17

Surface and Interface Physics

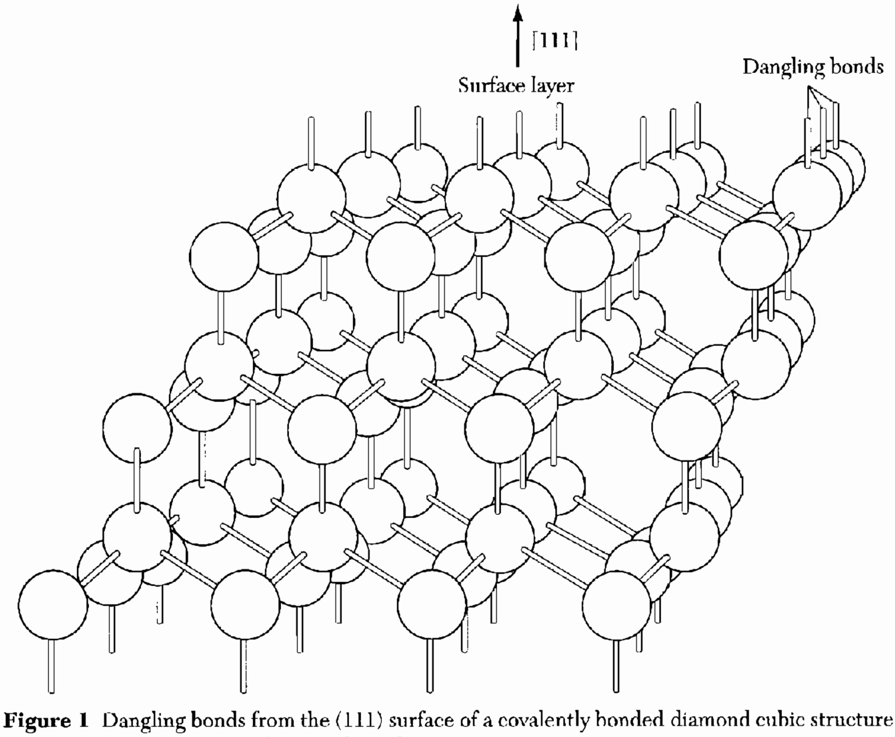


Figure 1 Dangling bonds from the (111) surface of a covalently bonded diamond cubic structure. (After M. Prutton, *Surface physics*, Clarendon, 1975.)

Reconstruction and Relaxation

The **surface** of a crystalline solid in vacuum is generally defined as the few, approximately three, outermost atomic layers of the solid that differ significantly from the bulk. The surface may be entirely clean or it may have foreign atoms deposited on it or incorporated in it. The bulk of the crystal is called the **substrate**.

If the surface is clean the top layer may be either **reconstructed** or, sometimes, unreconstructed. In unreconstructed surfaces the atomic arrangement is in registry with that of the bulk except for an interlayer spacing change (called multilayer relaxation) at the top surface.

The shrinking of the interlayer distance between the first and second layer of atoms with respect to subsequent layers in the bulk is a rather dominant phenomenon. The surface may be thought of as an intermediate between a diatomic molecule and the bulk structure. Because the interatomic distances in diatomic molecules are much smaller than in the bulk, there is a rationale for the surface relaxation. This may be contrasted with reconstruction where the relaxation of atoms yields new surface primitive cells. In relaxation the atoms maintain their structure in the surface plane as it was (according to the projection of the bulk cell on the surface); only their distance from the bulk changes.

Sometimes in metals, but most often in nonmetals, the atoms in the surface layer form superstructures in which the atoms in the layer are not in registry with the atoms in corresponding layers in the substrate. This is surface reconstruction; it can be a consequence of a rearrangement of broken covalent or ionic bonds at the surface. Under such conditions the atoms at the surface bunch into rows with alternately larger and smaller spacings than in the bulk. That is, for some crystals held together by valence bonds, creation of a surface would leave unsaturated bonds dangling into space (Fig. 1). The energy may then be lowered if neighboring atoms approach each other and form bonds with their otherwise unused valence electrons. Atomic displacements can be as large as 0.5 Å.

Reconstruction does not necessarily require formation of a superstructure. For example, on GaAs (110) surfaces a rotation of the Ga-As bond leaves the point group intact. The driving force is electron transfer from Ga to As, which fills the dangling bonds on As and depletes them on Ga.

Surfaces of planes nominally of high indices may be built up of low index planes separated by steps one (or two) atoms in height. Such terrace-step arrangements are important in evaporation and desorption because the attachment energy of atoms is often low at the steps and at kinks in the steps. The chemical activity of such sites may be high. The presence of periodic arrays of steps may be detected by double and triple beams of diffraction in **LEED** (see below) experiments.

SURFACE CRYSTALLOGRAPHY

The surface structure is in general periodic only in two dimensions. The surface structure can be the structure of foreign material deposited on the substrate or it can be the selvage of the pure substrate. In Chapter 1 we used the term Bravais lattice for the array of equivalent points in two or in three dimensions, that is, for diperiodic or triperiodic structures. In the physics of surfaces it is common to speak of a two-dimensional **lattice**. Further, the area unit may be called a **mesh**.

We showed in Fig. 1.7 four of the five nets possible for a diperiodic structure; the fifth net is the general oblique net, with no special symmetry relation between the mesh basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$. Thus the five distinct nets are the oblique, square, hexagonal, rectangular, and centered rectangular.

The substrate net parallel to the surface is used as the reference net for the description of the surface. For example, if the surface of a cubic substrate crystal is the (111) surface, the substrate net is hexagonal (Fig. 1.7b), and the surface net is referred to these axes.

The vectors $\mathbf{c}_1$, $\mathbf{c}_2$ that define the mesh of the surface structure may be expressed in terms of the reference net $\mathbf{a}_1$, $\mathbf{a}_2$ by a matrix operation $\mathbf{P}$:

$$\begin{pmatrix}\mathbf{c}_1\\ \mathbf{c}_2\end{pmatrix} = \mathbf{P}\begin{pmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{pmatrix} = \begin{pmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{pmatrix}\begin{pmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{pmatrix} . \tag{1}$$

Provided that the included angles of the two meshes are equal, the shorthand notation due to E. A. Wood may be used. In this notation, which is widely used, the relation of the mesh $\mathbf{c}_1$, $\mathbf{c}_2$ to the reference mesh $\mathbf{a}_1$, $\mathbf{a}_2$ is expressed as

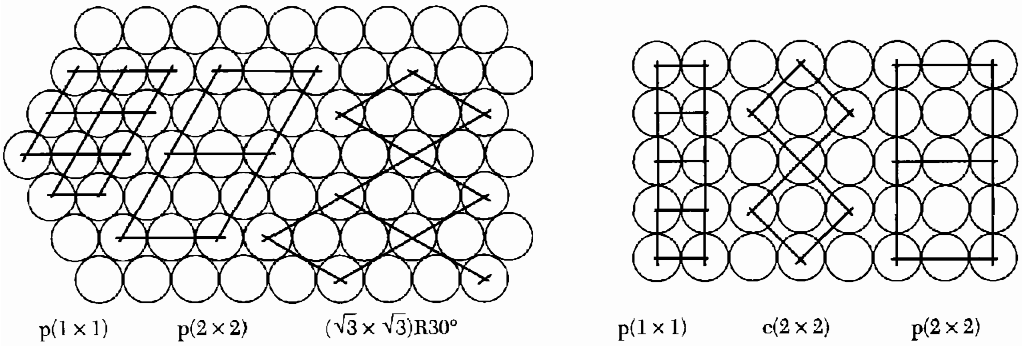
$$\left(\frac{c_1}{a_1} \times \frac{c_2}{a_2}\right) R\alpha \ , \tag{2}$$

in terms of the lengths of the mesh basis vectors and the angle α of relative rotation R of the two meshes. If $\alpha = 0$, the angle is omitted. Examples of the Wood notation are given in Fig. 2.

The reciprocal net vectors of the surface mesh may be written as $\mathbf{c}_1^*$, $\mathbf{c}_2^*$, defined by

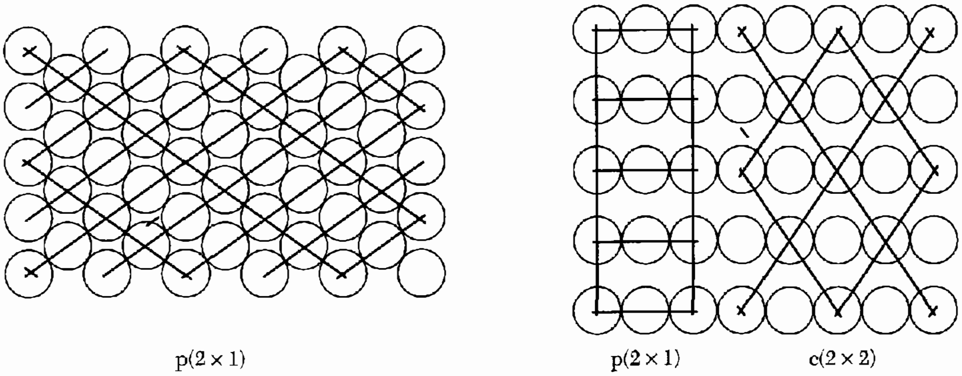
$$\mathbf{c}_1 \cdot \mathbf{c}_2^* = \mathbf{c}_2 \cdot \mathbf{c}_1^* = 0 \ ; \qquad \mathbf{c}_1 \cdot \mathbf{c}_1^* = \mathbf{c}_2 \cdot \mathbf{c}_2^* = 2\pi \text{ (or 1)} \ . \tag{3}$$

Here the 2π (or 1) indicates that two conventions are in use. The definitions (3) used in Fig. 3 may be compared with the definitions in Chapter 2 of the reciprocal lattice vectors of a triperiodic lattice.



fcc(111), hcp(0001)
(a)

fcc(100), bcc(100)
(b)



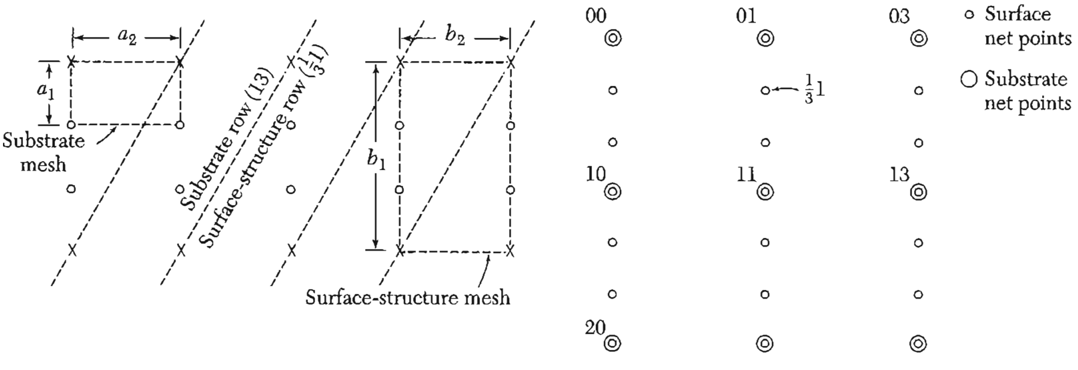
bcc(110)
(c)

fcc(110)
(d)

Figure 2 Surface nets of adsorbed atoms. The circles represent atoms in the top layer of the substrate. In (a) the designation fcc(111) means the (111) face of an fcc structure. This face determines a reference net. The lines represent ordered overlayers, with adatoms at the intersections of two lines. The intersection points represent diperiodic nets (lattices in two dimensions). The designation p(1 × 1) in (a) *is a primitive mesh unit for which the basis is identical with the basis of* the reference net. In (b) the c(2 × 2) mesh unit is a centered mesh with basis vectors twice as long as those of the reference net. Atomic adsorption on metals takes place most often into those surface sites (hollow sites) that maximize the number of nearest-neighbor atoms on the substrate. (After Van Hove.)

The reciprocal net points of a diperiodic net may be thought of—when we are in three dimensions—as rods. The rods are infinite in extent and normal to the surface plane, where they pass through the reciprocal net points. It may be helpful to think of the rods as generated by a triperiodic lattice which is expanded without limit along one of its axes. Then the reciprocal lattice points along this axis are moved closer together and in the limit form a rod.

The usefulness of the rod concept comes out with the Ewald sphere construction explained in Fig. 2.8. Diffraction occurs everywhere the Ewald



(a) (b)

Figure 3 A (3 × 1) surface structure, (a) real-space; and (b) reciprocal-space diagrams. (After E. A. Wood.)

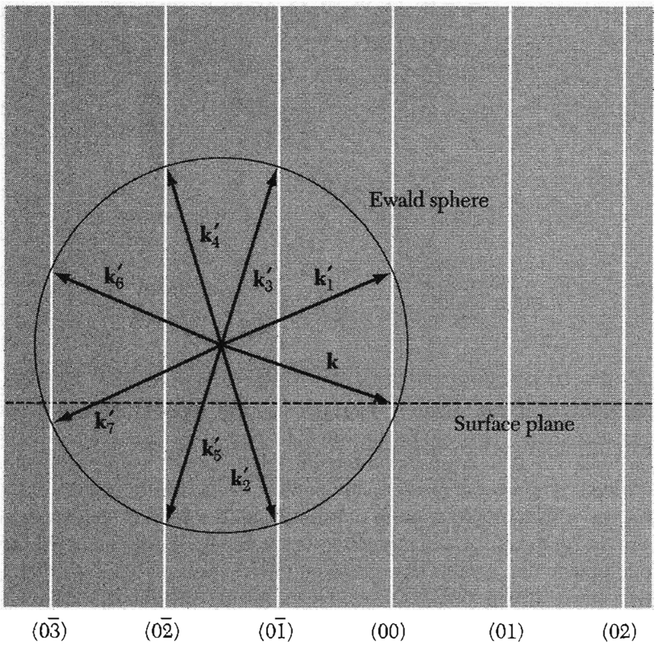


Figure 4 Ewald sphere construction for diffraction of incident wave **k** by a square net, when **k** is parallel to one axis of the mesh. The back scattered beams in the plane of the paper are $\mathbf{k}_4'$, $\mathbf{k}_5'$, $\mathbf{k}_6'$, $\mathbf{k}_7'$. Diffracted beams out of the plane of the paper will also occur. The vertical lines are the rods of the reciprocal net.

sphere intercepts a reciprocal net rod. Each diffracted beam is labelled with the indices *hk* of the reciprocal net vector

$$g = h\mathbf{c}_1^* + k\mathbf{c}_2^* \tag{4}$$

forming the beam.

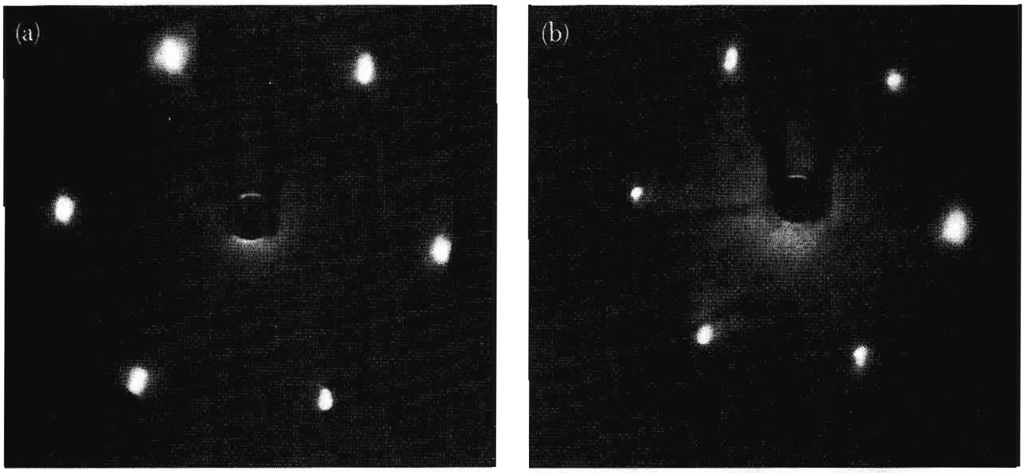


Figure 5 LEED patterns from a Pt(111) crystal surface for incident electron energies of 51 and 63.5 eV. The diffraction angle is greater at the lower energy. (After G. A. Somorjai, *Chemistry in two dimensions: surfaces*, Cornell, 1981.)

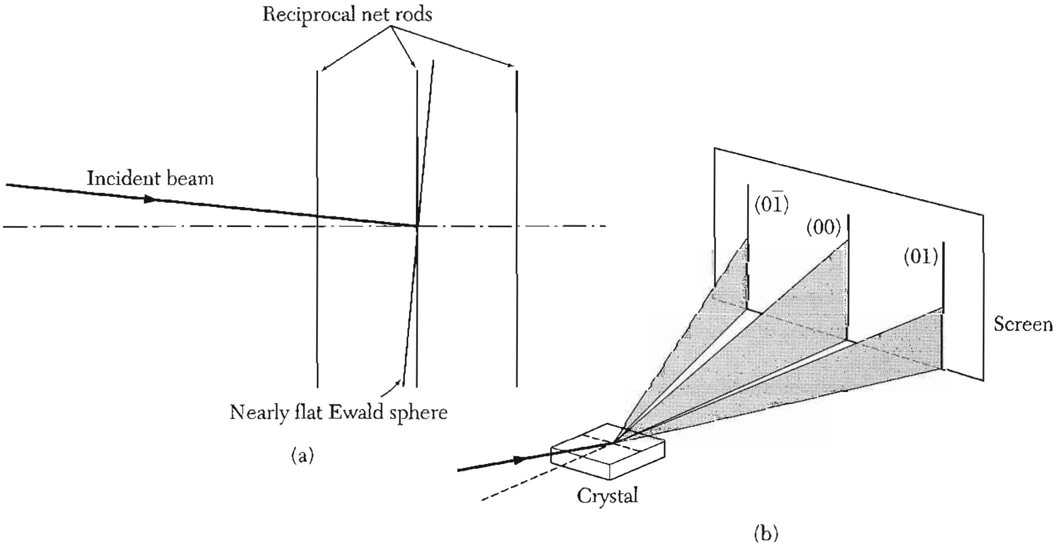


Figure 6 The RHEED method. In (a) the high-energy incident electron beam at a glancing angle to the crystal surface is associated with an Ewald sphere of large radius, so large that the surface is nearly flat in relation to the separation between adjacent rods of the reciprocal net. The formation of diffraction lines on a plane screen is shown in (b). (After Prutton.)

Low energy electron diffraction (LEED) is illustrated by Fig. 4. The electron energy is typically in the range 10–1000 eV. With this arrangement Davisson and Germer in 1927 discovered the wave nature of the electron. An experimental pattern is shown in Fig. 5.

Reflection High-Energy Electron Diffraction. In the RHEED method a beam of high-energy electrons is directed upon a crystal surface at grazing incidence. By adjustment of the angle of incidence one can arrange the normal component of the incoming wavevector to be very small, which will minimize the penetration of the electron beam and enhance the role of the crystal surface.

The radius k of the Ewald sphere for 100 keV electrons will be $\simeq 10^3$ Å^{-1}, which is much longer than the shortest reciprocal lattice vector $2\pi/a \approx 1$ Å^{-1}. It follows that the Ewald sphere will be nearly a flat surface in the central scattering region. The intercept of the rods of the reciprocal net with the nearly flat sphere will be nearly a line when the beam is directed at grazing incidence. The experimental arrangement is shown in Fig. 6.

SURFACE ELECTRONIC STRUCTURE

Work Function

The work function W of the uniform surface of a metal is defined as the difference in potential energy of an electron between the vacuum level and the Fermi level. The vacuum level is the energy of an electron at rest at a point sufficiently far outside the surface so that the electrostatic image force on the electron may be neglected—more than 100 Å from the surface. The Fermi level is the electrochemical potential of the electrons in the metal.

Typical values of electron work functions are given in Table 1. The orientation of the exposed crystal face affects the value of the work function

Table 1 Electron work functions[a]

(Values obtained by photoemission, except tungsten obtained by field emission.)

Element	Surface plane	Work function, in eV
Ag	(100)	4.64
	(110)	4.52
	(111)	4.74
Cs	polycrystal	2.14
Cu	(100)	4.59
	(110)	4.48
	(111)	4.98
Ge	(111)	4.80
Ni	(100)	5.22
	(110)	5.04
	(111)	5.35
W	(100)	4.63
	(110)	5.25
	(111)	4.47

[a] After H. D. Hagstrum.

because the strength of the electric double layer at the surface depends on the concentration of surface positive ion cores. The double layer exists because the surface ions are in an asymmetrical environment, with vacuum (or an adsorbed foreign atom layer) on one side and the substrate on the other side.

The work function is equal to the threshold energy for photoelectric emission at absolute zero. If $\hbar\omega$ is the energy of an incident photon, then the Einstein equation is $\hbar\omega = W + T$, where T is the kinetic energy of the emitted electron and W is the work function.

Thermionic Emission

The rate of emission of thermionic electrons depends exponentially on the work function. The derivation follows.

We first find the electron concentration in vacuum in equilibrium with electrons in a metal at temperature $\tau(=k_BT)$ and chemical potential μ. We treat the electrons in the vacuum as an ideal gas, so that their chemical potential is

$$\mu = \mu_{\text{ext}} + \tau \log(n/n_Q) \ , \tag{5}$$

by *TP*, Chapter 5. Here

$$n_Q = 2(m\tau/2\pi\hbar^2)^{3/2} \ , \tag{6}$$

for particles of spin 1/2.

Now $\mu_{\text{ext}} - \mu = W$, by the definition of the work function W. Thus, from (5),

$$n = n_Q \exp(-W/\tau) \ . \tag{7}$$

The flux of electrons that leaves the metal surface when all electrons are drawn off is equal to the flux incident on the surface from outside:

$$J_n = \tfrac{1}{4} n\bar{c} = n(\tau/2\pi m)^{1/2} \tag{8}$$

by *TP*(14.95) and (14.121). Here $\bar{c}$ is the mean speed of the electrons in the vacuum. The electric charge flux is eJ_n or

$$J_e = (\tau^2 me/2\pi^2\hbar^3)\exp(-W/\tau) \ . \tag{9}$$

This is called the Richardson-Dushman equation for thermionic emission.

Surface States

At the free surface of a semiconductor there often exist surface-bound electronic states with energies in the forbidden gap between the valence and conduction bands of the bulk semiconductor. We can obtain a good impression of the nature of the surface states by considering the wave functions in the weak binding or two-component approximation of Chapter 7, in one dimension. (The wave functions in three dimensions will have extra factors $\exp[i(k_y y + k_z z)]$ in the y, z plane of the surface.)

If the vacuum lies in the region $x > 0$, the potential energy of an electron in this region can be set equal to zero:

$$U(x) = 0 \ , \qquad x > 0 \ . \tag{10}$$

In the crystal the potential energy has the usual periodic form:

$$U(x) = \sum_G U_G \exp(iGx) \ , \qquad x < 0 \ . \tag{11}$$

In one dimension $G = n\pi/a$, where n is any integer, including zero.

In the vacuum the wave function of a bound surface state must fall off exponentially:

$$\psi_{\text{out}} = \exp(-sx) \ , \qquad x > 0 \ . \tag{12}$$

By the wave equation the energy of the state referred to the vacuum level is

$$\epsilon = -\hbar^2 s^2/2m \ . \tag{13}$$

Within the crystal the two-component wave function of a bound surface state will have the form, for $x < 0$,

$$\psi_{\text{in}} = \exp(qx + ikx)[C(k) + C(k-G)\exp(-iGx)] \ , \tag{14}$$

by analogy with (7.49), but with the addition of the factor $\exp(qx)$ which serves to bind the electron to the surface.

We now come to an important consideration that restricts the allowed values of the wavevector k. If the state is bound, there can be no current flow in the x direction, normal to the surface. This condition is assured in quantum mechanics if the wave function can be written as a real function of x, a condition already satisfied by the exterior wave function (12). But (14) can be a real function only if $k = \frac{1}{2}G$, so that

$$\psi_{\text{in}} = \exp(qx)\left[C(\tfrac{1}{2}G)\exp(iGx/2) + C(-\tfrac{1}{2}G)\exp(-iGx/2)\right] \ . \tag{15}$$

This is real provided $C^*(\frac{1}{2}G) = C(-\frac{1}{2}G)$. Thus k_x for a surface state does not have a continuum of values, but is limited to discrete states associated with Brillouin zone boundaries.

The state (15) is damped exponentially in the crystal. The constants s, q are related by the condition that ψ and $d\psi/dx$ are continuous at $x = 0$. The binding energy ϵ is determined by solving the two-component secular equation analogous to (7.46). The plot of Fig. 7.12 is helpful in this connection.

Tangential Surface Transport

We have seen that there may exist surface-bound electronic states with energies in the forbidden gap between the valence and conduction bands of the substrate crystal. These states may be occupied or vacant; their existence must affect the statistical mechanics of the problem. This means that the states modify the local equilibrium concentration of electrons and holes, as expressed as a shift of the chemical potential relative to the band edges. Because the chemical potential is independent of position in an equilibrium system, the energy bands must be displaced or bent, as in Fig. 7.

The thickness and carrier concentration in the surface layer may be changed by applying an electric field normal to the surface. The effect of an external field is utilized in the metal-oxide-semiconductor field-effect transistor (MOSFET). This has a metal electrode just outside the semiconductor surface and insulated from it by a layer of oxide. A voltage, the gate voltage V_g, is applied between the metal and semiconductor that modulates the n_s, the surface charge density per unit area:

$$\Delta n_s = C_g \Delta V_g \ ,$$

where C_g is the capacitance per unit area between the metallic gate and the semiconductor. This surface charge layer forms the conducting pathway of the MOSFET. The conductance of a surface layer of length L and width W between two electrical contacts is:

$$G = (W/L) n_s e \mu \ ,$$

where μ is the carrier mobility. The carrier density n_s, and hence the conductance, is controlled by the gate voltage. This three-terminal electronic valve is

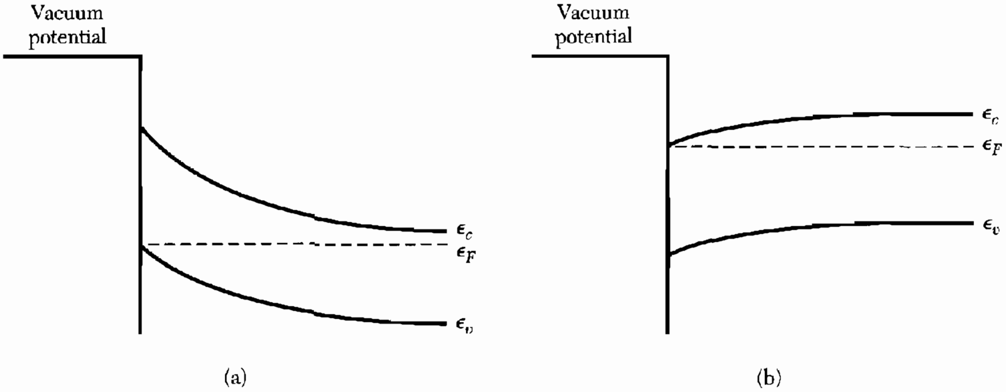


Figure 7 Band bending near a semiconductor surface that can give a highly conducting surface region. (a) Inversion layer on an n-type semiconductor. For the bending as shown, the hole concentration at the surface is far larger than the electron concentration in the interior. (b) Accumulation layer on an n-type semiconductor, with an electron concentration at the surface that is far higher than in the interior.

a principal component in microelectronic systems. The electronic states occupied by the carriers at the surface are quantized along the direction normal to the interface, as treated in Problem 2.

MAGNETORESISTANCE IN A TWO-DIMENSIONAL CHANNEL

The static magnetoconductivity tensor in 3D was found in Problem 6.9. Here we translate that result to a 2D surface conductance channel in the xy plane, with the static magnetic field in the z direction, normal to the MOS layer. We assume the surface density of electrons is $n_s = N/L^2$. The surface conductance is defined as the volume conductivity times the layer thickness. The surface current density is defined as the current crossing a line of unit length in the surface.

Thus, with (6.43) and (6.65) the surface tensor conductance components become

$$\sigma_{xx} = \frac{\sigma_0}{1 + (\omega_c\tau)^2} \; ; \qquad \sigma_{xy} = \frac{\sigma_0\omega_c\tau}{1 + (\omega_c\tau)^2} \; , \tag{16}$$

where $\sigma_0 = n_s e^2\tau/m$ and $\omega_c = eB/mc$ in CGS and eB/m in SI. The following discussion is written in CGS only, except where ohms are used.

These results apply specifically in the relaxation time approximation used in Chapter 6. When $\omega_c\tau \gg 1$, as for strong magnetic field and low temperatures, the surface conductivity components approach the limits

$$\sigma_{xx} = 0 \; ; \qquad \sigma_{xy} = n_s ec/B \; . \tag{17}$$

The limit for σ_{xy} is a general property of free electrons in crossed electric E_y and magnetic fields B_z. We establish the result that such electrons drift in the x direction with velocity $v_D = cE_y/B_z$. Consider the electrons from a Lorentz frame that moves in the x direction with this velocity. By electromagnetic theory there is in this frame an electric field $E_y' = -v_D B_z/c$ that will cancel the applied field E_y for the above choice of v_D. Viewed in the laboratory frame, all electrons drift in the x direction with velocity v_D in addition to any velocity components they had before E_y was applied.

Thus $j_x = \sigma_{xy}E_y = n_s e v_D = (n_s ec/B)E_y$, so that

$$\sigma_{xy} = n_s ec/B \tag{18}$$

as in (17). The experiments measure the voltage V in the y direction and the current I in the x direction (Fig. 8). Here $I_x = j_x L_y = (n_s ec/B)(E_y L_y) = (n_s ec/B)V_y$. The Hall resistance is

$$\rho_H = V_y/I_x = B/n_s ec \; . \tag{18a}$$

We see that j_x can flow with zero E_x, so that the effective conductance j_x/E_x can be infinite. Paradoxically, this limit occurs only when σ_{xx} and σ_{yy} are zero. Consider the tensor relations

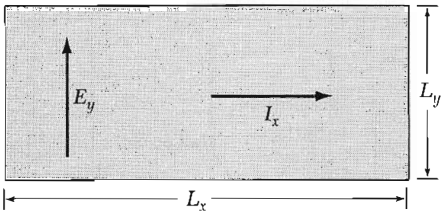


Figure 8 Applied field E_y and drift current I_x in a quantum Hall effect (IQHE) experiment.

$$j_x = \sigma_{xx}E_x + \sigma_{xy}E_y \ ; \qquad j_y = \sigma_{yx}E_x + \sigma_{yy}E_y \ . \tag{19}$$

In the Hall effect geometry $j_y = 0$, so that $E_y = (\sigma_{xy}/\sigma_{yy})E_x$, with $\sigma_{xy} = -\sigma_{yx}$. Thus

$$j_x = (\sigma_{xx} + \sigma_{xy}^2/\sigma_{yy})E_x = \sigma(\mathrm{eff})E_x \ , \tag{20}$$

and in the limit $\sigma_{xx} = \sigma_{yy} = 0$ the effective conductance is infinite.

Integral Quantized Hall Effect (IQHE)

The results of the original measurements[1] under quantum conditions of temperature and magnetic field are shown in Fig. 9. The results are remarkable: at certain values of the gate voltage the voltage drop in the direction of current flow goes essentially to zero, as if the effective conductance were infinite. Further, there are plateaus of the Hall voltage near these same values of gate voltage, and the values of the Hall resistivity V_H/I_x at these plateaus are accurately equal to (25,813/integer) ohms, where 25,813 is the value of h/e^2 expressed in ohms.

The IQHE voltage minima V_{pp} may be explained on a model that is, however, oversimplified. Later we give a general theory. Apply a strong magnetic field such that the separation $\hbar\omega_c \gg k_BT$. It is meaningful to speak of Landau levels that are completely filled or completely empty. Let the electron surface concentration (proportional to the gate voltage) be adjusted to any of the set of values that cause the Fermi level to fall at a Landau level: from (9.33) and (9.34),

$$seB_s/hc = n_s \ , \tag{21}$$

where s is any integer and n_s is the electron surface concentration.

When the above conditions are satisfied, the electron collision time is greatly enhanced. No elastic collisions are possible from one state to another state in the same Landau level because all possible final states of equal energy are occupied. The Pauli principle prohibits an elastic collision. Inelastic collisions to a vacant Landau level are possible with the absorption of the necessary energy from a phonon, but there are very few thermal phonons of energy greater than the interlevel spacing by virtue of the assumption $\hbar\omega_c \gg k_BT$.

---

[1] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

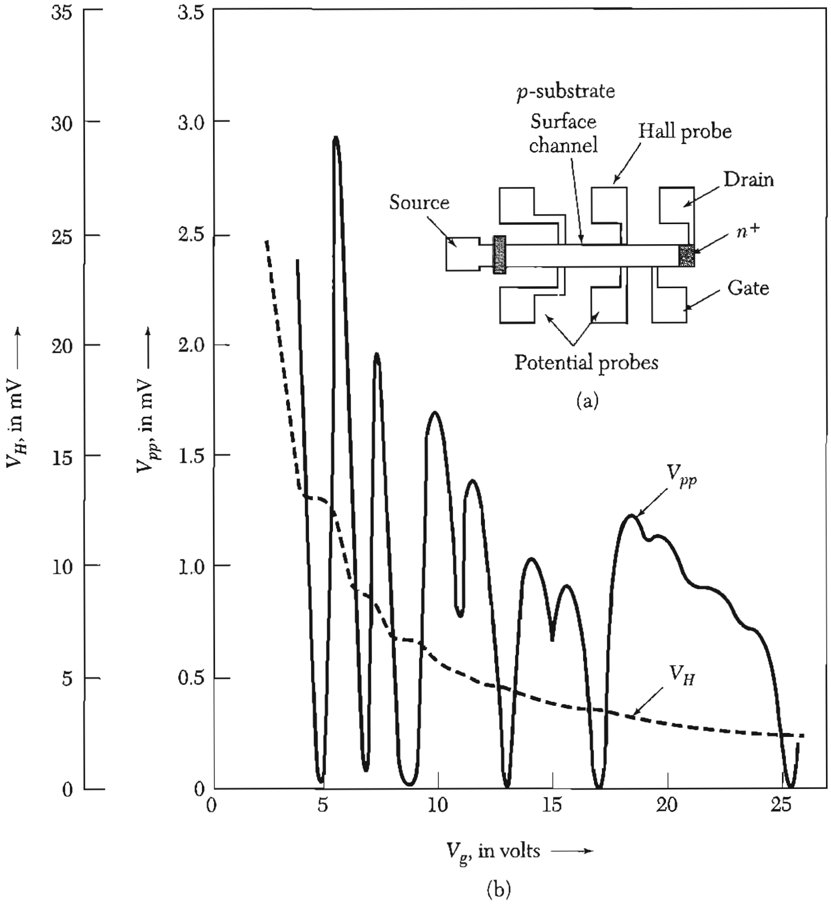


Figure 9 In the original IQHE measurements a magnetic field of 180 kG (18 T) points out of the paper. The temperature is 1.5 K. A constant current of 1 μA is made to flow between the source and the drain. Voltages V_{pp} and V_H are plotted versus the gate voltage V_g, which is proportional to the Fermi level. (After K. von Klitzing, G. Dorda, and M. Pepper.)

The quantization of the Hall resistance follows on combining (18a) and (21):

$$\rho_H = h/se^2 = 2\pi/sc\alpha \ , \tag{22}$$

where α is the fine structure constant $e^2/\hbar c \cong 1/137$, and s is an integer.

IQHE in Real Systems

The measurements (Fig. 9) suggest that the above theory of the IQHE is too good. The Hall resistivity is accurately quantized at 25,813/s ohms, whether or not the semiconductor is of very high purity and perfection. The sharp Landau levels (Fig. 10a) are broadened in the real crystal (Fig. 10b), but this does not affect the Hall resistivity. The occurrence of plateaus in the Hall

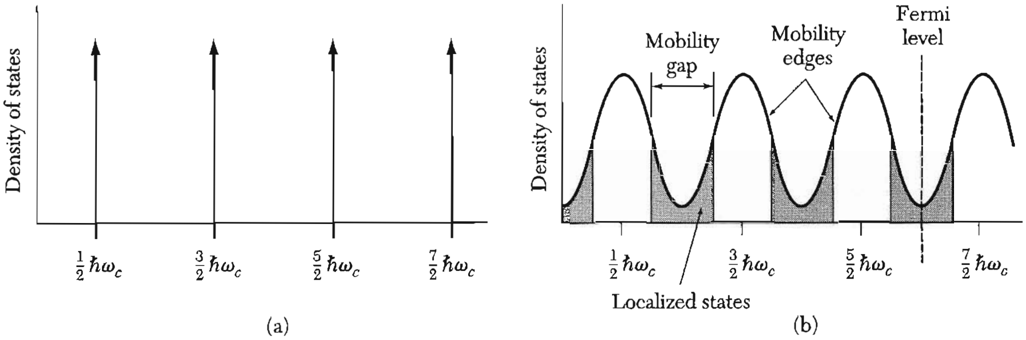


Figure 10 Density of states in a 2D electron gas in a strong magnetic field. (a) Ideal 2D crystal. (b) Real 2D crystal, with impurities and other imperfections.

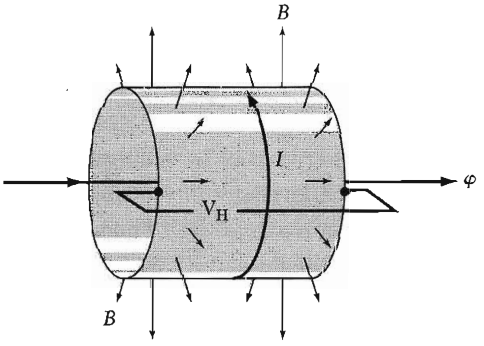


Figure 11 Geometry for Laughlin's thought-experiment. The 2D electron system is wrapped around to form a cylinder. A strong magnetic field B pierces the cylinder everywhere normal to its surface. A current I circles the loop, giving rise to the Hall voltage V_H and a small magnetic flux φ through the loop.

resistance, evident in the U_H curve of Fig. 9, is not expected in ideal systems because partially filled Landau levels will exist for all gate voltages except those for which the Fermi level exactly coincides with a Landau level. Yet the experiments show that a range of V_g values gives the exact Hall resistance.

Laughlin[2] interpreted the results for real systems as the expression of the general principle of gauge invariance. The argument is subtle and somewhat reminiscent of the flux quantization in a superconductor in Chapter 10.

In Laughlin's thought-experiment the 2D electron system is bent to form a cylinder (Fig. 11) whose surface is pierced everywhere by a strong magnetic field B normal to the surface. The current I (former I_x) circles the loop. The magnetic field B acts on the charge carriers to produce a Hall voltage V_H (former V_y) perpendicular to the current and to B; that is, V_H is developed between one edge of the cylinder and the other.

The circulating current I is accompanied by a small magnetic flux φ that threads the current loop. The aim of the thought-experiment is to find the

---

[2] R. B. Laughlin, Phys. Rev. B **23**, 5632 (1981); see also his article in the *McGraw-Hill year-book of science and technology*, 1984, pp. 209–214. A review is given by H. L. Stormer and D. C. Tsui, Science **220**, 1241 (1983).

relation between I and V_H. We start with the electromagnetic relation that relates I to the total energy U of a resistanceless system:

$$\frac{\partial U}{\partial t} = -V_x I_x = \frac{I}{c}\frac{\partial \varphi}{\partial t} \ ; \qquad I = c\frac{\delta U}{\delta \varphi} \ . \tag{23}$$

The value of I can now be found from the variation δU of the electronic energy that accompanies a small variation $\delta\varphi$ of the flux.

The carrier states divide into two classes:

- Localized states, which are not continuous around the loop.
- Extended states, continuous around the loop.

Localized and extended states cannot coexist at the same energy, according to our present understanding of localization.

The two classes of states respond differently to the application of the flux φ. The localized states are unaffected to first order because they do not enclose any significant part of φ. To a localized state a change in φ looks like a gauge transformation, which cannot affect the energy of the state.

The extended states enclose φ, and their energy may be changed. However, if the magnetic flux is varied by a flux quantum, $\delta\varphi = hc/e$, all extended orbits are identical to those before the flux quantum was added. The argument here is identical to that for the flux quantization in the superconducting ring treated in Chapter 10, but with the $2e$ of the Cooper pair replaced by e.

If the Fermi level falls within the localized states of Fig. 10b, all extended states (Landau levels) below the Fermi level will be filled with electrons both before and after the flux change $\delta\varphi$. However, during the change an integral number of states, generally one per Landau level, enter the cylinder at one edge and leave it at the opposite edge.

The number must be integral because the system is physically identical before and after the flux change. If the transferred state is transferred while occupied by one electron, it contributes an energy change eV_H; if N occupied states are transferred, the energy change is NeV_H.

This electron transfer is the only way the degenerate 2D electron system can change its energy. We can understand the effect by looking at a model system without disorder in the Landau gauge for the vector potential:

$$\mathbf{A} = -By\hat{\mathbf{x}} \ . \tag{24}$$

An increase δA that corresponds to the flux increase $\delta\varphi$ is equivalent to a displacement of an extended state by $\delta A/B$ in the y direction. By the Stokes theorem and the definition of the vector potential we have $\delta\varphi = L_x\delta A$. Thus $\delta\varphi$ causes a motion of the entire electron gas in the y direction.

By $\delta U = NeV_H$ and $\delta\varphi = hc/e$, we have

$$I = c(\Delta U/\Delta\varphi) = cNe^2V_H/hc = (Ne^2/h)V_H \ , \tag{25}$$

so that the Hall resistance is

$$\rho_H = V_H/I = h/Ne^2 \ , \tag{26}$$

as in (22).

Fractional Quantized Hall Effect (FQHE). A quantized Hall effect has been reported for similar systems at fractional values of the index s, by working at lower temperatures and higher magnetic fields. In the extreme quantum limit the lowest Landau level is only partially occupied, and the integral QHE treated above should not occur. It has been observed,[3] however, that the Hall resistance ρ_H is quantized in units of $3h/e^2$ when the occupation of the lowest Landau level is 1/3 and 2/3, and ρ_{xx} vanishes for these occupations. Similar breaks have been reported for occupations of 2/5, 3/5, 4/5, and 2/7.

p-n JUNCTIONS

A *p-n* junction is made from a single crystal modified in two separate regions. Acceptor impurity atoms are incorporated into one part to produce the *p* region in which the majority carriers are holes. Donor impurity atoms in the other part produce the *n* region in which the majority carriers are electrons. The interface region may be less than 10^{-4} cm thick. Away from the junction region on the *p* side there are (−) ionized acceptor impurity atoms and an equal concentration of free holes. On the *n* side there are (+) ionized donor atoms and an equal concentration of free electrons. Thus the majority carriers are holes on the *p* side and electrons on the *n* side, Fig. 12.

Holes concentrated on the *p* side would like to diffuse to fill the crystal uniformly. Electrons would like to diffuse from the *n* side. But diffusion will upset the local electrical neutrality of the system.

A small charge transfer by diffusion leaves behind on the *p* side an excess of (−) ionized acceptors and on the *n* side an excess of (+) ionized donors. This charge double layer creates an electric field directed from *n* to *p* that inhibits diffusion and thereby maintains the separation of the two carrier types. Because of this double layer the electrostatic potential in the crystal takes a jump in passing through the region of the junction.

In thermal equilibrium the chemical potential of each carrier type is everywhere constant in the crystal, even across the junction. For holes

$$k_BT \ln p(\mathbf{r}) + e\varphi(\mathbf{r}) = \text{constant} \tag{27a}$$

---

[3] D. C. Tsui, H. L. Stormer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1562 (1982); A. M. Chang et al., Phys. Rev. Lett. **53**, 997 (1984). For a discussion of the theory see R. Laughlin in G. Bauer et al., eds., *Two-dimensional systems, heterostructures, and superlattices*, Springer, 1984.

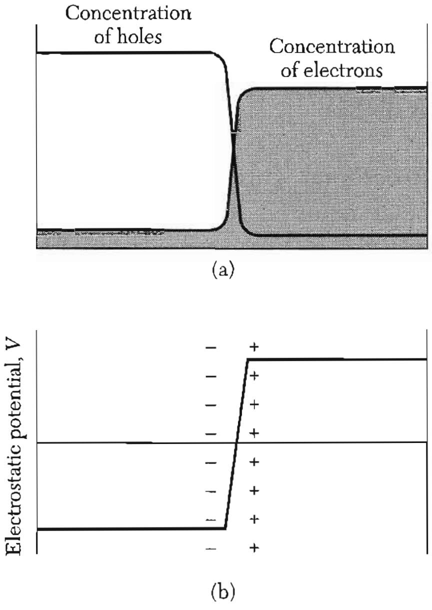


Figure 12 (a) Variation of the hole and electron concentrations across an unbiased (zero applied voltage) junction. The carriers are in thermal equilibrium with the acceptor and donor impurity atoms, so that the product pn of the hole and electron concentrations is constant throughout the crystal in conformity with the law of mass action. (b) Electrostatic potential from acceptor ($-$) and donor ($+$) ions near the junction. The potential gradient inhibits diffusion of holes from the p side to the n side, and it inhibits diffusion of electrons from the n side to the p side. The electric field in the junction region is called the built-in electric field.

across the crystal, where p is the hole concentration and φ the electrostatic potential. Thus p is low where φ is high. For electrons

$$k_BT \ln n(\mathbf{r}) - e\varphi(\mathbf{r}) = \text{constant} \ , \tag{27b}$$

and n will be low where φ is low.

The total chemical potential is constant across the crystal. The effect of the concentration gradient exactly cancels the electrostatic potential, and the net particle flow of each carrier type is zero. However, even in thermal equilibrium there is a small flow of electrons from n to p where the electrons end their lives by combination with holes. The recombination current J_{nr} is balanced by a current J_{ng} of electrons which are generated thermally in the p region and which are pushed by the built-in field to the n region. Thus in zero external applied electric field

$$J_{nr}(0) + J_{ng}(0) = 0 \ , \tag{28}$$

for otherwise electrons would accumulate indefinitely on one side of the barrier.

Rectification

A p-n junction can act as a rectifier. A large current will flow if we apply a voltage across the junction in one direction, but if the voltage is in the opposite direction only a very small current will flow. If an alternating voltage is applied across the junction the current will flow chiefly in one direction—the junction has rectified the current (Fig. 13).

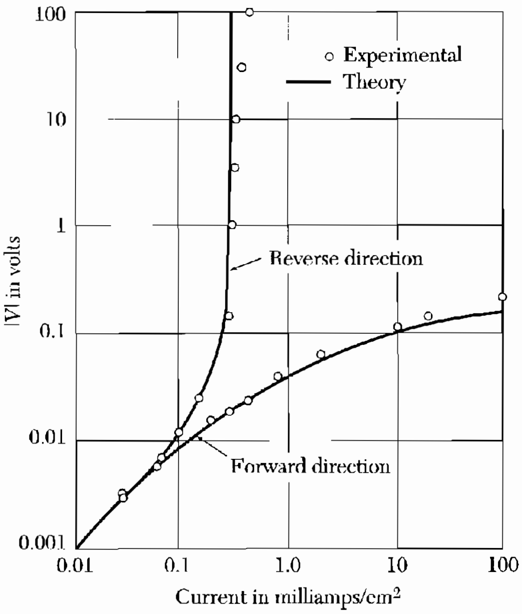


Figure 13 Rectification characteristic of a p-n junction in germanium, after Shockley.

For back voltage bias a negative voltage is applied to the p region and a positive voltage to the n region, thereby increasing the potential difference between the two regions. Now practically no electrons can climb the potential energy hill from the low side of the barrier to the high side. The recombination current is reduced by the Boltzmann factor:

$$J_{nr}(V\text{ back}) = J_{nr}(0)\exp(-e|V|/k_BT) \ . \tag{29}$$

The Boltzmann factor controls the number of electrons with enough energy to get over the barrier.

The thermal generation current of electrons is not particularly affected by the back voltage because the generation electrons flow downhill (from p to n) anyway:

$$J_{ng}(V\text{ back}) = J_{ng}(0) \ . \tag{30}$$

We saw in (28) that $J_{nr}(0) = -J_{ng}(0)$; thus the generation current dominates the recombination current for a back bias.

When a forward voltage is applied, the recombination current increases because the potential energy barrier is lowered, thereby enabling more electrons to flow from the n side to the p side:

$$J_{nr}(V\text{ forward}) = J_{nr}(0)\exp(e|V|/k_BT) \ . \tag{31}$$

Again the generation current is unchanged:

$$J_{ng}(V\text{ forward}) = J_{ng}(0) \ . \tag{32}$$

The hole current flowing across the junction behaves similarly to the electron current. The applied voltage which lowers the height of the barrier for electrons also lowers it for holes, so that large numbers of electrons flow from the n region under the same voltage conditions that produce large hole currents in the opposite direction.

The electric currents of holes and electrons are additive, so that the total forward electric current is

$$I = I_s[\exp(eV/k_BT) - 1] \ , \tag{33}$$

where I_s is the sum of the two generation currents. This equation is well satisfied for p-n junctions in germanium (Fig. 13), but not quite as well in other semiconductors.

Solar Cells and Photovoltaic Detectors

Let us shine light on a p-n junction, one without an external bias voltage. Each absorbed photon creates an electron and a hole. When these carriers diffuse to the junction, the built-in electric field of the junction separates them at the energy barrier. The separation of the carriers produces a forward voltage across the barrier: forward, because the electric field of the photoexcited carriers is opposite to the built-in field of the junction.

The appearance of a forward voltage across an illuminated junction is called the **photovoltaic effect**. An illuminated junction can deliver power to an external circuit. Large area p-n junctions of silicon are used as solar panels to convert solar photons to electrical energy.

Schottky Barrier

When a semiconductor is brought into contact with a metal, there is formed in the semiconductor a barrier layer from which charge carriers are severely depleted. The barrier layer is also called a depletion layer or exhaustion layer.

In Fig. 14 an n-type semiconductor is brought into contact with a metal. The Fermi levels are coincident after the transfer of electrons to the conduction band of the metal. Positively charged donor ions are left behind in this region that is practically stripped of electrons. Here the Poisson equation is

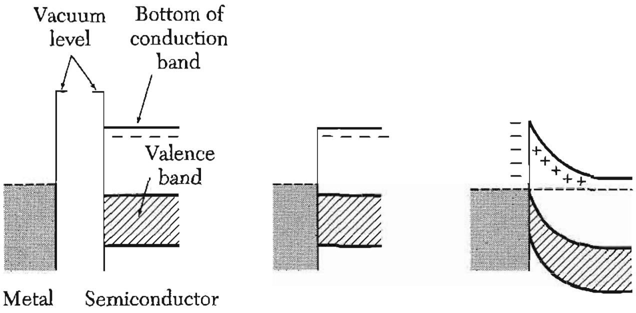
$$\text{(CGS)} \quad \operatorname{div} \mathbf{D} = 4\pi ne \qquad\qquad \text{(SI)} \quad \operatorname{div} \mathbf{D} = ne/\epsilon_0 \ , \tag{34}$$

where n is the donor concentration. The electrostatic potential is determined by

$$\text{(CGS)} \quad d^2\varphi/dx^2 = -4\pi ne/\epsilon \qquad\qquad \text{(SI)} \quad d^2\varphi/dx^2 = ne/\epsilon\epsilon_0 \ , \tag{35}$$

which has a solution of the form

$$\text{(CGS)} \quad \varphi = -(2\pi ne/\epsilon)x^2 \qquad\qquad \text{(SI)} \quad \varphi = -(ne/2\epsilon\epsilon_0)x^2 \ . \tag{36}$$



(a) Before contact (b) Just after contact (c) Equilibrium established

Figure 14 Rectifying barrier between a metal and an n-type semiconductor. The Fermi level is shown as a broken line.

The origin of x has been taken for convenience at the right-hand edge of the barrier. The contact is at $-x_b$, and here the potential energy relative to the right-hand side is $-e\varphi_0$, whence the thickness of the barrier is

$$\text{(CGS)}\quad x_b = (\epsilon|\varphi_0|/2\pi ne)^{1/2} \qquad\qquad \text{(SI)}\quad x_b = (2\epsilon\epsilon_0|\varphi_0|/ne)^{1/2}\ . \qquad (37)$$

With $\epsilon = 16$; $e\varphi_0 = 0.5$ eV; $n = 10^{16}\ \text{cm}^{-3}$, we find $x_b = 0.3\ \mu\text{m}$. This is a somewhat simplified view of the metal-semiconductor contact.

HETEROSTRUCTURES

Semiconductor heterostructures are layers of two or more different semiconductors grown coherently with one common crystal structure. Heterostructures offer extra degrees of freedom in the design of semiconductor junction devices, because both the impurity doping and the conduction and valence band offsets at the junction can be controlled. Because of this freedom many devices that utilize compound semiconductors incorporate heterostructures. Examples include semiconductor lasers in CD players and high-speed devices for cell-phone systems.

A heterostructure may be viewed as a single crystal in which the occupancy of the atomic sites changes at the interface. As an example, one side of the interface can be Ge and the other side GaAs: both lattice constants are 5.65 Å. One side has the diamond structure and the other side the cubic zinc sulfide structure. Both structures are built up from tetrahedral covalent bonds and fit together coherently as if they were a single crystal. There are a few edge dislocations (Chapter 21) to relieve the strain energy near the interface.

The band gaps, however, are different, and this difference is the source of the real interest in the heterostructure, apart from the technical virtuosity in

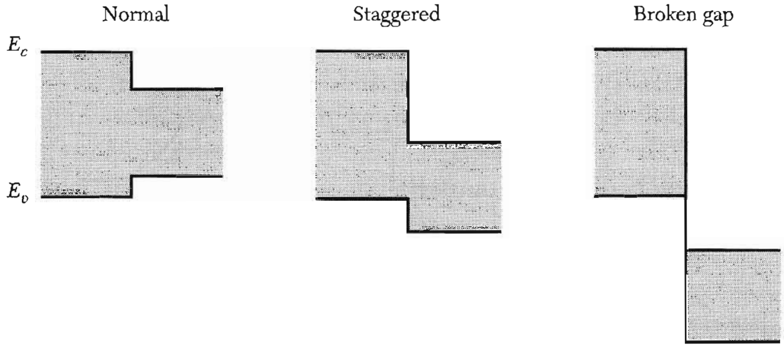


Figure 15 Three types of band edge offsets at hetero-interfaces. The forbidden gaps are shown shaded. The offset called normal occurs, for example, in GaAs/(Al,Ga)As. The "broken-gap" offset occurs in the GaSb/InAs heterojunction.

forming the structure. The band gaps are 0.67 eV for Ge and 1.43 eV for GaAs, at 300 K. The relative alignment of the conduction and valence band edges offers several possibilities, as shown in Fig. 15. Calculations suggest that the top of the valence band E_v in Ge should lie about 0.42 eV higher than in GaAs. The bottom of the conduction band E_c in Ge should lie about 0.35 eV lower than in GaAs, so that the offsets are classified as normal in the scheme of Fig. 15.

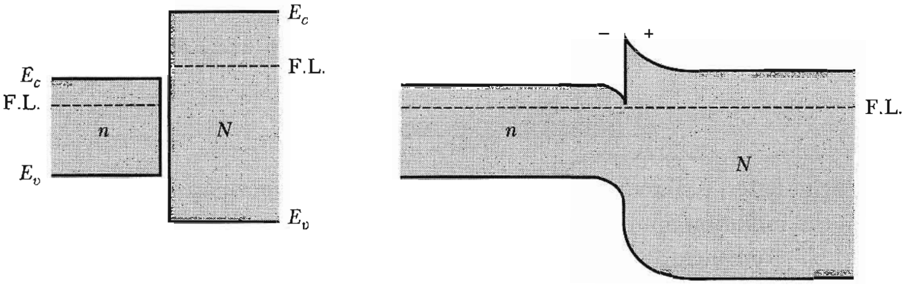
Band edge offsets act as potential barriers in opposite senses on electrons and holes. Recall that electrons lower their energy by "sinking" on an energy band diagram, whereas holes lower their energy by "floating" on the same diagram. For the normal alignment both electrons and holes are pushed by the barrier from the wide-gap to the narrow-gap side of the heterostructure.

Other important semiconductor pairs used in heterostructure are AlAs/GaAs, InAs/GaSb, GaP/Si, and ZnSe/GaAs. Good lattice matching in the range 0.1–1.0 percent is often accomplished by use of alloys of different elements, which may also adjust energy gaps to meet specific device needs.

n-N Heterojunction

As a practical example, consider two *n*-type semiconductors with a large offset of the two conduction bands, as sketched in Fig. 16a for a semiconductor pair with a normal band line-up. The *n*-type material with the higher conduction band edge is labeled with a capital letter as *N*-type, and the junction shown is called an *n-N* junction. The electron transport properties across the junction are similar to those across a Schottky barrier. Far from the interface the two semiconductors must be electrically neutral in composition. However, the two Fermi levels, each determined by the doping, must coincide if there is to be zero net electron transport in the absence of an external bias voltage.

These two considerations fix the "far-off" conduction band edge energies relative to the Fermi level, as in Fig. 16b. The combination of a specified band



(a) (b)

Figure 16 (a) Two semiconductors not in contact; the absolute band edge energies are labeled E_c for the conduction band edge and E_v for the valence band edge. An "absolute energy" means referred to infinite distance. The Fermi levels in the two materials are determined by the donor concentrations, as well as by the band structure. (b) The same semiconductors as a heterojunction, so that the two parts are in diffusive equilibrium. This requires that the Fermi level (F.L.) be independent of position, which is accomplished by transfer of electrons from the N-side to the n-side of the interface. A depletion layer of positively ionized donors is left behind on the N-side.

offset (determined by the host material composition) at the interface and the distant band energies (determined by the Fermi level) can be reconciled only if the bands bend near the interface, as in the figure. The necessary band bending is created by space charges consequent to the transfer of electrons from the N-side to the lower n-side. This transfer leaves behind on the N-side a positive donor space charge layer, which through the Poisson equation of electrostatics is the source of the positive second derivative (upward curvature) in the conduction band edge energy on that side.

On the n-side there is now a negative space charge because of the excess of electrons on that side. The layer of negative space charge gives a negative second derivative (downward curvature) in the conduction band edge energy. On the n-side the band as a whole bends down toward the junction. This differs from the usual p-n junction. The downward bending and the potential step form a potential well for electrons. The well is the basis for the new physical phenomena characteristic of heterostructure physics.

If the doping on the n-side (low E_c) is reduced to a negligible value, there will be very few ionized donors on that side in the electron-rich layer. The mobility of these electrons is largely limited only by lattice scattering, which falls off sharply as the temperature is lowered. Low-temperature mobilities as high as $10^7\ \mathrm{cm^2\ V^{-1} s^{-1}}$ have been observed in GaAs/(Al, Ga)As.

If now the thickness of the N-side semiconductor is reduced below the depletion layer thickness on that side, the N material will be entirely depleted of its low-mobility electrons. All of the electrical conduction parallel to the interface will be carried by the high-mobility electrons on the n-side, equal in number to the number of ionized N-side donors, but spatially separated from them by the potential step. Such high-mobility structures play a large role in solid state studies of 2D electron gases and also in new classes of high-speed field effect transistors for computer applications at low temperatures.

SEMICONDUCTOR LASERS

Stimulated emission of radiation can occur in direct-gap semiconductors from the radiation emitted when electrons recombine with holes. The electron and hole concentrations created by illumination are larger than their equilibrium concentrations. The recombination times for the excess carriers are much longer than the times for the conduction electrons to reach thermal equilibrium with each other in the conduction band, and for the holes to reach thermal equilibrium with each other in the valence band. This steady-state condition for the electron and hole populations is described by separate Fermi levels μ_c and μ_v for the two bands, called **quasi-Fermi levels**.

With μ_c and μ_v referred to their band edges, the condition for population inversion is that

$$\mu_c > \mu_v + \epsilon_g \ . \tag{38}$$

For laser action the quasi-Fermi levels must be separated by more than the band gap.

Population inversion and laser action can be achieved by forward bias of an ordinary GaAs or InP junction, but almost all practical injection lasers employ the double heterostructure proposed by H. Kroemer (Fig. 17). Here the lasing semiconductor is embedded between two wider-gap semiconductor regions of opposite doping, creating a quantum well that confines both electrons and holes. An example is GaAs embedded in (Al,Ga)As. In such a structure there is a potential barrier that prevents the outflow of electrons to the p-type region, and an opposite potential barrier that prevents the outflow of holes to the n-type region.

The value of μ_c in the optically active layer lines up with μ_n in the n contact; similarly, μ_v lines up with μ_p in the p contact. Inversion can be achieved if we apply a bias voltage larger than the voltage equivalent of the active layer energy gap. The diode wafer provides its own electromagnetic cavity, for the reflectivity at the crystal-air interface is high. Crystals are usually polished to provide two flat parallel surfaces; the radiation is emitted in the plane of the heterojunctions.

Crystals with direct band gaps are required normally for junction lasers. Indirect gaps involve phonons as well as photons; carriers recombine less efficiently because of competing processes, and no laser action has been observed in indirect gap semiconductors.

Gallium arsenide has been widely studied as the optically active layer. It emits in the near infrared at 8383 Å or 1.48 eV; the exact wavelength depends

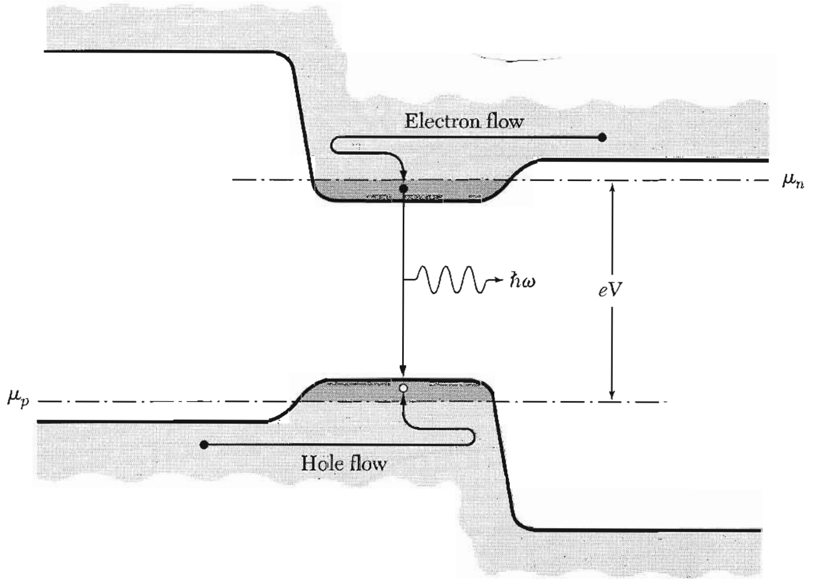


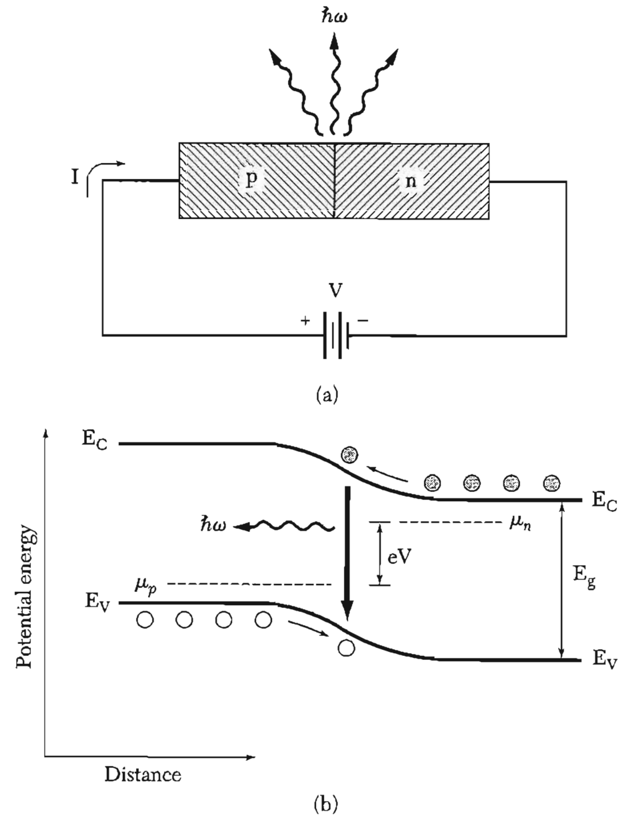
Figure 17 Double heterostructure injection laser. Electrons flow from the right into the optically-active layer, where they form a degenerate electron gas. The potential barrier provided by the wide energy gap on the p side prevents the electrons from escaping to the left. Holes flow from the left into the active layer, but cannot escape to the right.

on temperature. The gap is direct (Chapter 8). In a heterojunction the system is very efficient: the ratio of light energy output to dc electrical energy input is near 50 percent, and the differential efficiency for small changes is up to 90 percent.

The wavelength can be adjusted over a wide range in the alloy system $Ga_xIn_{1-x}P_yAs_{1-y}$, so that we can match the laser wavelength to the absorption minimum of optical fibers used as a transmission medium. The combination of double heterostructure lasers with glass fibers forms the basis of the new lightwave communication technology that is rapidly replacing transmission of signals over copper lines.

LIGHT-EMITTING DIODES

The efficiency of light-emitting diodes is now at the point of exceeding incandescent lamps. Consider a p-n junction with a voltage source V splitting the two chemical potentials μ_n and μ_p by eV, as in Figure 18. Electrons from the n side are injected into the p side, and holes from the p side are injected into the n side. These injected carriers annihilate each other across the junction, thus generating photons if the quantum efficiency is unity.



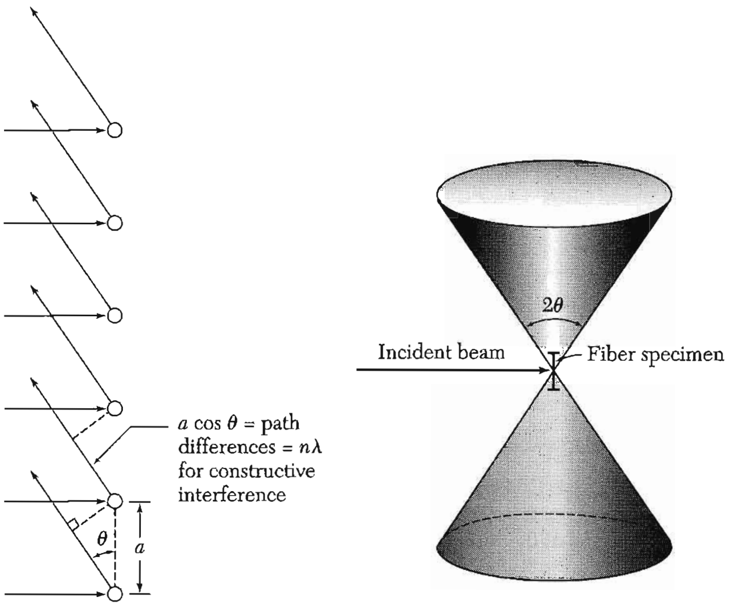
(a)

(b)

Figure 18 Electron-hole recombination into photons, across a *p-n* junction.

The generation or recombination process will be much stronger in a direct-gap semiconductor (Fig. 8.5a) than in an indirect gap semiconductor (Fig. 8.5b). In a direct-gap semiconductor such as GaAs, the band-to-band photons are absorbed in a distance ≈ 1 μm, which is strong absorption. The direct-gap ternary semiconductor GaAs$_{1-x}$ P$_x$ gives light tuned to shorter wavelengths as the composition variable x is increased. This composition was made by Holonyak into one of the first *p-n* diode lasers and into the first visible-spectrum (red) LED. Blue-emitting heterostructures have now been made, such as In$_x$Ga$_{1-x}$N − Al$_y$Ga$_{1-y}$N.

The performance of LEDs has increased markedly over the years, from about 0.1 lumens/watt in 1962 to about 40 lumens/watt in 2004, compared with 15 lumens/watt for a standard white unfiltered incandescent lamp. To quote Craford and Holonyak, "We are entering an entirely new era in lighting (illumination) with an ultimate form of lamp—a direct-gap III-V alloy *p-n* heterostructure."



(a) (b)

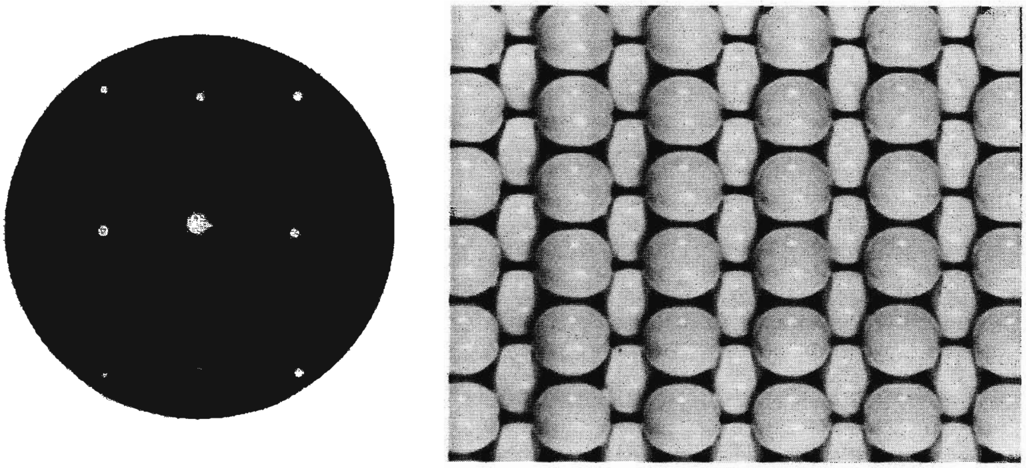
Figure 19 The diffraction pattern from a single line of lattice constant a in a monochromatic x-ray beam perpendicular to the line. (a) The condition for constructive interference is $a \cos \theta = n\lambda$, where n is an integer. (b) For given n the diffracted rays of constant λ lie on the surface of a cone.

Problems

1. ***Diffraction from a linear array and a square array.*** The diffraction pattern of a linear structure of lattice constant a is explained* in Fig. 19. Somewhat similar structures are important in molecular biology: DNA and many proteins are linear helices. (a) A cylindrical film is exposed to the diffraction pattern of Fig. 19b; the axis of the cylinder is coincident with the axis of the linear structure or fiber. Describe the appearance of the diffraction pattern on the film. (b) A flat photographic plate is placed behind the fiber and normal to the incident beam. Sketch roughly the appearance of the diffraction pattern on the plate. (c) A single plane of atoms forms a square lattice of lattice constant a. The plane is normal to the incident x-ray beam. Sketch roughly the appearance of the diffraction pattern on the photographic plate. Hint: The diffraction from a plane of atoms can be inferred from the patterns for two perpendicular lines of atoms. (d) Figure 20 shows the electron diffraction pattern in the backward direction from the nickel atoms on the (110) surface of a nickel crystal. Explain the orientation of the diffraction pattern in relation to the atomic positions of the surface atoms shown in the model. Assume that only the surface atoms are effective in the reflection of low-energy electrons.

---

*Another viewpoint is useful: for a linear lattice the diffraction pattern is described by the single Laue equation $\mathbf{a} \cdot \Delta\mathbf{k} = 2\pi q$, where q is an integer. The lattice sums which led to the other Laue equations do not occur for a linear lattice. Now $\mathbf{a} \cdot \Delta\mathbf{k}$ = constant is the equation of a plane; thus the reciprocal lattice becomes a set of parallel planes normal to the line of atoms.



(a) (b)

Figure 20 (a) Backward scattering pattern of 76 eV electrons incident normally on the (110) face of a nickel crystal; a model of the surface is shown in (b). (Courtesy of A. U. MacRae.)

2. ***Surface subbands in electric quantum limit.*** Consider the contact plane between an insulator and a semiconductor, as in a metal-oxide-semiconductor transistor or MOSFET. With a strong electric field applied across the SiO_2-Si interface, the potential energy of a conduction electron may be approximated by $V(x) = eEx$ for x positive and by $V(x) = \infty$ for x negative, where the origin of x is at the interface. The wavefunction is 0 for x negative and may be separated as $\psi(x,y,z) = u(x) \exp[i(k_y y + k_z z)]$, where $u(x)$ satisfies the differential equation

$$-(\hbar^2/2m)d^2u/dx^2 + V(x)u = \epsilon u \ .$$

With the model potential for $V(x)$ the exact eigenfunctions are Airy functions, but we can find a fairly good ground state energy from the variational trial function $x \exp(-ax)$. (a) Show that $\langle\epsilon\rangle = (\hbar^2/2m)a^2 + 3eE/2a$. (b) Show that the energy is a minimum when $a = (3eEm/2\hbar^2)^{1/3}$. (c) Show that $\langle\epsilon\rangle_{\min} = 1.89(\hbar^2/2m)^{1/3}\,(3eE/2)^{2/3}$. In the exact solution for the ground state energy the factor 1.89 is replaced by 1.78. As E is increased the extent of the wavefunction in the x direction is decreased. The function $u(x)$ defines a surface conduction channel on the semiconductor side of the interface. The various eigenvalues of $u(x)$ define what are called electric subbands. Because the eigenfunctions are real functions of x the states do not carry current in the x direction, but they do carry a surface channel current in the yz plane. The dependence of the channel on the electric field E in the x direction makes the device a field effect transistor.

3. ***Properties of the two-dimensional electron gas.*** Consider a two-dimensional electron gas (2DEG) with twofold spin degeneracy but no valley degeneracy. (a) Show that the number of orbitals per unit energy is given by: $D(\epsilon) = m/\pi h^2$. (b) Show that the sheet density is related to the Fermi wavevector by: $n_s = k_F^2/2\pi$. (c) Show that, in the Drude model, the sheet resistance, i.e., the resistance of a square segment of the 2DEG, can be written as: $R_s = (h/e^2)/(k_F\ell)$ where $\ell = v_\mu\tau$ is the mean free path.