edges of the first Brillouin zone, they have complicated forms at intermediate energy values, between the center and the edges of the first Brillouin zone. Due to the periodicity of electron energy in the ***k*** space, the iso-energetic surfaces are the same in all cells in the reciprocal space, so that the iso-energetic surfaces are multiple connected. The same general results are obtained in the weak-binding approximation.

## Electrons and holes

If the electron in the crystal, with an electric charge $-e$ and a positive (isotropic) effective mass $m_{eff} > 0$ is placed in an electric field ***E***, its acceleration

$$\boldsymbol{a} = d\boldsymbol{v} / dt = \boldsymbol{F} / m_{eff} = -e\boldsymbol{E} / m_{eff} \tag{53}$$

is similar in form to that of a free electron. On the other hand, in the neighborhood of the maximum energy value in the band $m_{eff} < 0$ and the electron equation of motion is given by

$$\boldsymbol{a} = d\boldsymbol{v} / dt = \boldsymbol{F} /(-|\, m_{eff}\,|) = -e\boldsymbol{E} /(-|\, m_{eff}\,|)\,. \tag{54}$$

The negative effective mass in the equation above has no analog for free particles. It has, as consequence, that an electric field will decelerate the electron, instead of accelerating it. To avoid such an awkward interpretation, it is considered that the motion of the electron with a negative effective mass in the electric field ***E*** is equivalent with the motion of a quasi-particle with electric charge $+e$ and a positive effective mass, equal to $|m_{eff}|$. This quasi-particle is called **hole**. The hole is a quasi-particle since, unlike the electron, it has no physical meaning in free space; holes exist only in crystals and reflect the behavior of valence electrons. In a generalized sense, even the electron in the crystal can be seen as a quasi-particle endowed with an effective mass, which is different from that of the free electron.

If the electrons in the upper part of an occupied energy band (the valence band) acquire sufficient energy from thermal vibrations, for example, to go into an unoccupied state in the next energy band, called conduction band, the remaining empty states can be considered as holes. Since the empty states can be occupied by other electrons, the holes can be seen as moving throughout the valence band, i.e. they can be regarded as free quasi-particles in the valence band. In other words, an energy band occupied with electrons with the exception of its upper part, can be seen as partially occupied with holes. Because the electrons that participate at electrical and thermal conduction are those able to move (quasi-)free in the crystal, the concept of holes allows a major simplification in the treatment of the system of electrons in the valence band: it is no longer necessary to deal with the motion of the entire system of electrons, but only with the motion of a much smaller number of holes.

In general, the holes in the valence band have not only an opposite electric charge, but also a different effective mass than the electrons in the conduction band, since their dispersion relation is different (the valence and conduction bands originate from different isolated atomic levels). The holes move in the direction of the applied electric field, whereas electrons move in the opposite direction!.

## Density of electron states

Real crystals have finite sizes and, therefore, the solution of the Schrödinger equation for the electrons in the crystal depends on the boundary conditions. As for phonons, we assume that in large crystals the surface phenomena do not influence significantly the electron dynamics inside the crystal, and hence use the cyclic (Born-Karman) boundary conditions. In the orthorhombic symmetry, for example, if the dimensions of the crystal along the orthogonal Cartesian coordinates $x_i$, $i$ = 1,2,3 ($x_1 = x$, $x_2 = y$, $x_3 = z$) are denoted by $L_i$, the cyclic conditions impose that

$$\begin{aligned}\psi_{\mathbf{k}}(x_1,x_2,x_3) &= \psi_{\mathbf{k}}(x_1+L_1,x_2,x_3) = \psi_{\mathbf{k}}(x_1,x_2+L_2,x_3) = \psi_{\mathbf{k}}(x_1,x_2,x_3+L_3) \\ &= \psi_{\mathbf{k}}(x_1+L_1,x_2+L_2,x_3+L_3)\end{aligned} \tag{55}$$

or, since $\psi_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r})\exp(i\boldsymbol{k}\cdot\boldsymbol{r})$ with $u_{\boldsymbol{k}}(\boldsymbol{r})$ the same in all lattice cells,

$$k_i L_i = 2\pi m_i, \qquad k_i = 2\pi m_i / L_i \tag{56}$$

with $m_i$ integer numbers. Similarly, in a finite crystal with $N_i$ atoms along the *i* direction, such that $L_i = N_i a_i$, with $a_i$ the respective lattice constants, it follows that

$$k_i = \frac{2\pi m_i}{N_i a_i}, \tag{57}$$

where, for wavevectors in the first Brillouin zone, the integers $m_i$ can take only $N_i$ values in the intervals

$$-N_i/2 \le m_i < N_i/2. \tag{58}$$

Thus, for a simple lattice (with one atom in the basis), the number of distinct energy states of electrons in an allowed energy band is $N = N_1 N_2 N_3$, and these states can be occupied by $2N$ electrons because, according to the Pauli principle for fermions, only two electrons with opposite spins can occupy an energy state characterized by a given ***k***. In large crystals the distance between energy levels is quite small and the energy band is approximated as a (quasi-)continuous function of ***k***, case in which any sum over states in the ***k*** space can be replaced (as in the case of phonons) by an integral over the first Brillouin zone. More precisely, for any function $F(\boldsymbol{k})$,

$$\sum_{\boldsymbol{k}} F(\boldsymbol{k}) = \frac{V}{(2\pi)^3}\sum_{\boldsymbol{k}} F(\boldsymbol{k})\Delta\boldsymbol{k} = \frac{V}{(2\pi)^3}\int_{1^{st}BZ} F(\boldsymbol{k})d\boldsymbol{k} \tag{59}$$

where

$$\Delta\boldsymbol{k} = \Delta k_1 \Delta k_2 \Delta k_3 = \frac{(2\pi)^3}{L_1 L_2 L_3} = \frac{(2\pi)^3}{V} \tag{60}$$

is the volume in ***k*** space occupied by a distinct electron state.

In particular, the density of states in the $\boldsymbol{k}$ space, defined as the number of electron states with a given spin orientation, $dN_{el}$, in the volume element $d\boldsymbol{k}$, is given by

$$\frac{dN_{el}}{d\boldsymbol{k}} = \frac{1}{\Delta\boldsymbol{k}} = \frac{V}{(2\pi)^3}, \tag{61}$$

or

$$N_{el} = \frac{V}{(2\pi)^3} \int_{1^{st}BZ} d\boldsymbol{k} .$$

The density of states per unit volume, defined as the number of states per unit volume with a given spin orientation in the energy interval $dE$ is then

$$D(E) = \frac{1}{V}\frac{dN_{el}}{dE} = \frac{d}{dE}\left( \frac{1}{(2\pi)^3} \int_{V_E} d\boldsymbol{k} \right), \tag{62}$$

where $V_E$ is the volume in $\boldsymbol{k}$ space between the iso-energetic surfaces $E(\boldsymbol{k})$ and $E(\boldsymbol{k}) + dE$. As for phonons, $d\boldsymbol{k} = dSdk_\perp = dS(dE/|\nabla_{\boldsymbol{k}} E|)$ with $dS$ the infinitesimal element on the $E(\boldsymbol{k})$ = const. surface and $dk_\perp$ normal to this surface, and the expression above simplifies to

$$D(E) = \frac{1}{(2\pi)^3} \int_{E(\boldsymbol{k})=const} \frac{dS}{|\nabla_{\boldsymbol{k}} E|}, \tag{63}$$

which becomes

$$D(E) = \frac{1}{(2\pi)^3} \frac{m_{eff}k}{\hbar^2} \int_{\Omega_k} d\Omega_k = \frac{m_{eff}k}{2\pi^2\hbar^2} = \frac{(2m_{eff})^{3/2}}{4\pi^2\hbar^3}(E - E_0)^{1/2} \tag{64}$$

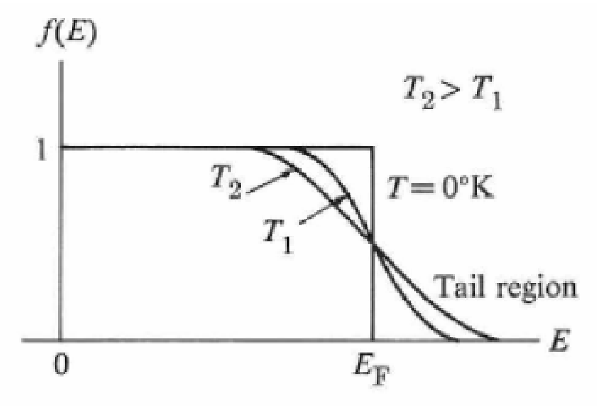
for spherical iso-energetic surfaces, for which $E(\boldsymbol{k}) = E_0 + \hbar^2 k^2 / 2m_{eff}$. In this case $|\nabla_{\boldsymbol{k}} E| = \hbar^2 k / m_{eff}$ and $dS = k^2 d\Omega_k$, with $d\Omega_k$ the element of solid angle.

## Classification of solid state materials

One of the most important achievements of the energy band theory is the possibility to classify solid state materials in metals, isolators, and semiconductors. This classification is not based on the structure of the energy bands, which is quite the same in all materials, but on the degree of occupation of these bands. The available number of energy states is occupied by electrons in agreement with the Pauli exclusion principle. More precisely, at low temperatures (in principle, at $T$ = 0 K) the states are occupied in the order of increasing energy value, such that only two electrons (with opposite spins) are allowed on an energy state $E$ with a given $\boldsymbol{k}$ value. The available states are occupied by electrons up to an energy level called Fermi energy, or Fermi level $E_F$. Because electrons are fermions, their quantum statistical distribution function at temperature $T$ is described by the Fermi-Dirac formula

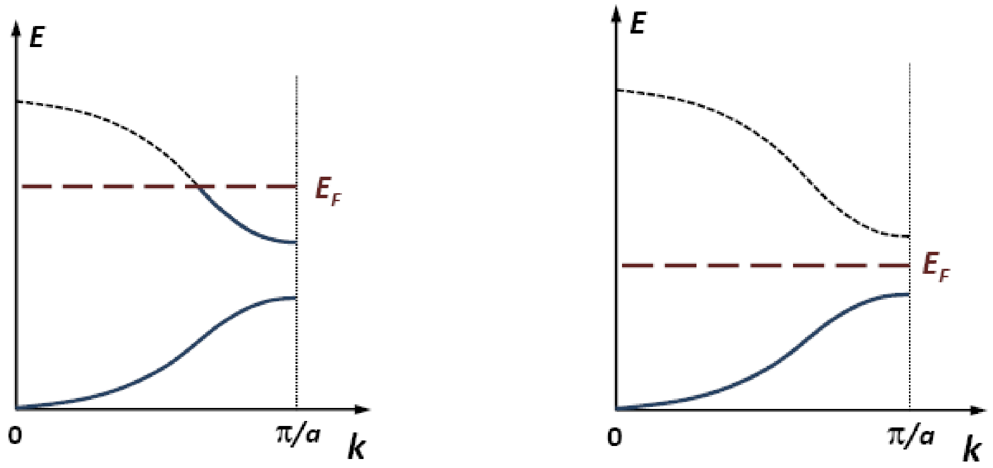
$$f(E) = \frac{1}{1 + \exp[(E - E_F)/k_B T]} \tag{65}$$

with $k_B$ the Boltzmann constant. The temperature dependence of the Fermi-Dirac distribution function is illustrated in the figure below. At $T$ = 0 K, the distribution is a step function, equal to 1 for energies smaller than the Fermi energy, and equal to 0 otherwise.



Two situations can exist:

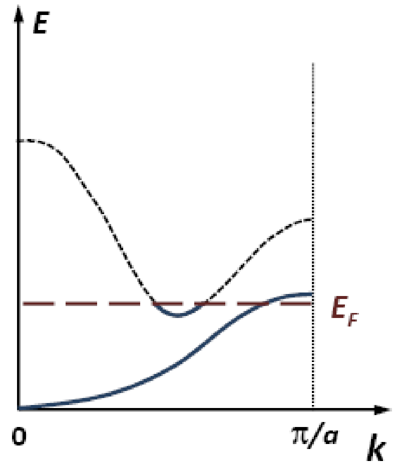
1) at low temperatures, $T \cong 0$ K, the Fermi level is inside an energy band (see the figure below, left), i.e. electrons occupy partially the last energy band. The material is then a metal and can easily conduct electricity since the electrons in the vicinity of $E_F$, accelerated by a small applied electric field, can occupy available empty states with higher energy.



2) at low temperatures, $T \cong 0$ K, the electrons occupy completely a number of energy bands, so that the Fermi level is inside the energy gap between the last occupied band, called valence band, and the following empty band, called conduction band (see the figure above, right). In this case a small electric field does not provide sufficient energy for the electron to jump into empty states in the higher energy band, and therefore no electric current can flow through the material. We deal in this situation with a dielectric.

At higher temperatures, $T \neq 0$ K, the dielectric materials, in turn, can be either isolators or semiconductors, depending on the width of the energy gap between the valence and conduction bands. If $E_g < 3$ eV, thermal fluctuations can excite electrons from the valence to the conduction band, where they can contribute to electrical conduction, and the material is in this case a semiconductor. As a result, in (undoped) semiconductors the number of electrons in the conduction band is equal to the number of holes in the valence band. When an applied electric field is applied, the current has two contributions: one from electrons and the other from holes, which are drifted in opposite directions. In isolators, $E_g > 3$ eV and no electrical conduction exists at moderate temperatures or electric fields.

Because the total number of electrons in a crystal with $N$ unit cells and $s$ atoms in the basis, each atom having $Z$ electrons, is $NsZ$, and an energy band can accommodate $2N$ electrons, it follows that the number of occupied energy bands is $NsZ/2N = sZ/2$. Then, in simple lattices with $s = 1$, there is always a partially occupied band for odd $Z$ values and these materials should be metals. This is the case of monovalent alkaline metals (Li, Na, K, Cs, Rb) and noble metals (Cu, Au, Ag), which have only one valence electron. However, it is not true that materials with $s = 1$ and even $Z$ are always dielectrics. For example, bivalent elements such as Be, Mg, Ca, Sr, B, are metals. The explanation is that, in these cases, the completely empty conduction band overlaps the completely occupied valence band over a small energy interval (see the figure below). The electrical conductivity is poorer, though, if the overlap is slight. Note that in genuine dielectric materials, at $T$ = 0 K there is no overlap between the conduction and valence bands.

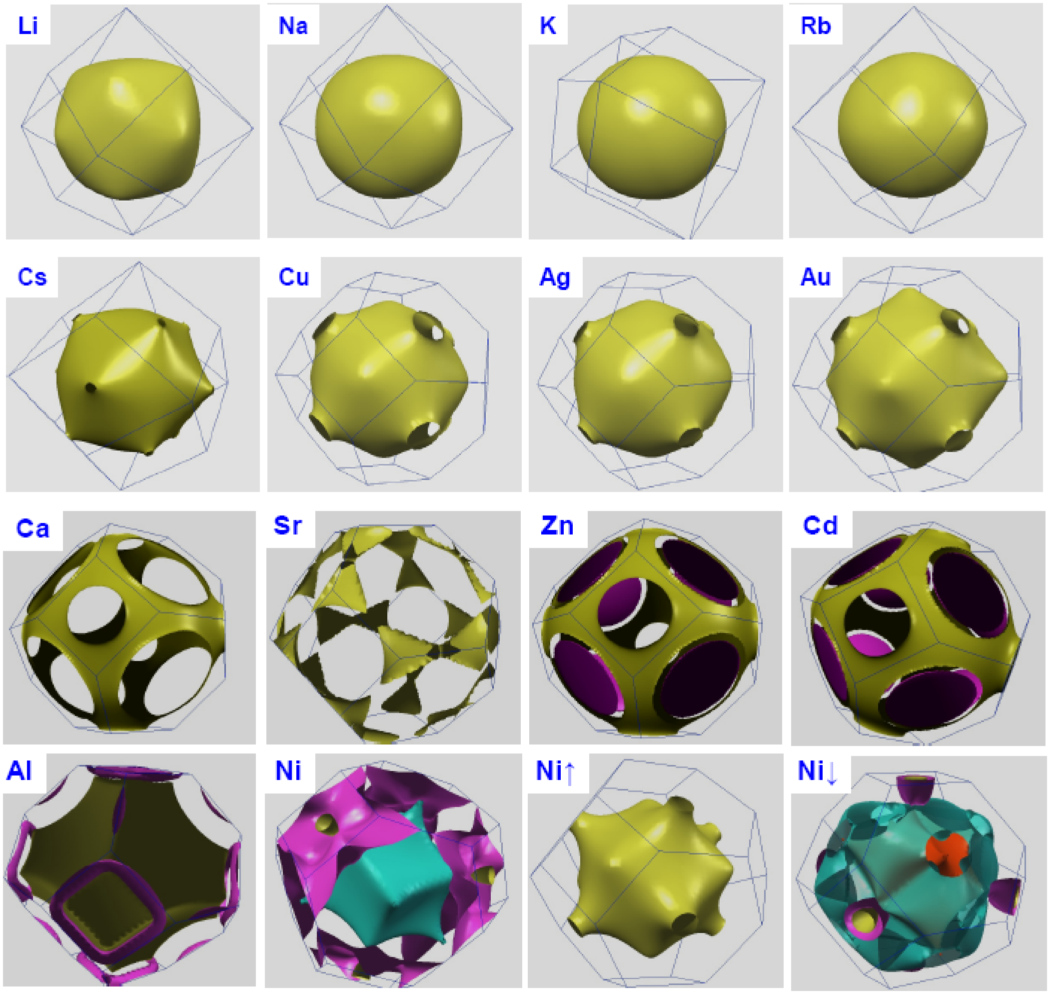


Because at the top of partially occupied energy bands the charge carriers are holes, a mixed conduction (electrons and holes) is expected when the conduction and valence bands overlap over a narrow energy range. The materials that have a mixed conduction at $T$ = 0 K are called semimetals; they differ from metals in that in metals the conduction is always due to electrons only. The semimetals have more complicated band structures than metals and dielectrics. Examples of semimetals are As, Sb, Bi; these materials have an electrical conductivity with up to four orders of magnitude smaller than in metals.

In metals, the iso-energetic surface in the reciprocal space defined at $T$ = 0 K through

$$E(\boldsymbol{k}) = E_F \tag{66}$$

is called Fermi surface. The Fermi surface separates the states with a low occupation density from those with a high occupation density and determines the physical properties of metals, especially the electrical properties, in which only electrons within an energy interval of the order of $k_B T$ around the Fermi energy participate. Examples of simple and complex Fermi surfaces are given in the figures below. Note that in Ni, which has strong magnetic properties, the Fermi surfaces for electrons with opposite spins are different.


Li
Na
K
Rb
Cs
Cu
Ag
Au
Ca
Sr
Zn
Cd
Al
Ni
Ni↑
Ni↓

# Statistics of charge carriers

The statistic properties of charge carriers are determined when these are in statistical/thermodynamical equilibrium with the crystalline lattice. To find these properties we need to know the distribution function and the density of states of electrons and holes.

For a general system of electrons characterized by a distribution function (probability of occupying available electronic states) $f(E)$, the concentration of electrons is

$$\begin{aligned} n &= \frac{1}{V}\sum_{\boldsymbol{k},\sigma} f(E(\boldsymbol{k})) = \frac{2}{V}\sum_{\boldsymbol{k}} f(E(\boldsymbol{k})) = \frac{2}{(2\pi)^3}\int f(E(\boldsymbol{k}))d\boldsymbol{k} \\ &= \frac{2}{(2\pi)^3}\int \frac{f(E)dSdE}{|\nabla_{\boldsymbol{k}}E|} = 2\int_{E_{\min}}^{E_{\max}} f(E)D(E)dE \end{aligned} \qquad (1)$$

where the factor 2 originates from the sum over the spin states with index $\sigma$. The density of electronic states for spherical iso-energetic surfaces is given by

$$D(E) = \frac{(2m_{eff})^{3/2}}{4\pi^2\hbar^3}(E-E_0)^{1/2}. \qquad (2)$$

Because there are differences in the distribution function of charge carriers in metals and semiconductors, in the following we treat these cases separately.

## Statistics of electrons in metals

In metals the Fermi energy is inside an energy band and one can define the Fermi surface. The density of electrons in metals with spherical iso-energetic surfaces is given by (2) with $E_0 = 0$ and, at a finite temperature *T*, the electrons occupy the energy states according to the Fermi-Dirac distribution function

$$f(E) = \frac{1}{1+\exp[(E-E_F)/k_BT]}. \qquad (3)$$

In this case, the electron concentration becomes

$$n = \frac{(2m_{eff})^{3/2}}{2\pi^2\hbar^3} \int_0^{\infty} \frac{E^{1/2} dE}{\exp[(E - E_F)/k_B T] + 1} \tag{4}$$

if the upper limit of the integral is approximated with ∞. This approximation is justified since the integrand decreases rapidly. With the change of variables $x = E/k_B T$ , $y = E_F / k_B T$ , we finally obtain

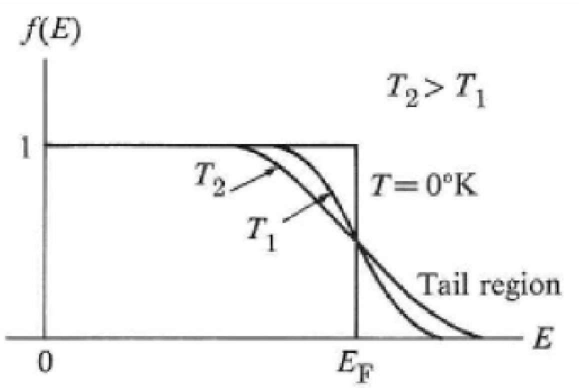
$$n = \frac{(2m_{eff} k_B T)^{3/2}}{2\pi^2\hbar^3} F_{1/2}(y) \tag{5}$$

where

$$F_\alpha(y) = \int_0^{\infty} \frac{x^\alpha dx}{\exp(x - y) + 1} \tag{6}$$

are the Fermi-Dirac integrals. They are evaluated numerically. Note that the same result is obtained for an arbitrary $E_0$ . In this case, in (4) $E_0$ appears in the numerator of the integrand and in the lower limit of the integral, but disappears in the final result (equations (5) and (6)), if the new variables are chosen as $(E - E_0)/k_B T = x$, $(E_F - E_0)/k_B T = y$ .

The temperature dependence of the Fermi-Dirac distribution function is represented in the figure below. At all temperatures, $f(E_F) = 1/2$ .



At $T$ = 0 K, where the Fermi-Dirac distribution is a Heaviside function, i.e.

$$f(E)=\begin{cases}1, & E<E_F\\ 0, & E>E_F\end{cases} \tag{7}$$

the concentration of electrons in metals is

$$n=\frac{(2m_{eff}k_BT)^{3/2}}{2\pi^2\hbar^3}\int_0^{E_F^0/k_BT}x^{1/2}dx=\frac{(2m_{eff}k_BT)^{3/2}}{2\pi^2\hbar^3}\frac{2}{3}\left(\frac{E_F^0}{k_BT}\right)^{3/2}=\frac{(2m_{eff})^{3/2}}{3\pi^2\hbar^3}(E_F^0)^{3/2}, \tag{8}$$

with $E_F^0$ the Fermi energy level at $T$ = 0 K, or

$$E_F^0=\frac{\hbar^2(3\pi^2n)^{2/3}}{2m_{eff}}=\frac{\hbar^2k_F^2}{2m_{eff}}, \tag{9}$$

where $k_F=(3\pi^2n)^{1/3}$ is the Fermi wavenumber. From (9) it follows that the Fermi energy at $T$ = 0 increases as the electron concentration increases, the system of electrons being in the fundamental state if the Fermi sphere is completely occupied and in an excited state if electrons occupy states with $|\boldsymbol{k}|>k_F$.

For finite but low temperatures, $y=E_F/k_BT>>1$, and spherical iso-energetic surfaces, the Fermi-Dirac integrals can be approximated with

$$F_\alpha(y)=\frac{y^{\alpha+1}}{\alpha+1}\left(1+\frac{\pi^2}{6}\frac{\alpha(\alpha+1)}{y^2}\right), \tag{10}$$

and so

$$n=\frac{(2m_{eff})^{3/2}}{3\pi^2\hbar^3}E_F^{3/2}\left[1+\frac{\pi^2}{8}\left(\frac{k_BT}{E_F}\right)^2\right]. \tag{11}$$

On the other hand, the electron concentration in metals does not depend on temperature because an increase in $T$ affects only the thermal excitation of electrons on higher energy levels. It follows thus that the Fermi level must depend on temperature and, from

$$E_F^{3/2} = \frac{(E_F^0)^{3/2}}{\left[1+\frac{\pi^2}{8}\left(\frac{k_B T}{E_F}\right)^2\right]} \cong \frac{(E_F^0)^{3/2}}{\left[1+\frac{\pi^2}{8}\left(\frac{k_B T}{E_F^0}\right)^2\right]}, \tag{12}$$

we obtain

$$E_F = E_F^0\left[1-\frac{\pi^2}{12}\left(\frac{k_B T}{E_F^0}\right)^2\right]. \tag{13}$$

This temperature dependence is very weak and the Fermi energy is, in the first approximation, almost constant as $T$ varies.

## Statistics of charge carriers in semiconductors

In semiconductors the Fermi level is situated between the valence and the conduction band, which are separated by an energy gap. At $T$ = 0 K, all energy states below $E_F$ are occupied with electrons according to the Pauli principle, i.e. according to the Fermi-Dirac distribution function, and all states above the Fermi level are empty. Because the holes in semiconductors can be viewed as states not occupied by electrons, the probability that a state $\boldsymbol{k}$, with energy $E(\boldsymbol{k})$ is occupied by a hole (i.e. is empty for electrons) is

$$1-f(E) = 1-\frac{1}{1+\exp[(E-E_F)/k_B T]} = \frac{1}{1+\exp[(E_F-E)/k_B T]}. \tag{14}$$

If $E_F / k_B T >> 1$ the Fermi-Dirac distribution function has almost the step-function form characteristic for $T$ = 0 K. In this case the distribution function is called **degenerate** and is encountered in materials with large concentrations of electrons, such as metals and heavily doped semiconductors. On the contrary, when $E_F / k_B T << -1$ or $-E_F / k_B T >> 1$, the exponential term in the Fermi-Dirac distribution function is much larger than unity and the distribution function resembles the classical Maxwell-Boltzmann distribution,

$$f_{cl}(E) = \exp(E_F / k_B T)\exp(-E / k_B T) << 1, \tag{15}$$

which can be seen as the “tail” of the Fermi-Dirac distribution. This case corresponds to low electron concentrations, in particular to undoped (intrinsic) semiconductors, in which the electron concentration is with few orders of magnitude smaller than in metals. The Maxwell-Boltzmann distribution function is called **nondegenerate**.

## Intrinsic semiconductors

The density of states for electrons in an intrinsic semiconductor is similar to that in metals (for an arbitrary $E_0$) and is found to be

$$D_n(E) = \frac{m_n k}{2\pi^2\hbar^2} = \frac{(2m_n)^{3/2}}{4\pi^2\hbar^3}(E - E_c)^{1/2} \tag{16}$$

for spherical iso-energetic surfaces $E(\boldsymbol{k}) = E_c + \hbar^2 k^2 / 2m_n$, where $E_c$ is the minimum energy in the conduction band, and $m_n$ is the effective mass of electrons in the vicinity of this minimum energy. In an analogous manner, the density of states for the holes with effective mass $m_p$ in the valence band is found to be

$$D_p(E) = \frac{m_p k}{2\pi^2\hbar^2} = \frac{(2m_p)^{3/2}}{4\pi^2\hbar^3}(E_v - E)^{1/2}, \tag{17}$$

where $E_v$ is the maximum energy in the valence band. This expression was found taking into account that the hole energy increases in the opposite direction as that of the electron. Both electron and hole density of states must be multiplied with 2 if the spin degeneracy is included.

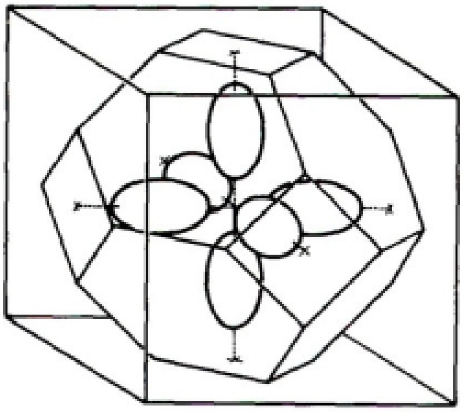
However, the dispersion relation is not always spherical. For example, the conduction band of Ge and Si are ellipsoidal iso-energetic surfaces with $N_{eq}$ equivalent minima arranged symmetrically in the first Brillouin zone ($N_{eq}$ = 4 for Ge and $N_{eq}$ = 6 for Si). This case can be reduced to that of spherical iso-energetic surfaces if, in the neighborhood of these minima situated at $\boldsymbol{k}_0$, the dispersion relation

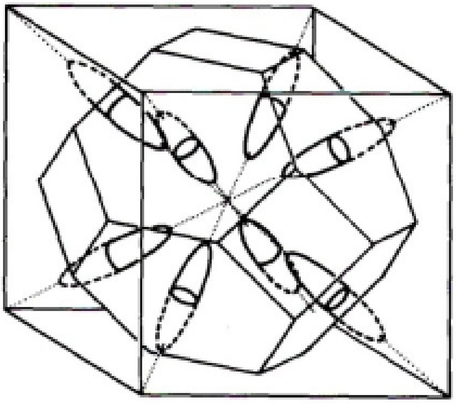
$$E_{\boldsymbol{k}} = E_c + \frac{\hbar^2}{2}\left[\frac{(k_1 - k_{01})^2}{m_{n1}} + \frac{(k_2 - k_{02})^2}{m_{n2}} + \frac{(k_3 - k_{03})^2}{m_{n3}}\right] \tag{18}$$

is transformed into a spherical dispersion relation

$$E_{k'} = E_c + \frac{\hbar^2}{2m_n'}(k_1'^2 + k_2'^2 + k_3'^2) \tag{19}$$

by a change of variables $k_i - k_{0i} = k_i'\sqrt{m_{ni}/m_n'}$. The figures below illustrate the iso-energetic surfaces in Si (left) and Ge (right). There are 4 complete ellipsoids (8 half-ellipsoids) in the first Brillouin zone in Ge.





Then, in the calculation of the density of states we must account for the fact that $d\mathbf{k} = dk_1 dk_2 dk_3 = [(m_{n1}m_{n2}m_{n3})^{1/2} / m_n'^{3/2}]d\mathbf{k}'$, from which it follows that the density of states becomes

$$D_n(E) = \frac{(2m_n')^{3/2}}{4\pi^2\hbar^3}\frac{(m_{n1}m_{n2}m_{n3})^{1/2}}{m_n'^{3/2}}(E - E_c)^{1/2}. \tag{20}$$

Because there are $N_{eq}$ equivalent minima/ellipsoids, the density of states is given by

$$D_c(E) = \frac{(2m_c)^{3/2}}{4\pi^2\hbar^3}(E - E_c)^{1/2} \tag{21}$$

where the effective mass of the density of states in the conduction band is defined as

$$m_c = N_{eq}^{2/3}(m_{n1}m_{n2}m_{n3})^{1/3}. \tag{22}$$

## Concentration of charge carriers and Fermi energy

In an intrinsic (undoped) semiconductor, the requirement of electrical neutrality imposes the same concentration for both electrons and holes,

$$n = p\,. \tag{23}$$

In intrinsic semiconductors the concentrations of charge carriers are small ($n$ and $p$ are smaller with several orders of magnitude than the electron concentration in metals), so that the available energy states are no longer occupied according to the Fermi-Dirac distribution function, as in metals, but according to the classical Maxwell-Boltzmann distribution. So, the electron an hole concentrations in intrinsic semiconductors with spherical iso-energetic surfaces are given by

$$n = 2\int_{E_{\min}}^{E_{\max}} f(E)D_n(E)dE = \frac{(2m_n k_B T)^{3/2}}{2\pi^2\hbar^3}F_{1/2}(y_c)\,, \qquad p = \frac{(2m_p k_B T)^{3/2}}{2\pi^2\hbar^3}F_{1/2}(y_v) \tag{24}$$

with $y_c = (E_F - E_c)/k_B T$, $y_v = (E_v - E_F)/k_B T$. The expressions are similar to that found in metals, with $x_c = (E - E_c)/k_B T$, $x_v = (E_v - E)/k_B T$, respectively, except that in this case the Fermi-Dirac integrals become

$$F_\alpha(y) = \exp(y)\int_0^\infty x^\alpha \exp(-x)dx\,, \tag{25}$$

since we use the Maxwell-Boltzmann distribution. Because $F_{1/2}(y) = (\sqrt{\pi}/2)\exp(y)$,

$$n = \frac{(2\pi m_n k_B T)^{3/2}}{4\pi^3\hbar^3}\exp\left(\frac{E_F - E_c}{k_B T}\right) = N_c \exp\left(\frac{E_F - E_c}{k_B T}\right) = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right), \tag{26}$$

$$p = \frac{(2\pi m_p k_B T)^{3/2}}{4\pi^3\hbar^3}\exp\left(\frac{E_v - E_F}{k_B T}\right) = N_v \exp\left(\frac{E_v - E_F}{k_B T}\right) = N_v \exp\left(-\frac{E_F - E_v}{k_B T}\right), \tag{27}$$

where

$$N_c = \frac{(2\pi m_n k_B T)^{3/2}}{4\pi^3\hbar^3}\,, \qquad N_v = \frac{(2\pi m_p k_B T)^{3/2}}{4\pi^3\hbar^3}\,, \tag{28}$$

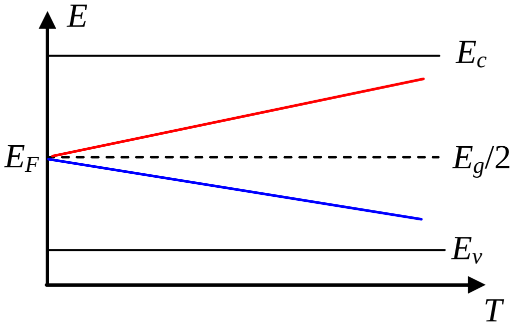
are the effective densities of states in the conduction and valence bands, respectively. From the relations above it follows that, unlike the degenerate case in metals, the concentrations of electrons and holes in non-degenerate semiconductors depend strongly on temperature.

The Fermi energy is determined from the neutrality condition, i.e. from

$$m_n^{3/2}\exp\left(\frac{E_F-E_c}{k_BT}\right)=m_p^{3/2}\exp\left(\frac{E_v-E_F}{k_BT}\right). \quad (29)$$

The result is

$$E_F=\frac{E_c+E_v}{2}+\frac{3}{4}k_BT\ln\left(\frac{m_p}{m_n}\right) \quad (30)$$



and, at $T$ = 0 K, the Fermi level is at the middle of the energy gap: $E_F=(E_c+E_v)/2$. As the temperature increases, the Fermi level shifts towards the energy band with the smaller effective mass (see the figure above); it remains at the centre of the energy gap, irrespective of temperature, only if $m_n=m_p$. A special situation is encountered in semiconductors with a small energy gap width $E_g=E_c-E_v$ and $m_p/m_n>>1$ or $m_p/m_n<<1$ (the first situation is much more common). In this case, the Fermi level can enter inside the band with the smaller effective mass, and the semiconductor becomes degenerate at even moderate temperatures. An example of such a material is InSb, for which $E_g$ = 0.18 eV and $m_p/m_n\cong 10$.

It should be emphasized that, starting from the expression of the electron concentration in semiconductors, it follows that the conditions of applicability of the Maxwell-Boltzmann statistics is

$$\exp\left(\frac{E_F-E_c}{k_BT}\right)=\frac{4\pi^3\hbar^3 n}{(2\pi m_n k_BT)^{3/2}}<<1. \quad (31)$$

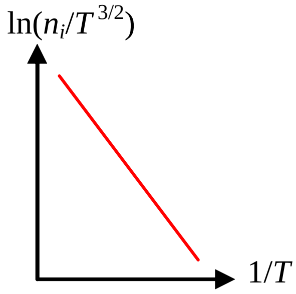
If the energy reference is chosen such that $E_c=0$, $\exp(E_F/k_BT)<<1$ for small electron concentrations, high temperature, and high effective masses. On the contrary, if these conditions are not satisfied and $\exp(E_F/k_BT)>>1$, i.e. if the electron concentration is high, the temperature is low, and the effective mass is small, the electron gas is degenerate, and the criterion above is known as the degeneracy criterion. Electrons in metals satisfy the degeneracy criterion.

The intrinsic concentration of carriers can be defined as

$$n_i = n = p = \sqrt{np} = \sqrt{N_c N_v}\exp\left(-\frac{E_g}{2k_B T}\right) = 2\left(\frac{k_B}{2\pi\hbar^2}\right)^{3/2}(m_n m_p)^{3/4}T^{3/2}\exp\left(-\frac{E_g}{2k_B T}\right). \quad (32)$$

This concentration does not depend on the Fermi level, and its temperature dependence is summarized in the formula (see figure)

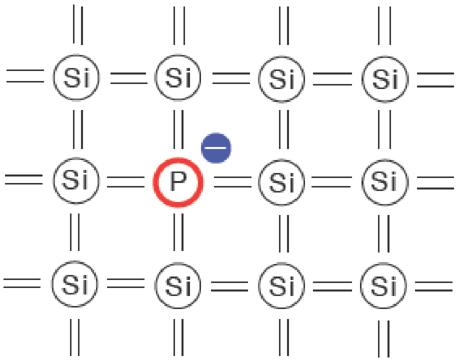
$$n_i(T) \propto T^{3/2}\exp\left(-\frac{E_g}{2k_B T}\right), \quad (33)$$

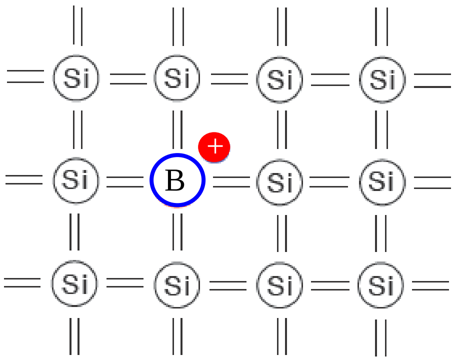


and the band gap can be determined from the slope of the figure at the right side. For a large class of semiconductors, the energy gap width depends on temperature as: $E_g(T) = E_{g0} - \alpha T$ .

## Extrinsic semiconductors

An extrinsic semiconductor is a doped semiconductor, i.e. it contains donor and/or acceptor impurities In donor impurities, the number of valence electrons is higher than in the host material, and the extra electrons are not involved in binding with the atoms of the host material; they are still localized around the donor impurity but can easily participate at the process of charge carrier transport when an electric field is applied. The number of free electrons that contribute to electrical conduction increases in the presence of donor impurities. On the contrary, acceptor impurities have a smaller number of valence electrons than the host material, and their stable binding with host atoms requires an additional electron from the host material. As a result, the number of holes increases when acceptor impurities are present.





For Si (see the schematic figures above), which forms covalent bonds with other four atoms, P is a donor impurity since it has five valence electrons, and B is an acceptor impurity because it has only three valence electrons. Donor and/or acceptor impurities are purposefully introduced into a host semiconductor to increase its electrical conduction. On the other hand, the unintentional doping of a crystal introduces defects into the crystalline lattice and degrades its electrical performances. Therefore, doping is a process that should be carefully controlled. The donor and acceptor impurities have discrete energy levels inside the bandgap of the host material, which can be occupied by electrons and, respectively, holes. The charge carriers on the impurity levels are localized around donors and acceptors and are not free to move around the host crystal unless the impurities are ionized, i.e. the electrons on donor energy levels are promoted (by thermal energy or applied electric fields) into the conduction band and the holes on acceptor levels are excited into the valence band of the host material.

Let us denote by $N_d$ and $N_a$ the concentration of donor and acceptor impurities, by $N_d^0$ and $N_a^0$ the concentrations of electrons and holes localized on the donors and acceptors, respectively, i.e. the concentrations of neutral donors and acceptors, and by $N_d^+ = N_d - N_d^0$ and $N_a^- = N_a - N_a^0$ the concentrations of the ionized impurities. The ionization of donors and acceptors leads to an increase of the number of free charge carriers with respect to the intrinsic semiconductor case. If $n$ and $p$ are the total concentrations of free electrons and holes, the condition of charge neutrality is

$$n + N_a^- = p + N_d^+ , \tag{34a}$$

or

$$n + N_a + N_d^0 = p + N_d + N_a^0 . \tag{34b}$$

Unlike in the case of free charge carriers, the distribution function of the electrons localized on donor impurities and that of the holes localized on acceptor impurities is not the Fermi-Dirac function, which is valid when two electrons with opposite spins can occupy an energy level, according to the Pauli principle. The reason is that only a single electron can occupy an energy level on an impurity atom. If another electron is brought on this level, its energy varies significantly due to the strong electrostatic interaction between electrons.

According to the Fermi-Dirac distribution function, the ratio between the probability that the state is occupied and the probability that the state is empty is,

$$\frac{f(E)}{1-f(E)}=\exp\left(-\frac{E-E_F}{k_BT}\right) . \tag{35}$$

On the contrary, since only one electron can exist on an impurity level, the same ratio is now

$$\frac{f_d(E)}{1-f_d(E)}=2\exp\left(-\frac{E-E_F}{k_BT}\right) \tag{36}$$

for electrons on donors (the level is occupied twice as fast), or

$$f_d(E)=\frac{1}{1+(1/2)\exp[(E-E_F)/k_BT]}, \tag{37}$$

and

$$f_a(E)=\frac{1}{1+(1/2)\exp[(E_F-E)/k_BT]} \tag{38}$$

for holes on acceptor impurities. In general, the factor $(1/2)$ in front of the exponential term in the denominator should be replaced by $(1/g)$, with $g$ the degeneracy of the energy level.

Taking into account that the density of states of the discrete donor/acceptor energy levels $E_i$ are

$$D_i(E)=N_i\delta(E-E_i), \qquad i=d,a \tag{39}$$

the concentrations of the electrons localized on the donors and of holes localized on acceptors are given by

$$N_d^0=\int f_d(E)D_d(E)dE=\frac{N_d}{1+\frac{1}{2}\exp\left(\frac{E_d-E_F}{k_BT}\right)}, \tag{40}$$

$$N_a^0 = \int f_a(E) D_a(E) dE = \frac{N_a}{1 + \frac{1}{2}\exp\left(\frac{E_F - E_a}{k_B T}\right)}, \tag{41}$$

and the neutrality condition becomes

$$\begin{aligned} N_a + \frac{N_d}{1 + \frac{1}{2}\exp\left(\frac{E_d - E_F}{k_B T}\right)} + N_c \exp\left(\frac{E_F - E_c}{k_B T}\right) = \\ N_d + \frac{N_a}{1 + \frac{1}{2}\exp\left(\frac{E_F - E_a}{k_B T}\right)} + N_v \exp\left(\frac{E_v - E_F}{k_B T}\right) \end{aligned} \tag{42}$$

from which one can determine the position of the Fermi energy level. The Fermi level in extrinsic semiconductors is different than in intrinsic semiconductors!

**Extrinsic semiconductors with only one impurity type**

Let us consider that the semiconductor is nondegenerate and has only donor impurities with concentration $N_d = N_d^0 + N_d^+$. In this case $N_a = 0$, $N_a^0 = 0$, and (34b) can be expressed as

$$n + N_d^0 = p + N_d, \tag{43}$$

with *n* and *p* determined as above, with the help of the Maxwell-Boltzmann statistics. The electrons in the conduction band are generated either through the ionization of donor impurities (through the transition from the donor level to an energy level in the conduction band), process that requires the energy $E_{gd} = E_c - E_d$ (equal to the ionization energy of the donor impurity), or through the ionization of the atoms in the crystal (the transition of an electron from the valence in the conduction band), process that requires an energy equal to $E_g$. Because $E_{gd} << E_g$, the contribution of the two processes differs as a function of the temperature. More precisely, at low temperatures the dominant process is the ionization of impurities, whereas at high temperatures the electron transitions between the valence and conduction bands prevail. We have extrinsic conduction at low temperatures, and intrinsic conduction at high temperatures.

**I) Extrinsic conduction regime**

At low temperatures, at which the conduction electrons originate from the ionization of donor impurities, for a given donor concentration $N_d >> p$ the neutrality condition becomes $n + N_d^0 = N_d$, or

$$N_c \exp\left(\frac{E_F - E_c}{k_B T}\right) = N_d - \frac{N_d}{1 + \frac{1}{2}\exp\left(\frac{E_d - E_F}{k_B T}\right)}. \tag{44}$$

This is a second order equation for $\exp(E_F / k_B T)$, which can be easily solved by introducing the variables $x = \exp[(E_F - E_d)/k_B T]$, $y = (N_d / N_c)\exp[(E_c - E_d)/k_B T]$. In terms of these variables (44) can be written as $2x^2 + x - y = 0$, from which it follows that

$$E_F = E_d + k_B T \ln\left[\frac{1}{4}\left(\sqrt{1 + 8\frac{N_d}{N_c}\exp\left(\frac{E_c - E_d}{k_B T}\right)} - 1\right)\right]. \tag{45}$$

For extremely low temperatures, for which $8(N_d / N_c)\exp[(E_c - E_d)/k_B T] >> 1$, (45) can be approximated as $E_F = E_d + k_B T \ln\sqrt{(N_d / 2N_c)\exp[(E_c - E_d)/k_B T]}$, or

$$E_F = \frac{E_c + E_d}{2} + \frac{k_B T}{2}\ln\left(\frac{N_d}{2N_c}\right), \tag{46}$$

which reduces to

$$E_F = \frac{E_c + E_d}{2} \tag{47}$$

at $T$ = 0 K. The temperature dependence of the Fermi level can be determined taking into account that $N_c = (2\pi m_n k_B T)^{3/2} / (4\pi^3 \hbar^3) \propto T^{3/2}$. At temperatures of only few K, when $2N_c < N_d$, $E_F$ shifts towards the conduction band but, as the temperatures increases until $2N_c = N_d$, the Fermi level takes again the value at $T$ = 0 K. Thus, in this temperature interval $E_F$ reaches a maximum value at a temperature

$$T_{\max} = \frac{N_d^{2/3} \pi \hbar^{4/3}}{2^{1/3} e k_B m_n}, \tag{48}$$

determined from the condition $dE_F / dT = 0$, or $\ln(N_d / 2N_c) = 3/2$. In (48) $e$ is not the electric charge, but the basis of the natural logarithm!

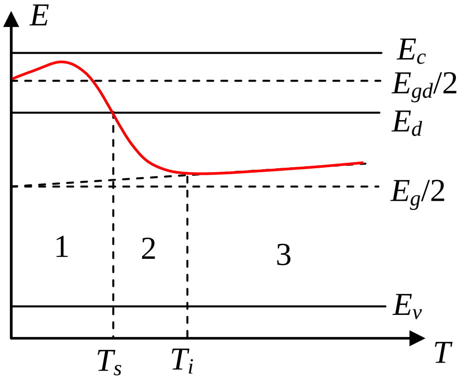
The maximum value of the Fermi energy is found to be

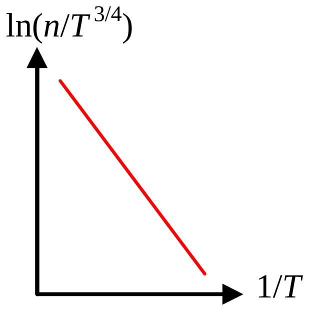
$$E_{F,\max} = E_F(T_{\max}) = \frac{E_c + E_d}{2} + \frac{3\pi}{2^{7/3}} \frac{\hbar^{4/3}}{e m_n} N_d^{2/3}. \tag{49}$$

The Fermi level can even reach the minimum value of the conduction band, i.e. $E_{F,\max} = E_c$ for a critical concentration impurity

$$N_{d,cr} = \frac{4}{\hbar^2} \left( \frac{e m_n}{3\pi} \right)^{3/2} E_{gd}^{3/2}. \tag{50}$$

At this critical concentration the semiconductor becomes degenerate. The temperature dependence of the Fermi level is represented in the figure below, left.





A further increase in temperature, which corresponds to $2N_c > N_d$, leads to a decrease in the Fermi level value towards $E_d$, until this value is reached for a so-called saturation temperature $T_s$. The temperature interval $0 < T < T_s$ is called the weak ionization region (see region 1 in the figure above). In this temperature interval, from (46) it follows that

$$n = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) = \sqrt{\frac{N_c N_d}{2}} \exp\left(\frac{E_d - E_c}{2k_B T}\right) = \sqrt{\frac{N_c N_d}{2}} \exp\left(-\frac{E_{gd}}{2k_B T}\right), \qquad (51)$$

i.e. $n \propto N_d^{1/2}$, and, since $N_c \propto T^{3/2}$, the temperature dependence of the electron concentration is $n \propto T^{3/4} \exp(-E_{gd} / 2k_B T)$. The ionization energy of the donor impurities, $E_{gd}$, can thus be determined from the slope of the $\ln(n / T^{3/4}) = f(1/T)$ plot (see the figure above, right).

At still higher temperatures, for which $8(N_d / N_c)\exp[(E_c - E_d)/k_B T] << 1$, (45) can be approximated as $E_F = E_d + k_B T \ln\{(N_d / N_c)\exp[(E_c - E_d)/k_B T]\}$, or

$$E_F = E_c + k_B T \ln(N_d / N_c), \qquad (52)$$

the logarithm being negative since $N_c >> N_d$. Thus, the Fermi energy decreases as the temperature increases and becomes lower than $E_d$, level reached at the saturation temperature

$$T_s = \frac{E_{gd}}{k_B \ln[N_c(T_s) / N_d]}. \qquad (53)$$

In this temperature interval the electron concentration is given by (see (52))

$$n = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) = N_d, \qquad (54)$$

result that shows that the donor impurities are totally ionized, and the electron concentration is independent of temperature for $T > T_s$. The regime is an exhausting regime for donor impurities and in the figure above is indicated as region 2.

**II) Intrinsic conduction regime**

For high-enough temperatures the hole concentration starts to increase and becomes comparable with the electron concentration. In particular, if $p >> N_d^0$ the neutrality condition (43) can be written as $n = p + N_d$. In this regime of high temperatures the donors are completely ionized, the charge carriers originating from the ionization of the host semiconductor material. For a nondegenerate semiconductor $p = n_i^2 / n$, which introduced in the neutrality condition leads to

$$n^2 - N_d n - n_i^2 = 0, \tag{55}$$

the solution of this equation being

$$n = \frac{N_d}{2}\left(1 + \sqrt{1 + 4\frac{n_i^2}{N_d^2}}\right), \quad p = n_i^2 / n = \frac{2n_i^2}{N_d\left(1 + \sqrt{1 + 4\frac{n_i^2}{N_d^2}}\right)}. \tag{56}$$

Because at high temperatures the host material is the main source of charge carriers, the expressions for electron and hole concentrations in intrinsic semiconductors apply, and the Fermi energy level, determined from $n = N_c \exp[(E_F - E_c)/k_B T]$, with $n$ from (56), is

$$E_F = E_c + k_B T \ln\left[\frac{N_d}{2N_c}\left(1 + \sqrt{1 + 4\frac{n_i^2}{N_d^2}}\right)\right] = E_c + k_B T \ln\left[\frac{N_d}{2N_c}\left(1 + \sqrt{1 + 4\frac{N_c N_v}{N_d^2}\exp\left(-\frac{E_g}{k_B T}\right)}\right)\right]. \tag{57}$$

The expression above can be studied in two extreme situations:

1) $4n_i^2 / N_d^2 << 1$, case in which

$$n = N_d, \; p = n_i^2 / N_d, \tag{58}$$

and (in agreement with our previous results for the totally ionized/exhausted impurities)

$$E_F = E_c + k_B T \ln\left(\frac{N_d}{N_c}\right), \tag{59}$$

2) $4n_i^2 / N_d^2 >> 1$, case in which

$$n = p = n_i \tag{60}$$

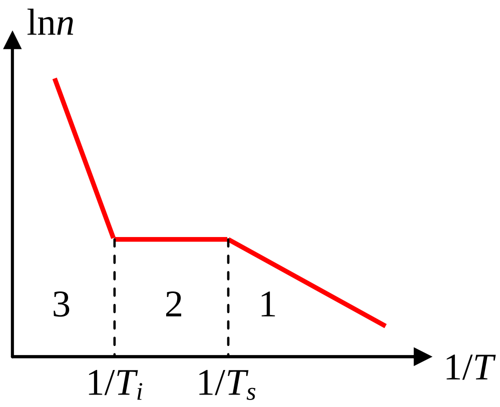
and (as for the intrinsic semiconductor)

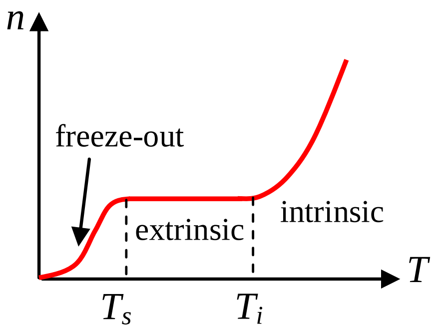
$$E_F = \frac{E_c + E_v}{2} + \frac{k_B T}{2} \ln\left(\frac{N_v}{N_c}\right) = \frac{E_c + E_v}{2} + \frac{3}{4} k_B T \ln\left(\frac{m_p}{m_n}\right). \tag{61}$$

The temperature dependence of the Fermi level in this region of intrinsic conduction is indicated in the figure above (see region 3). At high-enough temperatures the increase of the electron concentration in the conduction band originates from electron transitions from the valence band. The transition temperature from the exhausting regime of impurities to the region of intrinsic conduction can be determined from (58) and (60), i.e. from $n_i = N_d$, and is found to be

$$T_i = \frac{E_g}{k_B \ln[N_c(T_i) N_v(T_i) / N_d^2]}. \tag{62}$$

Summarizing, the temperature dependence of the electron concentration shows three distinct regions (see the figures below). The logarithmic dependence of the concentration on the inverse of the temperature can be approximated with a straight line in regions 1 and 3 (see figure below, left) if we neglect the influence of the factors $T^{3/2}$ and $T^{3/4}$, respectively, in comparison with the exponentials terms, and the parameters $E_g$ and $E_{gd}$ can be determined from the corresponding slopes. On the contrary, in region 2 the electron concentration is approximately constant, since the donor impurities are exhausted.





# Electronic specific heat

## Electronic specific heat in metals

In metals, the electronic specific heat per unit volume, calculated at constant volume, is defined as

$$C_{el} = \frac{dE_{el}}{dT} \tag{1}$$

where the energy per unit volume of the system of non-interacting electrons is given by

$$E_{el} = \frac{1}{V}\sum_{\boldsymbol{k},\sigma} E(\boldsymbol{k}) f(E(\boldsymbol{k})) = \frac{2}{V}\sum_{\boldsymbol{k}} E(\boldsymbol{k}) f(E(\boldsymbol{k})) = 2\int_0^\infty E f(E) D(E) dE\,, \tag{2}$$

with $D(E) = (2m_{eff})^{3/2} E^{1/2} / 4\pi^2\hbar^3$ (see the course on electron statistics in metals). In the normalized coordinates $E / k_B T = x$, $E_F / k_B T = y$ and for spherical iso-energetic surfaces the energy per unit volume becomes

$$E_{el} = \frac{(2m_{eff})^{3/2}}{2\pi^2\hbar^3}\int_0^\infty \frac{E^{3/2} dE}{\exp[(E - E_F)/k_B T] + 1} = \frac{(2m_{eff} k_B T)^{3/2} k_B T}{2\pi^2\hbar^3} F_{3/2}(y) = n k_B T \frac{F_{3/2}(y)}{F_{1/2}(y)}\,, \tag{3}$$

where

$$F_\alpha(y) = \int_0^\infty \frac{x^\alpha dx}{\exp(x - y) + 1} \tag{4}$$

are the Fermi-Dirac integrals. The last equality in (3) follows because (see the course on the statistics of electrons in metals)

$$n = \frac{(2m_{eff} k_B T)^{3/2}}{2\pi^2\hbar^3} F_{1/2}(y)\,. \tag{5}$$

So, taking into account that

$$dF_\alpha(y)/dy = \alpha F_{\alpha-1}(y)\,, \tag{6}$$

the electronic heat capacity can be expressed as

$$C_{el} = \frac{5}{2}\frac{(2m_{eff}k_BT)^{3/2}k_B}{2\pi^2\hbar^3}F_{3/2}(y) + \frac{(2m_{eff}k_BT)^{3/2}k_BT}{2\pi^2\hbar^3}\frac{dF_{3/2}}{dy}\frac{dy}{dT} = \frac{3}{2}nk_B\left(\frac{5}{3}\frac{F_{3/2}(y)}{F_{1/2}(y)} + T\frac{dy}{dT}\right). \tag{7}$$

At low temperatures, from $F_\alpha(y) = \frac{y^{\alpha+1}}{\alpha+1}\left(1+\frac{\pi^2}{6}\frac{\alpha(\alpha+1)}{y^2}\right)$, $E_F = E_F^0\left[1-\frac{\pi^2}{12}\left(\frac{k_BT}{E_F^0}\right)^2\right]$ and $y = E_F/k_BT >> 1$ we find that

$$\frac{F_{3/2}(y)}{F_{1/2}(y)} = \frac{3}{5}y\left(1+\frac{\pi^2}{2y^2}\right) \cong \frac{3}{5}\frac{E_F}{k_BT}\left[1+\frac{\pi^2}{2}\left(\frac{k_BT}{E_F^0}\right)^2\right] = \frac{3}{5}\frac{E_F^0}{k_BT}\left[1+\frac{5\pi^2}{12}\left(\frac{k_BT}{E_F^0}\right)^2\right] \tag{8a}$$

$$T\frac{dy}{dT} = \frac{1}{k_B}\left(\frac{dE_F}{dT} - \frac{E_F}{T}\right) = -\frac{E_F^0}{k_BT}\left[1+\frac{\pi^2}{12}\left(\frac{k_BT}{E_F^0}\right)^2\right] \tag{8b}$$

and so

$$C_{el} = \frac{3}{2}nk_B\frac{\pi^2}{3}\frac{k_BT}{E_F^0} = C_{el}^{cl}\frac{\pi^2}{3}\frac{T}{T_F}\,, \tag{9}$$

where the Fermi temperature is a parameter defined as $T_F = E_F^0/k_B$ and $C_{el}^{cl} = (3/2)nk_B$ is the classical electronic specific heat. $C_{el}^{cl}$ is obtained using the same general expression (7) as above, but with the Fermi-Dirac distribution function replaced by the Maxwell-Boltzmann distribution, case in which

$$F_\alpha(y) \cong \int_0^\infty x^\alpha \exp[-(x-y)]dx\,, \tag{10a}$$

$$F_{3/2}(y)/F_{1/2}(y) = 3/2\,, \tag{10b}$$

$$dE_F/dT = E_F/T - 3/2k_B\,, \tag{10c}$$

$$Tdy/dT = -3/2\,. \tag{10d}$$

The equality (10c) follows from (5) and the requirement that $dn/dT = 0$, considering that in the nondegenerate case $dF_\alpha(y)/dT = (dF_\alpha/dy)(dy/dT) = (dy/dT)F_\alpha(y)$. Then, (10d) is obtained from (10c) and (8b), so that, finally, $C_{el}^{cl} = (3/2)nk_B$.

The ratio $T/T_F$ can be seen as the fraction of excited electrons at temperature *T*, the other electrons being "frozen" due to the Pauli principle. The value of this ratio at room temperature is typically $10^{-2}$.

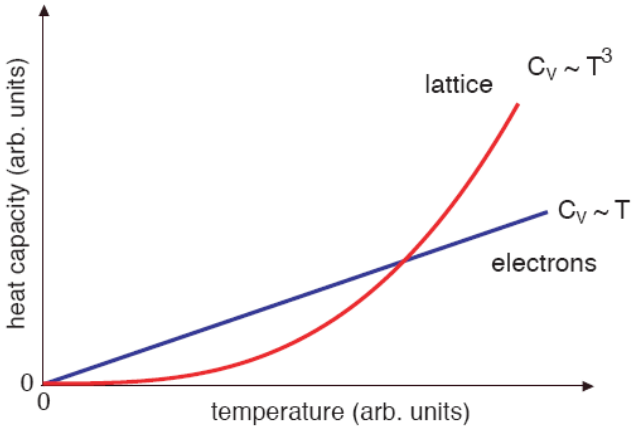
The linear relation between the electronic specific heat in metals and temperature is generally expressed as

$$C_{el} = \gamma T\,, \tag{11}$$

where $\gamma = \pi^2 nk_B^2 / 2E_F^0$ is known as the Sommerfeld constant. Although this constant has been derived using the approximation of spherical iso-energetic surfaces, its value remains the same for general surfaces.

Taking into account also the phononic contribution to the specific heat (see the lecture on phononic heat capacity), at low temperatures the specific heat is given by (see the figure below)

$$C_V = C_{el} + C_{ph} = \gamma T + aT^3\,, \tag{12}$$



where

$$a = \frac{12\pi^4}{5} n_{ion} \frac{k_B}{\Theta_D^3} \tag{13}$$

with $n_{ion}$ the ion concentration. The electronic term dominates at very low temperatures, for which $\gamma > aT^2$, i.e. for

$$\frac{T}{\Theta_D} < \frac{1}{2\pi}\sqrt{\frac{5}{6}\frac{n}{n_{ion}}\frac{\Theta_D}{T_F}}\,. \tag{14}$$

In particular, the Debye temperature can be determined from the slope of the curve $C_V/T = \gamma + aT^2 = f(T^2)$ at very low temperatures, while $\gamma$ is determined from the value of this dependence at $T = 0$. The values of $\gamma$ for several metals are given in the table below.

| Metal | $\gamma \cdot 10^{-4}$ (J/mol·K$^2$) | Metal | $\gamma \cdot 10^{-4}$ (J/mol·K$^2$) | Metal | $\gamma \cdot 10^{-4}$ (J/mol·K$^2$) |
|---|---|---|---|---|---|
| Li | 17 | Ag | 6.6 | Zn | 6.5 |
| Na | 17 | Au | 7.3 | Al | 13.5 |
| K | 20 | Be | 2.2 | Fe | 49.8 |
| Cu | 6.9 | Mg | 13.5 | Co | 47.3 |
| Ca | 27.3 | Ba | 27 | Ni | 70.2 |

## Carrier specific heat in intrinsic semiconductors

In a nondegenerate intrinsic semiconductor with spherical iso-energetic surfaces, the energies of the system of electrons and holes are given by, respectively (see (3) and (10b))

$$E_{el} = nk_BT\frac{F_{3/2}(y)}{F_{1/2}(y)} = \frac{3}{2}nk_BT\,, \qquad E_h = \frac{3}{2}pk_BT\,. \tag{15}$$

The total energy of charge carriers is however equal to

$$E_{carr} = \frac{3}{2}nk_BT + nE_g + \frac{3}{2}pk_BT \tag{16}$$

since the free electrons in the conduction band have an additional potential energy of $nE_g$. Because in an intrinsic semiconductor $n = p = n_i$, with $n_i$ the intrinsic carrier concentration, (16) can be written as

$$E_{carr} = n_i(3k_BT + E_g)\,, \tag{17}$$

and the carrier specific heat is

$$C_{carr} = \frac{dE_{carr}}{dT} = \frac{dn_i}{dT}(3k_BT + E_g) + 3n_ik_B\,, \tag{18}$$

when the weak temperature dependence of the bandgap is neglected. Because

$$n_i = \sqrt{N_cN_v}\exp\left(-\frac{E_g}{2k_BT}\right) \propto T^{3/2}\exp\left(-\frac{E_g}{2k_BT}\right), \tag{19}$$

it follows that

$$\frac{dn_i}{dT} = \frac{n_i}{2T}\left(3 + \frac{E_g}{k_BT}\right) \tag{20}$$

and

$$C_{carr} = \frac{n_ik_B}{2}\left(3 + \frac{E_g}{k_BT}\right)^2 + 3n_ik_B = n_ik_B\left[\frac{15}{2} + 3\frac{E_g}{k_BT} + \frac{1}{2}\left(\frac{E_g}{k_BT}\right)^2\right]. \tag{21}$$

This expression is valid if $E_g \geq k_BT$, since otherwise the degeneracy of the system of electrons and holes must be taken into account. From (21) it follows that at low temperatures the contribution of charge carriers to the specific heat in an intrinsic semiconductor can be neglected, due to the exponential temperature dependence of $n_i$.

In an extrinsic semiconductor the specific heat of charge carriers can be calculated in a similar manner. More precisely, in (16) one must introduce the correct concentrations of free carriers in all conduction regimes, and must account for their specific distribution function and temperature dependence. The carrier specific heat of free electrons and holes is found, then, to depend on both the concentration of donor and acceptor ions and of their energy levels. At low temperatures this contribution to the specific heat is, again, negligible.

# Kinetics of charge carriers in solids

## Boltzmann kinetic equation

When an electric or a magnetic field is applied on a crystal, the displacement of charge carriers induces transport (or kinetic) phenomena. The distribution function of charge carriers with energy $E(\boldsymbol{k}) = E_{\boldsymbol{k}}$ in equilibrium is described by the Fermi-Dirac function

$$f_0(E_{\boldsymbol{k}}) = \frac{1}{1 + \exp[(E_{\boldsymbol{k}} - E_F)/k_B T]} . \tag{1}$$

On the other hand, in the presence of external fields, the system of charge carriers is no longer in equilibrium and the corresponding distribution function $f(\boldsymbol{k}, \boldsymbol{r}, t)$ depends, in general, on spatial coordinates and time.

In a semiclassical treatment, the number of particles that follow a certain trajectory is conserved in the absence of scattering processes, so that $df/dt = 0$. However, scattering/collision processes of electrons on phonons, impurities or defects in the crystalline lattice are unavoidable, so that the total derivative of the distribution function does not vanish any more, but is equal to the variation of the distribution function due to collisions. More precisely,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \boldsymbol{r}} \cdot \dot{\boldsymbol{r}} + \frac{\partial f}{\partial \boldsymbol{k}} \cdot \dot{\boldsymbol{k}} = \left( \frac{\partial f}{\partial t} \right)_{coll} , \tag{2}$$

or

$$\frac{\partial f}{\partial t} = \left( \frac{\partial f}{\partial t} \right)_{coll} - \boldsymbol{v} \cdot \nabla_{\boldsymbol{r}} f - \frac{\boldsymbol{F}}{\hbar} \cdot \nabla_{\boldsymbol{k}} f \tag{3}$$

where $v$ is the electron velocity in the crystal and $\boldsymbol{F} = d\boldsymbol{p}/dt = \hbar d\boldsymbol{k}/dt$ is the external force. In a stationary state, when the distribution function is independent of time, $\partial f / \partial t = 0$, and, if we consider the effect of the Lorentz force $\boldsymbol{F} = -e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$ only, we obtain the kinetic Boltzmann equation

$$\boldsymbol{v}\cdot\nabla_{\boldsymbol{r}} f - \frac{e}{\hbar}(\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B})\cdot\nabla_{\boldsymbol{k}} f = \left(\frac{\partial f}{\partial t}\right)_{coll} . \tag{4}$$

To find the distribution function $f(\boldsymbol{k},\boldsymbol{r},t)$ from this equation it is necessary to know the collision term in the right-hand-side. This is a difficult problem, which can be simplified by introducing the relaxation time $\tau(\boldsymbol{k})$, which describes the return to equilibrium of the distribution function when the external fields are switched off:

$$-\left(\frac{\partial f}{\partial t}\right)_{coll} = \frac{f - f_0}{\tau(\boldsymbol{k})}, \tag{5}$$

or

$$f - f_0 = (f - f_0)_{t=0} \exp[-t/\tau(\boldsymbol{k})]. \tag{6}$$

The relaxation time is thus the interval after which the change in the equilibrium distribution function decreases *e* times after the external fields are turned off. The introduction of the relaxation time parameter is possible when the collision processes are elastic, i.e. when the energy of charge carriers is not modified at scattering, and act independently (there is no interference of electron states). Moreover, the inequality $\tau >> \hbar / k_B T$ must be satisfied, where $\hbar / k_B T = \tau_c$ is the collision time. This inequality expresses the fact that the collision time can be neglected, i.e. the collisions are instantaneous. In addition, the external fields must not modify the energy spectrum of electrons in the crystal; this condition prohibits intense magnetic fields, for example, which lead to the quantization of electron energy levels.

The quantum nature of electrons is apparent only in the collision term, through the electron quantum states that satisfy the Pauli principle. A detailed balance between the number of electrons in the state characterized by the wavevector $\boldsymbol{k}$ and those in the state $\boldsymbol{k}'$ leads to the collision term

$$\left(\frac{\partial f}{\partial t}\right)_{coll} = \sum_{\boldsymbol{k}'} P(\boldsymbol{k}',\boldsymbol{k}) f(\boldsymbol{k}')[1 - f(\boldsymbol{k})] - \sum_{\boldsymbol{k}'} P(\boldsymbol{k},\boldsymbol{k}') f(\boldsymbol{k})[1 - f(\boldsymbol{k}')] \tag{7}$$

where $P(\boldsymbol{k},\boldsymbol{k}')$ is the electron transition probability per unit time from state $\boldsymbol{k}$ into the state $\boldsymbol{k}'$. In the equilibrium state

$$P(\boldsymbol{k}',\boldsymbol{k})\, f_0(\boldsymbol{k}')[1 - f_0(\boldsymbol{k})] = P(\boldsymbol{k},\boldsymbol{k}')\, f_0(\boldsymbol{k})[1 - f_0(\boldsymbol{k}')]. \tag{8}$$

We consider distribution functions that can be approximated as perturbations of $f_0$, i.e. that can be expressed as $f(\boldsymbol{k}) = f_0(E_{\boldsymbol{k}}) + f_1(\boldsymbol{k})$, with

$$f_1(\boldsymbol{k}) = -\frac{df_0}{dE_{\boldsymbol{k}}} \boldsymbol{\chi}(E_{\boldsymbol{k}}) \cdot \boldsymbol{k} << f_0(E_{\boldsymbol{k}}), \tag{9}$$

where $\boldsymbol{\chi}(E_{\boldsymbol{k}})$ an, as yet, unknown vectorial function (it has specific forms for different scattering processes). Under these conditions we can express the collision term as

$$\begin{aligned}\left(\frac{\partial f}{\partial t}\right)_{coll} &= \sum_{\boldsymbol{k}'} P(\boldsymbol{k},\boldsymbol{k}')\, f_0(\boldsymbol{k})[1 - f_0(\boldsymbol{k}')]\left(\frac{f(\boldsymbol{k}')[1 - f(\boldsymbol{k})]}{f_0(\boldsymbol{k}')[1 - f_0(\boldsymbol{k})]} - \frac{f(\boldsymbol{k})[1 - f(\boldsymbol{k}')]}{f_0(\boldsymbol{k})[1 - f_0(\boldsymbol{k}')]}\right) \\ &= -\frac{1}{k_B T}\sum_{\boldsymbol{k}'} P(\boldsymbol{k},\boldsymbol{k}')\, f_0(E_{\boldsymbol{k}})[1 - f_0(E_{\boldsymbol{k}'})][\boldsymbol{\chi}(E_{\boldsymbol{k}}) \cdot \boldsymbol{k} - \boldsymbol{\chi}(E_{\boldsymbol{k}'}) \cdot \boldsymbol{k}']\end{aligned} \tag{10}$$

if only the linear terms in $f_1$ are retained and the identity

$$-k_B T \frac{df_0}{dE} = f_0(E)[1 - f_0(E)] \tag{11}$$

is employed. Then,

$$\frac{1}{\tau(\boldsymbol{k})} = -\frac{1}{f_1(\boldsymbol{k})}\left(\frac{\partial f}{\partial t}\right)_{coll} = \sum_{\boldsymbol{k}'} P(\boldsymbol{k},\boldsymbol{k}')\left[\frac{1 - f_0(E_{\boldsymbol{k}'})}{1 - f_0(E_{\boldsymbol{k}})}\right]\left[1 - \frac{\chi(E_{\boldsymbol{k}'}) \cdot \boldsymbol{k}'}{\chi(E_{\boldsymbol{k}}) \cdot \boldsymbol{k}}\right], \tag{12}$$

or $1/\tau(\boldsymbol{k}) = \sum_{\boldsymbol{k}'} P(\boldsymbol{k},\boldsymbol{k}')(1 - k'_\chi / k_\chi)$ for elastic collisions, when $E_{\boldsymbol{k}} = E_{\boldsymbol{k}'}$. Here $k'_\chi$, $k_\chi$ are the projections of $\boldsymbol{k}$, $\boldsymbol{k}'$ on the vector $\boldsymbol{\chi}$. The calculation of the relaxation time can be performed for different scattering mechanisms, the temperature dependence of this parameter being generally expressed as

$$\tau(E) = A(T)E^r, \tag{13}$$

where $A(T)$ is an energy-independent coefficient, and $r$ is characteristic for each collision type: $r = 3/2$ for collisions with acoustic phonons in metals, $r = -1/2$ for the same mechanism in semiconductors, $r = 1/2$ for scattering on optical phonons in polar semiconductors at high temperatures, $r = 0$ for scattering on neutral impurities in metals, and $r = 3/2$ for collisions with ionized impurities in semiconductors. If several scattering mechanisms coexist,

$$\frac{1}{\tau(E)} = \sum_i \frac{1}{\tau_i(E)} . \qquad (14)$$

## Electrical conductivity

The electrical conductivity in a crystal is characterized by the tensor $\hat{\sigma}$ that appears in the definition of the density of electric current:

$$\boldsymbol{j} = \hat{\sigma}\boldsymbol{E} , \qquad (15)$$

or on components $\mu,\ \nu = x, y, z$

$$j_\mu = \sum_\nu \sigma_{\mu\nu} E_\nu . \qquad (16)$$

In an isotropic solid the electrical conductivity is a scalar parameter and $\boldsymbol{j} = \sigma\boldsymbol{E}$ . On the other hand, for $f(\boldsymbol{k}) = f_0(E_{\boldsymbol{k}}) + f_1(\boldsymbol{k})$, the density of electric current per crystal volume can be expressed as

$$\boldsymbol{j} = \frac{1}{V}\sum_{\boldsymbol{k},\sigma} - e\boldsymbol{v}f(\boldsymbol{k}) = -\frac{2e}{V}\sum_{\boldsymbol{k}} \boldsymbol{v}f_1(\boldsymbol{k}) = -\frac{e}{4\pi^3}\int \boldsymbol{v}f_1(\boldsymbol{k})d\boldsymbol{k} \qquad (17)$$

since the equilibrium density function $f_0(E_{\boldsymbol{k}})$ does not bring any contribution if the sum above is performed over all positive and negative $\boldsymbol{k}$ values ($E_{\boldsymbol{k}}$, and hence $f_0(E_{\boldsymbol{k}})$, is an even function of $\boldsymbol{k}$, whereas $\boldsymbol{v} = \hbar^{-1}\nabla_{\boldsymbol{k}} E$ is an odd function of $\boldsymbol{k}$). The perturbation term of the equilibrium distribution function, $f_1(\boldsymbol{k})$, is determined from the kinetic Boltzmann equation.

More precisely, if only an electric field of intensity ***E*** is applied and there is no temperature gradient in the sample, i.e. if $\nabla_{\boldsymbol{r}} f = 0$, the kinetic Boltzmann equation can be written as

$$-\frac{e}{\hbar}\boldsymbol{E}\cdot\nabla_{\boldsymbol{k}} f_0 - \frac{e}{\hbar}\boldsymbol{E}\cdot\nabla_{\boldsymbol{k}} f_1 + \frac{f_1(\boldsymbol{k})}{\tau(\boldsymbol{k})} = 0\,. \tag{18}$$

The second term can be neglected at small electric fields, when only linear effects in ***E*** are considered, case in which the electrical conductivity is independent of the electric field, and the equation above can be solved to obtain

$$f_1(\boldsymbol{k}) = e\tau(\boldsymbol{k})\boldsymbol{E}\cdot\hbar^{-1}\nabla_{\boldsymbol{k}} E\frac{df_0}{dE} = e\tau(\boldsymbol{k})\boldsymbol{E}\cdot\boldsymbol{v}\frac{df_0}{dE}\,. \tag{19}$$

From (15), (17) and (19) it follows that the tensor of the electrical conductivity can be expressed as

$$\sigma_{\mu\nu} = \frac{e^2}{4\pi^3}\int\left(-\frac{df_0}{dE}\right)\tau(\boldsymbol{k})v_\mu v_\nu d\boldsymbol{k}\,. \tag{20}$$

In isotropic crystals, for an electric field along the *x* direction, the conductivity is scalar:

$$\sigma = \sigma_{xx} = \frac{e^2}{4\pi^3}\int\left(-\frac{df_0}{dE}\right)\tau(\boldsymbol{k})v_x^2 d\boldsymbol{k}\,. \tag{21}$$

For spherical iso-energetic surfaces, $E_{\boldsymbol{k}} = \hbar^2 k^2/2m_{eff}$, $v_\mu = \hbar k_\mu/2m_{eff}$, and in spherical coordinates with $\theta$ the polar angle and $\varphi$ the azimuthal angle, $k_x = k\sin\theta\cos\varphi$, $k_y = k\sin\theta\sin\varphi$, $k_z = k\cos\theta$, $d\boldsymbol{k} = k^2 dk\sin\theta d\theta d\varphi$, and

$$\sigma = \sigma_{xx} = \frac{e^2\hbar^2}{4\pi^3 m_{eff}^2}\int\left(-\frac{df_0}{dE}\right)\tau(\boldsymbol{k})k^4 dk\int_0^\pi \sin^3\theta d\theta\int_0^{2\pi}\cos^2\varphi d\varphi = \frac{e^2\hbar^2}{3\pi^3 m_{eff}^2}\int_{k_{\min}}^{k_{\max}}\left(-\frac{df_0}{dE}\right)\tau(\boldsymbol{k})k^4 dk\,. \tag{22}$$

Taking into account that $k^4 dk = (2m_{eff})^{3/2} m_{eff} E^{3/2} dE/\hbar^5$ and that the relaxation time depends on energy, such that its statistical average can be defined as

$$\langle\tau(E)\rangle = \frac{\int_0^\infty \left(-\frac{df_0}{dE}\right)\tau(E)E^{3/2}dE}{\int_0^\infty \left(-\frac{df_0}{dE}\right)E^{3/2}dE}, \tag{23}$$

the electrical conductivity becomes

$$\sigma = \frac{ne^2}{m_{eff}}\langle\tau(E)\rangle, \tag{24}$$

where we have used the fact that for metals the electron concentration can be written as

$$\begin{aligned} n &= \frac{(2m_{eff})^{3/2}}{2\pi^2\hbar^3}\int_0^\infty \frac{E^{1/2}dE}{\exp[(E-E_F)/k_BT]+1} \\ &= \frac{(2m_{eff})^{3/2}}{2\pi^2\hbar^3}\left[\frac{2}{3}\frac{E^{3/2}}{\exp[(E-E_F)/k_BT]+1}\Bigg|_0^\infty + \frac{2}{3}\int_0^\infty\left(-\frac{df_0}{dE}\right)E^{3/2}dE\right] = \frac{(2m_{eff})^{3/2}}{3\pi^2\hbar^3}\int_0^\infty\left(-\frac{df_0}{dE}\right)E^{3/2}dE \end{aligned} \tag{25}$$

Similar relations are obtained along the principal axes for a crystal with elliptical iso-energetic surfaces.

Alternatively, the density of electric current can be expressed as

$$\boldsymbol{j} = -en\boldsymbol{v} = en\mu\boldsymbol{E} \tag{26}$$

where the mobility $\mu$ of charge carriers/electrons is introduced through $\boldsymbol{v} = -\mu\boldsymbol{E}$. The minus sign indicates that the motion of electrons is opposite to the direction of the applied electric field $\boldsymbol{E}$. The relation between the electric conductivity and the mobility is thus

$$\sigma = ne\mu, \qquad \text{i.e.} \qquad \mu = \frac{e\langle\tau(E)\rangle}{m_{eff}}. \tag{27}$$

In anisotropic crystals the mobility, as the electric conductivity, is a tensor.