6 Subbands

In Chapter 5 we saw that the energy levels $E_b(\vec{k})$ in a periodic solid can labeled in terms of $\vec{k}$, with the number of branches b equal to the number of basis functions per unit cell. Strictly speaking, this requires us to assume periodic boundary conditions in all directions so that the periodicity is preserved everywhere even at the "ends." Real solids usually have "ends" where periodicity is lost, but this is commonly ignored as a surface effect that has no influence on bulk properties. The finite size of actual solids normally leads to no observable effects, but as we scale down the size of device structures, the discreteness of energy levels becomes comparable to the thermal energy $k_B T$ leading to experimentally observable effects. Our objective in this chapter is to describe the concept of subbands which is very useful in describing such "size quantization" effects. In Section 6.1 we will describe the effect of size quantization on the $E(\vec{k})$ relation using specific examples. We will then look at its effect on experimentally observable quantities, like the density of states (DOS), $D(E)$ in Section 6.2 and the number of subbands or modes, $M(E)$. In Section 6.3 we will see that the maximum conductance of a wire is proportional to the number of modes around the Fermi energy ($E = \mu$), the maximum conductance per mode being equal to the fundamental constant $G_0 \equiv q^2/h$ (Eq. (1.1)) as discussed in the introductory chapter. Finally in Section 6.4, I will discuss the matter of the appropriate velocity for an electron in a periodic solid. For free electrons, the wavefunction has the form of plane waves $\sim \exp(\pm ikx)$ and the corresponding velocity is $\hbar k/m$. Electrons in a periodic solid also have wavefunctions that can be labeled with a k, but they are not plane waves. So what is the quantity (if there is one!) that replaces $\hbar k/m$ and how do we obtain it from our knowledge of the bandstructure $E_b(\vec{k})$?

6.1 Quantum wells, wires, dots, and "nanotubes"

We saw in Chapter 5 that a good way to catalog the energy levels of a homogeneous periodic solid is in terms of the wavevector $\vec{k}$. How do we catalog the energy levels of a nanostructured device? As an example, consider the transistor structure discussed in Chapter 1, modified for convenience to include two gate electrodes symmetrically

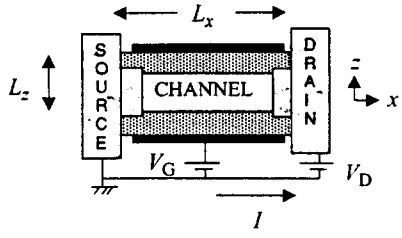


Fig. 6.1.1 Sketch of a dual-gate nanoscale field effect transistor (FET). The top gate is held at the same potential, V_G, as the bottom gate.

on either side of the channel (Fig. 6.1.1). The x-dimension (L) is getting quite short but since electrons can flow in and out at the contacts, one needs to enforce "open boundary conditions" which we will discuss in Chapter 8. But it is not too wrong to treat it as a closed system assuming periodic boundary conditions, at least in the absence of bias ($V_D = 0$). In the z-direction we have tight confinement leading to observable effects beyond what one might expect on the basis of periodic boundary conditions. The y-dimension (perpendicular to the plane of the paper) is typically a few microns and could be considered large enough to ignore surface effects, but as devices get smaller this may not be possible. So how do we label the energy levels of a structure like this?

In a homogeneous solid, electrons are free to move in all three directions and the energy levels can be classified as $E_b(k_x, k_y, k_z)$, where the subscript b refers to different bands. By contrast, the transistor structure shown in Fig. 6.1.1 represents a quantum well where the electrons are free to move only in the x–y plane. We could estimate the energy levels in this structure from our knowledge of the energy levels $E_b(k_x, k_y, k_z)$ of the homogeneous solid, by modeling the z-confinement as a ring of circumference L_z, so that the resulting periodic boundary conditions restrict the allowed values of k_z to a coarse lattice given by $k_z = p2\pi/L_z$:

$$E_{b,p}(k_x, k_y) \approx E_b(k_x, k_y, k_z = p2\pi/L_z)$$

where the additional subscript p can be called a *subband* index. This works quite well for ring-shaped structures like carbon nanotubes, but most low-dimensional structures involve more complicated confinement geometries and in general it takes considerably more work to compute the subband energy levels of a low-dimensional solid from the bulk bandstructure. For the transistor structure shown in Fig. 6.1.1 the insulator layers act like infinite potential walls (see Fig. 2.1.3b) and we can obtain fairly accurate estimates by assuming that the resulting box restricts the allowed values of k_z to a coarse lattice given by $k_z = p\pi/L_z$. The energy levels can then be classified as

$$E_{b,p}(k_x, k_y) \approx E_b(k_x, k_y, k_z = p\pi/L_z)$$

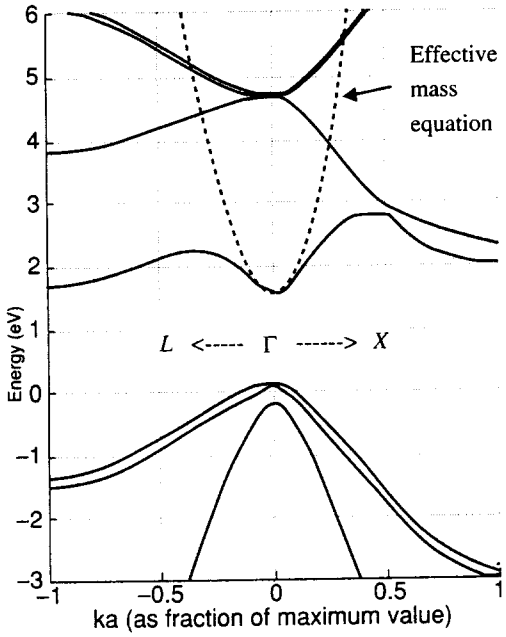


Fig. 6.1.2 Solid curves show the full bandstructure obtained from the sp^3s^* model described in Chapter 5. The dashed curve shows the dispersion obtained from a one-band effective mass model (Eq. (6.1.1)) with parameters adjusted for best fit: $E_c = 1.55$ eV and $m_c = 0.12m$ (m is the free electron mass). Actually the accepted value for the effective mass for GaAs is $0.07m$, but the sp^3s^* model parameters that we use are optimized to give the best fit over the entire band, and are not necessarily very accurate near the band edge.

As we have explained, this is only approximate, but the main point I am trying to make is that *quantum wells* have *1D subbands*, each having a *2D dispersion relation*, $E(k_x, k_y)$.

How small does the dimension L_z have to be in order for the structure to qualify as a quantum well? Answer: when it leads to experimentally observable effects. This requires that the discrete energy levels corresponding to the quantized values of $k_z = p\pi/L_z$ be less than or comparable to the thermal energy $k_B T$, since all observable effects tend to be smoothed out on this energy scale by the Fermi function. To obtain a "back-of-an-envelope" estimate, let us assume that the dispersion relation $E_b(\vec{k})$ in the energy range of interest is described by a parabolic relation with an effective mass, m_c:

$$E(\vec{k}) \approx E_c + \frac{\hbar^2 (k_x^2 + k_y^2 + k_z^2)}{2m_c} \qquad \textit{Bulk solid} \tag{6.1.1}$$

where E_c and m_c are constants that can be determined to obtain the best fit (see Fig. 6.1.2). These parameters are referred to as the conduction band edge and the conduction band effective mass respectively and are well-known for all common semiconductors.

The z-confinement then gives rise to subbands (labeled p) such that

$$E_p(k_x, k_y) \approx E_c + p^2\varepsilon_z + \frac{\hbar^2(k_x^2 + k_y^2)}{2m_c} \qquad \textit{Quantum well}$$

$$\varepsilon_z = \frac{\hbar^2\pi^2}{2m_cL_z^2} = \frac{m}{m_c}\left(\frac{10\text{ nm}}{L_z}\right)^2 \times 3.8\text{ meV} \tag{6.1.2}$$

A layer 10 nm thick would give rise to subband energies ~4 meV if the effective mass m_c were equal to the free electron mass m. Materials with a smaller effective mass (like GaAs with $m_c = 0.07m$) lead to a larger energy separation and hence more observable effects of size quantization.

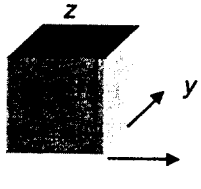
We could take this a step further and consider structures where electrons are free to move only in the x-direction and are confined in both y- and z-directions, as shown below.

This would be the case if, for example, the width of the FET in Fig. 6.1.1 in the direction perpendicular to the plane of the paper were made really small, say less than 100 nm. Such *quantum wires* have *2D subbands*, each having a *1D dispersion relation* that can be estimated by quantizing both the y- and z-components of the wavevector $\vec{k}$:

$$E_{n,p}(k_x) \approx E_c + n^2\varepsilon_y + p^2\varepsilon_z + \frac{\hbar^2k_x^2}{2m_c} \qquad \textit{Quantum wire}$$

where ε_y is related to the y-dimension L_y by a relation similar to Eq. (6.1.2). Finally, one could consider structures that confine electrons in all three dimensions (as shown below) leading to discrete levels like atoms that can be estimated from

$$E_{m,n,p} \approx E_c + \frac{m^2\hbar^2\pi^2}{2m_cL_x^2} + \frac{n^2\hbar^2\pi^2}{2m_cL_y^2} + \frac{p^2\hbar^2\pi^2}{2m_cL_z^2} \qquad \textit{Quantum dot}$$



Such quantum dot structures are often referred to as artificial atoms.

Carbon nanotubes: Carbon nanotubes provide a very good example for illustrating the concept of subbands (Fig. 6.1.3). We saw in Section 5.2 that the energy levels of a sheet of graphite can be found by diagonalizing the (2 × 2) matrix

$$h(\vec{k}) = \begin{bmatrix} 0 & h_0 \\ h_0^* & 0 \end{bmatrix} \tag{6.1.3}$$

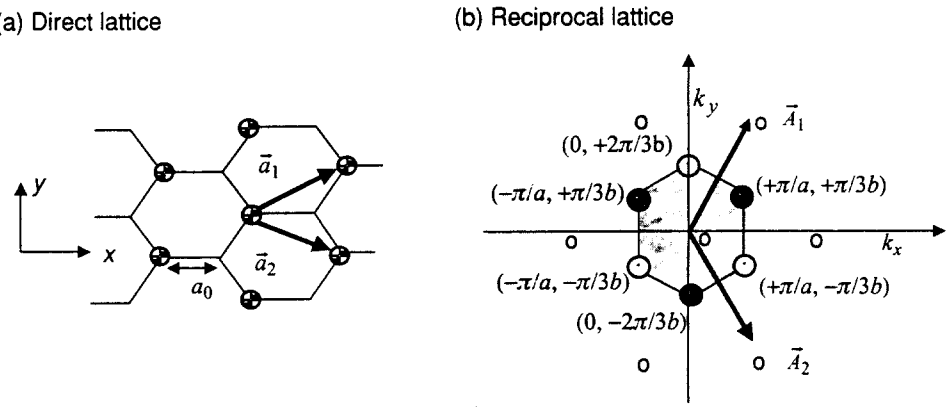


Fig. 6.1.3 (a) Arrangement of carbon atoms on the surface of graphite. (b) Reciprocal lattice showing Brillouin zone (shaded).

where

$$h_0 \equiv -t\left(1 + e^{i\vec{k}\cdot\vec{a}_1} + e^{i\vec{k}\cdot\vec{a}_2}\right) = -t\left(1 + 2e^{ik_x a}\cos k_y b\right)$$

The eigenvalues are given by

$$E = \pm|h_0| = \pm t\sqrt{1 + 4\cos k_y b \cos k_x a + 4\cos^2 k_y b} \tag{6.1.4}$$

Since each unit cell has two basis functions, the total number of states is equal to $2N$, N being the number of unit cells. Each carbon atom contributes one electron to the π-band, giving a total of $2N$ electrons that fill up exactly half the states. Since the energy levels are symmetrically disposed about $E = 0$, this means that all states with $E < 0$ are occupied while all states with $E > 0$ are empty, or equivalently one could say that the Fermi energy is located at $E = 0$.

Where in the k_x–k_y plane are these regions with $E = 0$ located? Answer: wherever $h_0(\vec{k}) = 0$. It is easy to see that this occurs at the six corners of the Brillouin zone:

$$e^{ik_x a}\cos k_y b = -1/2 \qquad \begin{aligned} k_x a &= 0, \quad k_y b = \pm 2\pi/3 \\ k_x a &= \pi, \quad k_y b = \pm\pi/3 \end{aligned}$$

These six points are special as they provide the states right around the Fermi energy and thus determine the electronic properties. They can be put into two groups of three:

$$(k_x a, k_y b) = (0, -2\pi/3), \quad (-\pi, +\pi/3), \quad (+\pi, +\pi/3) \tag{6.1.5a}$$

$$(k_x a, k_y b) = (0, +2\pi/3), \quad (-\pi, -\pi/3), \quad (+\pi, -\pi/3) \tag{6.1.5b}$$

All three within a group are equivalent points since they differ by a reciprocal lattice vector. Each of the six points has *one-third* of a valley around it within the first Brillouin zone (shaded area in Fig. 6.1.3b). But we can translate these by appropriate reciprocal lattice vectors to form two *full* valleys around two of these points, one from

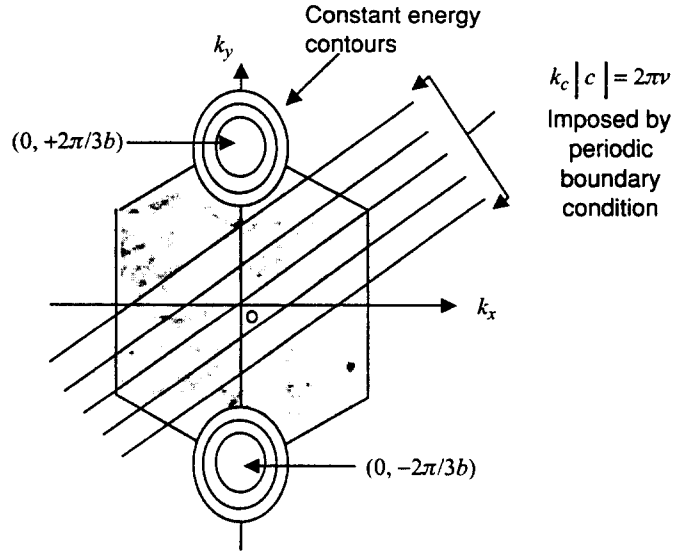


Fig. 6.1.4 Reciprocal lattice of graphite showing straight lines $k_c\,|c| = 2\pi\,v$ representing the constraint imposed by the nanotube periodic boundary conditions.

each group:

$$(k_x a, k_y b) = (0, \pm 2\pi/3)$$

Once a sheet of graphite is rolled up into a nanotube, the allowed values of k are constrained by the imposition of periodic boundary conditions along the circumferential direction. Note that this periodic boundary condition is a *real* one imposed by the physical structure, rather than a conceptual one used to facilitate the counting of states in a large structure whose exact boundary conditions are unimportant. Defining a circumferential vector

$$\vec{c} = m\vec{a}_1 + n\vec{a}_2 = \hat{x}(m+n)a + \hat{y}(m-n)b \tag{6.1.6}$$

that joins two equivalent points on the x–y plane that connect to each other on being rolled up, we can express the requirement of periodic boundary condition as

$$\vec{k}\cdot\vec{c} \equiv k_c|c| = k_x a(m+n) + k_y b(m-n) = 2\pi\,v \tag{6.1.7}$$

which defines a series of parallel lines, each corresponding to a different integer value for v (Fig. 6.1.4). We can draw a one-dimensional dispersion relation along any of these lines, giving us a set of dispersion relations $E_v(k)$, one for each *subband* v.

Whether the resulting subband dispersion relations will show an energy gap or not depends on whether one of the lines defined by Eq. (6.1.7) passes through the center of one of the valleys

$$(k_x a, k_y b) = (0, \pm 2\pi/3)$$

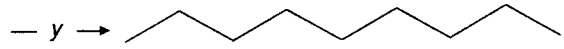


Fig. 6.1.5

where the energy levels lie at $E = 0$. It is easy to see from Eq. (6.1.7) that in order for a line to pass through $k_x a = 0$, $k_y b = 2\pi/3$ we must have

$$(m - n)/3 = \nu$$

Since ν is an integer this can only happen if $(m - n)$ is a multiple of three: nanotubes satisfying this condition are metallic.

Zigzag and armchair nanotubes: Consider a specific example: a nanotube with a circumferential vector along the y-direction, $\vec{c} = \hat{y}2bm$, which is called a zigzag nanotube because the edge (after rolling) looks zigzag (Fig. 6.1.5). The periodic boundary condition then requires the allowed values of k to lie parallel to the k_x-axis described by (the circumference is $2bm$)

$$k_y 2bm = 2\pi\nu \rightarrow k_y = \frac{2\pi}{3b}\frac{3\nu}{2m} \tag{6.1.8}$$

as shown in Fig. 6.1.6a.

Figure 6.1.7 shows the two "lowest" subbands corresponding to values of the subband index ν that give rise to the smallest gaps around $E = 0$. If $m = 66$ (i.e. a multiple of three), one of the subbands will pass through $(k_x a, k_y b) = (0, \pm 2\pi/3)$ and the dispersion relation for the lowest subbands look as shown in Fig. 6.1.7a, with no gap in the energy spectrum. But if $m = 65$ (not a multiple of three), then no subband will pass through $(k_x a, k_y b) = (0, \pm 2\pi/3)$ giving rise to a gap in the energy spectrum as shown in Fig. 6.1.7b.

A nanotube with a circumferential vector along the x-direction, $\vec{c} = \hat{x}2am$, is called an armchair nanotube because the edge (after rolling) looks like an armchair as shown in Fig. 6.1.8 (this requires some imagination!). The periodic boundary condition then requires the allowed values of k to lie parallel to the k_y-axis described by (the circumference is again $2bm$)

$$k_x 2am = 2\pi\nu \rightarrow k_x = \frac{2\pi\nu}{2ma} \tag{6.1.9}$$

as shown in Fig. 6.1.6b. The subband with $\nu = 0$ will always pass through the special point $(k_x a, k_y b) = (0, \pm 2\pi/3)$ giving rise to dispersion relations that look metallic (Fig. 6.1.7a) regardless of the value of m.

A useful approximation: Electrical conduction is determined by the states around the Fermi energy and so it is useful to develop an approximate relation that describes the regions of the E–k plot around $E = 0$. This can be done by replacing

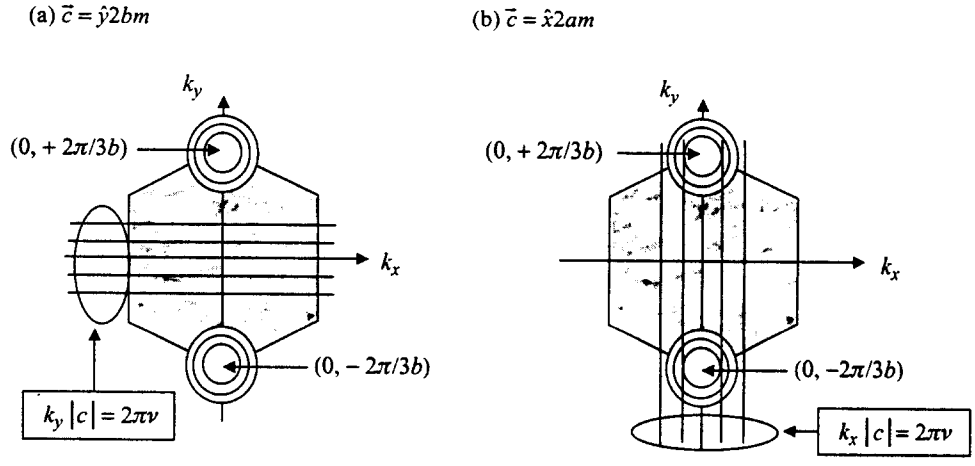


Fig. 6.1.6 (a) A zigzag nanotube obtained by rolling a sheet of graphite along the y-axis with $\vec{c} = \hat{y}2bm$ has its allowed k-values constrained to lie along a set of lines parallel to the k_x-axis as shown. One of the lines will pass through $(0, 2\pi/3b)$ only if m is a multiple of three. (b) An armchair nanotube obtained by rolling a graphite sheet along the x-axis with $\vec{c} = \hat{x}2am$ has its k-values constrained to lie along lines parallel to the k_y-axis as shown. One of the lines will always pass through $(0, 2\pi/3b)$ regardless of the value of m.

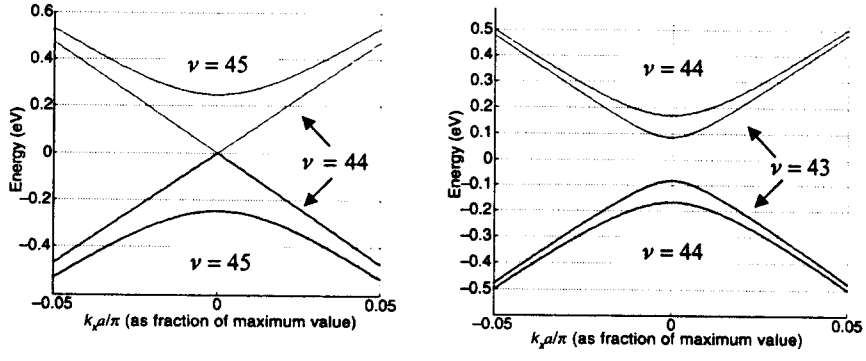


Fig. 6.1.7 Dispersion relation for the two "lowest" subbands of a zigzag nanotube with (a) $D = 5.09$ nm, $m = 66$ showing metallic character (no gap in energy spectrum) and (b) $D = 5.02$ nm, $m = 65$ showing semiconducting character (gap in energy spectrum).

the expression for $h_0(\vec{k}) = -t(1 + 2e^{ik_x a}\cos k_y b)$ with a Taylor expansion around $(k_x a, k_y b) = (0, \pm 2\pi/3)$ where the energy gap is zero (note that $h_0 = 0$ at these points):

$$h_0 \approx k_x \left[\frac{\partial h_0}{\partial k_x}\right]_{k_x a=0,\, k_y b=\pm 2\pi/3} + \left(k_y \mp \frac{2\pi}{3b}\right)\left[\frac{\partial h_0}{\partial k_y}\right]_{k_x a=0,\, k_y b=\pm 2\pi/3}$$

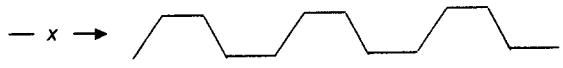


Fig. 6.1.8

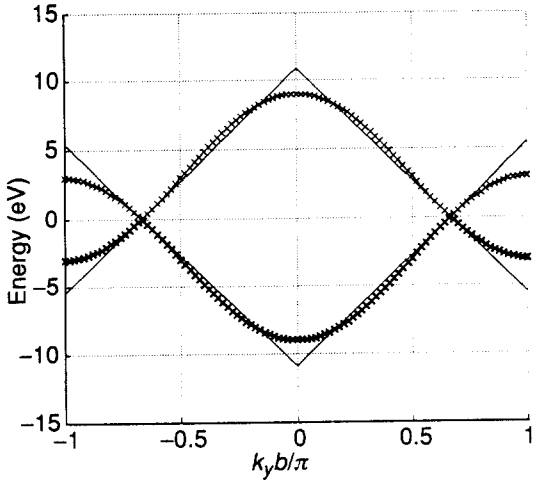


Fig. 6.1.9 Energy dispersion relation plotted as a function of $k_y b$ along the line $k_x a = 0$. The solid curve is obtained from Eq. (6.1.11), while the crosses are obtained from Eq. (6.1.4).

It is straightforward to evaluate the partial derivatives:

$$\frac{\partial h_0}{\partial k_x} = \left[-2iat\, e^{ik_x a} \cos k_y b\right]_{k_x a=0,\ k_y b=\pm 2\pi/3} = iat = i3a_0 t/2$$

$$\frac{\partial h_0}{\partial k_y} = \left[2bt\, e^{ik_x a} \sin k_y b\right]_{k_x a=0,\ k_y b=\pm 2\pi/3} = \pm bt\sqrt{3} = \pm 3a_0 t/2$$

so that we can write

$$h_0(\vec{k}) \approx i\frac{3a_0 t}{2}(k_x \mp i\beta_y)$$

where

$$\beta_y \equiv k_y \mp (2\pi/3b) \tag{6.1.10}$$

The corresponding energy dispersion relation (cf. Eq. (6.1.4)) can be written as

$$E(\vec{k}) = \pm|h_0| = \pm\frac{3ta_0}{2}\sqrt{k_x^2 + \beta_y^2} \tag{6.1.11}$$

This simplified approximate relation (obtained from a Taylor expansion of Eq. (6.1.4) around one of the two valleys) agrees with the exact relation fairly well over a wide range of energies, as is evident from Fig. 6.1.9. Within this approximation the

constant-energy contours are circles isotropically disposed around the center of each valley, $(0, +2\pi/3b)$ or $(0, -2\pi/3b)$.

How large is the energy gap of a semiconducting nanotube? The answer is independent of the specific type of nanotube, as long as $(m - n)$ is not a multiple of three so that the gap is non-zero. But it is easiest to derive an expression for the energy gap if we consider a zigzag nanotube. From Eqs. (6.1.8), (6.1.10), and (6.1.11) we can write

$$E(k_x) = \pm \frac{3ta_0}{2}\sqrt{k_x^2 + \left[\frac{2\pi}{3b}\left(\frac{3\nu}{2m} - 1\right)\right]^2} \tag{6.1.12}$$

so that the energy gap for subband ν can be written as the difference in the energies between the + and − branches at $k_x = 0$:

$$E_{g,\nu} = 3ta_0\frac{2\pi}{2mb}\left(\nu - \frac{2m}{3}\right)$$

This has a minimum value of zero corresponding to $\nu = 2m/3$. But if m is not a multiple of three then the minimum value of $(\nu - 2m/3)$ is equal to $1/3$. This means that the minimum energy gap is then given by

$$E_g = ta_0\frac{2\pi}{2mb} = \frac{2ta_0}{d} \approx \frac{0.8\,\text{eV}}{d} \tag{6.1.13}$$

where d is the diameter of the nanotube in nanometers, so that πd is equal to the circumference $2mb$.

6.2 Density of states

In the last section we discussed how size quantization effects modify the $E(\vec{k})$ relationship leading to the formation of subbands. However, it should be noted that such effects do not appear suddenly as the dimensions of a solid are reduced. It is not as if a bulk solid abruptly changes into a quantum well. The effect of reduced dimensions shows up gradually in experimental measurements and this can be appreciated by looking at the density of states (DOS), $D(E)$, which is reflected in conductance measurements.

The DOS tells us the number of energy eigenstates per unit energy range and it clearly depends on the $E(\vec{k})$ relationship. To be specific, let us assume for the moment that we are near one of the valleys in the conduction band where the energy levels can be described by a parabolic relation with some effective mass m_c:

$$E(\vec{k}) = E_c + \frac{\hbar^2 k^2}{2m_c} \tag{6.2.1}$$

What is the corresponding DOS if the vector $\vec{k}$ is constrained to one dimension (a quantum wire), two dimensions (a quantum well), or three dimensions (bulk solid)? The standard procedure for counting states is to assume a rectangular box of size

$L_x L_y L_z$ with periodic boundary conditions (see Fig. 2.1.4) in all three dimensions (cf. Eq. (2.1.17)):

$$k_x = \frac{2\pi}{L_x}\nu_x \qquad k_y = \frac{2\pi}{L_y}\nu_y \qquad k_z = \frac{2\pi}{L_z}\nu_z \tag{6.2.2}$$

where ν_x, ν_y, and ν_z are integers. We then assume that the box is so large that the allowed k-values are effectively continuous and we can replace any summations over these indices with integrals:

$$\sum_{k_x} \to \int_{-\infty}^{+\infty} \frac{\mathrm{d}k_x}{2\pi/L_x} \qquad \sum_{k_y} \to \int_{-\infty}^{+\infty} \frac{\mathrm{d}k_y}{2\pi/L_y} \qquad \sum_{k_z} \to \int_{-\infty}^{+\infty} \frac{\mathrm{d}k_z}{2\pi/L_z} \tag{6.2.3}$$

In other words, the allowed states in k-space are distributed with a density of $(L/2\pi)$ per unit k in each k-dimension. Hence the total number of allowed states $N(k)$, up to a certain maximum value k, is given by

$$\frac{L}{2\pi}2k = \frac{kL}{\pi} \qquad \text{1D with } L \equiv L_x$$

$$\frac{L_x L_y}{4\pi^2}\pi k^2 = \frac{k^2 S}{4\pi} \qquad \text{2D with } S \equiv L_x L_y$$

$$\frac{L_x L_y L_z}{8\pi^3}\frac{4\pi k^3}{3} = \frac{k^3 \Omega}{6\pi^2} \qquad \text{3D with } \Omega \equiv L_x L_y L_z$$

Using the dispersion relation Eq. (6.2.1) we can convert $N(k)$ into $N(E)$, which tells us the total number of states having an energy less than E. The derivative of this function gives us the density of states (DOS):

$$D(E) = \frac{\mathrm{d}}{\mathrm{d}E}N(E) \tag{6.2.4}$$

The results for one, two, and three dimensions are summarized in Table 6.2.1. It can be seen that a parabolic $E(\vec{k})$ relation (Eq. (6.2.1)) gives rise to a DOS that varies as $E^{-1/2}$ in 1D, E^0 (i.e. constant) in 2D, and $E^{1/2}$ in 3D. Note, however, that the 1D or 2D results do not give us the total DOS for a quantum wire or a quantum well. They give us the DOS *only for one subband* of a quantum wire or a quantum well. We have to sum over all subbands to obtain the full DOS. For example, if a quantum well has subbands p given by (see Eq. (6.1.1))

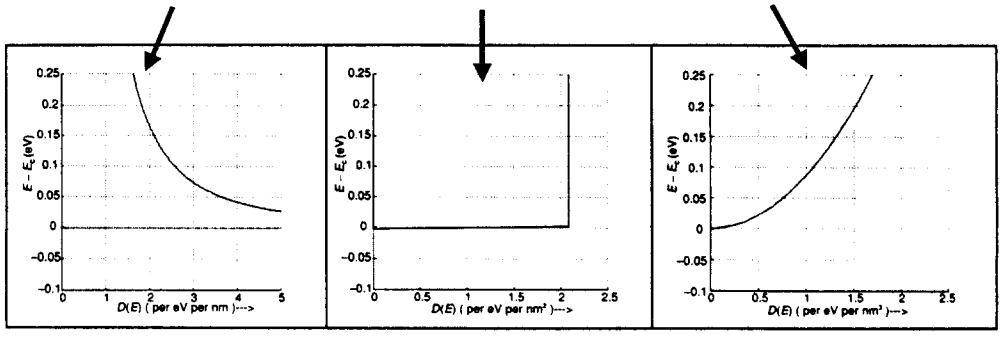
$$E_p(k_x, k_y) \approx E_c + p^2\varepsilon_z + \frac{\hbar^2(k_x^2 + k_y^2)}{2m_c}$$

then the DOS would look like a superposition of many 2D DOS:

$$D(E) = \frac{m_c S}{2\pi\hbar^2}\sum_p \vartheta(E - E_c - p^2\varepsilon_z) \tag{6.2.5}$$

Table 6.2.1 *Summary of DOS calculation in 1D, 2D, and 3D for parabolic isotropic dispersion relation (Eq. (6.2.1)). Plots assume $m_c = m$ (free electron mass) and results are for one spin only*

1D	2D	3D
$N(k) = \frac{L}{2\pi}2k = \frac{kL}{\pi}$	$\frac{S}{4\pi^2}\pi k^2 = \frac{k^2 S}{4\pi}$	$\frac{\Omega}{8\pi^3}\frac{4\pi k^3}{3} = \frac{k^3\Omega}{6\pi^2}$
$N(E) = \frac{L[2m_c(E-E_c)]^{1/2}}{\pi\hbar}$	$\frac{S\,2m_c(E-E_c)}{4\pi\hbar^2}$	$\frac{\Omega[2m_c(E-E_c)]^{3/2}}{6\pi^2\hbar^3}$
$D(E) = \frac{m_c L}{\pi\hbar}\left(\frac{1}{2m_c(E-E_c)}\right)^{1/2}$	$\frac{S m_c}{2\pi\hbar^2}$	$\frac{\Omega m_c}{2\pi^2\hbar^3}[2m_c(E-E_c)]^{1/2}$



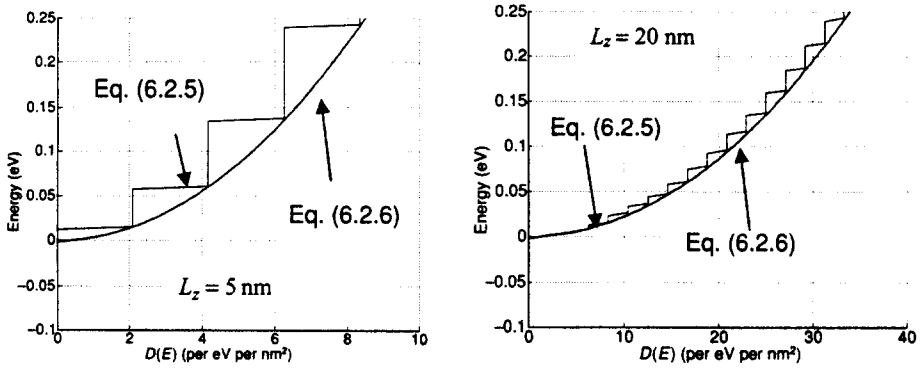


Fig. 6.2.1 Density of states $D(E)$ for a 2D box calculated from Eq. (6.2.5) (2D with quantized subbands) and compared with that obtained from the 3D relation in Eq. (6.2.6). As the width L_z of the box is increased from 5 to 20 nm, the DOS approaches the 3D result (the conduction band effective mass m_c is assumed equal to the free electron mass m).

Figure 6.2.1 shows the density of states calculated from Eq. (6.2.5) with $\varepsilon_z = \pi^2\hbar^2/2m_c L_z^2$ as given in Eq. (6.1.2). It is apparent that as the width L_z is increased from 5 to 20 nm, the DOS approaches the result obtained from the 3D DOS with the volume Ω set equal to SL_z:

$$D_{3D}(E) = \frac{Sm_c}{2\pi^2\hbar^3}[2m_c(E-E_c)]^{1/2}L_z \tag{6.2.6}$$

Table 6.2.2 *Summary of DOS per spin per valley*

	Graphite Eq. (6.2.7)	Zigzag nanotube Eq. (6.2.8)
$N(k) =$	$\frac{S}{4\pi}k^2$	$\frac{L}{\pi}k_x$
$N(E) =$	$\frac{SE^2}{4\pi a^2t^2}$	$\sum_\nu \frac{L}{\pi at}\sqrt{E^2-E_\nu^2}$
$D(E) =$	$\frac{SE}{2\pi a^2t^2}$	$\sum_\nu \frac{L}{\pi at}\frac{E}{\sqrt{E^2-E_\nu^2}}$

From graphite to a nanotube : The evolution of the DOS from a sheet of graphite to a nanotube also provides an instructive example of how size quantization effects arise as the dimensions are reduced. For a sheet of graphite we can approximate the $E(\vec{k})$ relation for each of the two valleys centered at $(k_x a, k_y b) = (0, \pm 2\pi/3)$ as (see Eq. (6.1.11), $a = 3a_0/2$)

$$E(\vec{k}) = \pm\frac{3ta_0}{2}\sqrt{k_x^2+\beta_y^2} = \pm ta|\vec{k}| \tag{6.2.7}$$

As we have seen, the energy subbands for a zigzag nanotube are given by (see Eq. (6.1.12))

$$E(k_x) = \pm ta\sqrt{k_\nu^2+k_x^2} = \pm\sqrt{E_\nu^2+(tak_x)^2} \tag{6.2.8}$$

where

$$k_\nu \equiv \frac{2\pi}{3b}\left(\frac{3\nu}{2m}-1\right) \quad \text{and} \quad E_\nu \equiv tak_\nu$$

In calculating $D(E)$ the steps are similar to those shown in Table 6.2.1, though the details are somewhat different because the dispersion relation is different (Eqs. (6.2.7), (6.2.8)), as summarized in Table 6.2.2. Figure 6.2.2 compares the density of states for a zigzag nanotube of length L and diameter d (note: circumference $= \pi d = 2mb$):

$$D_{\text{ZNT}}(E) = \sum_\nu \frac{2L}{\pi at}\frac{E}{\sqrt{E^2-E_\nu^2}} \quad \text{with} \quad E_\nu \equiv \frac{2at}{d}\left(\nu-\frac{2m}{3}\right) \tag{6.2.9}$$

with the density of states for a graphite sheet of area πLd:

$$D_{\text{G}}(E) = \frac{Ld}{a^2t^2}|E| \tag{6.2.10}$$

for a nanotube with $m = 200$ (corresponding to $d = 15.4$ nm) and a nanotube with $m = 800$ ($d = 61.7$ nm).

It is apparent that the smaller nanotube has a DOS that is distinctly different from graphite. But the larger nanotube is less distinguishable, especially if we recall that experimental observations are typically convolved with the thermal broadening function

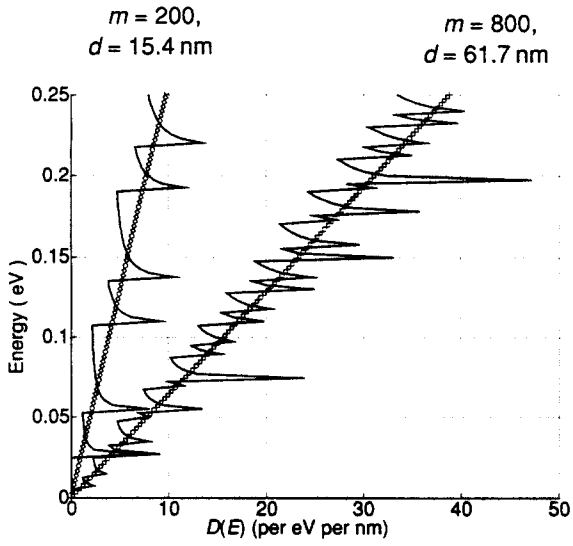


Fig. 6.2.2 Density of states $D(E)$ for a zigzag nanotube calculated from Eq. (6.2.9) (solid curves) compared with that obtained from the result for graphite (Eq. (6.2.10), crosses).

which has a width of $\sim k_B T \sim 0.026$ eV at room temperature (to be discussed later; see Fig. 7.3.4).

It is easy to see analytically that the DOS for zigzag nanotubes with the summation index ν replaced by an integral (cf. Eq. (6.2.9))

$$D_{ZNT}(E) \approx \int 2\mathrm{d}\nu \frac{2L}{\pi at} \frac{|E|}{\sqrt{E^2 - E_\nu^2}} \quad \text{with} \quad \mathrm{d}E_\nu \equiv \frac{2at}{d}\,\mathrm{d}\nu$$

$$= \int_0^E \mathrm{d}E_\nu \frac{2Ld}{\pi a^2 t^2} \frac{|E|}{\sqrt{E^2 - E_\nu^2}} = \frac{Ld}{a^2 t^2}|E|$$

becomes identical to the DOS for graphite (cf. Eq. (6.2.10)).

Anisotropic dispersion relation: We have used a few examples to illustrate the procedure for converting an $E(\vec{k})$ relation to a density of states $D(E)$. This procedure (see Tables 6.2.1 and 6.2.2) works well when the dispersion relation $E(\vec{k})$ is isotropic. However, the details become more complicated if the relation is anisotropic. For example, silicon has six separate conduction band valleys, each of which is *ellipsoidal*:

$$E = E_c + \frac{\hbar^2 k_x^2}{2m_{xx}} + \frac{\hbar^2 k_y^2}{2m_{yy}} + \frac{\hbar^2 k_z^2}{2m_{zz}} \tag{6.2.11}$$

For a given energy E, the constant energy contour (Fig. 6.2.3) looks like an ellipsoid whose major axes are given by $\sqrt{2m_{xx}(E - E_c)}/\hbar$, $\sqrt{2m_{yy}(E - E_c)}/\hbar$,

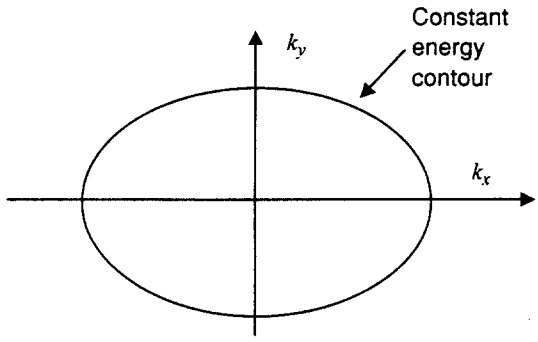


Fig. 6.2.3

and $\sqrt{2m_{zz}(E-E_c)}/\hbar$. The volume of this ellipsoid is

$$\frac{4\pi}{3}\frac{\sqrt{2m_{xx}(E-E_c)}}{\hbar}\frac{\sqrt{2m_{yy}(E-E_c)}}{\hbar}\frac{\sqrt{2m_{zz}(E-E_c)}}{\hbar}$$

so that

$$N(E)=\frac{\Omega}{8\pi^3}\frac{4\pi}{3}\frac{\sqrt{2m_{xx}(E-E_c)}}{\hbar}\frac{\sqrt{2m_{yy}(E-E_c)}}{\hbar}\frac{\sqrt{2m_{zz}(E-E_c)}}{\hbar}$$

and

$$D(E)\equiv\frac{\mathrm{d}}{\mathrm{d}E}N(E)=\frac{\Omega}{2\pi^2\hbar^3}\sqrt{2m_{xx}m_{yy}m_{zz}(E-E_c)} \tag{6.2.12}$$

which reduces to the earlier result (see Table 6.2.1) if the mass is isotropic.

Formal expression for DOS: In general if we have a system with eigenstates labeled by an index α, then we can write the total number of states, $N_T(E)$ with energy less than E as

$$N_T(E)=\sum_\alpha \vartheta(E-\varepsilon_\alpha) \tag{6.2.13}$$

where $\vartheta(E)$ denotes the unit step function which is equal to zero for $E<0$, and equal to one for $E>0$. The derivative of the unit step function is a delta function, so that

$$D(E)\equiv\frac{\mathrm{d}}{\mathrm{d}E}N_T(E)=\sum_\alpha \delta(E-\varepsilon_\alpha) \tag{6.2.14}$$

This expression represents a sequence of spikes rather than a continuous function. Formally we can obtain a continuous DOS from Eq. (6.2.14) by letting the size of the system get very large and replacing the summation by an integral as we have been doing. For example, if

$$E(\vec{k})=E_c+\frac{\hbar^2k^2}{2m^*}$$

$$\text{then } D(E) = \sum_{\vec{k}} \delta(E - \varepsilon_{\vec{k}})$$

$$= \sum_{k_x} \sum_{k_y} \sum_{k_z} \delta\left(E - E_c - \frac{\hbar^2(k_x^2 + k_y^2 + k_z^2)}{2m^*}\right)$$

where

$$k_x = \frac{2\pi}{L_x} \nu_x \qquad k_y = \frac{2\pi}{L_y} \nu_y \qquad k_z = \frac{2\pi}{L_z} \nu_z$$

and ν_x, ν_y, and ν_z are integers. We then let the volume get very large and replace the summations over these indices with integrals:

$$\sum_{k_x} \to \int_{-\infty}^{+\infty} \frac{dk_x}{2\pi/L_x} \qquad \sum_{k_y} \to \int_{-\infty}^{+\infty} \frac{dk_y}{2\pi/L_y} \qquad \sum_{k_z} \to \int_{-\infty}^{+\infty} \frac{dk_z}{2\pi/L_z}$$

$$D(E) = \frac{\Omega}{8\pi^3} \int_0^{+\infty} dk_x\, dk_y\, dk_z\, \delta\left(E - E_c - \frac{\hbar^2 k^2}{2m^*}\right)$$

$$= \frac{\Omega}{8\pi^3} \int_0^{+\infty} 4\pi k^2\, dk\, \delta\left(E - E_c - \frac{\hbar^2 k^2}{2m^*}\right)$$

$$= \frac{\Omega}{2\pi^2} \int_0^{+\infty} \frac{m^* dE}{\hbar^2} \frac{\sqrt{2m^*(E - E_c)}}{\hbar} \delta\left(E - E_c - \frac{\hbar^2 k^2}{2m^*}\right)$$

$$= \frac{\Omega m^*}{2\pi^2 \hbar^3} [2m^*(E - E_c)]^{1/2}$$

as we obtained earlier (cf. 3D DOS in Table 6.2.1). This procedure is mathematically a little more involved than the previous procedure and requires integrals over delta functions.

The real value of the formal expression in Eq. (6.2.14) is that it is generally valid regardless of the $E(\vec{k})$ relation or whether such a relation even exists. Of course, in general it may not be easy to replace the summation by an integral, since the separation between energy levels may not follow a simple analytical prescription like Eq. (6.2.2). But we can still calculate a continuous DOS from Eq. (6.2.14) by broadening each spike into a Lorentzian (see Eq. (1.3.2)):

$$\delta(E - \varepsilon_\alpha) \to \frac{\gamma/2\pi}{(E - \varepsilon_\alpha)^2 + (\gamma/2)^2} \tag{6.2.15}$$

The DOS will look continuous if the individual energies ε_α are spaced closely compared to the broadening γ, which is usually true for large systems.

Separable problems: An interesting result that can be proved using Eq. (6.2.14) is that if the eigenvalue problem is separable, then the overall DOS is given by the convolution of the individual densities. For example, suppose we have a 2D problem that separates into x- and y-components (as shown in Eq. (2.3.6) for 3D problems) such that the overall energies are given by the sum of the x-energy and the y-energy:

$$\varepsilon(n, m) = \varepsilon_x(n) + \varepsilon_y(m) \tag{6.2.16}$$

We could define an x-DOS and a y-DOS:

$$D_x(E) = \sum_n \delta[E - \varepsilon_x(n)] \tag{6.2.17a}$$

$$D_y(E) = \sum_m \delta[E - \varepsilon_y(m)] \tag{6.2.17b}$$

and it is straightforward to show that the total DOS

$$D(E) = \sum_n \sum_m \delta[E - \varepsilon_x(n) - \varepsilon_y(m)] \tag{6.2.17c}$$

can be written as a convolution product of the x-DOS and the y-DOS:

$$D(E) = \int_{-\infty}^{+\infty} dE' D_x(E')\, D_y(E - E') \tag{6.2.18}$$

6.3 Minimum resistance of a wire

Now that we have discussed the concept of subbands, we are ready to answer a very interesting fundamental question. Consider a wire of cross-section $L_y L_z$ with a voltage V applied across it (Fig. 6.3.1). What is the conductance of this wire if the contacts were perfect, and we reduce its length to very small dimensions? Based on Ohm's law, we might be tempted to say that the conductance will increase indefinitely as the length of the wire is reduced, since the resistance (inverse of conductance) is proportional to the length. However, as I pointed out in Section 1.3 the maximum conductance $G = I/V$ for a *one-level conductor* is a fundamental constant given by

$$G_0 \equiv q^2/h = 38.7\ \mu\text{S} = (25.8\ \text{k}\Omega)^{-1} \tag{6.3.1}$$

We are now ready to generalize this concept to a wire with a finite cross-section. It has been established experimentally that once the length of a wire has been reduced sufficiently that an electron can cross the wire without an appreciable chance of scattering (ballistic transport) the conductance will approach a constant value given by (assuming "perfect" contacts)

$$G = [M(E)]_{E=E_F}\, G_0 \tag{6.3.2}$$

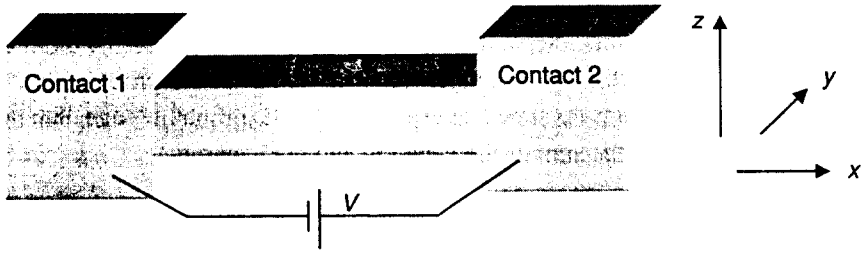


Fig. 6.3.1 A wire of cross-section L_yL_z with a voltage V applied across large contact.

where $M(E)$ is the number of "modes" or subbands at an energy E. The actual number of modes $M(E)$ at a given energy depends on the details of the wire, but the maximum conductance per mode is G_0 independent of these details. This can be shown as follows.

Maximum current in a single-moded wire: Consider a mode or subband ν with a dispersion relation $E_\nu(k_x)$. The current carried by the states having a positive group velocity can be written as

$$I = \frac{-q}{L} \sum_{v_x(k_x)>0} v_x(k_x)$$

We have seen earlier that for free electrons with $E = \hbar^2k^2/2m$, the velocity is given by $\hbar k/m$, which is equal to the momentum $\hbar k$ divided by the mass m. But what is the appropriate velocity for electrons in a periodic solid having some dispersion relation $E_\nu(\vec{k})$? The answer requires careful discussion which we will postpone for the moment (see Section 6.4) and simply state that the correct velocity is the group velocity generally defined as the gradient of the dispersion relation

$$\hbar\vec{v}(\vec{k}) = \vec{\nabla}_k E_\nu(\vec{k})$$

so that in our 1D example we can write $v_x(k_x) = \partial E_\nu(k_x)/\partial k_x$ and the current is given by

$$I = \frac{-q}{L} \sum_{v(k_x)>0} \frac{1}{\hbar} \frac{\partial E_\nu(k_x)}{\partial k_x}$$

$$= -q \int \frac{dk_x}{2\pi} \frac{1}{\hbar} \frac{\partial E_\nu(k_x)}{\partial k_x} = \frac{-q}{h} \int dE_\nu \qquad (6.3.3)$$

showing that each mode of a wire carries a current of (q/h) per unit energy. At equilibrium there is no net current because states with positive and negative velocities are all equally occupied. An applied voltage V changes the occupation of the levels over an energy range $E_F \pm (qV/2)$ creating a non-equilibrium situation whose details depend on the nature of the coupling to the contacts. But regardless of all these details, it is

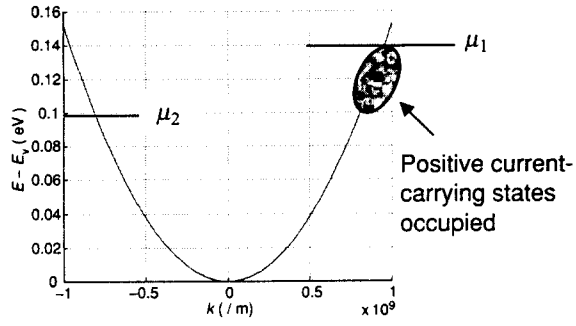


Fig. 6.3.2

apparent that the maximum net current will be established if the positive velocity states are occupied up to $\mu_1 = E_F + (qV/2)$ while the negative velocity states are occupied up to $\mu_2 = E_F - (qV/2)$, so that in the energy range

$$E_F - (qV/2) < E < E_F + (qV/2)$$

only the positive velocity states are occupied (Fig. 6.3.2). From Eq. (6.3.3) we can write the current carried by these states belonging to mode ν as

$$I = \frac{-q}{h}(\mu_1 - \mu_2) \rightarrow \frac{q^2}{h}V$$

so that the maximum conductance of mode ν is equal to q^2/h as stated earlier (see Eq. (6.3.1)). Note that this result is independent of the dispersion relation $E_\nu(k_x)$ for the mode ν.

Number of modes: How many modes $M(E)$ we actually have at a given energy, however, is very dependent on the details of the problem at hand. For example, if the relevant energy range involves the bottom of the conduction band then we may be able to approximate the band diagram with a parabolic relation (see Fig. 6.1.2, Eq. (6.1.1)). The subbands can then be catalogued with two indices (n, p) as shown in Fig. 6.3.3

$$E_{n,p}(k_x) \approx E_c + n^2\varepsilon_y + p^2\varepsilon_z + \frac{\hbar^2 k_x^2}{2m_c} \tag{6.3.4}$$

where $\varepsilon_y = \pi^2\hbar^2/2m_cL_y^2$ and $\varepsilon_z = \pi^2\hbar^2/2m_cL_z^2$, assuming that the electrons are confined in the wire by infinite potential wells of width L_y and L_z in the y- and z-directions respectively. The mode density $M(E)$ then looks as shown in Fig. 6.3.2, increasing with energy in increments of one every time a new subband becomes available.

The details of the subbands in the valence band are much more complicated, because of the multiple bands that are coupled together giving rise to complicated dispersion

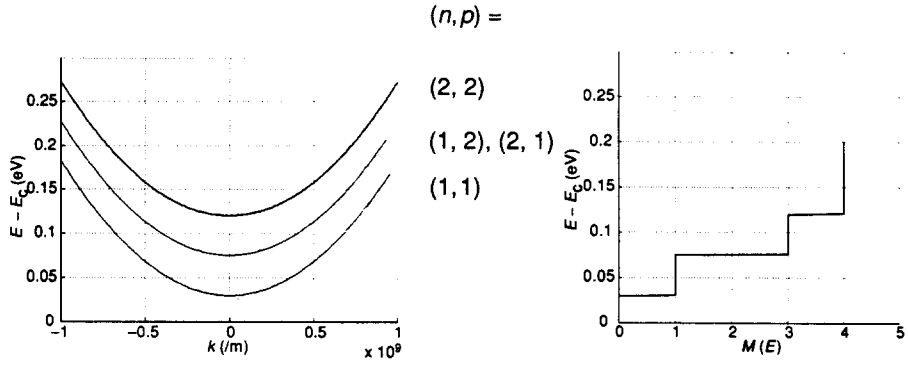


Fig. 6.3.3 Energy dispersion relation showing the four lowest conduction band subbands (see Eq. (6.3.4)) in a rectangular wire with $L_y = L_z = 10\,\text{nm}$, $m_c = 0.25m$.

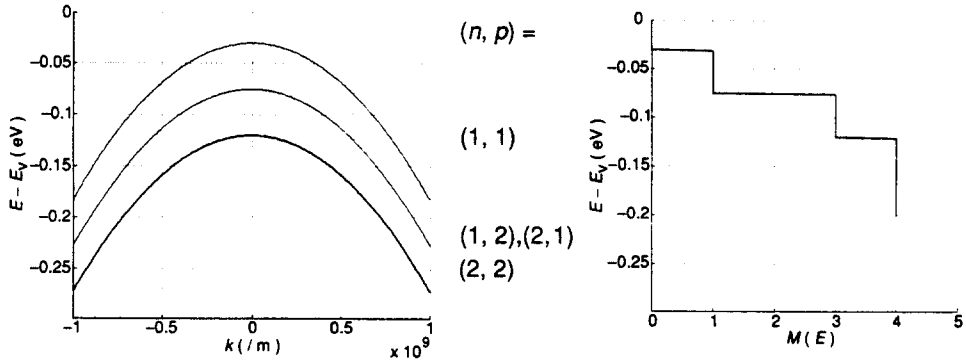


Fig. 6.3.4 Energy dispersion relation showing the four lowest valence band subbands (see Eqs. (6.3.5) and (6.3.6)) in a rectangular wire with $L_y = L_z = 10\,\text{nm}$, $m_v = 0.25m$.

relations. For simple back-of-an-envelope estimates we could approximate with a simple *inverted* parabola

$$h(\vec{k}) = E_v - \frac{\hbar^2 k^2}{2m_v} \tag{6.3.5}$$

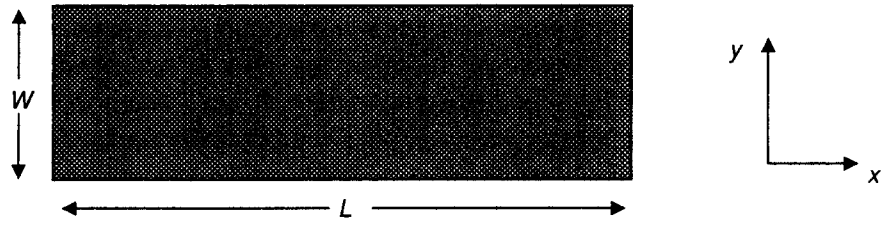
We then get inverted dispersion relations for the different subbands

$$E_{n,p}(k_x) \approx E_v - n^2\varepsilon_y - p^2\varepsilon_z - \frac{\hbar^2 k_x^2}{2m_v} \tag{6.3.6}$$

with a mode density $M(E)$ that increases with decreasing energy as shown in Fig. 6.3.4.

Minimum resistance of a wide conductor: We have see that there is a minimum resistance (h/q^2) that a wire can have per mode. An interesting consequence of this is that there is a minimum resistance that a given conductor could have. For example, we could model a field effect transistor (FET) as a two-dimensional conductor with

a width W and a length L as shown in the figure below. As the length L is reduced, it would approach a ballistic conductor whose resistance is dominated by the contact resistance. This is a well-known fact and device engineers work very hard to come up with contacting procedures that minimize the contact resistance. What is not as well-recognized is that there is a fundamental limit to how small the contact resistance can possibly be – even with the best conceivable contacts.



To estimate this minimum contact resistance, let us assume an n-type semiconductor where conduction takes place through the conduction band states described by a parabolic dispersion relation of the form given in Eq. (6.3.4). Assuming that it has only one subband along the z-direction we can write the electron density per unit area at $T = 0$ K as (see Table 6.2.1)

$$n_s = \frac{m_c}{\pi\hbar^2}(E_F - E_c) \tag{6.3.7}$$

The maximum conductance is given by

$$G_{max} = \frac{2q^2}{h}\,\mathrm{Int}\sqrt{\frac{E_F - E_c}{\hbar^2\pi^2/2m_cW^2}} \approx W\,\frac{2q^2}{h}\sqrt{\frac{2n_s}{\pi}}$$

where we have made use of Eq. (6.3.7) and used the symbol Int(x) to denote the largest integer smaller than x. The minimum resistance is given by the inverse of G_{max}:

$$R_{min}W = \frac{h}{2q^2}\sqrt{\frac{\pi}{2n_s}} = \frac{16.28\,\mathrm{k\Omega}}{\sqrt{n_s}} \tag{6.3.8}$$

With a carrier density of $n_s = 10^{12}/\mathrm{cm}^2$, this predicts $R_{min}W \approx 160\ \Omega\,\mu\mathrm{m}$. Note that we have assumed only one subband arising from the z-confinement. For silicon, the six conduction valleys give rise to six sets of subbands, of which the lowest two are degenerate. This means that if the L_z-dimension is small enough, there will be two degenerate subbands arising from the z-confinement and the corresponding minimum contact resistance will be half our estimate which was based on one z-subband: $R_{min}W \approx 80\ \Omega\,\mu\mathrm{m}$.