Chapter 1

1.1 Importance of mobile carriers

Any performing system has to carry some form of energy from one spatial point to other. For this purpose, mobile members are needed in the system. It is electrical energy that is being carried inside the semiconductor system. Also heat energy can be carried by mobile carriers. If no mobile carrier is available inside the solid state semiconductor, it would not be useful as far as the system performance is concerned. It is very obvious that fluid can flow and while flowing, it can carry some energy. In contrast, it is amazing to realize the fact that there are some mobile members inside a solid like metal or semiconductor to conduct electrical or heat energy!

In metals, carriers are electrons. In semiconductors, two types of carriers are there, electrons and holes. It is very important to control the numbers and properties of these carriers to make the system performing, since they are the ones to carry energy. Importance of the mobile carriers can be felt from Table 1.1 that shows and inverse relationship between the density of mobile electrons in a system and its electrical resistivity. It is the semiconductor that shows highest amount of flexibility in the electrical resistivity (or its inverse, electrical conductivity). The controllability of the resistivity of semiconductor materials using today's technology drives the success in integrated circuit (IC) technology industries. Other than the structure of the device, it is the numbers and properties of mobile carriers that determine the system performance.

Table 1.1: Conduction or mobile electron density and resistivity for various kinds of material systems.

Material Type (examples)	Mobile Electron Density (cm^{-3})	Resistivity ($\Omega - cm$)
Superconductor (Sn, Pb at 0.1K to 4K)	10^{23}	0
Good Conductor (Metals: K, Na, Cu, Au)	$10^{22} - 10^{23}$	$10^{-6} - 10^{-5}$
Conductor (Semi-metals: As, B, Graphite)	$10^{17} - 10^{22}$	$10^{-5} - 10^{-2}$
Semiconductor (Ge, Si, GaAs, GaP, InP, etc.)	$10^{6} - 10^{17}$	$10^{-2} - 10^{9}$
Semi-insulator (Amorphous Si)	$10^{1} - 10^{5}$	$10^{10} - 10^{14}$
Insulator ($\text{SiO}_2, Si_3N_4, etc.$)	1 - 10	$10^{14} - 10^{22}$

1.2 Means of obtaining mobile carriers

If some sort of energy, such as heat or light, is supplied from outside, a semiconductor generates electrons and holes, the mobile carriers of the system. Essentially we learn from quantum mechanics that the electrons can stay only in some energy levels. Precisely, these energy levels are discrete in nature, and the electrons tend to stay in the lower most level if it is not occupied. With the supply of energy from outside, the electrons gain energy and tend to move towards higher energy levels and become mobile since they are not tightly bound by the atoms at higher energy levels. However, the most efficient way to obtain mobile carriers within a semiconductor is by doping which we shall learn later.

1.3 Towards estimating the number of mobile carriers

Let's discuss the procedure for estimating the number of electrons. Essentially we estimate the carrier density instead of estimating the total carrier number. Since electrons can stay in certain energy levels, it makes sense to carry out the calculation in energy domain. Intuitively, the electrons density can be figured out if we can estimate the available number of energy levels within a unit volume and multiply it with a factor that is a measure of occupation of those energy levels by electrons. Note that the availability of the energy levels does not ensure their occupancy by electrons. Calculation of the available number of energy levels per unit volume requires the knowledge of how the energy levels are distributed in energy space, that is to say, whether the density of energy levels is uniform or not. If the density of available energy levels is uniform, the total available energy levels can be obtained by multiplying the level density with the energy span of interest. However, in practice for 3D crystal system we find that both density of energy levels or energy states and occupation factor are nonlinear functions of energy itself. Hence, we can get the density of mobile electrons as

$$n = \int_{E_1}^{E_2} D(E) f(E) dE \tag{1.1}$$

where $D(E)$ is the density of energy states (DOS) in per eV per cm^3 and $f(E)$ is the occupation factor. The lower limit in (1.1), i.e., E_1 signifies a minimum amount of energy required to obtain a mobile electron. Note that we are not interested to obtain the density of total electrons available in the crystal system, but in the density of mobile electrons only. To become mobile, it has to gain a minimum amount of energy (which is E_1) over and above which the electrons can gain kinetic energy. The upper limit is the maximum energy that the electrons can obtain being within the semiconductor. Before we discuss on $f(E)$, we need to discuss more on $D(E)$. To understand and obtain an expression for $D(E)$, we need to look into a crystal system, at least, to some extent.

1.4 Crystal and lattice

Figs. 1.1 (a) and (b) show a 1D crystal and the corresponding 1D lattice with lattice constant a, respectively. A lattice is a mathematical abstraction of the corresponding crystal system. A lattice point represents the set of atomic arrangement that gets repeated in the corresponding crystal system. The 1D lattice and the corresponding crystal systems are extended from $-\infty$ to $+\infty$ to comply with translational symmetry. A 2D or 3D lattice and the corresponding crystal system also comprise regular arrangement and are extended

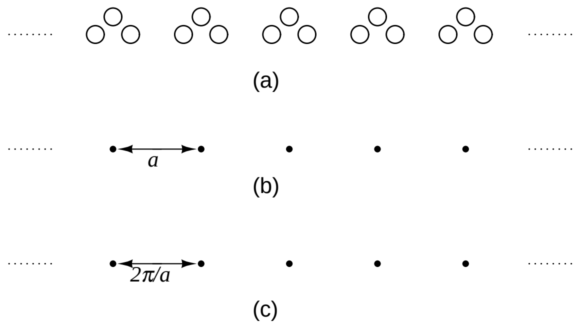

Figure 1.1: (a) One dimensional real crystal where a group of three atoms are repeating in x-direction from $-\infty$ to $+\infty$ (although only five repeatations are shown), (b) Real space lattice of the 1D crystal, (c) Corresponding reciprocal or Fourier space lattice.

infinitely, in principle, to comply with translational symmetry. Let's again concentrate on 1D lattice with lattice constant a. Following translational symmetry, the spatial distribution of lattice point can be expressed as

$$f(x) = f(x \pm na) = \Sigma_{n=-\infty}^{+\infty} F_n e^{in(2\pi/a)x}. \tag{1.2}$$

Fig. 1.1(c) shows the distribution of the real space point of Fig. 1.1(b) in Fourier space. Here the spacing between the two points is $2\pi/a$. We obtain another lattice in Fourier space or k-space once we express real space lattice function $f(x)$ in terms of Fourier series in (1.2). Similarly Figs. 1.2(a) and (b) show the real space 2D lattice and the corresponding 2D Fourier or k-space lattice, respectively. The lattice spacing in x-direction is a and in y-direction is b, which in Fourier space become $2\pi/a$ in k_x-direction and $2\pi/b$ in k_y-direction. Similarly one can imagine that a 3D lattice having lattice spacing a, b, c in x, y, z-directions, respectively, will have a Fourier space 2D lattice with lattice spacing $2\pi/a$, $2\pi/b$, and $2\pi/c$ in k_x, k_y, and k_z directions, respectively. The Fourier space lattice is popularly known as reciprocal space lattice. If we consider the mobile electrons to be weakly bound and nearly free, up to certain values for kinetic energy, we can assume them to have a wave function like a plane wave

$$\psi = \psi_0 e^{i(\omega t + kx)} \tag{1.3}$$

where ω, t, k and x are angular frequency, time, wave vector and real space respectively. Note that ω and t are Fourier pair and similarly k and x. Now

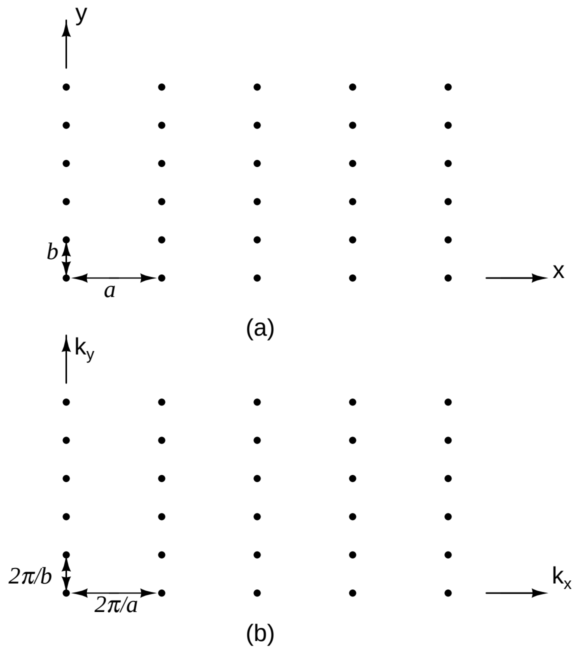

Figure 1.2: (a) Two dimensional real space lattice of a 2D crystal, (b) Corresponding reciprocal or Fourier space lattice.

according to quantum mechanics, momentum is an operator given by

$$p = \frac{\hbar}{i} \frac{\partial}{\partial x} \tag{1.4}$$

in 1D. If we apply the operator (1.4) on to the wave function (1.3), we obtain

$$\frac{\hbar}{i} \frac{\partial \psi}{\partial x} = p\psi = \hbar k \psi \tag{1.5}$$

yielding electron momentum

$$p = \hbar k = mv \tag{1.6}$$

where m and v are mass and velocity of electron. If E_1 is potential energy for electron and the energy excess to E_1 generates its kinetic energy, then total energy can be written as

$$E = E_1 + \frac{1}{2} m v^2 = E_1 + \frac{\hbar^2 k^2}{2m}. \tag{1.7}$$

Strictly speaking, the relation in (1.7) is valid for free electrons. Although electrons are not absolutely free in semiconductors, we can assume them to be nearly free within the bulk of the semiconductor crystal and therefore, we find (1.7) to predict fairly well the actual energy of electrons within the semiconductor if it is only few k_BT energy levels higher than E_1 (with k_B as Boltzmann constant and T as temperature in Kelvin). At room temperature, $k_BT = 0.0259eV$. Since most of the mobile electrons remain within a few k_BT energy levels from E_1, (1.7) is very much useful. To describe the number of mobile carriers, it is imperative that one requires to count the available energy levels at first. If we integrate the density of states, $D(E)$ appearing in (1.1) without the factor, $f(E)$, we can obtain the total number of energy levels available within E_1 and E_2. Let's denote this variable as $N_V(E)$. The suffix $'V'$ indicates that it is a measure per unit volume. Therefore,

$$N_V(E) = \int_{E_1}^{E_2} D(E)dE \tag{1.8}$$

or,

$$D(E) = \frac{dN_V(E)}{dE}. \tag{1.9}$$

It is clear that to obtain an expression of $D(E)$, we require an expression for $N_V(E)$. Since E and k are related by (1.7), we can start for obtaining an expression for $N_V(k)$ and then convert it into $N_V(E)$ using (1.7). To obtain an expression of $N_V(k)$, i.e., the number of k-states per unit volume within certain energy range, we need to understand periodic boundary condition.

1.5 Periodic boundary condition

Fig. 1.3(a) shows a lattice specimen of length L with $(N + 1)$ lattice points. Periodic boundary condition implies that the 1st point is equivalent with the $(N + 1)$th point. Such a condition ensures the translational symmetry (1.2) even for the finite lattice system. If we assume a time-independent version of (1.3), i.e., $\psi(kx)$, as the electron wave function within the lattice, it must obey the periodic boundary condition, i.e.,

$$\psi(kx)|_{x=0} = \psi(kx)|_{x=L}. \tag{1.10}$$

Figs. 1.3(b) and (c) show the possible wavefunctions that satisfy (1.10). It is clear that if the wave function is zero at $x = 0$ and $x = L$, then it takes the form of $\sin(kx)$ (Fig. 1.3(b)) and if it is non-zero at these points, it takes the form of $\cos(kx)$ (Fig. 1.3(c)). Now if it has to satisfy (1.10) simultaneously for $\sin(kx)$ and $\cos(kx)$, the following relation must hold:

$$k_n L = n2\pi \tag{1.11}$$

for any integer value of n and k_n signifies various discrete values of k (depending on the value of n) that are allowed in the lattice system. It means that only some discrete k values are allowed for electrons staying within the lattice system. Let's now try to count the discrete k-states. If a is the lattice constant, $L = Na$. Therefore, $k_1 = \frac{2\pi}{Na}$, $k_2 = 2\frac{2\pi}{Na}$, ... and $k_N = \frac{2\pi}{a}$. Again $k_{N+1} = \frac{2\pi}{a} + k_1$ and with respect to the sine or cosine function, k_{N+1} appears to be equivalent with k_1 because

$$\begin{aligned}\sin(k_{N+1}x_m) &= \sin(\frac{2\pi}{a}ma + k_1 ma) = \sin(k_1 x_m),\\ \cos(k_{N+1}x_m) &= \cos(\frac{2\pi}{a}ma + k_1 ma) = \cos(k_1 x_m).\end{aligned} \tag{1.12}$$

Similarly k_{N+2} is equivalent with k_2 and so on. Therefore we can limit our investigation for k-states from 0 to $\frac{2\pi}{a}$ since the k-states are repeating after that. The spacing between two k-states appears to be $\Delta k = k_2 - k_1 = k_3 - k_2 = ... = \frac{2\pi}{L}$. If the specimen size is quite large, Δk reduces and it reduces to zero if L tends to infinity. Therefore, finite size of the lattice system yields discreteness in the available k values. It would be continuous if the size of the lattice is infinite. Since the lattice points are spaced by a, the k-states are limited within $(0, \frac{2\pi}{a})$ range. Therefore, discreteness in the real space lattice yields a finite size in the reciprocal k-space.

1.6 Density of states

Now if we define $N(k_r)$ as the number of states having k-values less than equal to $|k_r|$, the volume of interest in k-space must be a sphere in 3D case with k_r as the radius of the sphere. In 2D case, it would be a circle with k_r as the radius of the circle and in 1D case, it would be a line of length $2k_r$, with $+k_r$ in $+ve$ direction and $-k_r$ in $-ve$ direction. In 1D case, we obtained $\Delta k = \frac{2\pi}{L}$ which signifies that if we move in k-space by an amount of $\frac{2\pi}{L}$, we encounter only one allowed k-state. Similarly in a rectangular 2D specimen of size $L_x L_y$, if we cover an area of $\frac{2\pi 2\pi}{L_x L_y}$, we shall encounter only one allowed

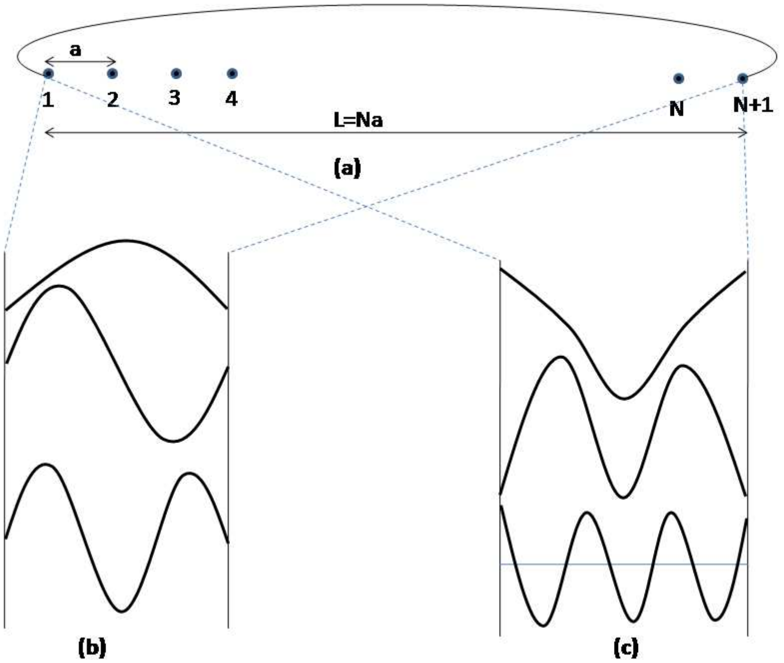

Figure 1.3: (a) One dimensional finite lattice with N+1 lattice points and lattice constant a. With periodic boundary condition, 1st lattice is equivalent with $(N+1)$th lattice point and the length of the specimen, $L = Na$. (b) $\sin(kx)$ like wave functions for electrons within the lattice obeying (1.10). (c) $\cos(kx)$ like wave functions for electrons within the lattice obeying (1.10).

k-state. Finally in a 3D lattice specimen having a size of $L_xL_yL_z$, one has to cover a volume of $\frac{2\pi 2\pi 2\pi}{L_xL_yL_z}$ to encounter one allowed k-state. Therefore, if one covers unit k-volume in 3D lattice, an amount of $\frac{L_xL_yL_z}{8\pi^3}$ number of k-states will be encountered. Within a spherical volume of radius k_r, therefore, the total number of k-states will be $\frac{4}{3}\pi k_r^3 \frac{L_xL_yL_z}{8\pi^3}$. So following the definition of $N(k_r)$, one writes

$$N(k_r) = \frac{4}{3}\pi k_r^3 \frac{L_xL_yL_z}{8\pi^3} = \frac{L_xL_yL_z}{6\pi^2}k_r^3. \tag{1.13}$$

Using (1.7), one can easily find this number in terms of energy, E, i.e., the number of energy states having energy less than equal to $E - E_1$, as

$$N(E - E_1) = \frac{L_x L_y L_z}{6\pi^2 \hbar^3} \{2m(E - E_1)\}^{3/2}. \tag{1.14}$$

Therefore, the corresponding value per unit volume of the lattice specimen appears to be

$$N_V(E - E_1) = \frac{1}{6\pi^2 \hbar^3} \{2m(E - E_1)\}^{3/2}. \tag{1.15}$$

Following (1.9), the density of energy states appears to be

$$D(E - E_1) = 2\frac{dN_V(E - E_1)}{dE} = \frac{m}{\pi^2 \hbar^3} \sqrt{2m(E - E_1)} \tag{1.16}$$

where a factor of 2 is multiplied to consider the spin degeneracy.

1.7 Energy bands and bandgap

Fig. 1.4(a) shows a band structure of a semiconductor for the energy levels of its valence electrons. The upper branch follows the trend of (1.7) if the origin is shifted to a k value where the E-k diagram shows a minima. The lower branch also follows (1.7), but in an inverted manner. Note that (1.7) is useful here to predict the electron's energy for limited k-values which are not too high. More precisely, such a pair of energy branches at different k-values are obtained as the solution for energy levels from *Schrodinger*'s equation if *Bloch* wave function is used. The *Bloch* wave function takes the form of (1.3) with a ψ_0 taking the periodicity of the lattice into consideration. Here we need not bother on how such a solution is obtained for different semiconductor. Analysis of such a solution is more important here.

From (1.7), one can obtain a relation for electron mass as

$$m^* = \frac{\hbar}{\frac{d^2E}{dk^2}}. \tag{1.17}$$

Here instead of mass m, we have equated the right hand side with m^*, which is referred to as effective mass of electron. As the curvature $\frac{d^2E}{dk^2}$ increases, effective mass reduces and this way the effective mass of mobile electrons in the upper branch of Fig. 1.4(a) can vary from free electron mass, m (=

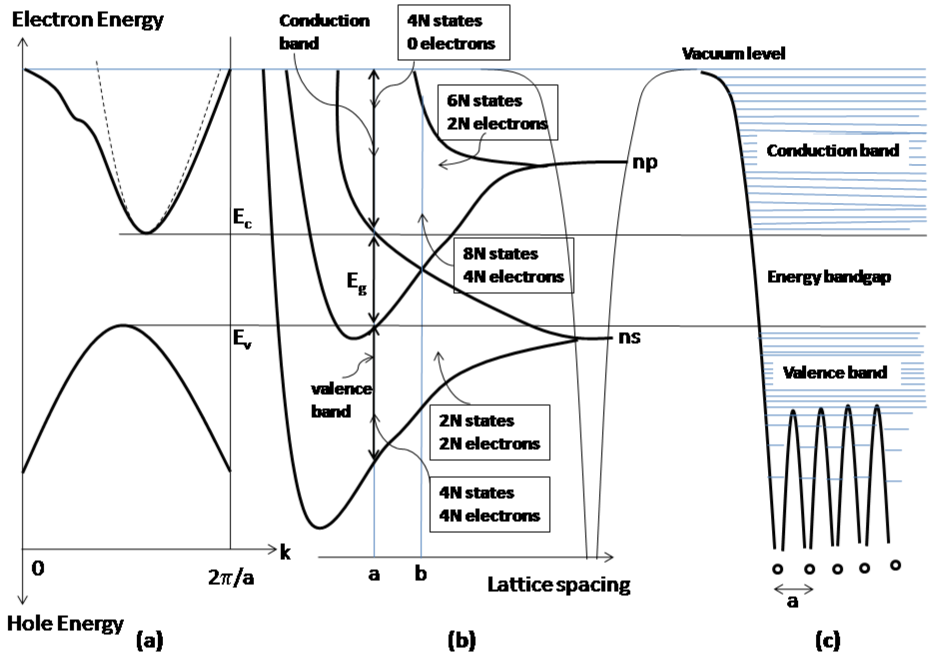

Figure 1.4: (a) Energy (E) versus wave vector (k) relationship in a semiconductor crystal with lattice spacing a. The dotted line in the upper branch shows the limitation and suitability of (1.7) to capture the actual E-k behavior. (b) Formation of the band structure when the lattice spacing is reduced towards left. (c) Positional energy band diagram (E-x) for the same semiconductor crystal with lattice constant, a.

$9.1.10^{-31}kg$). Sometimes it is more than m ($m^* = 1.1m$), sometimes it is less ($m^* = 0.25m$). Now if we consider the lower branch in Fig. 1.4(a), one sees the curvature to be negative leading to a negative effective mass! Instead of assigning a negative mass to a negative charge (electron), one can assume this energy branch for a positive charge with positive effective mass, energy of which increases downward in Fig. 1.4(a). Such a positive charge (magnitude is same as that of electron) with positive effective mass is a hole. Note that electron (negative charge) energy increases in upward direction in Fig. 1.4(a) whereas hole (positive charge) energy increases in downward direction. The upper branch corresponds to conduction band and the lower branch is called valence band. Minimum separation between the two branches is known as energy gap or bandgap.

In a crystal with N lattice points (assume large value for N), if we gradually reduce the lattice constant, the valence electrons energy levels (ns and np) are broadened into energy bands as shown in Fig. 1.4(b). This formation of energy bands is due to many energy levels of many atoms comprising the N lattice points. Energy levels of each atom are discrete as long as their in-between spacings are large. As the spacing reduces, energy lines of similar levels adjust themselves within a very small space and form energy bands. At certain energy spacing (b in Fig. 1.4(b)), the two broadened energy bands may merge depending upon the nature of the crystal or lattice. If N lattice points are made of N atoms, the total energy states corresponding to the valence energy levels are $2N + 6N = 8N$. However, total number of valence electrons are $2N + 2N = 4N$. Therefore, $4N$ empty states will be there. At a particular lattice constant (a in Fig. 1.4(b)), the two energy bands are separated by a finite energy gap, known as bandgap, E_g. The bottom energy band (known as valence band) have $4N$ energy levels and the upper band (called as conduction band) have $4N$ levels. Fig. 1.4(c) shows the position dependent energy band (E-x) diagram for the particular semiconductor with lattice constant a. Note that Fig. 1.4(a) depicts the same energy band diagram with respect to the wave vector (E-k diagram) for the same real-space lattice having lattice constant, a. Bottom of the conduction band is written as E_c, top of the valence band as E_v, and energy bandgap, $E_g = E_c - E_v$.

Lower the energy gap of a system, larger the mobile carriers at a given temperature (since bound electrons in valence band can easily move to the conduction band), hence more electrical conductivity as per Table 1.1. For metals, there is no energy gap whereas for insulators, the energy gap is too high (around $10eV$). A moderate value of energy gap in semiconductors (around 0.5 to $3eV$) opens up the possibility to use them as near-metals or as near-insulators at our will by changing the number of mobile carriers in the system using certain technology and operating conditions.