

TOPIC 11: EXCESS CARRIERS & RECOMBINATION-GENERATION PROCESSESREADING ASSIGNMENT

"Excess Carriers in Semiconductors", from *Solid State and Semiconductor Physics*, J.P. McKelvey

SUPPLEMENTARY REFERENCE

"Recombination-Generation Processes", from *Advanced Semiconductor Fundamentals*, R.F. Pierret

LECTURE PROGRAM

1. Continuity Equations (10.1-10.2)
2. Ambipolar Diffusion (10.2)
3. Diffusive Spread of Injected Carriers (10.3)
 - a. $E=0$
 - b. $E \neq 0$
4. Surface Recombination (10.5)
 - a. Flux interchange between bulk and surface
 - b. Surface recombination velocity
5. Recombination Processes (10.8)
 - a. Direct
 - b. In-direct
 - c. Shockley-Read theory of trap recombination (center recombination)
 - d. 'Normal' and 'Slow' trapping centers
 - e. Recombination via shallow levels (see 5.1 of Pierret)
 - f. Recombination involving excitons (see 5.1 of Pierret)
 - g. Auger recombination (see 5.1 of Pierret)
6. Generation Processes (see 5.1 of Pierret)

CHAPTER 10

EXCESS CARRIERS IN SEMICONDUCTORS

10·1 INTRODUCTION

In a metal it is practically impossible to alter the bulk concentration of free charge carriers in any way whatever. Excess carriers can be introduced, but only by inducing an electrical charge on the specimen. Under these conditions, the excess charge density resides essentially at the surface of the metal, the bulk concentration of charge carriers being completely unaffected. No change in the bulk transport properties can be observed in this situation.

In a semiconductor, however, it is possible to profoundly alter the charge carrier concentration in the bulk material *without* the introduction of any significant electric charge density. This can be done because it is possible for two types of charge carriers, holes and electrons, to be present simultaneously; if excess holes and electrons are introduced in *pairs*, large deviations from the thermal equilibrium values of bulk carrier concentration can be created without the formation of any net charge density. The formation of these carrier densities in excess of equilibrium is accompanied by a significant modulation of the bulk conductivity of the material, as determined by (9.7-6). It is this possibility which permits the use of semiconductor elements as electronic devices such as rectifiers, transistors and switching units. By *extracting* electrons and holes in pairs one may obtain bulk carrier concentrations *less* than the equilibrium values quite as easily as excess concentrations are created by the introduction or injection of charge carrier pairs.

Excess carriers can be created in semiconductors by illuminating the material with light of frequency such that the photon energy $\hbar\omega_0$ equals or exceeds the forbidden gap energy $\Delta\varepsilon$. Under these conditions the incident photons have enough energy to break the covalent electron pair lattice bonds, liberating free electrons and leaving holes behind at the excitation sites. The excess electrons and holes which are thus created in pairs contribute to the conductivity of the crystal, which is then increased under illumination, the increase being proportional to the light intensity. This phenomenon is known as photoconductivity and is quite characteristic of all semiconductors. Since a photon is absorbed in the creation of each electron-hole pair, light of wavelength sufficiently short that the photon energy exceeds $\Delta\varepsilon$ will be quite strongly absorbed within the crystal; on the other hand, light of longer wavelength, whose photon energy is less than $\Delta\varepsilon$ cannot create electron-hole pairs and is absorbed hardly at all. The absorption spectrum of a semiconductor is therefore always

characterized by a very rapid change of absorption coefficient at a wavelength corresponding to photon energy $\Delta\epsilon$, the crystal being strongly absorptive for shorter wavelengths and nearly transparent for longer wavelengths, as shown in Figure 10.1(a). The region of sharp transition from opaqueness to transparency is called the *absorption edge*. In germanium the absorption edge occurs in the near infrared at about $\lambda = 1.75 \mu$, corresponding to $\Delta\epsilon = 0.7 \text{ eV}$, in silicon at about 1.13μ , corresponding to

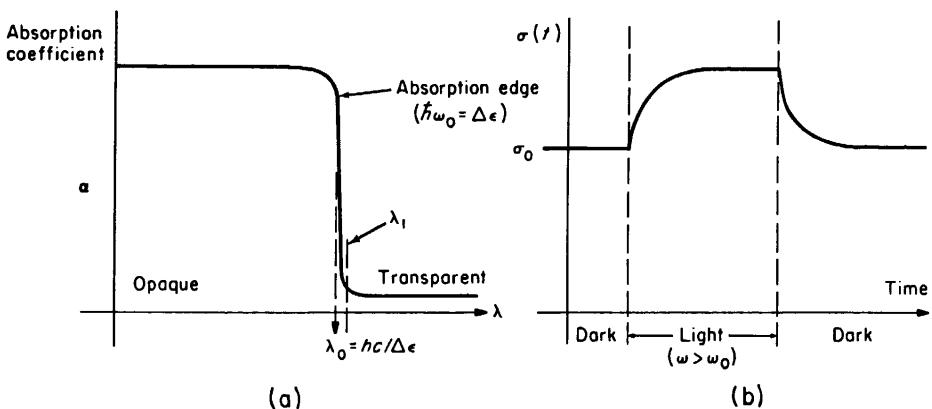


FIGURE 10.1. (a) Absorption spectrum of a typical semiconductor in the neighborhood of the fundamental absorption edge, (b) Photoconductive response of a semiconductor crystal to light of wavelength sufficiently short to excite excess electron-hole pairs.

$\Delta\epsilon = 1.12 \text{ eV}$. The forbidden energy gap $\Delta\epsilon$ can be accurately determined from the absorption spectrum by measuring the wavelength at which the absorption edge occurs. The photoconductive response of a semiconductor to light of wavelength short enough to create electron-hole pairs is shown in Figure 10.1(b).

Excess carriers can also be created in semiconductor crystals by high energy particle beams (electrons, protons, α -particles), X-rays, γ -radiation, and by suitably biased metal-semiconductor contacts or $p-n$ junctions, about which we shall have more to say later.

10.2 TRANSPORT BEHAVIOR OF EXCESS CARRIERS: THE CONTINUITY EQUATIONS

In any semiconductor crystal electron-hole pairs are continually being generated by thermal or other means and are continually recombining. In thermal equilibrium the only generation process is thermal generation, and the rate at which pairs are generated and the rate at which they recombine must be equal; otherwise, the concentration of electrons and holes would build up or decline as a function of time. The equilibrium thermal generation rate $g_0 (\text{cm}^{-3} \text{ sec}^{-1})$ is the number of electron-hole pairs generated per unit volume per unit time from thermal breakage of covalent bonds. This quantity is a function of temperature and certain crystal parameters, but is independent of electron and hole concentration. The recombination rate is related to the mean time

which elapses between the generation of an electron or hole and its subsequent recombination. This quantity is called the mean carrier *lifetime*. If there were only one carrier of a given species per unit volume of the crystal at any given time, then the number of recombination events per unit volume per second involving that type of carrier would be $1/\tau$, where τ is the lifetime of that carrier species.¹ When there are n carriers per unit volume, then the recombination rate (number of recombination events per unit volume per unit time) is n times as great or n/τ . In thermal equilibrium, the concentration of electrons is n_0 and the concentration of holes is p_0 ; since the generation and recombination rates of holes and of electrons must be equal under these circumstances, then we must have

$$g_{0n} = \frac{n_0}{\tau_{n_0}} \quad \text{and} \quad g_{0p} = \frac{p_0}{\tau_{p_0}} \quad (10.2-1)$$

where g_{0n} and g_{0p} are the equilibrium thermal generation rates for electrons and holes, respectively, and where τ_{n_0} and τ_{p_0} are the average lifetimes for the respective carrier species under equilibrium conditions. In addition, since thermal generation always results in the production of electron-hole pairs, one hole being generated for each electron, g_{0n} and g_{0p} must be the same, whereby

$$g_{0n} = \frac{n_0}{\tau_{n_0}} = \frac{p_0}{\tau_{p_0}} = g_{0p}. \quad (10.2-2)$$

The equality of electron and hole generation rates as well as the equality of electron and hole recombination rates must be preserved even in systems which are not in a steady-state condition, because an electron generation (or recombination) is inevitably accompanied by a hole generation (or recombination). In a non-steady-state system, however, the generation and recombination rates for electrons (or holes) may be different. In any event, we may *always* write

$$g_n = g_p \quad \text{and} \quad \frac{n}{\tau_n} = \frac{p}{\tau_p}, \quad (10.2-3)$$

where g_n and g_p are the actual generation rates, n and p the local concentrations and τ_n and τ_p the lifetime of electrons and holes, respectively. In these equations the generation rates g_n and g_p as well as the concentrations n and p and the lifetimes τ_n and τ_p must in general be regarded as *functions of space coordinates and time* within the crystal.

Consider now a region of the crystal of dimensions dx, dy, dz , as shown in Figure 10.2, and suppose that particle flux densities J_p and J_n of holes and electrons, respectively, are flowing into and out of this region. The x -component of J_p at $x + dx$ can

¹ We have heretofore used the symbols τ_n , τ_p and τ to refer to relaxation times connected with scattering processes. Unfortunately, the *same* symbols are *universally* used in the literature to represent recombinative lifetimes as well. Since this practice has been sanctified by many years of usage, and since we do not wish to adopt a notation which is very complex, we shall also use the same symbols for both sets of quantities. From this point forward, however, we shall understand that τ_n , τ_p and τ , always refer to *recombinative lifetimes* unless otherwise specified.

SEC. 10.2

SEC. 10.2

EXCESS CARRIERS IN SEMICONDUCTORS

323

sequent re-
are only one
ie, then the
hat type of
in there are
combination
equilibrium,
; since the
qual under

(10.2-1)

and holes,
tive carrier
ion always
d for each

(10.2-2)

quality of
ms which
recombi-
on). In a
; for elec-

(10.2-3)

tions and
itions the
etimes τ_n
ne within

in Figure
s, respec-
+ dx can

ected with
e to repre-
s of usage,
the same
nd that τ ,

be expressed in terms of J_p at x by making a Taylor expansion of $J_{px}(x + dx)$; in the limit where dx becomes small only the first two terms will be significant, whereby

$$J_{px}(x + dx) = J_{px}(x) + \frac{\partial J_{px}}{\partial x} dx. \quad (10.2-4)$$

The net increase in the number of holes within the region per unit time arising from a difference in the x -components of J_p at the two faces $ABCD$ and $EFGH$ is then

$$[J_{px}(x) - J_{px}(x + dx)] dy dz = -\frac{\partial J_p}{\partial x} dx dy dz. \quad (10.2-5)$$

There will be similar terms arising from the difference of the y -components of J_p at

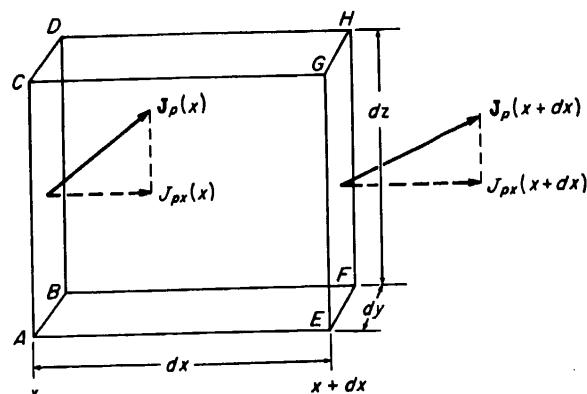


FIGURE 10.2. Carrier flow into and out of a small volume element of the crystal.

the faces $ACGE$ and $BDHF$, and from the difference of the z -components of J_p at $ABFE$ and $CDHG$. The total net increase per unit time in the number of holes within the region arising from these terms is $-\nabla \cdot J_p dx dy dz$. If generation and recombination processes are active within the volume, there will be $g_p dx dy dz$ holes generated and $(p/\tau_p) dx dy dz$ holes lost by recombination per unit time within the region of interest. The algebraic sum of all these quantities represents the total net increase of the number of holes inside the volume $dx dy dz$ per unit time, which is clearly $(\partial p/\partial t) dx dy dz$. Taking the algebraic sum of the various contributions, equating it to $(\partial p/\partial t) dx dy dz$ and canceling the volume element on both sides of the equation, one may easily show that

$$-\nabla \cdot J_p + g_p - \frac{p}{\tau_p} = \frac{\partial p}{\partial t}, \quad (10.2-6)$$

while a similar calculation for electrons within the same volume element will obviously yield

$$-\nabla \cdot J_n + g_n - \frac{n}{\tau_n} = \frac{\partial n}{\partial t}. \quad (10.2-7)$$

Equations (10.2-6) and (10.2-7) are referred to as the equations of continuity for

electrons and holes. The solutions of these equations, under appropriate boundary conditions, describe the distribution of electron and hole concentration as a function of space coordinates and time, and furnish a complete description of the transport behavior of electrons and holes in the semiconductor under nonequilibrium conditions. In order to arrive at explicit solutions of the continuity equations, it is necessary to express the current in terms of the concentration. This is easily done by writing the particle flux density as the sum of a diffusion flux density and a drift current density arising from any electric field which might be present. We may in this way write

$$\mathbf{J}_p = -D_p \nabla p + p\mu_p \mathbf{E} \quad (10.2-8)$$

and

$$\mathbf{J}_n = -D_n \nabla n - n\mu_n \mathbf{E}. \quad (10.2-9)$$

The origin of the current component due to the field in the above equations is quite clear;² the diffusion term states, in effect, that whenever there is a *gradient* of concentration ∇p , a net current of particles flows from regions of high concentration to regions of lower concentration, proportional to the gradient of concentration at any point. This diffusive flow is analogous to the flow of heat in the presence of a temperature gradient. The constant of proportionality is called the *diffusion coefficient* or *diffusivity*. That the diffusive component of current should have this form can be proved rigorously by the methods developed in Chapter 7 starting with the Boltzmann equation, assuming the relaxation time approximation (7.2-4). We shall not reproduce the details of this proof here, but it is assigned as an exercise for the reader; it affords good practice in utilizing the transport theory discussed previously. In the case where the mean free path is velocity-independent, which is quite a good approximation in semiconductors where lattice scattering is dominant, it turns out that the diffusion coefficients for electrons and holes can be expressed as

$$D_n = \lambda_n \bar{c}_n / 3 \quad \text{and} \quad D_p = \lambda_p \bar{c}_p / 3 \quad (10.2-10)$$

where λ_n and λ_p are electron and hole mean free paths and \bar{c}_n and \bar{c}_p are the mean thermal velocities for electrons and holes, respectively. In the circumstances under which (10.2-10) is valid, the mean free paths and relaxation times for electrons and holes are related by (7.3-18). If the equations (10.2-10) are expressed in terms of relaxation times, and if the thermal velocities are represented by (5.4-20), using the appropriate effective masses, it can readily be seen with the aid of (7.3-13) that the diffusion coefficients can be represented in terms of the mobilities as

$$D_n = \frac{\mu_n k T}{e} \quad \text{and} \quad D_p = \frac{\mu_p k T}{e}. \quad (10.2-11)$$

These relations between diffusion coefficients and mobilities are known as the *Einstein relations*. Although derived above under fairly restrictive assumptions, they can be shown to be true in all systems which obey Boltzmann statistics. As a matter of fact,

² It is assumed in (10.2-8) and (10.2-9) that no *magnetic* fields are present, or at least that their effect on the carriers is small compared to that of the electric field. The symbols \mathbf{J}_n and \mathbf{J}_p will be used to represent particle flux density, while $\mathbf{I}_n (= -e\mathbf{J}_n)$ and $\mathbf{I}_p (= e\mathbf{J}_p)$ will be used for electrical current density.

SEC. 10.2

SEC. 10.2

EXCESS CARRIERS IN SEMICONDUCTORS

325

boundary
a function
transport
conditions.
cessary to
writing the
nt density
y write

(10.2-8)

(10.2-9)

is is quite
nt of con-
tration to
on at any
temperature-
efficient or
m can be
loltzmann
reproduce
it affords
ase where
mation in
diffusion

(10.2-10)

the mean
ces under
trons and
terms of
using the
) that the

(10.2-11)

: Einstein
y can be
r of fact,

: that their
will be used
cal current

if an appropriate numerical constant is included, they may be shown to apply to Fermi systems as well.

If the current equations (10.2-8) and (10.2-9) are substituted into the continuity equations (10.2-6) and (10.2-7), the latter can be written as

$$D_p \nabla^2 p - \mu_p \nabla \cdot (p \mathbf{E}) + g_p - \frac{p}{\tau_p} = \frac{\partial p}{\partial t} \quad (10.2-12)$$

$$\text{and} \quad D_n \nabla^2 n + \mu_n \nabla \cdot (n \mathbf{E}) + g_n - \frac{n}{\tau_n} = \frac{\partial n}{\partial t}. \quad (10.2-13)$$

The divergence terms in the above equations can be transformed to a slightly more tractable form by the use of the vector identity

$$\nabla \cdot (\phi \mathbf{A}) = \mathbf{A} \cdot \nabla \phi + \phi \nabla \cdot \mathbf{A} \quad (10.2-14)$$

which is true for any vector \mathbf{A} and any scalar function ϕ . Also, it is advantageous to express the generation rates as the sum of the thermal generation rate plus the rate at which carriers are generated in *excess* of the thermal generation rate, hence

$$g_n = g_{0n} + g'_n \quad \text{and} \quad g_p = g_{0p} + g'_p, \quad (10.2-15)$$

where g'_n and g'_p are the excess generation rates. The thermal generation rates may now be expressed in terms of equilibrium concentrations and lifetimes by (10.2-1). After making these transformations, the continuity equations (10.2-12) and (10.2-13) take the form

$$D_p \nabla^2 p - \mu_p (\mathbf{E} \cdot \nabla p + p \nabla \cdot \mathbf{E}) + g'_p - \left(\frac{p}{\tau_p} - \frac{p_0}{\tau_{p_0}} \right) = \frac{\partial p}{\partial t} \quad (10.2-16)$$

$$D_n \nabla^2 n + \mu_n (\mathbf{E} \cdot \nabla n + n \nabla \cdot \mathbf{E}) + g'_n - \left(\frac{n}{\tau_n} - \frac{n_0}{\tau_{n_0}} \right) = \frac{\partial n}{\partial t}. \quad (10.2-17)$$

The field \mathbf{E} can be expressed as the sum of the applied field and the *internal* field which arises because of the fact that the diffusing particles are charged, whereby

$$\mathbf{E} = \mathbf{E}_{int} + \mathbf{E}_{app}. \quad (10.2-18)$$

The internal field has its origin in the fact that the mean free paths (and thus the diffusion coefficients) of electrons and holes may be different. If the electrons and holes could move entirely independently of each other, the faster diffusing species would tend to go ahead of the slower diffusing species, leaving the latter behind altogether. When this actually begins to happen, however, since the diffusing particles are charged, a separation of positive and negative charges begins to occur, and an internal electric field which tends to retard the faster diffusing particles and to drag the slower ones on more rapidly is set up. This is the field \mathbf{E}_{int} of Equation (10.2-18). Eventually this field becomes strong enough to counteract entirely any tendency for the positive and negative charge distributions to separate further; the positive and

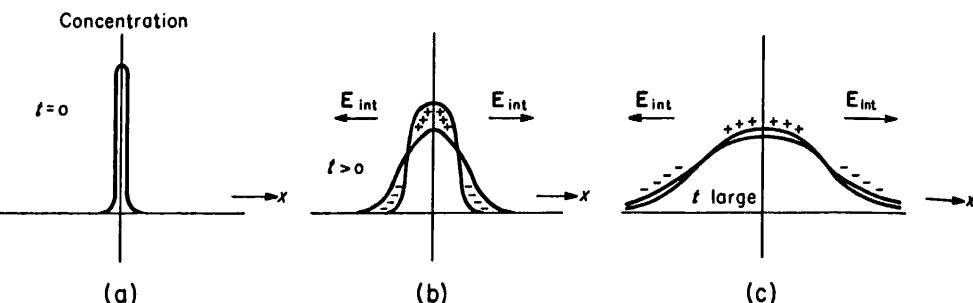


FIGURE 10.3. Successive stages in the diffusive spreading of a distribution of excess electron-hole pairs, which initially has the form of a δ -function, in an intrinsic semiconductor.

negative particle distributions then diffuse *together*, the effective diffusion coefficient being larger than that of the more slowly diffusing species, but smaller than that of the more rapidly diffusing type. This phenomenon, which is known as *ambipolar* diffusion, is illustrated in Figure 10.3. As we shall soon see, a similar effect alters the character of the drift motion which takes place when an applied electric field is present.

In Equations (10.2-16) and (10.2-17), it is clear from the preceding discussion that there are really *three* unknown quantities, n , p , and \mathbf{E} . To arrive at a complete solution for these quantities another equation is needed. The set of equations may be completed by writing Poisson's equation relating the electric field and the net charge density $e(p - n + N_d - N_a + p_a - n_d)$. According to this,

$$\nabla \cdot \mathbf{E} = \frac{4\pi\rho}{\kappa} = \frac{4\pi e(p - n + N_d - N_a + p_a - n_d)}{\kappa} \quad (10.2-19)$$

where κ is the dielectric constant. We should note that since the applied field \mathbf{E}_{app} has no internal sources or sinks, its divergence is zero, whence from (10.2-18)

$$\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{E}_{int}. \quad (10.2-20)$$

Equations (10.2-16), (10.2-17), and (10.2-19) now form a set of three equations which can in principle be solved for the three unknowns n , p , and \mathbf{E}_{int} .

Unfortunately, there seems to be no way of arriving at the solution of these equations in a straightforward analytical manner. We must, therefore, make some reasonable physical approximation which will allow us to arrive at a solution which, although not exact, will be adequate for most cases of practical importance. The approximation which we shall use for this purpose is the *electrical neutrality* condition, or *charge balance* assumption. We shall assume in (10.2-16) and (10.2-17) that the excess electron density $\delta n = n - n_0$ is just balanced by an excess hole density $\delta p = p - p_0$. It is clear that this cannot be precisely correct, for then no internal field would ever be set up, and it is the internal field arising from the basic disparity in electron and hole mean free paths which tends to keep the diffusing electron and hole density distributions together to begin with. However, such a small difference between electron and hole density (compared with the excess carrier densities themselves) is ordinarily required to create this internal field, that the charge balance approximation is usually a very good one insofar as the carrier densities of Equations (10.2-16) and (10.2-17) are concerned. We shall *not*, however, expect to use the charge neutrality assumption in (10.2-19). On the contrary, we may *calculate* the field \mathbf{E}_{int} from

SEC. 10.2

(10.2-16) and (10.2-17) using the charge neutrality assumption, since there will then be only the two unknowns, δp and E_{int} , in these two equations. This calculated value of E_{int} may then be used in (10.2-19) to determine the source density (thus the disparity in electron and hole densities) required to maintain the field. If the calculated source density is only a small fraction of the excess electron and hole densities calculated from (10.2-16) and (10.2-17) we may be confident of the consistency of our calculation and the validity of the assumptions upon which it is based; otherwise, we must conclude that our results are inconsistent with our original hypotheses and our calculations are in error. We shall see later that the former state of affairs is usually the one which is encountered in actual practice.³

Following this line of reasoning, then, we shall assume initially that

$$\delta n = n - n_0 = p - p_0 = \delta p. \quad (10.2-21)$$

We shall limit ourselves to a discussion of homogeneous samples, where the impurity density is uniform throughout. For such samples n_0 and p_0 are constants and the gradients and time derivatives of n and p are simply equal to the gradients and time derivatives of δn and δp , respectively. Also, from (10.2-3) it is readily seen that the generation and recombination terms in (10.2-12) are equal to those in (10.2-13). The generation and recombination terms in (10.2-16), which arise directly from the corresponding terms in (10.2-12) are therefore equal to the generation and recombination terms in (10.2-17) which arise in the same way from those terms in (10.2-13). Employing these results in (10.2-16) and (10.2-17) it may easily be shown that

$$D_p \nabla^2(\delta p) - \mu_p (\mathbf{E} \cdot \nabla(\delta p) + p \nabla \cdot \mathbf{E}) + g' - \left(\frac{p_0 + \delta p}{\tau_p} - \frac{p_0}{\tau_{p_0}} \right) = \frac{\partial(\delta p)}{\partial t} \quad (10.2-22)$$

$$D_n \nabla^2(\delta p) + \mu_n (\mathbf{E} \cdot \nabla(\delta p) + n \nabla \cdot \mathbf{E}) + g' - \left(\frac{p_0 + \delta p}{\tau_p} - \frac{p_0}{\tau_{p_0}} \right) = \frac{\partial(\delta p)}{\partial t} \quad (10.2-23)$$

where

$$g'_n = g'_p = g'. \quad (10.2-24)$$

The term involving $\nabla \cdot \mathbf{E}$ may be eliminated by multiplying (10.2-22) by $n\mu_n$, (10.2-23) by $p\mu_p$ and adding the two equations (noting from (10.2-21) that $n - p = n_0 - p_0$) to obtain

$$\frac{n\mu_n D_p + p\mu_p D_n}{n\mu_n + p\mu_p} \nabla^2(\delta p) - \frac{\mu_n \mu_p (n_0 - p_0)}{n\mu_n + p\mu_p} \mathbf{E} \cdot \nabla(\delta p) + g' - \left(\frac{p_0 + \delta p}{\tau_p} - \frac{p_0}{\tau_{p_0}} \right) = \frac{\partial(\delta p)}{\partial t}. \quad (10.2-25)$$

Now in the coefficient of the first term above the mobilities can be expressed in terms of the diffusion coefficients by the Einstein relations (10.2-11). Equation (10.2-25) may thus be written as

$$D^* \nabla^2(\delta p) - \mu^* \mathbf{E} \cdot \nabla(\delta p) + g' - \frac{\delta p}{\tau} = \frac{\partial(\delta p)}{\partial t} \quad (10.2-26)$$

³This line of approach for the semiconductor case was first taken by W. van Roosbroeck [Phys. Rev., 91, 282 (1953)], but similar methods were used much earlier to treat problems dealing with the transport of electrons and ions in gas discharges, which are quite similar.

where

$$D^* = \frac{(n + p)D_n D_p}{n D_n + p D_p} = \frac{(n_0 + p_0 + 2\delta p)D_n D_p}{(n_0 + \delta p)D_n + (p_0 + \delta p)D_p} \quad (10.2-27)$$

$$\mu^* = \frac{(n_0 - p_0)\mu_n \mu_p}{n \mu_n + p \mu_p} = \frac{(n_0 - p_0)\mu_n \mu_p}{(n_0 + \delta p)\mu_n + (p_0 + \delta p)\mu_p} \quad (10.2-28)$$

and where τ is an "excess carrier lifetime" defined by

$$\frac{\delta p}{\tau} = \frac{p_0 + \delta p}{\tau_p} - \frac{p_0}{\tau_{p_0}} = \frac{n_0 + \delta p}{\tau_n} - \frac{n_0}{\tau_{n_0}}. \quad (10.2-29)$$

The excess concentration of the other carrier species, δn , is, of course, equal to δp in accord with the charge balance assumption (10.2-21).

Equation (10.2-26) is the *ambipolar* transport equation which is obeyed by the excess electron and hole density.⁴ The ambipolar diffusion coefficient D^* and the ambipolar mobility μ^* are seen from (10.2-27) and (10.2-28) to be, in general, dependent upon the concentration δp of excess carriers. It is impossible to solve (10.2-26) under these circumstances, and one must in general proceed from this point using approximate or numerical methods. However, if the excess carrier density δp is much less than the larger of the two quantities (n_0, p_0), D^* and μ^* are substantially constant and analytical solutions of (10.2-26) are quite easily obtained. We shall restrict ourselves to the discussion of situations where this condition is fulfilled; we shall refer to this state of affairs as the *low level* case.

Let us now turn our attention to the case of strongly extrinsic *n*-type material, where $n_0 \gg p_0$ or δp . Suppose, for example, we have in equilibrium 10^{15} electrons and 10^{11} holes per cm^3 . There will be a certain equilibrium electron lifetime τ_{n_0} and an equilibrium hole lifetime τ_{p_0} , which will be related by (10.2-2) or (10.2-3). It is clear that τ_{p_0} is much smaller (by a factor of 10^4 in this case) than τ_{n_0} . Suppose now that a uniform excess density of electron-hole pairs is created such that $\delta p = \delta n = 10^{11} \text{ cm}^{-3}$. The concentration of electrons is now $n_0 + \delta n = 1.0001 \times 10^{15} \text{ cm}^{-3}$, an increase of 0.01 percent over the equilibrium density, while the concentration of holes is 2×10^{11} , double the equilibrium density. Since the density of electrons *relative to the equilibrium density* is hardly changed, the probability per unit time that any given hole will encounter an electron is practically unchanged. Therefore, the *hole lifetime* will be essentially unaffected, and will be practically independent of δp for such small values of δp . On the other hand, since the density of holes has doubled, the probability that any given electron will encounter a hole is essentially doubled, and the electron lifetime will be reduced to about half what it was in the equilibrium state. The electron lifetime must thus vary rapidly with δp even though that quantity may be quite small. Clearly the situation will be just the opposite in strongly extrinsic *p*-type material. In general, we observe that the lifetime of the *minority carrier* is essentially independent of δp for small values of δp while the majority carrier lifetime is not. For a strongly *n*-type semiconductor, then, $\tau_p = \tau_{p_0}$ for $\delta p \ll n_0$, and the excess carrier lifetime τ , as

⁴ This equation *resembles* the continuity equation for independently diffusing particles, which would have the form (10.2-13), but it is not exactly the same, the term $\nabla \cdot (p\mathbf{E})$ being replaced by $\mathbf{E} \cdot \nabla p$. If the field is constant the terms are the same, but otherwise the equations and their solutions will be different.

SEC. 10.2

SEC. 10.2

EXCESS CARRIERS IN SEMICONDUCTORS

329

10.2-27)

defined by (10.2-29) reduces to

$$\tau = \tau_{p_0} \quad (\text{strongly } n\text{-type material}). \quad (10.2-30)$$

10.2-28)

Likewise, for a strongly extrinsic *p*-type crystal, it is readily seen that (10.2-29) reduces to

$$\tau = \tau_{n_0} \quad (\text{strongly } p\text{-type material}). \quad (10.2-31)$$

0.2-29)

δp in
by the
and the
dependent
) under
approxim-
ach less
int and
rselves
to this

aterial,
electrons
 n_0 and
. It is
e now
 $= 10^{11}$
 -3 , an
holes
tive to
given
fetime
small
ability
electron
electron
small.
terial.
ndent
ongly
: τ , as

which
xed by
lutions

For the low-level case, then, the excess carrier lifetime τ simply reduces to the lifetime of the *minority* carrier. The terms *excess carrier lifetime* and *minority carrier lifetime* are often used interchangeably, despite the fact that the two are not always the same. In particular, when the conditions for the low level case are not fulfilled, or when the crystal is nearly intrinsic, the excess carrier lifetime τ differs from the minority carrier lifetime. It is the excess carrier lifetime which is ordinarily measured in experiments designed to detect excess carrier photoconductivity.

Under low-level conditions, it is clear from (10.2-27) and (10.2-28) that in the case of a strongly extrinsic *n*-type semiconductor D^* and μ^* reduce to D_p and μ_p , and for strongly extrinsic *p*-type material, D^* and μ^* reduce to D_n and $-\mu_n$. Combining these results with those expressed in (10.2-30) and (10.2-31), we may write for the extrinsic *n*-type samples

$$D_p \nabla^2(\delta p) - \mu_p \mathbf{E} \cdot \nabla(\delta p) + g' - \frac{\delta p}{\tau_{p_0}} = \frac{\partial(\delta p)}{\partial t} \quad (n_0 \gg p_0, \delta p) \quad (10.2-32)$$

and for extrinsic *p*-type samples

$$D_n \nabla^2(\delta n) + \mu_n \mathbf{E} \cdot \nabla(\delta n) + g' - \frac{\delta n}{\tau_{n_0}} = \frac{\partial(\delta n)}{\partial t} \quad (p_0 \gg n_0, \delta n). \quad (10.2-33)$$

In either case the concentration of the other species is obtained from the relation $\delta p = \delta n$. These equations look very much like continuity equations for *minority* carrier flow. The transport and recombination parameters are in all cases those of the minority carrier. It is often stated that in the above situation one writes down and solves the equation for diffusive flow of *minority* carriers, although this is not precisely true: Equations (10.2-32) and (10.2-33) contain within them in addition an implicit assumption about the behavior of the *majority* carriers.

In a sample for which n_0 and p_0 are not vastly different, and where δp is much less than the *smaller* of (n_0, p_0) , the ambipolar transport coefficients, (10.2-27) and (10.2-28), are constant and can be written as

$$D^* = \frac{(n_0 + p_0)D_n D_p}{n_0 D_n + p_0 D_p} = \frac{(p_0^2 + n_i^2)D_n D_p}{n_i^2 D_n + p_0^2 D_p} \quad (10.2-34)$$

and $\mu^* = \frac{(n_0 - p_0)\mu_n \mu_p}{n_0 \mu_n + p_0 \mu_p} = \frac{(n_i^2 - p_0^2)\mu_n \mu_p}{n_i^2 \mu_n + p_0^2 \mu_p}. \quad (10.2-35)$

The ambipolar diffusion coefficient approaches D_p for $p_0 \ll n_i$ and D_n for $p_0 \gg n_i$.

For intermediate values of p_0 , D^* lies somewhere between these two extreme values. Likewise for $p_0 \ll n_i$, μ^* approaches μ_p , and for $p_0 \gg n_i$, μ^* approaches $-\mu_n$, lying between these limits for intermediate values of p_0 . For material which is precisely intrinsic, $p_0 = n_i$ and the above equations reduce to

$$D^* = D_i^* = \frac{2D_n D_p}{D_n + D_p} \quad \text{and} \quad \mu_i^* = 0, \quad (10.2-36)$$

the continuity Equation (10.2-26) then becoming

$$D_i^* \nabla^2(\delta p) + g' - \frac{\delta p}{\tau} = \frac{\partial(\delta p)}{\partial t}. \quad (10.2-37)$$

Actually, for precisely intrinsic material, it is easily seen from (10.2-27) and (10.2-28) that this is the correct continuity equation for δp regardless of how large δp may be compared with n_0 or p_0 . In the same way it may be seen that this continuity equation is *approximately* correct for near-intrinsic material whenever the density of excess carriers δp is greatly in *excess* of the equilibrium majority carrier density since then μ^* becomes very small. As we shall see in a later chapter, this fact greatly simplifies the analysis of semiconductor device structures under high-current conditions. The fact that μ^* approaches zero in intrinsic samples does not mean that electrons and holes do not acquire a drift velocity when an electric field is applied, but only that an excess carrier concentration distribution does not drift under these circumstances. Electrons and holes drift into and out of such a distribution, although the distribution itself may remain stationary, or nearly so.

We must be careful to note that the electric field \mathbf{E} in Equations (10.2-26), (10.2-32), and (10.2-33) refers to the total electric field as given by (10.2-18), which *includes* the internal field set up by the ambipolar diffusion and drift of the charged particles of the system. We may derive an expression for this field by noting that the total electrical current density \mathbf{I} may be written, with the aid of (10.2-8), (10.2-9), and (10.2-21), as

$$\mathbf{I} = e(\mathbf{J}_p - \mathbf{J}_n) = \sigma \mathbf{E} + e(D_n - D_p) \nabla(\delta p), \quad (10.2-38)$$

where σ is the conductivity as given by (9.7-5). Solving for the field, using the Einstein relations to express the diffusion constants in terms of mobilities, and introducing the mobility ratio b , we may readily obtain

$$\mathbf{E} = \frac{\mathbf{I}}{\sigma} - \frac{kT}{e} \frac{b-1}{nb+p} \nabla(\delta p) = \mathbf{E}_{app} + \mathbf{E}_{int}. \quad (10.2-39)$$

The first term above is clearly the field arising from the application of an external potential source; the second represents the internal field. It is obvious that the internal field, as expressed in the second term, vanishes if there is no gradient of carrier concentration, and also if the hole and electron mobilities are identical ($b = 1$). In the latter instance, (10.2-27) and (10.2-28) give $D^* = D$ and $\mu^* = [(n_0 - p_0)/(n + p)]\mu$, where D and μ are the common hole and electron diffusion coefficient and mobility. Using these values, one may proceed to solve (10.2-26) using for \mathbf{E} the value of the applied field, wasting no further concern about the internal field.

SEC. 10.2

SEC. 10.2

EXCESS CARRIERS IN SEMICONDUCTORS

331

me values.
 $-\mu_n$, lying
; precisely

(10.2-36)

For materials in which the electron and hole mobilities are different, it is clear that the internal field will contribute to the total field E of (10.2-26). We shall see, however, that in the great majority of cases, this contribution can be neglected. The reason for this is that, aside from the effects which have been embodied in the modified transport coefficients D^* and μ^* , the internal field ordinarily exerts only a small *direct* influence upon the spatial distribution of excess carriers, compared to that of diffusion and applied fields. This can be demonstrated by a comparison of the solutions of (10.2-26) where there is diffusion but no field, and when there is a field, but essentially no diffusion. In the former case (assuming a one-dimensional geometry in which δp varies only along the x -direction, that the excess bulk generation rate g' is zero, and that steady-state conditions prevail, so that $\partial(\delta p)/\partial t = 0$) Equation (10.2-26) reduces to

$$\frac{d^2(\delta p)}{dx^2} - \frac{\delta p}{D^*\tau} = 0. \quad (10.2-40)$$

If we seek only solutions which approach zero as x becomes large, we must write

$$\delta p = A e^{-x/(D^*\tau)^{1/2}} \quad (10.2-41)$$

as the solution to (10.2-40). On the other hand, if there is a constant field E , and if diffusive transport may be neglected (10.2-26) may be written, approximately, as

$$\frac{d(\delta p)}{dx} = - \frac{\delta p}{\mu^* E \tau}, \quad (10.2-42)$$

whose solution is

$$\delta p = A' e^{-x/\mu^* E \tau}. \quad (10.2-43)$$

From these equations it is evident that there is a characteristic length $(D^*\tau)^{1/2}$ associated with purely diffusive transport in the steady state, and another characteristic length $\mu^* E \tau$ in this case where the effect of a field is predominant. If the field is zero, the latter length is likewise zero, and it is quite evident from the preceding results that the distribution of excess carriers resulting from purely diffusive transport will not be much affected by the presence of an electric field unless the field is sufficiently large that $\mu^* E \tau$ is comparable in magnitude with the characteristic diffusion length $(D^*\tau)^{1/2}$. In other words, if

$$\mu^* E \tau \ll \sqrt{D^* \tau}, \quad \text{or} \quad E \ll \sqrt{D^*/\mu^{*2} \tau} \quad (10.2-44)$$

then the effect of the electric field on the distribution of excess carrier concentration will be negligible. Inserting the values of D^* and μ^* from (10.2-27) and (10.2-28) and the value of the internal field as given by (10.2-39) into this expression, we may obtain as a condition on the concentration gradient that

$$|\nabla(\delta p)| \ll \frac{1}{\sqrt{D^* \tau}} \frac{(n + p)(nb + p)}{(b - 1)(n_0 - p_0)} \quad (10.2-45)$$

in order that the effect of the internal field be neglected in (10.2-26). In practice one might proceed initially by neglecting the internal field and calculating δp from (10.2-26) substituting for E the applied field. The gradient of δp could then be calculated from the resulting solution and checked against (10.2-45) to insure that that condition is satisfied by the solution everywhere within the region of interest. If (10.2-45) were violated, then the solution would have to be rejected, and a new calculation made in which the internal field is incorporated from the outset. In purely diffusive systems, we shall see that the solution often has the form $\delta p = (\text{const}) e^{-x/L}$ where L is approximately (if not precisely) equal to $(D^* \tau)^{1/2}$. In such instances the gradient of δp is given by $-\delta p/L$, and it is readily seen from the form of (10.2-45) that the condition expressed by this equation is never seriously violated regardless of what values n_0 , p_0 , δp and b may have. Larger values of $\nabla(\delta p)$ may be obtained if an applied field is present, but in such instances, although condition (10.2-45) may no longer be satisfied, it will generally be found that the internal field which is generated is now small compared to the *applied* field which is necessary to maintain this state of affairs. One may thus arrive at the conclusion that it is usually a very good approximation to regard the field E in Equation (10.2-26) as the *applied* field only, neglecting the explicit effect of the internal field. It should be noted, however, that the effect of the internal field is *implicitly* embodied in the modified transport coefficients D^* and μ^* .

The continuity Equations (10.2-16) and (10.2-17) lead directly to the single differential Equation (10.2-26) for δp when the electrical neutrality assumption is made. When an expression for δp has been obtained as a solution of (10.2-26), the validity of the electrical neutrality assumption may easily be checked by computing the *departure* from neutrality required to produce the internal field (10.2-39) and comparing the disparity in charge densities thus obtained with the total excess carrier density obtained from the solution of (10.2-26). Using (9.5-12), we may write Poisson's Equation (10.2-19) in terms of excess densities δn and δp ; using (10.2-20) and (10.2-39) to express the electric field, we obtain

$$\nabla \cdot E_{\text{int}} = \frac{4\pi e(\delta p - \delta n)}{\kappa} = \frac{kT}{e} \frac{b-1}{nb+p} \nabla^2(\delta p). \quad (10.2-46)$$

The ratio of the disparity in densities $\delta p - \delta n$ required to set up the internal field to the actual density δp may then be written as

$$\frac{\delta p - \delta n}{\delta p} = - \frac{\kappa k T}{4\pi e^2} \frac{b-1}{nb+p} \frac{\nabla^2(\delta p)}{\delta p}. \quad (10.2-47)$$

The expression on the right-hand side of this equation must be much less than unity for the charge balance assumption (10.2-21) to be valid. This condition may be expressed in a slightly different way by multiplying numerator and denominator above by n_i . The condition for validity of the charge balance assumption then becomes

$$\left| \frac{\delta p - \delta n}{\delta p} \right| = L_{D_i}^2 \frac{(b-1)n_i}{nb+p} \frac{\nabla^2(\delta p)}{\delta p} \ll 1 \quad (10.2-48)$$

where

$$L_{D_i} = \sqrt{\frac{\kappa k T}{4\pi e^2 n_i}} \quad (10.2-49)$$

SEC. 10.2

SEC. 10.3

EXCESS CARRIERS IN SEMICONDUCTORS

333

practice one
om (10.2-26)
culated from
condition is
0.2-45) were
ion made in
ive systems,
 L is approxi-
ent of δp is
e condition
t values n_0 ,
plied field
ger be satis-
s now small
ffairs. One
on to regard
t implicit effect
ternal field

ingle differ-
in is made.
the validity
puting the
comparing
ier density
Poisson's
d (10.2-39)

(10.2-46)

al field to

(10.2-47)

than unity
ay be ex-
for above
omes

(10.2-48)

(10.2-49)

is a parameter with the dimensions of length which is called the "intrinsic Debye length" associated with the material. At 300°K, the intrinsic Debye length is about 34μ for silicon and 0.96μ for germanium. If, as is very frequently the case, the excess concentration density has the form $\delta p = (\text{const}) e^{\pm x/L}$ where L is a characteristic decay distance, which in purely diffusive systems has the value $(D^* \tau)^{1/2}$, then $\nabla^2(\delta p)/\delta p = 1/L^2$ and (10.2-48) may be expressed as

$$\left| \frac{\delta p - \delta n}{\delta p} \right| = \frac{L_{D_i}^2}{L^2} \frac{(b-1)n_i}{nb+p} \ll 1. \quad (10.2-50)$$

It is clear from this that the assumption of electrical neutrality should be a good one when the characteristic distance L is much larger than the intrinsic Debye length, and where $nb + p$ is much larger than $(b-1)n_i$. These conditions are almost always met in intrinsic or extrinsic samples of silicon and germanium unless the excess carrier lifetime τ is extremely short. In the compound semiconductors of the III-V type, such as InSb and GaAs, carrier lifetimes are usually quite short; for these materials the condition (10.2-50) is often, though not always satisfied. In semiinsulating substances where $nb + p$ as well as L may be very small, condition (10.2-50) is violated more often than not, and other methods of analyzing added carrier problems must be used. The situation then is usually quite complex, although a very general theory for the analysis of such problems has been worked out by van Roosbroeck.⁵ Since the charge balance assumption is valid for the great majority of cases involving the more familiar semiconducting materials, we shall not consider explicitly situations where it is violated.

10.3 SOME USEFUL PARTICULAR SOLUTIONS OF THE CONTINUITY EQUATION

We shall now find it advantageous to examine the form which the solution of the continuity equation takes in certain particularly simple and important cases. To begin, let us examine what happens in a very large uniform crystal, whose boundaries may for practical purposes be assumed to be at infinity, in which there is no applied field nor bulk excess electron-hole pair generation. At time $t = 0$ we shall assume that a *uniform* spatial distribution of excess carrier density has somehow been created. Under these circumstances $\nabla(\delta p) = \nabla^2(\delta p) = 0$, and since the effect of diffusion is always to reduce rather than create concentration gradients, these quantities will remain zero at all later times. Equation (10.2-26) then takes the form

$$\frac{d(\delta p)}{dt} = -\frac{\delta p}{\tau}. \quad (10.3-1)$$

If τ is independent of δp this may easily be solved to give

$$\delta p = p - p_0 = Ae^{-t/\tau} \quad (10.3-2)$$

⁵ W. van Roosbroeck, *Phys. Rev.*, **123**, 474 (1961).

where A is an arbitrary constant, equal here to the value of δp at $t = 0$. The excess carrier density in this case dies away everywhere in an exponential fashion with a time constant τ equal to the excess carrier lifetime.

Another case of great importance is the steady-state solution for an infinite sample wherein excess carriers are created by a uniform plane generation source, which, for convenience, we shall assume coincides with the yz -plane. There is no applied field nor any excess carrier generation in the bulk. Under these conditions, the excess carrier density will vary only along the x -direction. For simplicity, we shall discuss the case of a strongly extrinsic n -type sample in which the excess carrier density is everywhere much less than the majority carrier concentration; under these conditions the transport coefficients D^* , μ^* and τ reduce to the constant coefficients D_p , μ_p and τ_p related to the minority carrier, and (10.2-26) becomes

$$\frac{d^2(\delta p)}{dx^2} - \frac{\delta p}{L_p^2} = 0 \quad (10.3-3)$$

where

$$L_p = \sqrt{D_p \tau_p}. \quad (10.3-4)$$

The solutions to (10.3-3) must everywhere have the general form

$$\delta p = p(x) - p_0 = A e^{x/L_p} + B e^{-x/L_p} \quad (10.3-5)$$

where A and B are arbitrary constants. Since there is a plane source of carriers located at $x = 0$ in the yz -plane, there will be some sort of singularity in the solution there, and it is best to consider the solution for δp in the region $x < 0$ (which we shall call $\delta p_-(x)$), and the solution in the region $x > 0$ (which we shall call $\delta p_+(x)$) separately. Accordingly, then, from (10.3-5), these must have the form

$$\delta p_+(x) = p_+(x) - p_0 = A_+ e^{x/L_p} + B_+ e^{-x/L_p} \quad (x > 0) \quad (10.3-6)$$

$$\text{and} \quad \delta p_-(x) = p_-(x) - p_0 = A_- e^{x/L_p} + B_- e^{-x/L_p} \quad (x < 0). \quad (10.3-7)$$

Since there is a nonzero bulk recombination rate, the excess carrier densities must approach zero far from the source. These boundary conditions require that $A_+ = B_- = 0$. Furthermore, from the geometrical symmetry of the situation it is clear that $\delta p(x)$ must be an even function of x , whereby $A_- = B_+ = A_0$ and

$$\delta p_+(x) = p_+(x) - p_0 = A_0 e^{-x/L_p} \quad (x > 0) \quad (10.3-8)$$

$$\text{and} \quad \delta p_-(x) = p_-(x) - p_0 = A_0 e^{x/L_p} \quad (x < 0). \quad (10.3-9)$$

The excess *electron* concentration, by (10.2-21) is simply equal to the excess hole concentration. The resulting concentration profiles are shown in Figure 10.4. The excess carrier concentration dies off exponentially on either side of the source plane, with a characteristic decay distance of L_p .

The minority carrier diffusion current associated with the excess carrier distribution can be calculated from (10.2-8) to be

$$J_{p+}(x) = -D_p \frac{d(\delta p_+)}{dx} = \frac{A_0 D_p}{L_p} e^{-x/L_p} \quad (x > 0) \quad (10.3-10)$$

SEC. 10.3

The excess
with a time

an infinite
urce, which,
no applied
i, the excess
hall discuss
r density is
conditions
 D_p , μ_p and

(10.3-3)

(10.3-4)

(10.3-5)

ers located
there, and
ll $\delta p_-(x)$,
y. Accord-

(10.3-6)

(10.3-7)

ties must
at $A_+ =$
t is clear

(10.3-8)

(10.3-9)

cess hole
0.4. The
ce plane,
distribu-

(10.3-10)

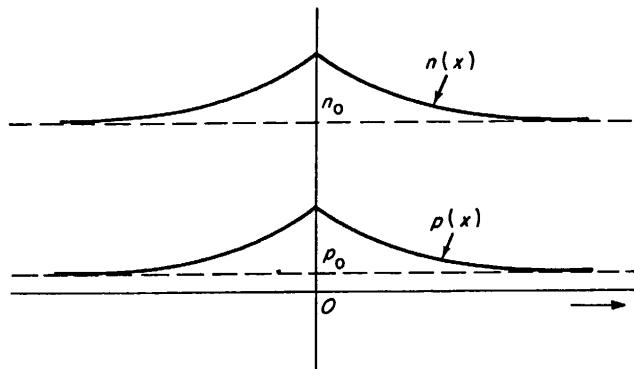


FIGURE 10.4. Steady-state electron and hole concentrations within an infinite crystal in which a plane source of excess carrier generation exists at $x = 0$.

$$J_{p-}(x) = -D_p \frac{d(\delta p_-)}{dx} = -\frac{A_0 D_p}{L_p} e^{x/L_p} \quad (x < 0). \quad (10.3-11)$$

As x approaches zero through positive values, J_{p+} approaches $A_0 D_p / L_p$, while as x approaches zero through negative values, J_{p-} approaches $-A_0 D_p / L_p$. The difference between these quantities is the total flux strength of the source, or the number of minority carriers per cm^2 per second emitted by the source. In other words

$$F_{0p} = J_{p+}(0) - J_{p-}(0) = \frac{2A_0 D_p}{L_p} \quad (10.3-12)$$

where F_0 is the flux source strength. This equation enables one to express the coefficient A_0 of (10.3-8) and (10.3-9) in terms of the flux strength F_0 when the source is described in terms of the initial emitted flux. For a sample which is not strongly extrinsic but in which the excess carrier density is everywhere sufficiently small, the solutions (10.3-8) and (10.3-9) are, of course, still good, provided that L_p is replaced by $L^* = (D^* \tau)^{1/2}$.

It is of some value in understanding the physical nature of ambipolar charge transport to examine in detail the electron and hole fluxes which are present in this particular example. For this purpose we shall consider a sample in which the equilibrium carrier density is arbitrary rather than an extrinsic n -type material. In general, there will be both hole and electron fluxes arising from diffusion and also fluxes of both carrier species which are due to the presence of the internal field as given by (10.2-39). Noting that in the region ($x > 0$) $\delta p_+ = A_0 e^{-x/L^*}$, whereby $d(\delta p_+)/dx = -\delta p_+/L^*$, and using the Einstein relations (10.2-11), the hole and electron fluxes due to diffusion may be written as

$$J_{p+}^d = -D_p \frac{dp_+}{dx} = -D_p \frac{d(\delta p_+)}{dx} = \frac{\mu_p k T}{e} \frac{\delta p_+}{L^*} \quad (10.3-13)$$

$$\text{and} \quad J_{n+}^d = -D_n \frac{dn_+}{dx} = -D_n \frac{d(\delta p_+)}{dx} = \frac{\mu_n k T}{e} \frac{\delta p_+}{L^*}. \quad (10.3-14)$$

The fluxes of holes and electrons which are created by the internal field (10.2-39) are

$$J_{p+}^f = p_+ \mu_p E_{\text{int}} = \frac{kT}{e} \frac{p_+ \mu_p (b - 1)}{n_+ b + p_+} \frac{\delta p_+}{L^*} \quad (10.3-15)$$

and

$$J_{n+}^f = -n_+ \mu_n E_{\text{int}} = -\frac{kT}{e} \frac{n_+ \mu_n (b - 1)}{n_+ b + p_+} \frac{\delta p_+}{L^*}. \quad (10.3-16)$$

The total hole and electron fluxes may now be found simply by adding the diffusion and internal field components together. The result is

$$J_{p+} = J_{n+} = \frac{\mu_p kT}{e} \frac{b(n_+ + p_+)}{n_+ b + p_+} \frac{\delta p_+}{L^*}. \quad (10.3-17)$$

From these equations, the fraction of the total flux of either species of carrier which is due to diffusion and to the presence of the internal field may be calculated very easily. It is found that

$$\begin{aligned} \frac{J_{p+}^d}{J_{p+}} &= \frac{n_+ b + p_+}{b(n_+ + p_+)} & \frac{J_{n+}^d}{J_{n+}} &= \frac{n_+ b + p_+}{n_+ + p_+} \\ \frac{J_{p+}^f}{J_{p+}} &= \frac{p_+ (b - 1)}{b(n_+ + p_+)} & \frac{J_{n+}^f}{J_{n+}} &= -\frac{n_+ (b - 1)}{n_+ + p_+} \end{aligned} \quad (10.3-18)$$

It is most interesting to examine these expressions in three specific cases; extrinsic *n*-type material, where $n_+ \gg p_+$, extrinsic *p*-type material, where $p_+ \gg n_+$ and intrinsic, where $n_+ = p_+$. In these three instances the Equations (10.3-18) become

<i>n</i> -type ($n_+ \gg p_+$)	<i>p</i> -type ($p_+ \gg n_+$)	intrinsic ($n_+ = p_+$)
$\frac{J_{p+}^d}{J_{p+}} = 1;$	$\frac{1}{b};$	$\frac{b+1}{2b}$
$\frac{J_{p+}^f}{J_{p+}} = 0;$	$1 - \frac{1}{b};$	$\frac{b-1}{2b}$
$\frac{J_{n+}^d}{J_{n+}} = b;$	$1;$	$\frac{b+1}{2}$
$\frac{J_{n+}^f}{J_{n+}} = 1 - b;$	$0;$	$\frac{1-b}{2}.$

(10.3-19)

It is clear from these results that in extrinsic samples the flux of *minority* carriers due to the internal field is *negligible in comparison with diffusion flux*. The *majority* carrier flux due to the field, however, is significant. This flux must, in fact, be large enough to assure quasineutrality at all points, and therefore must compensate for the imbalance between minority and majority carrier diffusion fluxes arising from the

SEC. 10.3

SEC. 10.3

EXCESS CARRIERS IN SEMICONDUCTORS

337

[0.2-39) are

(10.3-15)

(10.3-16)

the diffusion

(10.3-17)

es of carrier
e calculated

(10.3-18)

s; extrinsic
id intrinsic,

(10.3-19)

ty carriers
e majority
, be large
nsate for
from the

inequality of electron and hole diffusion coefficients. The situation may be understood by noting from (10.2-39) that in extrinsic samples the magnitude of the external field becomes very small because of the factor $nb + p$ in the denominator, which becomes quite large in extrinsic samples. The majority carrier flux arising from the internal field, however, is still appreciable, since it is proportional to the *product* of the very small internal field and the very large majority carrier density. The minority carrier flux due to the internal field, of course, is extremely small indeed, since both internal field and minority carrier density are small. The diffusion fluxes of majority and minority carriers, on the other hand, are dependent only upon the concentration gradients and the inherent electron and hole diffusion coefficients, and since the concentration gradients of electrons and holes are equal, must be comparable in magnitude, differing only by an amount proportional to the difference between the inherent diffusivities of the two carriers.

An understanding of this situation leads to the explanation of why, in extrinsic samples, the ambipolar diffusion coefficient is simply the inherent diffusion coefficient of the minority carrier. Under such conditions the internal field is small; it may create relatively large majority carrier fluxes, but its influence upon minority carrier flux is negligible. The minority carrier flux is therefore exclusively due to the *diffusion* of the minority carriers and is essentially unaffected by the internal field. The *observed* minority carrier flux is then simply that which would be produced by *independently diffusing minority carriers*. In order that the quasineutrality condition be satisfied, the majority carrier flux must be *the same*. Some of this flux is supplied by the diffusion of the majority carriers, but since the inherent diffusivities of the two carrier species are unequal, there will be an inevitable deficit (or excess). This deficit or excess may be accounted for by *majority* carrier flux whose source is the internal field. The net effect is that electron and hole fluxes are equal to each other and to the flux which would be produced by a distribution of independently diffusing minority carriers; the diffusion coefficient of the excess carrier distribution is thus that associated with the minority carrier. In extrinsic systems, it is therefore usually necessary to consider explicitly only the behavior of the minority carriers. In samples which are intrinsic or nearly intrinsic, of course, the effect of the internal field upon both carrier types is important. This fact is illustrated by (10.3-19) and by the dependence of the ambipolar transport coefficients upon both electron and hole diffusivity or mobility. Although these results have been discussed in connection with a rather restricted specific example, it is clear that the physical principles involved are the same in all systems and that the conclusions which are drawn are of general applicability.

The minority carrier diffusion fluxes (10.3-10) and (10.3-11) are now seen to be the *total* minority carrier fluxes in the case of an *extrinsic n-type sample*. In a crystal which is not extrinsic, the total electron or hole fluxes can be obtained from (10.3-17), by writing $b = \mu_n/\mu_p$ and using the Einstein relations, in the form

$$J_{p+} = J_{n+} = \frac{(n_+ + p_+)D_n D_p}{n_+ D_n + p_+ D_p} \cdot \frac{\delta p_+}{L^*} = -D^* \frac{d(\delta p_+)}{dx}. \quad (10.3-20)$$

In the example at hand, of course, since there is no applied field, the total electrical current $I = e(J_{p+} - J_{n+})$ is zero.

Next let us consider the extension of the preceding solution to the case where there is a *constant* applied field E_0 in the positive x -direction. The differential equation

(10.2-26) then may be written for extrinsic *n*-type material as

$$\frac{d^2(\delta p)}{dx^2} - \frac{\mu_p E_0}{D_p} \frac{d(\delta p)}{dx} - \frac{\delta p}{L_p^2} = 0. \quad (10.3-21)$$

This equation may be solved using the standard techniques applicable to linear differential equations with constant coefficients to give

$$\delta p(x) = p(x) - p_0 = A e^{\gamma_p + x/L_p} + B e^{\gamma_p - x/L_p} \quad (10.3-22)$$

where

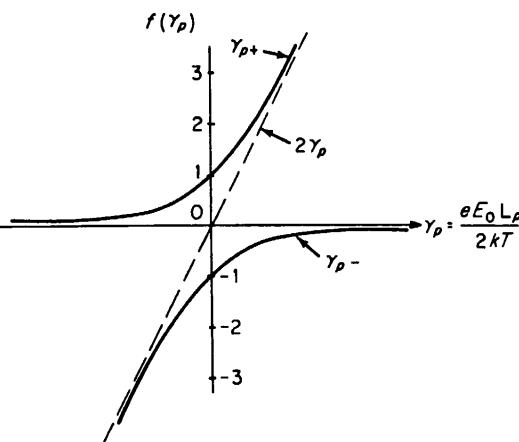
$$\gamma_{p\pm} = \gamma_p \pm \sqrt{1 + \gamma_p^2} \quad (10.3-23)$$

and

$$\gamma_p = \mu_p E_0 L_p / 2 D_p = e E_0 L_p / 2 k T. \quad (10.3-24)$$

This solution is the same as (10.3-5), except that the arguments of the exponential factors are multiplied by the factors γ_{p+} and $-\gamma_{p-}$, which depend upon the electric field. From the definition given above, it is readily seen that whatever value E_0 may have, γ_{p+} must always be positive while γ_{p-} must always be negative; when γ_{p+} and γ_{p-} are plotted as a function of field, the two quantities are represented by the two branches of a hyperbola, as shown in Figure 10.5. For $\gamma_p = 0$, corresponding to zero

FIGURE 10.5. A plot of the functions γ_{p+} and γ_{p-} .



field, (10.3-22) reduces to Equation (10.3-5) of the preceding section. For large positive values of γ_p (hence for fields such that $E_0 \gg 2kT/eL_p$), it may be shown with the aid of the binomial expansion and (10.3-24) that

$$\gamma_{p-} \approx -\frac{1}{2\gamma_p} = -\frac{L_p}{\mu_p E_0 \tau_p}, \quad (10.3-25)$$

while it is obvious that under these circumstances

$$\gamma_{p+} \approx 2\gamma_p = \mu_p E_0 \tau_p / L_p. \quad (10.3-26)$$

It is easily seen from (10.3-23) that γ_{p+} and γ_{p-} are always related by

$$\gamma_{p+}\gamma_{p-} = -1. \quad (10.3-27)$$

SEC. 10.3

SEC. 10.3

EXCESS CARRIERS IN SEMICONDUCTORS

339

We may now repeat the development of the zero-field case as given by Equations (10.3-6)–(10.3-12), assuming as before two separate expressions for $\delta p_+(x)$ in the region $x > 0$ and $\delta p_-(x)$ in the region $x < 0$, with coefficients A_+ , B_+ , A_- and B_- , and showing in the same way that A_+ and B_- must be taken as zero in order to insure proper behavior at $\pm\infty$. The resulting solution may be written

(10.3-21)

$$\delta p_+(x) = p_+(x) - p_0 = A_0 e^{\gamma_p - x/L_p} \quad (x > 0) \quad (10.3-28)$$

le to linear

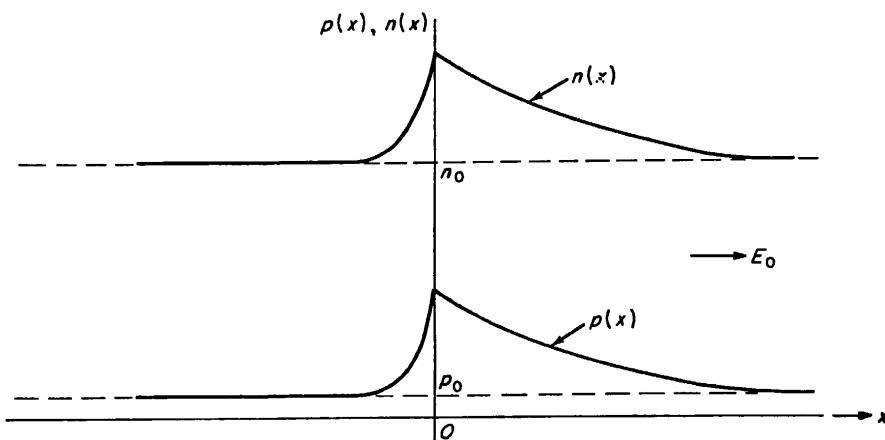


FIGURE 10.6. Steady-state electron and hole concentrations within a crystal in which a plane source of excess carriers exists at $x = 0$, and within which a constant electric field in the x -direction is present.

value of E_0 . Since the transport characteristics of the excess carrier distribution are essentially those of the minority carriers, we find that the transport of the excess carriers to the right (in the direction of motion of holes under the applied field) is aided by the action of the field, while the transport of carriers in the opposite direction is retarded. From (10.3-25) one may verify that (10.3-28) approaches the previously obtained result (10.2-43) for large applied fields, as might be expected. The minority carrier particle fluxes $J_{p+}(x)$ and $J_{p-}(x)$ may, as before, be calculated from (10.2-8); contributions are now obtained from both diffusion and electric field terms. Part of the resulting particle flux will arise from the drift of the holes which are present in thermal equilibrium in the applied field, and part from the diffusion and drift of the excess carrier distribution. The flux arising from the latter effect, evaluated at the source plane, is readily shown to be

(10.3-26)

$$F_{op} = J_{p+}(0) - J_{p-}(0) = A_0 \left[2\mu_p E_0 + \frac{D_p}{L_p} (\gamma_{p+} - \gamma_{p-}) \right]. \quad (10.3-30)$$

(10.3-27)

This equation is the analog of (10.3-12), which relates the amplitude coefficient A_0

to the total number of carriers per unit area per unit time, F_{0p} , generated at the source plane. When the sample is not strongly extrinsic, and when the added carrier density is sufficiently small, the above solutions will still be correct, provided that D_p , μ_p , L_p and τ_p are replaced by D^* , μ^* , L^* and τ , respectively.

The same remarks regarding the relation of fluxes due to diffusion and to the *internal* fields in connection with the field-free case apply also in the present situation; for example, in an extrinsic sample, the flux of minority carriers due to the internal field is negligible in comparison with the corresponding diffusion flux, while the majority carrier fluxes attributable to either source are of comparable magnitude. There are now, in addition, however, large fluxes of both carrier types, comprising both excess and equilibrium carriers, due to the *applied* field. These fluxes are superposed upon the fluxes arising from diffusion and internal fields; as a result the total electric current density no longer vanishes but is instead represented by σE_{appl} . It must be noted in this connection that the present example deals only with a situation wherein δp is everywhere much less than the equilibrium majority carrier density. Since the total electric current density must be everywhere the same, a situation wherein δp is large enough to appreciably modulate the conductivity would imply a field which could not be constant, but which would instead be large where δp is small and *vice versa*. For small values of δp the same difficulty is still present, but the corresponding variations in the field are so small that the constant-field treatment is a good approximation. There are also, of course, numerous other difficulties which arise when δp becomes comparable to the equilibrium carrier density.

As a final example, let us consider a uniform one-dimensional system in which N hole-electron pairs are generated instantaneously at a point $x = x'$ at time $t = t'$. The sample will be assumed to extend to infinity in both directions along the x -axis, and a constant applied electric field E_0 is assumed to be present. For an extrinsic *n*-type sample, Equation (10.2-26) then becomes

$$D_p \frac{\partial^2(\delta p)}{\partial x^2} - \mu_p E_0 \frac{\partial(\delta p)}{\partial x} - \frac{\delta p}{\tau_p} = \frac{\partial(\delta p)}{\partial t}. \quad (10.3-31)$$

A solution to this equation valid for all $t > t'$ which meets all the requirements of the problem is

$$\delta p(x,t) = \frac{N}{\sqrt{4\pi D_p(t-t')}} \exp\left[-\frac{(x-x')-\mu_p E_0(t-t')}{4D_p(t-t')} - \frac{t-t'}{\tau_p}\right]. \quad (10.3-32)$$

The fact that (10.3-32) satisfies the differential Equation (10.3-31) can be established by direct substitution.⁶ It is readily established that the initial value $\delta p(x,0)$ is zero everywhere except at $x = x'$, at which point it is infinite. The initial concentration distribution thus corresponds to a Dirac δ -function. For later times the distribution has a Gaussian shape, whose half-width increases with time and whose maximum amplitude decreases with time. The maximum concentration point also moves along the field direction with velocity $\mu_p E_0$. The expression (10.3-32) can be integrated over x between the limits $-\infty$ and ∞ (by letting $u = (x-x') - \mu_p E_0(t-t')$, whence $du = dx$, t being regarded as constant) to obtain the total number of excess holes δP

⁶ The solution (10.3-32) can be obtained by Fourier integral or Laplace transform techniques.

re source
r density
 ν_p, μ_p, L_p

d to the
tuation;
internal
while the
gnitude.
nprising
e super-
he total
 E_{appl} . It
tuation
density,
where-
a field
all and
corres-
a good
h arise

which
 $t = t'$.
x-axis,
intrinsic

3.3-31)

of the

3.3-32)

lished
; zero
ation
ution
num
along
over
ience
es δP

in the distribution as a function of time. The result is

$$\delta P(t) = \int_{-\infty}^{\infty} \delta p(x,t) dt = N e^{-(t-t')/\tau_p}. \quad (10.3-33)$$

The total number of carriers, initially N , thus falls off exponentially with time constant τ_p , as might be expected. The behavior of the excess carrier concentration as a function of x and t is shown in Figure 10.7.

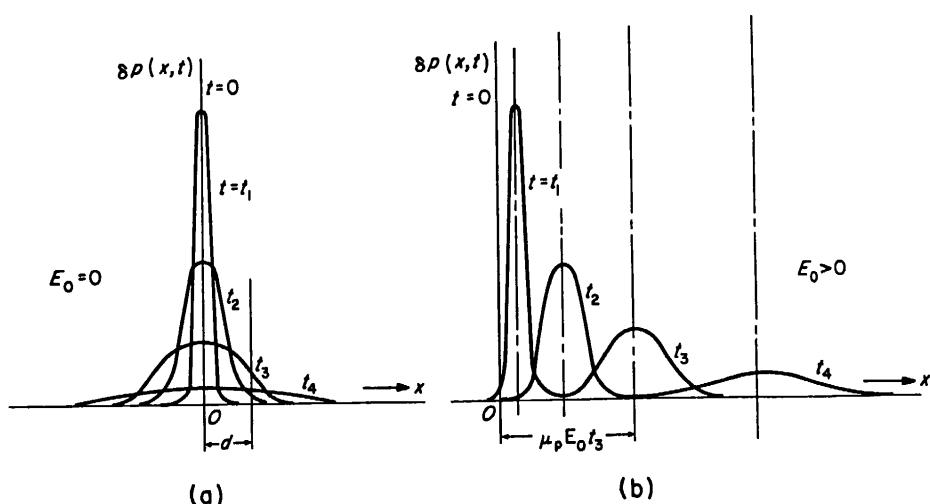


FIGURE 10.7. (a) Diffusive spread of excess carriers injected initially as a δ -function distribution at $x = 0$, with no applied electric field. (b) Diffusive spread and drift of the distribution in (a) with a constant electric field in the x -direction. In both cases $0 < t_1 < t_2 < t_3 < t_4$.

If, now, we revise our definition of N , the initial source strength, by regarding $N(x',t')$ $dx' dt'$ as the number of excess holes or electrons generated in a time interval dt' about $t = t'$ within a space element dx' about $x = x'$, then the subsequent concentration due to this *element* of generation as a function of x and t will be obtained by simply replacing N in (10.3-32) with $N(x',t') dx' dt'$; the resulting concentration distribution should then be labeled as a function of x' and t' as well as x and t , that is, it should be written as $\delta p(x,t;x',t')$. The concentration $\delta p(x,t)$ arising from a *known arbitrary distribution of generation* whose strength may vary with space and time coordinates x' and t' may then be calculated by *superposing* solutions of this form, each of which is characterized by the *local* generation strength $N(x',t') dx' dt'$. This superposition of infinitesimal elements, each with its own local source strength leads to an integral over dx' and dt' of the form

$$\delta p(x,t) = \int_{-\infty}^t \int_{-\infty}^{\infty} \frac{N(x',t')}{\sqrt{4\pi D_p(t-t')}} \exp\left[-\frac{(x-x') - \mu_p E_0(t-t')}{4D_p(t-t')} - \frac{t-t'}{\tau_p}\right] dx' dt'. \quad (10.3-34)$$

The time integral is carried only up to $t' = t$ because, of course, *future* generation does

not contribute to the concentration observed at time t . The above solution, naturally, applies only to an infinite sample, in which the concentration distribution satisfies the boundary condition that the concentration must approach zero far from the source. For finite samples, one would first have to find a solution for a δ -function initial source distribution, such as (10.3-32), but satisfying in addition appropriate boundary conditions at the surfaces of the sample.⁷ The same superposition procedure could then be used to obtain the solution for an arbitrary distribution of generation, with respect to the space and time coordinates, for samples of this sort.

This procedure can be fairly easily generalized to a three-dimensional system. The solution for an instantaneous generation of $N(\mathbf{r}', t') dv' dt'$ in the volume element dv' about the point $\mathbf{r} = \mathbf{r}'$ at the time $t = t'$, describing the subsequent concentration distribution arising from this initial generation element in an infinite sample, is

$$\delta p(\mathbf{r}, t; \mathbf{r}', t') = \frac{N(\mathbf{r}', t') dv' dt'}{[4\pi D_p(t-t')]^{\frac{3}{2}}} \exp\left[-\frac{\{(\mathbf{r} - \mathbf{r}') - \mu_p \mathbf{E}_0(t-t')\}^2}{4D_p(t-t')} - \frac{t-t'}{\tau_p}\right]. \quad (10.3-35)$$

It is possible in this case also to verify that (10.3-35) is a solution to the three-dimensional diffusion equation corresponding to (10.3-31) by direct substitution. The superposition of solutions of the form (10.3-35) may now be made in the same way as before, resulting in an integral over the generation source distribution of the form

$$\delta p(\mathbf{r}, t) = \int_{-\infty}^t \int_V \frac{N(\mathbf{r}', t')}{[4\pi D_p(t-t')]^{\frac{3}{2}}} \exp\left[-\frac{\{(\mathbf{r} - \mathbf{r}') - \mu_p \mathbf{E}_0(t-t')\}^2}{4D_p(t-t')} - \frac{t-t'}{\tau_p}\right] dv' dt'.$$

We shall find most of the particular solutions of the continuity equation which have been described in this section useful in discussing experimental work involving excess carrier distributions or semiconductor device structures of technological importance.

10.4 DRIFT MOBILITY AND THE HAYNES-SHOCKLEY EXPERIMENT

The Haynes-Shockley experiment was first performed in order to measure accurately and directly the drift mobility of holes and electrons in semiconductor crystals.⁸ Although it is of great interest because it was the first really successful *direct* measurement⁹ of charge carrier drift velocity, it is also very instructive in illustrating the transport behavior of excess electrons and holes in semiconductors, and also in demonstrating the basic features of transistor action, since the Haynes-Shockley circuit is in effect a type of transistor.

In the Haynes-Shockley experiment, as illustrated in Figure 10.8, excess minority carriers are injected into and collected from the crystal by metallic point probe

⁷ The solution for a δ -function initial source distribution is ordinarily referred to as the *Green's function* appropriate to a given system and a given set of boundary conditions.

⁸ J. R. Haynes and W. Shockley, *Phys. Rev.*, **81**, 835 (1951).

⁹ As contrasted with the Hall effect, which allows only an indirect determination of mobility.

naturally, it satisfies the he source. tial source dary con could then th respect

il system.
e element
entrance
, is

(10.3-35)

ie three-
ion. The
e way as
orm

$dv'dt'$.

ich have
g excess
ortance.

curately
ystals.⁸
measure-
e trans-
also in
ockley

minority
probe

Green's

bility.

contacts. Although we must postpone a detailed account of what happens at a rectifying metal point contact until later, a general idea of their behavior may be obtained by referring to Figure 10.9. In this illustration a semiconductor sample is depicted, to which a point probe contact and a large area contact have been made. The large area contact, which may be a suitably fabricated soldered or alloyed region or simply a pressure contact acts simply as an *ohmic* contact to the sample, and does not affect

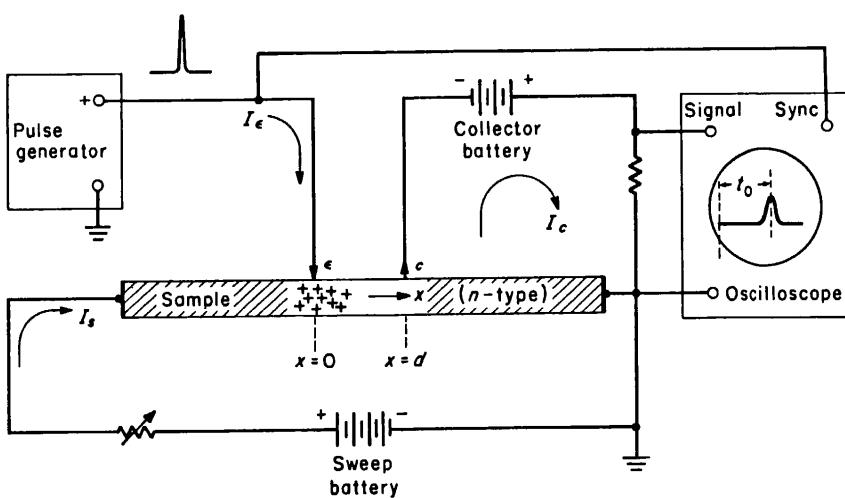


FIGURE 10.8. A schematic diagram of the Haynes-Shockley experiment.

the distribution of carriers in the sample in any appreciable way. The point probe, however, acts as a rectifying contact; when biased in such a way as to attract minority carriers to it, it will act as a collector of minority carriers, depleting the region of the crystal adjacent to it of these carriers until a steady state is reached wherein minority carriers diffuse from the interior of the crystal just rapidly enough to supply the deficit caused by their disappearance at the probe point. Under these conditions (called *reverse bias*) this minority carrier current is the only current which can flow through the probe, and since the supply of minority carriers in the crystal is quite small, the current flow is severely limited, and is relatively independent of the bias voltage because the supply of minority carriers is unaffected by changing the reverse bias voltage. This situation is illustrated in Figure 10.9(a) and (c) for *n*- and *p*-type semiconducting materials. When the bias voltage is reversed, so as to attract majority carriers to the probe, it is found that in addition to the flow of majority carriers into the probe, excess *minority* carriers are *injected* into the sample at the point contact. In this mode of operation current flows quite easily, and relatively large currents can be made to flow with small bias voltages. The current is found to increase approximately exponentially as the bias voltage is increased, and it is not difficult to produce a sufficient number of injected carriers to modulate the conductivity of the semiconductor quite strongly.¹⁰ This forward bias condition is illustrated in Figure 10.9(b) and (d). The

¹⁰ It should be remembered that an excess concentration of majority carriers must always accompany the excess minority carrier distribution. The majority carriers necessary to maintain electrical neutrality are in plentiful supply in the bulk of the crystal, whence they are drawn as the excess minority carrier distribution is formed. Any resulting deficit in the majority carrier concentration in the bulk is immediately made up by majority carriers drawn from the metallic ohmic contact.

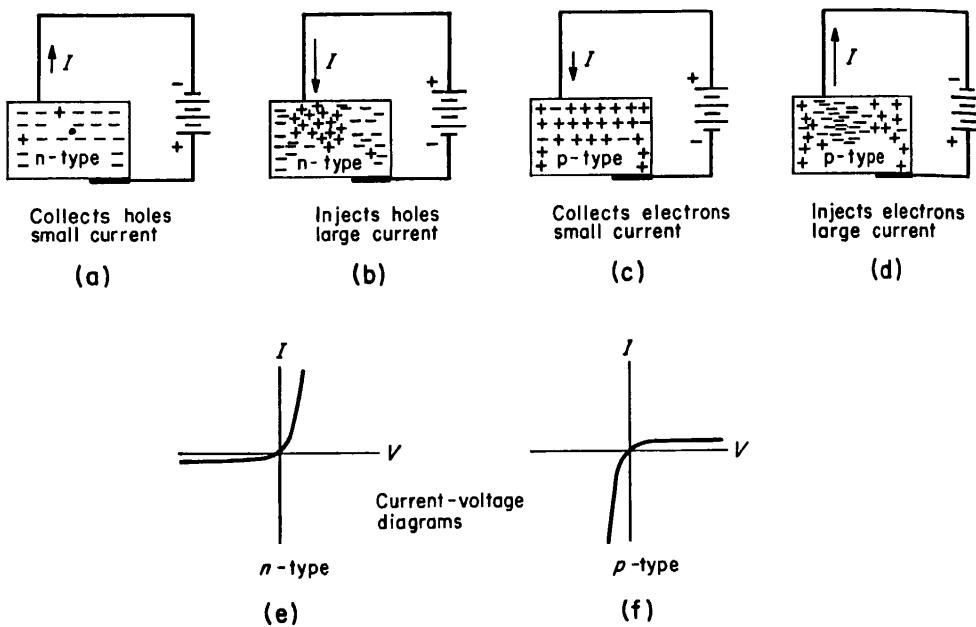


FIGURE 10.9. A point contact semiconductor rectifier (a) *n*-type, reverse bias, (b) *n*-type, forward bias, (c) *p*-type, reverse bias, (d) *p*-type, forward bias. (e) Current-voltage characteristic of *n*-type semi-conductor point contact rectifier, (f) Current-voltage characteristic of *p*-type semiconductor point contact rectifier.

current-voltage relationships for *n*-type and *p*-type crystals are shown in the same set of diagrams at (e) and (f), respectively. The point contact rectification effect, as shown at (e) and (f), in fact, provides quite a simple test for distinguishing between *n*-type and *p*-type semiconductor crystals.

In the Haynes-Shockley experiment a pulse of minority carriers is injected at the emitter probe e of Figure 10.8. The sample is shaped in the form of a long thin bar, and an electric field is set up by an external battery, which, of course, causes a sweep current I_s to flow in the sample. This electric field sweeps the injected minority carriers down the sample past a second electrode, positioned a known distance d away from the first, which is reverse-biased so that it acts as a collector of minority carriers. When there are no excess carriers present at the collector, the only current flowing in that electrode is a small saturation current which is due to the collection of equilibrium minority carriers by the probe point. As the distribution of injected minority carriers flows past the collector, the concentration of minority carriers in its neighborhood increases, and the number of minority carriers collected there per unit time goes up proportionally. If the oscilloscope trace is initiated at the time when the pulse of carriers is originally injected, nothing will be observed on the screen until the excess carrier distribution has arrived at the collector, at which time the collector current will increase and a signal pulse will be observed on the oscilloscope.

Since the excess carrier concentration as a function of the distance x along the sample is essentially that given by (10.3-32), and shown graphically in Figure 10.7(b) it is clear that the maximum of the concentration distribution moves along with velocity equal to $\mu^* E_0$, where μ^* is the ambipolar mobility and E_0 the applied field. (In (10.3-32) conditions are assumed to be such that D^* and μ^* equal D_p and μ_p , but this may not be generally true.) It is apparent that the time t_0 required for the maxi-

mum of the concentration pulse to traverse the known distance d between emitter and collector is just

$$t_0 = \frac{d}{v} = \frac{d}{\mu^* E_0} \quad (10.4-1)$$

giving $\mu^* = \frac{d}{t_0 E_0}$. (10.4-2)

If conditions are such that μ^* equals $-\mu_n$ or μ_p (i.e., if the sample is strongly extrinsic) this experiment allows one to measure accurately the minority carrier drift mobility.

Unfortunately, the pulse that arrives at the collector is not perfectly sharp, but is rather spread out due to diffusion, as illustrated in Figure 10.7(b). This causes no difficulty in the measurements, but leads to certain errors in the interpretation of the data unless great care is taken. The reason for this is that what is actually displayed on the screen of the oscilloscope is *not* a plot of the concentration *versus* distance at some fixed time, but rather a plot of the concentration at some fixed point as a function of time, $\delta p(d,t)$. A measurement to the maximum of this curve does not accurately represent the drift time unless the conditions of the experiment are arranged so that a minimum of diffusive spread takes place as the distribution drifts between emitter and collector. For example, suppose that the applied field were zero. Then the distribution of excess carriers would diffuse as shown in Figure 10.7(a). Even though there is no applied field, however, it is readily seen from the figure that a concentration maximum would be observed as a function of time at a point $x = d$; in the example shown there the maximum would occur at about $t = t_3$. This maximum is observed simply because diffusive transport alone is sufficient to move the excess carriers belonging to the distribution from emitter to collector, and then beyond. This effect leads to the observation of a finite drift time in the absence of a field, which, according to (10.4-2) would correspond to infinite mobility. Obviously, then, if what is *meant* by the drift time t_0 is the time elapsed between injection and the observation of the maximum value attained by $\delta p(d,t)$ as given by (10.3-32) (and which is what is in fact measured from the oscilloscope trace) then Equation (10.4-2) cannot be correct as it stands.

Actually, Equation (10.4-2) is approximately correct so long as the diffusive spread of the carrier distribution is reasonably small during the transit from emitter to collector. These conditions can be realized experimentally by the use of very large applied fields, which, however, lead to serious heating effects unless rather cumbersome pulse techniques are resorted to. Alternatively, one may derive a correct expression for the transit time from (10.4-2) which takes diffusion fully into account, and thus arrive at a modified expression¹¹ of the form

$$\mu^* = \frac{d}{t_0 E_0} (\sqrt{1 + x^2} - x) \quad (10.4-3)$$

where $x = \frac{2kT}{eE_0 d} \left(\frac{t_0}{\tau} + \frac{1}{2} \right)$. (10.4-4)

¹¹ J. P. McKelvey, *J. Appl. Phys.*, 27, 341 (1956).

It is necessary, of course, to know the excess carrier lifetime τ to use this formula, but, as we shall see, that quantity is easily determined by a simple independent measurement. The details involved in the calculation of (10.4-3) are assigned as an exercise for the reader.

It will be shown in Chapter 13 how transistor action can be explained and understood on the basis of the Haynes-Shockley experiment. By measuring both the diffusive spread and the transit time, it is possible to determine the diffusion coefficient D^* as well as the drift mobility by the Haynes-Shockley technique. This possibility has been exploited¹² to verify the Einstein relation experimentally in strongly extrinsic samples of germanium.

10.5 SURFACE RECOMBINATION AND THE SURFACE BOUNDARY CONDITION

If the effect of the surfaces of a finite crystal were solely to confine the charge carriers to the interior of the sample, then the boundary condition which the excess charge carrier distribution would have to satisfy at the sample surfaces would simply be that the electron and hole currents must vanish at the surfaces. Unfortunately, the situation is not quite as simple as this, because charge carriers may *recombine* at the surface, by mechanisms which are quite independent of those which regulate charge carrier recombination rates in the interior of the sample. Under these circumstances, the surface acts as a partial absorber for electrons and holes, and there may be net current flow toward the surfaces of the sample.

At first thought one might be tempted to conclude that since electrons and holes may recombine at the surface of a crystal, there should be a deficiency of charge carrier concentration near the surface which would result in a diffusive current of carriers toward the surface *even in the thermal equilibrium state*. This, however, is not true, because in addition to recombination, *thermal generation* of electron-hole pairs takes place at the surface, and in thermal equilibrium the generation rate precisely equals the rate at which carrier pairs recombine at the surface. There is, therefore, no net flux toward the surface, nor any change in concentrations in the surface region under thermal equilibrium conditions.¹³ This situation is a specific example of a very general principle of statistical mechanics, called the principle of *detailed balancing* or the principle of *microscopic reversibility*. This principle states that *in the thermal equilibrium condition* any given microscopic process and the reverse process must proceed at the same rate. We have already encountered an example of the validity of this principle when we ascertained that the equilibrium bulk recombination rate of electron-hole pairs and the bulk thermal generation rate were equal. The present

¹² Transistor Teachers Summer School (Bell Telephone Laboratories), *Phys. Rev.*, **88**, 1368 (1952).

¹³ This statement is true for the simple model of the surface which we are now discussing. There are changes in the equilibrium carrier concentration near the surface, associated with a surface space charge layer which arises because of the presence of *surface states*, at the actual physical surface of the sample, as we shall see later. The simple viewpoint adopted here is generally adequate for a phenomenological discussion of surface recombination.

SEC. 10.4

SEC. 10.5

EXCESS CARRIERS IN SEMICONDUCTORS

347

ula, but,
measure-
exercise

I under-
stand diffu-
sion D^*
ility has
xtrinsic

carriers
charge
be that
ly, the
at the
charge
tances,
be net

d holes
carrier
carriers
it true,
s takes
equals
no net
under
a very
sing or
hermal
must
alidity
rate of
resent

8, 1368

There
e space
face of
e for a

situation, in which the surface recombination rate and the surface thermal generation rate are equal at equilibrium, is simply illustrative of the same general law. It should be noted, of course, that the principle of detailed balance applies *only* to the thermal equilibrium state. In the present example, the surface thermal generation rate is a function only of the temperature, and is quite independent of local charge carrier concentrations, while the recombination rate clearly depends directly upon the local carrier concentrations. If the local charge carrier concentration exceeds the thermal equilibrium value, then the surface recombination rate will exceed the surface generation rate, and a net absorption of carriers by the surface will result, which in turn will set up a diffusive flow of carriers toward the surface. A depletion of excess carrier concentration in the neighborhood of the surface will likewise result in a diffusive flow of excess thermally generated carriers away from the surface. Similar remarks can, of course, be made regarding the flow of carriers in the bulk arising from local excesses or deficiencies in bulk carrier concentrations. It is important to realize that these general conclusions are valid independent of the precise mechanisms involved in the generation or recombination processes.

The effect of surface recombination upon the charge carrier distribution within the sample may now be investigated by considering the flux interchange between the surface and interior region of the sample, as illustrated by Figure 10.10. We shall

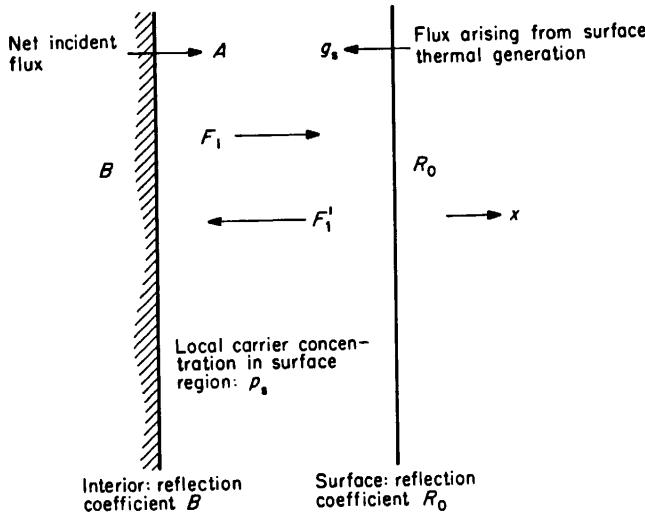


FIGURE 10.10. Flux diagram illustrating flux interchange between bulk and surface.

begin by defining the *surface reflection coefficient* R_0 as the probability that a single carrier in a single collision with the surface will be sent back to the bulk rather than absorbed by recombination. The corresponding recombination probability is thus $1 - R_0$. Likewise, the *bulk reflection coefficient* B will be defined as the probability that a carrier, upon entering the interior of the crystal from the surface, will reappear at the surface in the course of its random wandering through the material before recombining in the interior. These coefficients are assumed to be independent of carrier fluxes or concentrations. We may then consider the system of fluxes set up

between surface and interior shown in Figure 10.10. A net flux of particles of magnitude A is assumed to originate from the interior, and the surface generation flux g_s is directed from the surface toward the interior of the crystal. The surface of the crystal in Figure 10.10 is shown separated from the interior to clearly illustrate the flux interchange between the two, but, of course, actually the two regions are physically contiguous. The total flux F_1 flowing from interior to surface, and the total reverse flux F'_1 can be obtained by observing that F_1 is made up of the incident flux A plus that part of F'_1 which is reflected by the bulk, whereby

$$F_1 = A + BF'_1. \quad (10.5-1)$$

Similarly, $F'_1 = g_s + R_0 F_1.$ (10.5-2)

These two equations can be solved for F_1 and F'_1 , giving

$$F_1 = \frac{A + Bg_s}{1 - R_0 B} \quad (10.5-3)$$

and $F'_1 = \frac{g_s + AR_0}{1 - R_0 B}.$ (10.5-4)

Now it is easily proved that for a distribution of free particles which obey Boltzmann statistics, the number of particles per unit time crossing a plane surface of unit area in either direction is just $p\bar{c}/4$ where p is the local particle concentration.¹⁴ In the thermal equilibrium state, the particle concentration is p_0 everywhere and

$$F_1 = F'_1 = p_0\bar{c}/4. \quad (10.5-5)$$

This value of F_1 and F'_1 may be substituted into (10.5-3) and (10.5-4) and the resulting pair of equations solved for g_s and A_0 , which is the value of the incident flux A under conditions of thermal equilibrium, to yield

$$g_s = \frac{p_0\bar{c}}{4} (1 - R_0) \quad (10.5-6)$$

and $A_0 = \frac{p_0\bar{c}}{4} (1 - B).$ (10.5-7)

We have seen in the preceding argument that in the thermal equilibrium state F_1 and F'_1 have the common value $p\bar{c}/4$. If there is a departure from thermal equilibrium resulting from a diffusive flow set up by a concentration gradient, and if this departure from the equilibrium state is small enough that the Boltzmann distribution is still approximately correct at every point in the system, then F_1 and F'_1 will no longer be equal. Under these circumstances, we must regard the positively directed flux arriving at a point x_0 as having originated, on the average, a distance of the order of the mean free path "upstream" and the negatively directed flux arriving there as having originated on the average a distance of the order of the mean free path "downstream."

¹⁴ See, for example, Exercise 7 at the end of Chapter 5.

SEC. 10.5

SEC. 10.5

EXCESS CARRIERS IN SEMICONDUCTORS

349

magni-
lux g_s is
; crystal
ix inter-
ysically
reverse
 A plus

Then the positively directed flux F'_1 may be written

$$F'_1 = \left[p(x_0) - \alpha \lambda \left(\frac{\partial p}{\partial x} \right)_{x_0} \right] \frac{\bar{c}}{4}. \quad (10.5-8)$$

where α is a numerical constant of the order of unity, because the quantity in brackets represents, approximately, the local concentration at the place where the flux arriving at x_0 originated. Likewise, the negatively directed flux F'_1 may be written as

10.5-1)

$$F'_1 = \left[p(x_0) + \alpha \lambda \left(\frac{\partial p}{\partial x} \right)_{x_0} \right] \frac{\bar{c}}{4}. \quad (10.5-9)$$

10.5-2)

It is clear that the *sum* of the fluxes then must be

10.5-3)

$$F_1 + F'_1 = \frac{p\bar{c}}{2}, \quad (10.5-10)$$

10.5-4)

where p is the local concentration, independent of the concentration gradient, so long as the distribution function is not seriously disturbed from its equilibrium form.

zmann

it area

In the

0.5-5)

Using (10.5-10) to express the sum of the fluxes F_1 and F'_1 in the surface region, and using the expressions (10.5-3), (10.5-4) and (10.5-6) to represent F_1 , F'_1 and g_s , we may obtain

$$\frac{p_s \bar{c}}{2} = F_1 + F'_1 = \frac{A(1 + R_0)}{1 - R_0 B} + \frac{p_0 \bar{c}}{4} \frac{(1 - R_0)(1 + B)}{1 - R_0 B}, \quad (10.5-11)$$

ulting

under

0.5-6)

where p_s represents the concentration in the neighborhood of the surface. We must, of course, remember that while g_s has the same value as in the thermal equilibrium state, this is *not* true of A , whose magnitude depends upon the nature of the excess carrier distribution which may be present in the bulk. Equation (10.5-11) may, in fact, be solved for A to give

0.5-7)

$$A = \frac{p_s \bar{c}}{2} \frac{1 - R_0 B}{1 + R_0} - \frac{p_0 \bar{c}}{4} \frac{(1 - R_0)(1 + B)}{1 + R_0}. \quad (10.5-12)$$

and

rium

rture

; still

er be

iving

near

iving

am."

The *difference* between F_1 and F'_1 is the *net* flux of carriers, which, if there is no electric field present, as we shall assume, is simply equal to the net diffusion current $-D_p(\partial(p)/\partial x)$ evaluated at the surfaces. Writing the expression for $F_1 - F'_1$, using (10.5-3) and (10.5-4) to represent the fluxes, and using (10.5-6) and (10.5-12) to represent the values of g_s and A , we may obtain after some tedious but straightforward algebra,

$$F_1 - F'_1 = \frac{A(1 - R_0)}{1 - R_0 B} - \frac{g_s(1 - B)}{1 - R_0 B} = \frac{(p_s - p_0)\bar{c}}{2} \frac{1 - R_0}{1 + R_0} = -D_p \left(\frac{\partial(p)}{\partial x} \right)_s. \quad (10.5-13)$$

This equation is, in fact, a statement of the surface boundary condition which must be

applied to the continuity equation. It can conveniently be written in the form

$$-D_p \left(\frac{\partial(\delta p)}{\partial x} \right)_s = s \cdot (\delta p)_s \quad (10.5-14)$$

where

$$s = \frac{\bar{c} 1 - R_0}{2 1 + R_0}. \quad (10.5-15)$$

The s subscripts indicate that the quantities concerned are evaluated at the surface. The constant s is usually referred to as the *surface recombination velocity*; its relation to the more fundamental reflection coefficient is given by (10.5-15). The surface boundary condition (10.5-14) can be stated in a more general vector form as

$$-D_p [\mathbf{n} \cdot \nabla(\delta p)]_s = s \cdot (\delta p)_s, \quad (10.5-16)$$

where \mathbf{n} is a unit outward vector normal to the surface.

If there is no surface recombination, then $R_0 = 1$ and $s = 0$, from (10.5-5). In this case, (10.5-14) gives $\partial(\delta p)/\partial x = 0$, corresponding to the condition of no net diffusive flow to the surface. If carriers inevitably recombine upon striking the surface, then $R_0 = 0$ and $s = \bar{c}/2$; this value is clearly an upper limit for the surface recombination velocity. Nevertheless, it is quite common in this limiting case to set $s = \infty$ in (10.5-14), which is tantamount to taking the excess carrier concentration at the surface $(\delta p)_s$ to be zero. This approximation, which may greatly simplify calculation, can be shown to be a good one provided $D_p/L_p \ll \bar{c}/2$. For germanium at 300°K, $D_p \approx 50$ cm²/sec, $\bar{c} \approx 10^7$ cm/sec, and L_p is usually greater than 10^{-3} cm; the condition is therefore very well satisfied in this quite typical example. As a matter of fact, circumstances under which the requirement is violated arise only very infrequently, and the adoption of the boundary condition $s = \infty$ or $(\delta p)_s = 0$ in the high surface recombination limit is thus nearly always justified. An example, which shows clearly the origin of the condition discussed above, is assigned as an exercise. In all intermediate cases, where the recombination probability is neither zero nor unity, the boundary condition in the form (10.5-14) must be used. All the above calculations have been carried out with regard to excess holes in *n*-type material, but, of course, exactly similar considerations apply to excess electrons in a *p*-type semiconductor.

In germanium, quite high reflection coefficients (~ 0.99999) corresponding to surface recombination velocities of the order of 100 cm/sec may be obtained in samples whose surfaces have been prepared by careful chemical etching. For samples whose surfaces have been heavily damaged by lapping or other abrasive action, there are many dislocations and other lattice imperfections at the surfaces which may act as recombination centers, and the surface recombination velocity may be in excess of 10^5 cm/sec, corresponding to a value of less than 0.99 for the reflection coefficient.¹⁵ In silicon, the surface recombination velocity is usually much higher on etched surfaces, a typical value being perhaps 2000 cm/sec.

¹⁵ The reflection coefficient R_0 is usually quite close to unity for germanium and silicon, even in samples whose surfaces are quite heavily abraded. Nevertheless, in many cases, the effect of a recombination velocity of 10^5 cm/sec, corresponding to $R_0 = 0.99$, may be hardly distinguishable from the condition where $s = c/2$ and $R_0 = 0$.

SEC. 10.5

SEC. 10.6

EXCESS CARRIERS IN SEMICONDUCTORS

351

10.6 STEADY-STATE PHOTOCONDUCTIVITY

In order to illustrate the application of some of the principles which were discussed in the preceding sections, we shall now proceed to calculate the steady-state photoconductive response of a uniform semiconductor sample which is illuminated with radiation of wavelength long enough so as to be only slightly absorbed in passing through the crystal, but at the same time sufficiently short to create a measurable concentration of electron-hole pairs. A suitable wavelength would thus lie just on the long wavelength side of the absorption edge, at the position λ_1 in Figure 10.1(a). We shall assume that our sample is in the form of a rectangular solid, of dimensions x_0, y_0, z_0 , that the thickness x_0 is much less than the other dimensions y_0 and z_0 , and that the illumination is incident along the x -direction, as shown in Figure 10.11. Both

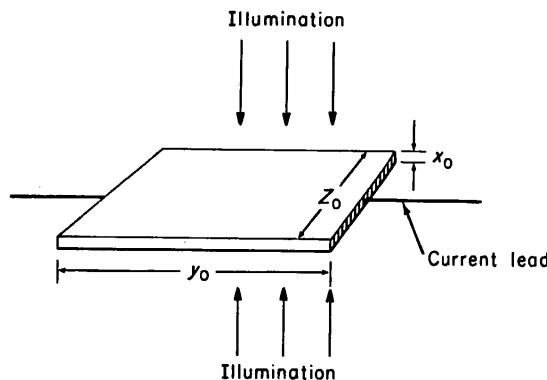


FIG. 10.11. Geometry of dc photoconductivity experiment discussed in Section 10.6.

of the large surfaces of the sample will have been prepared in the same way, so as to produce the same surface recombination velocity s on both sides. If the absorption coefficient is relatively small, the light intensity may be considered to be approximately uniform throughout the crystal, and this will lead to a generation rate of excess carriers g' which is constant and proportional to light intensity. In a large thin sample of the type shown, the excess carrier concentration varies essentially only along the x -direction, allowing one to use the one-dimensional form of the continuity equation. Since there are no fields in the x -direction, and since in the steady-state condition $\partial(\delta p)/\partial t = 0$, Equation (10.2-26) takes the form

$$\frac{d^2(\delta p)}{dx^2} - \frac{\delta p}{L_p^2} = -\frac{g'}{D_p} \quad (10.6-1)$$

where g' is constant. In writing the equation this way, we are assuming that we are dealing with extrinsic n -type material and that δp is everywhere small compared to the majority carrier density.

The solution to this equation may be expressed as the general solution to the homogeneous Equation (10.3-3) plus a particular solution to (10.6-1). It is clear that $\delta p = g'L_p^2/D_p = g'\tau_p$ is just such a particular solution to (10.6-1), and that the general solution to (10.6-1) can be written

$$\delta p(x) = A \cosh \frac{x}{L_p} + B \sinh \frac{x}{L_p} + g'\tau_p. \quad (10.6-2)$$

By the symmetry of the sample geometry shown in Figure 10.11, $\delta p(x)$ must be an even function of x , whereby $B = 0$ and (10.6-2) becomes

$$\delta p(x) = A \cosh \frac{x}{L_p} + g' \tau_p. \quad (10.6-3)$$

Applying the surface boundary condition (10.5-16) at either surface ($x = \frac{\pm x_0}{2}$) to this equation, we may evaluate the constant A to obtain

$$A = \frac{-sg' \tau_p}{s \cosh \frac{x_0}{2L_p} + \frac{D_p}{L_p} \sinh \frac{x_0}{2L_p}} \quad (10.6-4)$$

and

$$\delta p(x) = g' \tau_p \left[1 - \frac{s \cosh \frac{x}{L_p}}{s \cosh \frac{x_0}{2L_p} + \frac{D_p}{L_p} \sinh \frac{x_0}{2L_p}} \right]. \quad (10.6-5)$$

The concentration profile $\delta p(x)$ is shown in Figure 10.12(a) for several values of s .

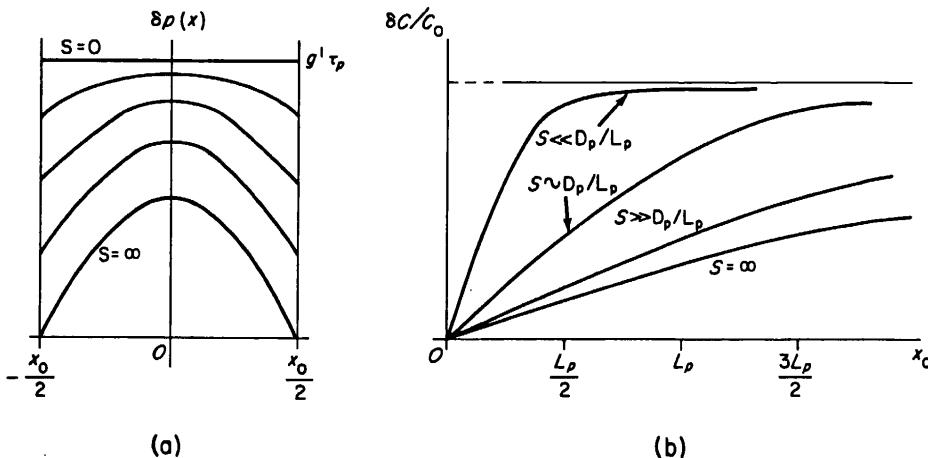


FIGURE 10.12. (a) Concentration profiles within the sample illustrated in Figure 10.11 for $s = 0$, $s = \infty$ and several intermediate values of s . (b) Photo-conductive response $\delta C/C_0$ as a function of thickness x_0 plotted for several values of s .

The change in conductance due to the presence of excess carriers will be proportional to the total number of excess carriers present in the sample. If we denote by δP the total number of excess holes and by δN the total number of excess electrons, then

$$\delta P = \delta N = y_0 z_0 \int_{-x_0/2}^{x_0/2} \delta p(x) dx, \quad (10.6-6)$$

where $y_0 z_0$ is the surface area of the illuminated face of the sample. If (10.6-5) is used

SEC. 10.6

SEC. 10.6

EXCESS CARRIERS IN SEMICONDUCTORS

353

must be an

(10.6-3)

 $\frac{\pm x_0}{2}$) toto represent $\delta p(x)$ in (10.6-6), the integral can easily be evaluated to give

$$\delta P = \delta N = 2y_0 z_0 g' \tau_p \left[\frac{1}{2} x_0 - \frac{s L_p \sinh \frac{x_0}{2L_p}}{s \cosh \frac{x_0}{2L_p} + \frac{D_p}{L_p} \sinh \frac{x_0}{2L_p}} \right]. \quad (10.6-7)$$

The differential element of conductance between the end electrodes associated with a thin sheet of material of thickness dx is given by

(10.6-4)

$$dC = \sigma(x) \frac{da}{y_0} = \frac{z_0 \sigma(x) dx}{y_0}. \quad (10.6-8)$$

But

(10.6-5)

$$\sigma(x) = \sigma_0 + \delta\sigma(x) = \sigma_0 + e\mu_p \delta p(x)(b+1), \quad (10.6-9)$$

whereby $dC = \frac{z_0}{y_0} [\sigma_0 + e\mu_p(b+1)\delta p(x)] dx. \quad (10.6-10)$

Integrating this equation between the limits $x = \pm x_0/2$, one may easily obtain

$$C = \frac{z_0 x_0}{y_0} \sigma_0 + \frac{z_0}{y_0} e\mu_p (b+1) \int_{-x_0/2}^{x_0/2} \delta p(x) dx. \quad (10.6-11)$$

But the first term above is the equilibrium conductance ($= \sigma_0 \cdot \text{area}/\text{length}$), while in the second, the integral can be expressed in terms of δP by (10.6-6). Making these substitutions (10.6-11) can be written as

$$C = C_0 + \frac{e\mu_p(b+1)\delta P}{y_0^2} = C_0 + \delta C. \quad (10.6-12)$$

The relative conductance change may then be expressed as $\delta C/C_0$, where $C_0 = \sigma_0 z_0 x_0 / y_0$, whence

$$\frac{\delta C}{C_0} = \frac{e\mu_p(b+1)\delta P}{\sigma_0 V}. \quad (10.6-13)$$

where δP is given by (10.6-7) and $V = x_0 y_0 z_0$ is the volume of the sample. The photoconductance $\delta C/C_0$ is shown plotted as a function of sample thickness for several values of s in Figure 10.12(b). For $s=0$, of course, the quantity $\delta C/C_0$ is independent of the sample thickness x_0 . For $s > 0$, the photoconductance is very small for thin samples, since then carriers may diffuse very rapidly to the surface and recombine there. For thicker samples, carriers in the interior reach the surface only after much diffusion, and surface recombination has much less effect on the overall conductance. In the limit $x_0 \rightarrow \infty$ the effect of the surface becomes negligible. It is clear that if the sample dimensions, D_p , the generation rate g' , and one of the two

be pro-
denote
ctrons,

10.6-6)

is used

quantities (s, L_p) are known, then either s or L_p , whichever is unknown, may be measured by this experiment. If two different samples, identical but for thickness, are used, then both s and L_p can be determined. The chief difficulty with the use of the steady-state photoconductivity effect as a technique for the determination of s or L_p is that the absolute generation rate g' must be known, and this is quite difficult to determine for any given experiment. This difficulty is eliminated if the *transient* photoconductivity is measured instead. The analysis of this effect is given in the next section.

10.7 TRANSIENT PHOTOCONDUCTIVITY; EXCESS CARRIER LIFETIME

In this section we shall consider the transient decay of photoconductivity produced by penetrating radiation in a sample exactly like that which was discussed in the previous section. It will be assumed that the light source has generated a uniform excess carrier density p_1 everywhere within the sample at time $t = 0$, at which instant the exciting radiation is abruptly turned off. The subsequent decay of the excess carrier distribution to equilibrium is then described by (10.2-26), which for the case at hand takes the form

$$D_p \frac{\partial^2(\delta p)}{\partial x^2} - \frac{\delta p(x,t)}{\tau_p} = \frac{\partial(\delta p)}{\partial t}. \quad (10.7-1)$$

The surface boundary conditions¹⁶ are

$$-D_p \left(\frac{\partial(\delta p)}{\partial x} \right)_{x_0/2} = s\delta p(x_0/2,t) \quad \text{and} \quad D_p \left(\frac{\partial(\delta p)}{\partial x} \right)_{-x_0/2} = s\delta p(-x_0/2,t), \quad (10.7-2)$$

while in addition, we must require that

$$\delta p(x,0) = p_1 = \text{const} \quad (10.7-3)$$

and

$$\lim_{t \rightarrow \infty} \delta p(x,t) = 0. \quad (10.7-4)$$

By making the substitution

$$\delta p(x,t) = e^{-t/\tau_p} u(x,t) \quad (10.7-5)$$

¹⁶ Although the steady state condition was assumed in arriving at the form (10.5-16) for the boundary condition, it is clear that this boundary condition can be used for time-dependent situations as well, provided that the carrier concentration in the surface region does not change much over the time needed for the flux equilibrium between surface and volume to establish itself. We shall always assume that this condition is realized.

SEC. 10.6

SEC. 10.7

EXCESS CARRIERS IN SEMICONDUCTORS

355

, may be
kness, are
use of the
if s or L_p
ifficult to
transient
the next

Equation (10.7-1) can be transformed into a differential equation for $u(x,t)$ of the form

$$D_p \frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \quad (10.7-6)$$

while the boundary conditions transform to

$$-D_p \left(\frac{\partial u}{\partial x} \right)_{x_0/2} = su(x_0/2, t); \quad D_p \left(\frac{\partial u}{\partial x} \right)_{-x_0/2} = su(-x_0/2, t) \quad (10.7-7)$$

and

$$u(x, 0) = p_1 = \text{const} \quad (-x_0/2 < x < x_0/2). \quad (10.7-8)$$

Now let us seek product solutions of the form

$$u(x, t) = X(x)T(t). \quad (10.7-9)$$

Substituting this form for the solution back into (10.7-6) it is readily seen that

$$\frac{1}{X(x)} \frac{d^2 X}{dx^2} = \frac{1}{D_p T(t)} \frac{dT}{dt} = -\alpha^2 = \text{const.} \quad (10.7-10)$$

Both sides of the equation above must be separately equal to a constant, since only in this way can a function of x alone and a function of t alone be equal for all possible values of x and t . The constant is written as $-\alpha^2$ so that for any real α it will be negative; this is necessary (as will soon become evident) to insure that (10.7-4) be satisfied. For the time dependence of the above equation, it is obvious that

$$\frac{dT}{dt} = -\alpha^2 D_p T(t), \quad (10.7-11)$$

whence

$$T(t) = e^{-\alpha^2 D_p t}. \quad (10.7-12)$$

For the spatial dependence, we have

$$\frac{d^2 X}{dx^2} = -\alpha^2 X(x), \quad (10.7-13)$$

whereby

$$X(x) = A \cos \alpha x + B \sin \alpha x. \quad (10.7-14)$$

By the symmetry of the problem, it is clear that the spatial dependence of the excess carrier concentration *must always be an even function of x*. A sine function, or any superposition of sine functions, however, is invariably odd, while a cosine or any superposition of cosines is even. It is apparent that this physical requirement can be fulfilled only by a cosine function or by a superposition of cosine functions, and that we must therefore set B equal to zero in (10.7-14). From (10.7-9), (10.7-12) and

(10.7-14), it is evident that a suitable solution may be written as

$$u(x,t) = X(x)T(t) = Ae^{-\alpha^2 D_p t} \cos \alpha x. \quad (10.7-15)$$

This solution, by itself, satisfies the differential equation (10.7-6), but does *not* satisfy the boundary conditions (10.7-7) and (10.7-8). However, we may form a linear superposition of solutions of this form such as

$$u(x,t) = \sum_n u_n(x,t) = \sum_n A_n e^{-\alpha_n^2 D_p t} \cos \alpha_n x, \quad (10.7-16)$$

which still satisfies the differential equation, and we are at liberty to choose the values A_n and α_n so that the boundary conditions are satisfied by the superposition. If *each term* of the summation in (10.7-16) is required to satisfy the surface boundary condition

$$-D_p(\partial u_n / \partial x)_{x_0/2} = s u_n(x_0/2, t), \quad (10.7-17)$$

then, of course, the surface boundary conditions (10.7-7) will automatically be satisfied by the superposition.¹⁷ Substituting a solution of the form (10.7-15) for u_n into (10.7-17), we find that in order to satisfy (10.7-17) we must have

$$\alpha_n A_n D_p e^{-\alpha_n^2 D_p t} \sin \frac{\alpha_n x_0}{2} = s A_n e^{-\alpha_n^2 D_p t} \cos \frac{\alpha_n x_0}{2}, \quad (10.7-18)$$

whereby α_n must be chosen in such a way that the equation

$$\operatorname{ctn} \frac{\alpha_n x_0}{2} = \frac{\alpha_n D_p}{s} = \frac{\alpha_n x_0}{2} (2D_p/sx_0) \quad (10.7-19)$$

is satisfied. The roots of this equation may be obtained numerically or graphically as the intersection of the curves $f(\alpha x_0) = \operatorname{ctn}(\frac{1}{2}\alpha x_0)$ and $f(\alpha x_0) = (\frac{1}{2}\alpha x_0)(2D_p/sx_0)$, as shown in Figure 10.13.

Now we must try to satisfy the boundary condition (10.7-8). We shall try to do this by choosing the values A_n so that at $t = 0$ the sum of all the $u_n(x,0)$ add to form a Fourier-like representation of the required initial concentration profile. To accomplish this, suppose that $u(x,0)$ is an *arbitrary even function*¹⁸ $f(x)$, so that

$$u(x,0) = \sum_{n=0}^{\infty} A_n \cos \alpha_n x = f(x). \quad (10.7-20)$$

If both sides of this equation are multiplied by $\cos \alpha_m x$ and integrated over the interval $(-x_0/2 < x < x_0/2)$, the result is

$$\sum_n \int_{-x_0/2}^{x_0/2} A_n \cos \alpha_n x \cos \alpha_m x dx = \int_{-x_0/2}^{x_0/2} f(x) \cos \alpha_m x dx. \quad (10.7-21)$$

¹⁷ Note that if $u(x,t)$ is an even function of x , the satisfaction of one of the boundary conditions (10.7-7) implies the satisfaction of the other.

¹⁸ If $f(x)$ were not even, we would have to admit solutions of the form $B_n \exp(-\alpha_n^2 D_p t) \sin \alpha_n x$ into our superposition as well as the cosine function we are using.

SEC. 10.7

SEC. 10.7

EXCESS CARRIERS IN SEMICONDUCTORS

357

(10.7-15)

but does not
form a linear

(10.7-16)

se the values
tion. If each
ry condition

(10.7-17)

✓ be satisfied
for u_n into

(10.7-18)

(10.7-19)

graphically
)($2D_p/sx_0$),
ll try to do
i to form a
accomplish

(10.7-20)

he interval

(10.7-21)

✓ conditions

 $D_p t) \sin \alpha_n x$

If the set of functions $\{\cos \alpha_n x\}$ is orthogonal on the interval $(-x_0/2 < x < x_0/2)$, then all the integrals on the left-hand side of this equation will vanish except that for which $n = m$, which is easily evaluated. Under these circumstances, solving for A_m , one may easily show that

$$A_m = \frac{2\alpha_m}{\alpha_m x_0 + \sin \alpha_m x_0} \int_{-x_0/2}^{x_0/2} f(x) \cos \alpha_m x \, dx. \quad (10.7-22)$$

It is clear that since this is a special case of the Sturm-Liouville problem, the functions $\{\cos \alpha_n a\}$ must form an orthogonal set.¹⁹ In any case, it is not difficult to show directly

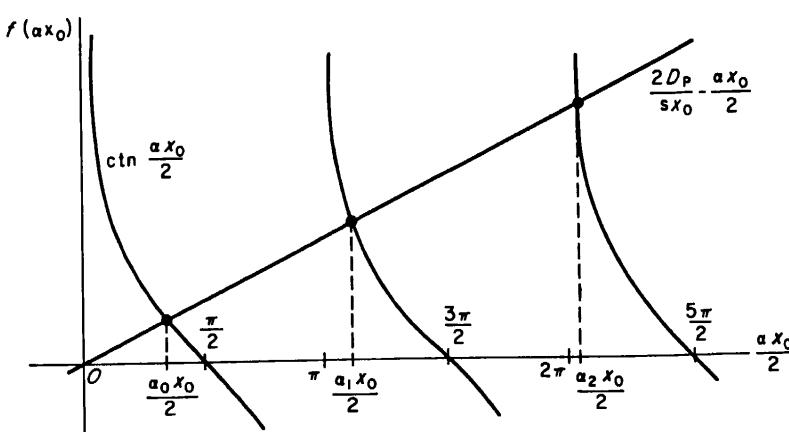


FIGURE 10.13. Diagram illustrating the determination of the roots of (10.7-19) by the graphical solution of the transcendental equation.

that provided the α_n are chosen so as to satisfy (10.7-19) the functions $\{\cos \alpha_n a\}$ are orthogonal. The details of this proof are left as an exercise for the reader.

For the case at hand, we must choose $f(x) = p_1 = \text{const}$, and with this choice, the integral in (10.7-22) can easily be evaluated giving

$$A_n = \frac{4p_1 \sin \frac{\alpha_n x_0}{2}}{\alpha_n x_0 + \sin \alpha_n x_0} \quad (10.7-23)$$

whereby

$$f(x) = p_1 = 4p_1 \sum_{n=0}^{\infty} \frac{\sin \frac{\alpha_n x_0}{2} \cos \alpha_n x}{\alpha_n x_0 + \sin \alpha_n x_0}. \quad (10.7-24)$$

Now $u(x,t)$ can be found by substituting the above value for A_n in (10.7-16), and $\delta p(x,t)$

¹⁹ R. V. Churchill, *Fourier Series and Boundary Value Problems*. New York: McGraw-Hill (1941), pp. 46-52.

can then be obtained from (10.7-5). The final result is

$$\delta p(x,t) = 4p_1 e^{-t/\tau_p} \sum_{n=0}^{\infty} \frac{\sin \frac{\alpha_n x_0}{2} \cos \alpha_n x}{\alpha_n x_0 + \sin \alpha_n x_0} e^{-\alpha_n^2 D_p t}. \quad (10.7-25)$$

When $s \rightarrow 0$, the slope of the straight line shown in Figure 10.13 becomes extremely large, and the eigenvalues $\{\alpha_n a\}$ approach $(0, 2\pi, 4\pi, 6\pi, \dots)$. For this set of values, $\sin(\alpha_n a/2) = 0$ for all α_n , and all the terms in the sum in (10.7-25) vanish, with the exception of the first. In the case of the first term, both numerator and denominator vanish, giving an indeterminate form. Using L'Hôpital's rule, it is readily shown that

$$\lim_{\alpha_0 x_0 \rightarrow 0} \frac{\sin \frac{\alpha_0 x_0}{2}}{\alpha_0 x_0 + \sin \alpha_0 x_0} = \lim_{\alpha_0 x_0 \rightarrow 0} \frac{\frac{1}{2} \cos \frac{\alpha_0 x_0}{2}}{1 + \cos \alpha_0 x_0} = \frac{1}{4}, \quad (10.7-26)$$

whence

$$\delta p(x,t) = p_1 e^{-t/\tau_p} \quad (s = 0). \quad (10.7-27)$$

The concentration profile for this case is plotted for several values of t in Figure 10.14(a). It is clear that the concentration remains the same at any given time at all points in the sample.

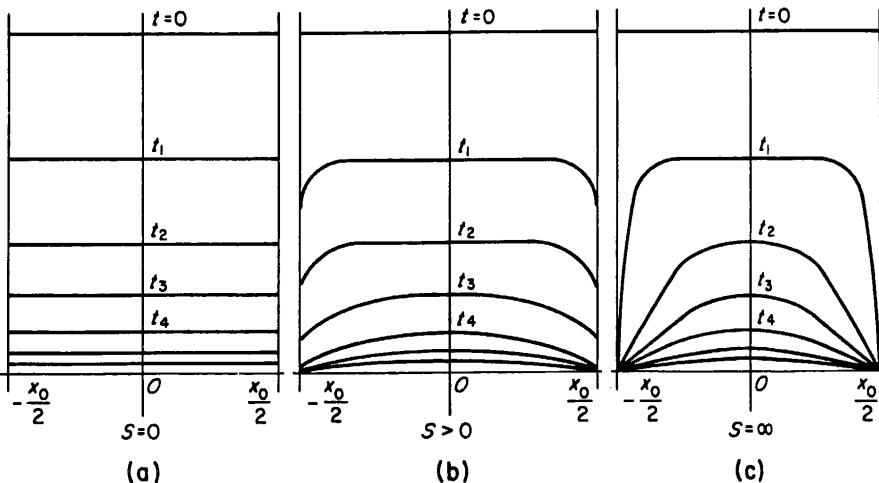


FIGURE 10.14. Concentration profiles within the sample for several values of t , (a) for $s = 0$, (b) for $s > 0$, (c) for $s = \infty$. In all cases $0 < t_1 < t_2 < t_3 < t_4$.

If $s \rightarrow \infty$, the slope of the straight line in Figure 10.13 approaches zero, the eigenvalues $\{\alpha_n a\}$ approaching $(\pi, 3\pi, 5\pi, 7\pi, \dots)$ in the limit. In this case (10.7-25) becomes

$$\delta p(x,t) = 4p_1 e^{-t/\tau_p} \sum_{n=0}^{\infty} \frac{(-1)^n \cos \frac{(2n+1)\pi x}{x_0}}{(2n+1)\pi} \exp \left[-\frac{(2n+1)^2 \pi^2 D_p t}{x_0^2} \right], \quad (10.7-28)$$

which is an ordinary Fourier series. In this instance, when $x = \pm x_0/2$, δp is zero, since $\cos(n + \frac{1}{2})\pi = 0$. The concentration profile for this case is plotted in Figure 10.14(c). For intermediate values of s , the values $\{\alpha_n\}$ must be determined numerically or graphically, and the results, when plotted, yield concentration profiles such as those shown in Figure 10.14(b).

(10.7-25)

The actual photoconductance can be obtained from $\delta p(x,t)$ by means of (10.6-12) and (10.6-6), whereby

$$\frac{\delta C(t)}{C_0} = \frac{e\mu_p(b+1)}{\sigma_0 V} \cdot y_0 z_0 \int_{-x_0/2}^{x_0/2} \delta p(x,t) dx, \quad (10.7-29)$$

or, using (10.7-25) and evaluating the integral,

(10.7-26)

$$\frac{\delta C(t)}{C_0} = \frac{8p_1 e\mu_p(b+1)}{\sigma_0} e^{-t/\tau_p} \sum_{n=0}^{\infty} \frac{\sin^2 \frac{\alpha_n x_0}{2}}{\alpha_n x_0 (\alpha_n x_0 + \sin \alpha_n x_0)} e^{-\alpha_n^2 D_p t}. \quad (10.7-30)$$

(10.7-27)

This expression is a summation of terms each of which has its own characteristic amplitude and each of which decays exponentially with time with a different time constant. It may be written in the form

$$\frac{\delta C(t)}{C_0} = \sum_m C_m e^{-t/\tau_m} \quad (10.7-31)$$

where C_m is the amplitude of the m th decay mode²⁰ defined by (10.7-30), and where each exponential time constant may be expressed as

$$\frac{1}{\tau_m} = \frac{1}{\tau_p} + \alpha_m^2 D_p. \quad (10.7-32)$$

Since α_0 is the smallest member of the set $\{\alpha_n\}$, the time constant τ_0 must be longer than any of the others; furthermore, the form of the amplitude factor C_m multiplying each term is such that the amplitudes of the higher terms are all smaller than that of the first. The result is that the higher order modes die out more rapidly than the zeroth, and after a sufficiently long time the decay can be represented as a simple exponential with the single *principal mode* decay constant τ_0 , as shown in Figure 10.15. It is clear that if the logarithm of the photoconductive response is plotted as a function of time, a straight line is obtained after the effects of the higher mode terms in (10.7-30) have died out. The slope of this line is $1/\tau_0$, where

$$\frac{1}{\tau_0} = \frac{1}{\tau_p} + \alpha_0^2 D_p. \quad (10.7-33)$$

There are two terms on the right-hand side of this equation; the first represents the effect of bulk recombination upon the observed principal mode time constant, the

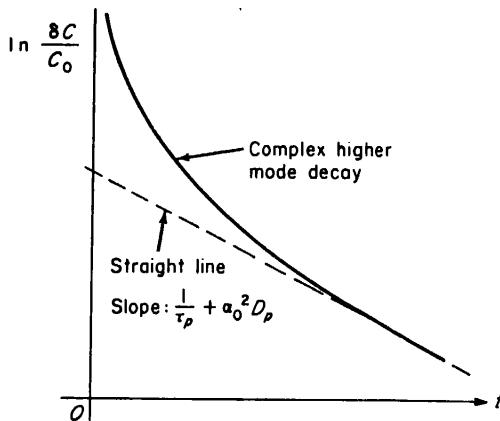
es of t ,
s zero, the
e (10.7-25)

(10.7-28)

²⁰ The index m is adopted here in order to avoid possible confusion with the electron lifetime τ_n .

second the effect of surface recombination. The *observed lifetime* τ_0 is shorter than the true bulk lifetime τ_p , unless the surface recombination velocity is zero, in which case, as we have already noted, $\alpha_0 = 0$. Although it is ordinarily impossible to reduce s to zero, experimentally, the same effect can be produced, as is evident from Figure 10.13, by arranging the conditions of the experiment so that $2D_p/sx_0$ is much less than unity. This may be accomplished by using a very thick sample. Under these circumstances carriers generated in the interior must diffuse a long way to the surface, whose role is then much less important than for a thin sample, and the observed lifetime τ_0 approximates the true bulk lifetime τ_p .

FIGURE 10.15. Logarithm of the transient photoconductive response *versus* time.



If the true bulk lifetime τ_p and the observed lifetime τ_0 are both known, then the surface recombination velocity can be evaluated by solving (10.7-33) for α_0 , inserting the value so obtained into (10.7-19) and solving for s . The result is

$$s = \sqrt{D_p \left(\frac{1}{\tau_0} - \frac{1}{\tau_p} \right)} \tan \frac{x_0}{2} \sqrt{\frac{1}{D_p} \left(\frac{1}{\tau_0} - \frac{1}{\tau_p} \right)}. \quad (10.7-34)$$

This formula may also be used to obtain the true bulk lifetime if the observed lifetime and surface recombination velocity are known.

If both τ_p and s are unknown to begin with, they may be determined by working first with a thick sample whose surface is treated in such a way as to minimize surface recombination; the transient photoconductive decay associated with this type of sample will approximate τ_p . The thick sample may then be cut into thin ones, the surfaces of these being treated so as to obtain the surface conditions under which the surface recombination velocity is to be measured. The decay constant τ_0 is then measured and s follows directly from (10.7-34), using the value for τ_p obtained previously. Alternatively, two samples of different thickness can be cut from material of *uniform* lifetime. The surfaces of both samples are treated so as to produce the conditions under which a measurement of s is desired, and the observed lifetime associated with each sample measured. Equation (10.7-34) can then be written for each sample, and the two resulting equations may then be solved numerically for the two unknown quantities, τ_p and s . A third method is to use only a single sample, and to adjust the values of τ_p and s in (10.7-30) so as to reproduce not only the observed principal mode time constant τ_0 (which can, after all, be obtained from many different combinations of τ_p and s) but *also* the observed initial higher-mode decay scheme. Although

ster than
in which
to reduce
n Figure
less than
these cir-
surface,
observed

the required mathematical analysis is complicated, this method has the advantage that only a single sample is used, and difficulties arising from inhomogeneities in bulk lifetime and surface recombination velocities are thereby minimized.

10.8 RECOMBINATION MECHANISMS: THE SHOCKLEY-READ THEORY OF RECOMBINATION

We shall now investigate briefly the physical aspects of electron-hole recombination. It is important to understand the physical processes involved in *direct* electron-hole recombination as well as in recombination by *trapping*, although the latter is found to be the dominant mechanism in the majority of situations involving covalent or intermetallic III-V semiconductors.

In direct recombination an electron in the conduction band and a hole in the valence band recombine without the assistance of any intermediate state. If both conduction and valence bands have energy minima at $k = 0$, then vertical transitions in which a hole near the top of the valence band may recombine with an electron near the bottom of the conduction band are possible. Since momentum must be conserved in such a transition, only electrons and holes whose \mathbf{k} -vectors satisfy the relation

$$\mathbf{k}_n + \mathbf{k}_p = 0 \quad (10.8-1)$$

may interact in this way. A photon of frequency ω given by

$$\hbar\omega = \epsilon_n - \epsilon_p \cong \epsilon_c - \epsilon_v \quad (10.8-2)$$

is emitted in the process.²¹ This *direct transition* process is illustrated in Figure 10.16(a). The absorption of a photon, accompanied by the production of an electron-hole pair may take place by a process which is simply the reverse of this one.

In substances such as germanium and silicon, where the minimum of the conduction band occurs for some value of the crystal momentum \mathbf{k}_0 other than zero, the recombination of an electron at the conduction band minimum and a hole at the top of the valence band requires (in addition to the production of a photon) the emission of a phonon of propagation constant $\mathbf{k}' = \mathbf{k}_0$ or the absorption of an already present phonon of propagation constant $\mathbf{k}' = -\mathbf{k}_0$ in order that crystal momentum be conserved. Such a process is called an *indirect transition*. Since some energy which would have been carried away by the photon in a direct transition is now carried away (or contributed) by the phonon, the frequency of the radiation associated with such transitions is given by

$$\hbar\omega = \epsilon_n - \epsilon_p \pm \epsilon(\mathbf{k}_0) = \epsilon_c - \epsilon_v \pm \epsilon'(\mathbf{k}_0) \quad (10.8-3)$$

²¹ In (10.8-1) the momentum carried away by the emitted photon has been neglected. This is possible because the electron wavelength (for an average energy kT) as given by (8.4-1) is much smaller than the photon wavelength for $\omega \cong \Delta\epsilon/\hbar$. The momenta are, of course, related to the respective wavelength by the de Broglie relation (4.9-7).

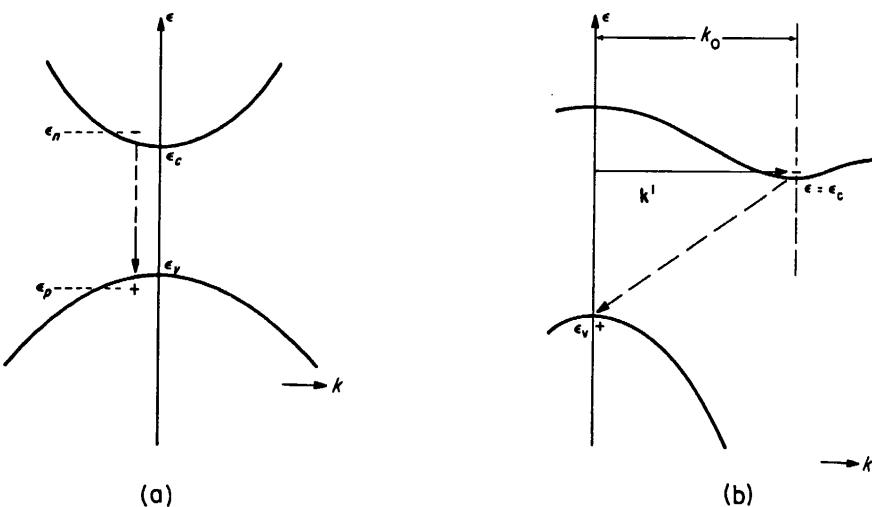


FIGURE 10.16. Electron-hole recombination via (a) direct transition, (b) indirect transition.

where $\epsilon'(\mathbf{k}_0)$ is the energy associated with a phonon of wave number \mathbf{k}_0 as obtained from the dispersion relation $\omega(\mathbf{k})$ for the lattice vibrations by

$$\epsilon'(\mathbf{k}_0) = \hbar\omega(\mathbf{k}_0). \quad (10.8-4)$$

The energy $\epsilon'(\mathbf{k}_0)$, although small compared with the forbidden gap energy $\Delta\epsilon$, is perceptible through its influence on the form of the *absorption* versus wavelength curve in the neighborhood of the absorption edge. Again, the absorption of a phonon to produce an electron-hole pair may be regarded as the inverse of this process. Actually, since the emitted or absorbed phonon may belong to either acoustical or optical branch, and may be either a transverse or longitudinal vibration, there are *four* separate phonons which can participate in such a process, each with a different frequency and hence a different energy ϵ' for $\mathbf{k} = \mathbf{k}_0$. The effects of all four of these phonons can be observed in the absorption spectra of this type of semiconductor near the fundamental absorption edge. Of course, direct transitions can also occur in materials where the conduction band minimum is not at $k = 0$, but if the energy difference at the conduction band edge between the point of minimum energy and $k = 0$ is at all appreciable, the electron population for energies corresponding to $k = 0$ will normally be so small that such transitions will be quite infrequent. Likewise, direct absorptive transitions will require incident photon energies much higher than those required to promote indirect transitions. Indirect transitions are predominant under normal conditions in all the elemental covalent semiconductors and III-V compound semiconductors with the single exception of InSb, in which both band minima are at $k = 0$.

It is possible, in either case, to calculate the rate at which transitions take place from the observed optical absorption spectrum of the material,²² and therefore to

²² In the steady state, the generation and recombination rate of electron-hole pairs must be equal; for this reason a calculation of the generation rate will also yield the recombination rate. For direct optical excitation of electron-hole pairs, however, the generation rate for any given incident photon wavelength must be proportional to the absorption coefficient for that wavelength. The optical absorption curve and the recombination rate arising from direct recombinative processes are therefore closely related.

infer the recombination rate and the excess carrier lifetime from the absorption spectrum. This calculation was first carried out by van Roosbroeck and Shockley,²³ who found that in germanium, for example, the excess carrier lifetime at room temperature should be about 0.75 sec, independent of donor or acceptor impurity density. This result was contradicted by experimental evidence; observed excess carrier lifetimes rarely exceeded 10^{-3} sec, and were found to be quite sensitive to the donor and acceptor impurity concentration. It was also well established experimentally that excess carrier lifetimes in germanium and silicon could be drastically reduced by the introduction of certain impurities, such as copper, in extremely tiny concentrations, and by the introduction of certain structural imperfections, such as dislocations. In addition, the van Roosbroeck-Shockley calculation predicted a temperature dependence of excess carrier lifetime which was not observed experimentally. It was apparent that in most materials, excess carrier lifetimes were not limited by radiative transitions of the type discussed above, but rather by some other process which was intimately associated with the presence of certain special types of impurity atoms and structural imperfections in small concentrations. These results led to a consideration of recombination by *trapping* of electrons and holes by localized energy levels lying deep within the forbidden energy gap, which were thought to be associated with certain "trapping" impurities and structural defects. This mechanism was first investigated by Shockley and Read.²⁴

The Shockley-Read theory of recombination involves a consideration of the statistics of occupation of such trapping levels. We shall consider in detail the behavior of an energy level at an energy ϵ_t within the forbidden energy region, which is neutral when empty and which may be occupied by an electron, thus acquiring a negative charge. These *traps* may promote electron-hole recombination by capturing electrons from the conduction band and subsequently transferring them to the valence band whenever a hole appears near the trap to recombine with the trapped electron. The net effect is to do away with an electron-hole pair, the trapping level being returned in the end to its original state. Since the trapping center is usually tightly coupled to

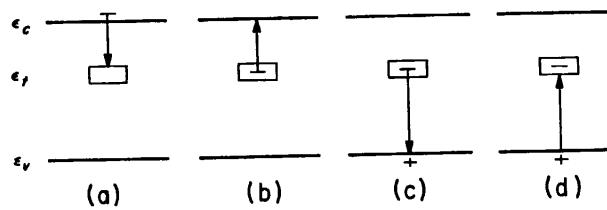


FIGURE 10.17. Four fundamental processes involved in recombination through traps; (a) the capture of a conduction electron by an empty trap, (b) the emission of an electron from the trap to the conduction band, (c) the capture of a hole from the valence band by a trap containing an electron, and (d) the promotion of a valence electron into an initially empty trap.

the lattice, the energy and momentum initially belonging to the electron and hole may be converted into lattice vibrations, with little or no electromagnetic energy being emitted in the process. There are four basic processes involved in electron-hole

²³ W. van Roosbroeck and W. Shockley, *Phys. Rev.*, **94**, 1558 (1954).

²⁴ W. Shockley and W. T. Read, Jr., *Phys. Rev.*, **87**, 835 (1952).

recombination through trapping centers, as illustrated in Figure 10.17; (a) the capture of an electron from the conduction band by an initially neutral empty trap, (b) the inverse of (a), involving the emission of an electron initially occupying a trapping level to the conduction band, (c) the capture of a hole in the valence band by a trap containing an electron, and (d), the inverse of (c), wherein a valence electron is promoted into an initially empty trap (this can equally well be regarded as the emission of a hole from the trap to the valence band).

Under these circumstances, the rate at which electrons from the conduction band are captured by traps will clearly be proportional to the number of electrons in the conduction band and to the number of *empty* traps available to receive them. We may write this as

$$R_{cn} = C_n N_t (1 - f(\varepsilon_t)) n \quad (10.8-5)$$

where C_n is a constant and N_t represents the total concentration of trapping centers in the crystal. Likewise, the rate at which electrons are emitted from filled traps to the conduction band will be proportional to the number of filled traps, and can thus be expressed as

$$R_{en} = E_n N_t f(\varepsilon_t) \quad (10.8-6)$$

where E_n is another proportionality constant. In equilibrium, in accord with the principle of detailed balancing, these two rates must be equal, whereby

$$E_n = n_0 C_n \frac{1 - f_0(\varepsilon_t)}{f_0(\varepsilon_t)}. \quad (10.8-7)$$

By substituting the explicit value of the Fermi function f_0 from (5.5-20) into this expression, we may easily obtain the very simple relation

$$E_n = n_0 C_n e^{(\varepsilon_t - \varepsilon_f)/kT}. \quad (10.8-8)$$

Substituting the value given by (9.3-6) for n_0 , (10.8-8) may finally be written as

$$E_n = n_1 C_n \quad (10.8-9)$$

where

$$n_1 = U_c e^{-(\varepsilon_c - \varepsilon_t)/kT}. \quad (10.8-10)$$

We are assuming, in writing (10.8-9) and (10.8-10), that the distribution of electrons in the conduction band is Maxwellian. The quantity n_1 is the electron concentration which would be present in the conduction band if the Fermi level were to coincide with the trap level ε_t . The *net* rate at which electrons are captured from the conduction band is simply the difference between (10.8-5) and (10.8-6), which, using (10.8-9) to express E_n in terms of C_n , is

$$R_n = R_{cn} - R_{en} = C_n N_t [n(1 - f(\varepsilon_t)) - n_1 f(\varepsilon_t)]. \quad (10.8-11)$$

In the same way, the rate at which holes are captured from the valence band must be proportional to the concentration of holes and to the number of traps which may

SEC. 10.8

SEC. 10.8

EXCESS CARRIERS IN SEMICONDUCTORS

365

the capture
ap, (b) the
a trapping
1 by a trap
ron is pro-
e emission

ction band
ons in the
them. We

(10.8-5)

ng centers
d traps to
1 can thus

(10.8-6)

with the

(10.8-7)

into this

(10.8-8)

as

(10.8-9)

(10.8-10)

electrons
ntration
coincide
nduction
0.8-9) to

(10.8-11)

nd must
ich may

receive holes (thus the number which are occupied by electrons) whence

$$R_{cp} = C_p N_t f(\epsilon_t) p. \quad (10.8-12)$$

The rate at which holes are emitted to the valence band from trapping centers (process (d) in Figure 10.17) is proportional to the number of traps occupied by holes (i.e., *empty*), from which

$$R_{ep} = E_p N_t (1 - f(\epsilon_t)). \quad (10.8-13)$$

As before, in the equilibrium state, R_{cp} must be equal to R_{ep} , and by reasoning which is exactly analogous to that used in deriving (10.8-9), one may easily show that

$$E_p = p_1 C_p \quad (10.8-14)$$

$$\text{where } p_1 = U_v e^{-(\epsilon_t - \epsilon_v)/kT}. \quad (10.8-15)$$

Again, p_1 may be visualized as the concentration of holes which would be found in the valence band if the Fermi level were at the energy of the trapping level. The *net* rate at which holes are captured from the valence band may be expressed as

$$R_p = R_{cp} - R_{ep} = C_p N_t [p f(\epsilon_t) - p_1 (1 - f(\epsilon_t))]. \quad (10.8-16)$$

When trapping centers are present it is no longer possible in general to assume that $\delta n = \delta p$, despite the fact that overall electrical neutrality is maintained, because some electrons which would otherwise be present in the conduction band as a part of δn may now be immobilized in traps, unable to function as conduction electrons. This situation causes some difficulty when the density of trapping centers is large, and has been treated in detail by Shockley and Read.²⁴ If N_t is quite small, however, the approximation $\delta n = \delta p$ is still very good.²⁵ We shall assume that this is the case, and in the majority of cases of practical interest in the covalent semiconductors this condition is indeed satisfied. In a sample in which excess carriers are recombining, under these conditions, we must have $R_p = R_n$ since every electron recombination is accompanied by the recombination of a hole. The common recombination rate is then related to the lifetime τ by

$$R_n = R_p = \frac{\delta p}{\tau}. \quad (10.8-17)$$

Equating (10.8-11) and (10.8-16) and solving for $f(\epsilon_t)$, one may obtain

$$f(\epsilon_t) = \frac{C_p p_1 + C_n n}{C_n (n + n_1) + C_p (p + p_1)}. \quad (10.8-18)$$

Substituting this value into either (10.8-11) or (10.8-16) and noting that $n_1 p_1 = n_i^2$, it

²⁵ It is shown by Shockley and Read that $\delta n \approx \delta p$ whenever any one of the four quantities n_0 , p_0 , n_1 and p_1 is large compared with N_t .

is clear that

$$R_n = R_p = \frac{\delta p}{\tau} = \frac{N_t C_n C_p (n p - n_i^2)}{C_n (n + n_1) + C_p (p + p_1)}. \quad (10.8-19)$$

If we now write $n = n_0 + \delta p$, $p = p_0 + \delta p$, we may express the above equation in the form

$$\frac{1}{\tau} = \frac{N_t (p_0 + n_0 + \delta p)}{\frac{1}{C_p} (n_0 + n_1 + \delta p) + \frac{1}{C_n} (p_0 + p_1 + \delta p)}. \quad (10.8-20)$$

In the limit where $n_0 \gg \delta p$, p_0 , n_1 , p_1 it is evident that (10.8-20) reduces to

$$\tau = \tau_{p0} = \frac{1}{C_p N_t}. \quad (10.8-21)$$

The lifetime of holes in a strongly extrinsic *n*-type sample having a given trapping center concentration is thus just the reciprocal of $C_p N_t$. Likewise, in the limit where $p_0 \gg \delta p$, n_0 , n_1 , p_1 , (10.8-20) reduces to

$$\tau = \tau_{n0} = \frac{1}{C_n N_t} \quad (10.8-22)$$

showing that in a strongly extrinsic *p*-type sample the electron lifetime is the reciprocal of $C_n N_t$. We may then write (10.8-20) as

$$\frac{1}{\tau} = \frac{p_0 + n_0 + \delta p}{\tau_{p0}(n_0 + n_1 + \delta p) + \tau_{n0}(p_0 + p_1 + \delta p)} \quad (10.8-23)$$

where (τ_{p0}, τ_{n0}) represent the lifetime of excess (holes, electrons) in highly extrinsic (*n*-type, *p*-type) material having the particular trapping characteristics which are being considered. The lifetime for materials intermediate between these two extremes is

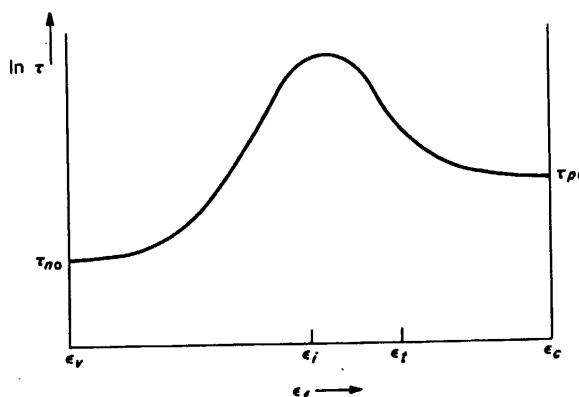


FIGURE 10.18. The dependence of lifetime upon the Fermi energy, as predicted by the Shockley-Read theory.

given by (10.8-23). For small values of δp (10.8-23) can be written as

$$\tau = \tau_{p0} \left(\frac{n_0 + n_1}{p_0 + n_0} \right) + \tau_{n0} \left(\frac{p_0 + p_1}{p_0 + n_0} \right). \quad (10.8-24)$$

From this equation the lifetime can be plotted as a function of the Fermi level, and thus as a function of n_0 and p_0 . It is found that the lifetime is a maximum when the Fermi level is at some point within the forbidden gap whose exact location depends upon the values τ_{p0} , τ_{n0} and n_1 , as illustrated in Figure 10.18. If τ_{n0} and τ_{p0} are equal, then the maximum will come at the intrinsic condition.

The variation of the lifetime with the injection level δp can be studied by solving (10.8-23) for τ , writing the result in the form

$$\tau = \tau_0 \left[\frac{1 + \frac{(\tau_{p0} + \tau_{n0})\delta p}{\tau_{p0}(n_0 + n_1) + \tau_{n0}(p_0 + p_1)}}{1 + \frac{\delta p}{n_0 + p_0}} \right] \quad (10.8-25)$$

where

$$\tau_0 = \frac{\tau_{p0}(n_0 + n_1) + \tau_{n0}(p_0 + p_1)}{n_0 + p_0}. \quad (10.8-26)$$

From this it is clear that if δp is sufficiently small, the lifetime will have the value τ_0 independent of δp . For larger values of δp , the lifetime depends upon δp , and may either increase or decrease with increasing values of δp , depending upon the relative values of τ_{p0} , τ_{n0} , n_0 and n_1 . For large values of δp it is easily seen that (10.8-25) gives

$$\tau = \tau_\infty = \tau_{p0} + \tau_{n0}. \quad (10.8-27)$$

If τ_∞ is different from τ_0 (as is generally the case), there is a monotonic variation of lifetime between the limits τ_0 and τ_∞ as δp is increased.

It is rather difficult to make precise and meaningful comparisons between the Shockley-Read theory and experiment, due to the difficulty of obtaining samples with precisely controlled concentrations of a single type of trapping center, and due to the number of unknown parameters whose value must be determined from (or adjusted to fit) experimental data.²⁶ Nevertheless, the theory seems to be consistent with all the experimental results to which it might be expected to apply to date.²⁷ It explains the marked observed variations of carrier lifetime with donor and acceptor density, injection level, concentration of trapping centers, and temperature. It may also be extended to provide a theory of surface recombination based upon trapping levels associated with the surface, but having properties similar to the bulk centers discussed herein insofar as the statistics of capture and emission are concerned.

Physically, the bulk trapping centers in semiconductor crystals often arise from

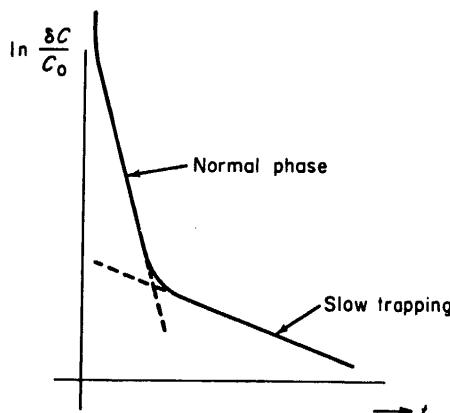
²⁶ The difficulty is not so much fitting experimental data, but in having so many adjustable parameters that the data can be explained by many possible sets of values for these parameters.

²⁷ See, for example, J. A. Burton *et al.*, *J. Phys. Chem.*, **57**, 853 (1953).

the presence of certain "trapping impurities." In the covalent semiconductors silicon and germanium these trapping impurities are often divalent *double* acceptors, such as Cu and Ni, which provide an ordinary acceptor level very close to the valence band and a second level deep within the forbidden region which is much less easily ionized, and which functions as a trapping level. The donor and acceptor levels arising from the presence of substitutional group III and group V impurities are almost completely ionized at all but very low temperatures and therefore do not ordinarily play an important role in recombination processes.²⁸ The presence of structural imperfections, particularly edge-type dislocations, has also been shown to give rise to acceptor trapping levels within the forbidden region, which may in many instances provide an important contribution to electron-hole recombination.

It sometimes happens that *two* distinct types of trapping centers are present in the same sample. Usually one of these is the normal recombination center as discussed above; the second set of trapping centers, however, quite often has associated with it a very small hole capture probability when filled with electrons. This leads to a very large value for τ_{p0} for that set of levels. Although naturally the Shockley-Read analysis must be repeated, considering both sets of trapping levels from the outset, to describe the details of the resulting situation, it is fairly easy to see intuitively that what must happen initially in a transient photoconductivity measurement such as that discussed in Section 10.7 is that some electrons recombine through the normal recombination levels, while others fall into the second set of traps. Before long, however, the normal recombination process has exhausted all excess electrons except those in the slow second set of traps, which then slowly capture the remaining excess holes. The resulting photoconductivity decay has two separate exponential sections with different time constants, as shown in Figure 10.19. There is an initial "normal"

FIGURE 10.19. Transient photoconductive response of a sample which contains a large number of "slow trapping centers."



recombination phase and a long "trapping tail" due to the slow traps. A certain confusion in terminology has arisen around this effect, in that the "normal" initial part of the decay is frequently referred to as "recombination" and the slow final decay is then ascribed to "trapping." Actually *both* phases of the process are governed by trapping phenomena, and it is more accurate to refer to the earlier as due to *normal* trapping and the latter as due to slow trapping effects.

²⁸ Except, of course, insofar as they influence the position of the Fermi level.

EXERCISES

ctors silicon
tors, such as
valence band
easily ionized,
arising from
t completely
rily play an
perfections,
ceptor trap-
provide an

is present in
as discussed
ated with it
is to a very
ckley-Read
the outset,
itively that
uch as that
he normal
long, how-
ons except
ing excess
al sections
“normal”

ing

certain con-
initial part
l decay is
erned by
o normal

1. Starting with the Boltzmann equation and assuming that the relaxation time approximation holds, show that in a steady state system containing free particles wherein a *gradient of concentration* along the x -direction exists, a diffusion flux $-D(\partial n/\partial x)$ is set up. Show that the diffusion coefficient D is equal to $\lambda\bar{c}/3$ where λ is the mean free path and \bar{c} is the mean thermal speed. You may assume that the mean free path λ is independent of velocity, and that the system obeys Maxwell-Boltzmann statistics in equilibrium. Hint: Assume that the distribution function $f(\mathbf{r}, \mathbf{v})$ can be written in the form $n(\mathbf{r})f(\mathbf{v})$; this is reasonable if the departure from the equilibrium state is not too great.

2. The results derived from the calculations of Problem 1 have been applied in the text to nonsteady state systems as well as to systems which are in the steady state. Discuss (qualitatively) under what circumstances this extension should be allowable.

3. Prove that if $f(x, t)$ satisfies the differential equation

$$D \frac{\partial^2 f}{\partial x^2} - \frac{f}{\tau} = \frac{\partial f}{\partial t}$$

then $f(\xi, t)$ ($\xi = x - \mu E_0 t$, with E_0 const) must be a solution to

$$D \frac{\partial^2 f}{\partial x^2} - \mu E_0 \frac{\partial f}{\partial x} - \frac{f}{\tau} = \frac{\partial f}{\partial t}.$$

4. A steady-state excess carrier distribution is created in an extrinsic n -type sample by a plane source of generation at the origin. The resulting diffusion and drift of the carrier distribution can be regarded as one-dimensional, along the x -direction; the extent of the sample along the x -axis is essentially infinite. There is a constant electric field E_0 in the $+x$ -direction. The excess carrier density is measured by measuring the reverse saturation current in two collector probes located at two different points, $x = a$ and $x = b$ ($b > a$). It is desired to calculate the diffusion length L_p from the ratio of these two measured excess carrier densities. Show that

$$L_p = \frac{d}{\sqrt{\ln K_0 \left(\ln K_0 + \frac{eE_0 d}{kT} \right)}}$$

where $K_0 = \delta p(a)/\delta p(b)$ and $d = b - a$.

5. Starting with the solution (10.3-32) to the continuity equation and defining the transit time as that time when the oscilloscope trace reaches a maximum, derive the results shown in (10.4-3) and (10.4-4).

6. Consider a uniform semiinfinite semiconductor whose surface coincides with the yz -plane. The material has a bulk excess carrier lifetime τ and there is a surface recombination velocity s associated with the surface. There is a constant uniform bulk generation rate of excess carriers g' everywhere within the sample. Assuming that a steady state has been reached, calculate the excess carrier density $\delta p(x)$ at all points within the sample.

7. From the results of Problem 6, show that provided $D/L \ll \bar{c}/2$ (D = diffusion coefficient, $L = (D\tau)^{1/2}$ = diffusion length) it makes little difference in the carrier concentration in the bulk or in the diffusive flux of carriers to the surface, whether s is taken to be $\bar{c}/2$ or infinity. Show that this condition can also be stated in the form $L \gg \frac{2}{3}\lambda$.

8. Prove explicitly that the set of functions $\{\cos \alpha_n a\}$ with α_n defined by (10.7-19) is orthogonal. Find the normalization constant, thus deriving (10.7-22).

9. The observed photoconductive decay constant associated with a very thick sample

of *n*-type Germanium ($D_p = 45 \text{ cm}^2/\text{sec}$, $\mu_p = 1800 \text{ cm}^2/\text{V}\cdot\text{sec}$) is $500 \mu\text{sec}$. The sample is cut into slices 0.1 cm thick, the surfaces are etched chemically, and the photoconductive decay constant is then observed to be $300 \mu\text{sec}$. What is the surface recombination velocity associated with the etched surfaces of the thin samples?

10. Show from the results of the Shockley-Read theory for a material with trapping centers for which $\tau_{n0} = \tau_{p0}$ that the maximum possible lifetime occurs when ε_i is at the intrinsic point, and that under these circumstances the lifetime is given by

$$\tau = \tau_{p0} \left[1 + \cosh \frac{\varepsilon_i - \varepsilon_t}{kT} \right]$$

where ε_t is the position of the Fermi level for the intrinsic condition.

11. Discuss the variation of lifetime with the injection level δp for a material in which $\tau_{p0} = \tau_{n0}$ and in which $\varepsilon_t = \varepsilon_v + \frac{1}{2}\Delta\varepsilon$, in terms of the Shockley-Read theory.

12. Show from the results and techniques developed in Section 10.5 that the numerical constant α in Equations (10.5-8) and (10.5-9) must have the value $2/3$.

13. It is sometimes stated that since the equilibrium flux of particles across a plane in a Boltzmann distribution is $p\bar{c}/4$, the maximum obtainable surface recombination velocity (associated with a surface of zero reflection coefficient) should be $s = \bar{c}/4$. This conclusion is contradicted by (10.5-15), which predicts that the maximum obtainable value for s is $\bar{c}/2$. Point out the physical error in this argument, and show that the value $\bar{c}/2$ given by (10.5-15) is really to be expected on physical grounds.

14. Consider a bar of extrinsic *n*-type germanium, of constant cross-section, extending along the x -axis. A constant current is made to flow along the positive x -direction by connecting a battery to the ends of the sample. A uniform density of excess electron-hole pairs is created at $t = 0$ in the region $x_1 < x < x_2$. This excess carrier distribution drifts with velocity $\mu_p E_0$ in the positive x -direction, according to the results of Section 10.2. Explain, using *physical* arguments only, why the excess carrier distribution moves to the right, along the $+x$ -axis, despite the fact that the excess electrons within the distribution must be moving to the *left* in the applied field. Draw a diagram showing (schematically) the internal field and the space charge density which causes it. You may assume that the bulk excess carrier lifetime is infinite, that the surface recombination velocity is zero, and that the excess carrier density is small compared with the equilibrium majority carrier density. *Hint:* Begin by noting that due to the constancy of current density along the sample and the modulation of conductivity by the excess carriers, the electric field must be *smaller* within the excess carrier pulse than outside it.

GENERAL REFERENCES

- A. Many and R. Bray, "Lifetime of Excess Carriers in Semiconductors," in *Progress in Semiconductors*, Heywood and Co., London (1958), Vol. 3, pp. 117-151.
- Allen Nussbaum, *Semiconductor Device Physics*, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1962).
- W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Company, Inc., New York (1950).
- R. A. Smith, *Semiconductors*, Cambridge University Press, London (1961).
- E. Spenke, *Electronic Semiconductors*, McGraw-Hill Book Co., Inc., New York (1958).

CHAPTER 5

Recombination–Generation Processes

When a semiconductor is perturbed from the equilibrium state there is typically an attendant modification in the carrier numbers inside the semiconductor. Recombination–generation (R–G) is nature's order-restoring mechanism, the means whereby the carrier excess or deficit inside the semiconductor is stabilized (if the perturbation is maintained) or eliminated (if the perturbation is removed). Since nonequilibrium conditions prevail during device operation, recombination–generation more often than not plays a major role in shaping the characteristics exhibited by a device. In this chapter we first provide basic R–G information and include a survey of recombination–generation processes. The majority of the chapter is devoted to a detailed examination of the often-dominant R–G center process; both bulk and surface recombination–generation are analyzed. The chapter concludes with a brief presentation of practical R–G facts and supplemental information.

5.1 INTRODUCTION

5.1.1 Survey of R–G Processes

In semiconductor work the terms *Recombination* and *Generation* are defined as follows:

Recombination: A process whereby electrons and holes (carriers) are annihilated or destroyed.

Generation: A process whereby electrons and holes are created.

These definitions are clearly of a very general nature and actually encompass a number of function-related processes. Herein we survey the more common R–G processes, using the energy band diagram as the prime visualization aid. Because of its particular relevance in optical applications, special note is made of how energy is added to or removed from the system during each of the R–G events.

Recombination Processes

- (1) **Band-to-Band Recombination.** Band-to-band recombination, also referred to as direct thermal recombination, is conceptually the simplest of all recombination processes. As pictured in Fig. 5.1(a), it merely involves the direct annihilation of a conduction band electron and a valence band hole, the electron falling from an allowed conduction band state into a vacant valence band state. This process is typically *radiative*, with the excess energy released during the process going into the production of a photon (light).
- (2) **R–G Center Recombination.** In Subsection 4.3.3 it was noted that certain impurity atoms can introduce allowed energy levels (E_T) into the midgap region of a semiconductor. Crystal defects, particularly defects "decorated" with impurity atoms, can also give rise to deep-level states. As shown in Fig. 5.1(b), the R–G centers thereby created act as intermediaries in the envisioned recombination process. First one type of carrier and then the other type of carrier is attracted to the R–G center. The capture of an electron and a hole at the same site leads to the annihilation of the electron-hole pair. Alternatively, the process may be described in terms of the state-to-state transitions of a single carrier: a carrier is first captured at the R–G site and then makes an annihilating transition to the opposite carrier band. R–G center recombination, also called indirect thermal recombination,¹ is characteristically non-radiative. Thermal energy (heat) is released during the process, or equivalently, lattice vibrations (phonons) are produced.
- (3) **Recombination via Shallow Levels.** Like R–G centers, donor and acceptor sites can also function as intermediaries in the recombination process (see Fig. 5.1(c)). If an electron is captured at a donor site, however, it has a high probability at room temperature of being re-emitted into the conduction band before completing the remaining step(s) of the recombination process. A similar statement can be made for holes captured at acceptor sites. For this reason donor and acceptor sites may be likened to extremely inefficient R–G centers, and the probability of recombination occurring via shallow levels is usually quite low at room temperature. It should be noted, nevertheless, that the largest energy step in shallow-level recombination is typically radiative and that the probability of observing shallow-level processes increases with decreasing system temperature.
- (4) **Recombination Involving Excitons.** Normally, electrons and holes may be viewed as individual particles that respond independently to applied forces. However, it is possible for an electron and a hole to become bound together into a hydrogen-atom-like arrangement which moves as a unit in response to applied forces. This coupled electron-hole pair is called an *exciton*. It is also possible for one of the exciton components to be trapped at a shallow-level site; the resulting configuration is called a *bound exciton*. Since a certain amount of energy goes into the

¹In the older device literature, R–G centers and recombination-generation via R–G centers are sometimes referred to as SRH (Shockley, Read, Hall) centers and SRH recombination-generation, respectively. W. Shockley and W. T. Read, Jr.^[1], and independently R. N. Hall^[2,3], were the first to model and investigate this process.

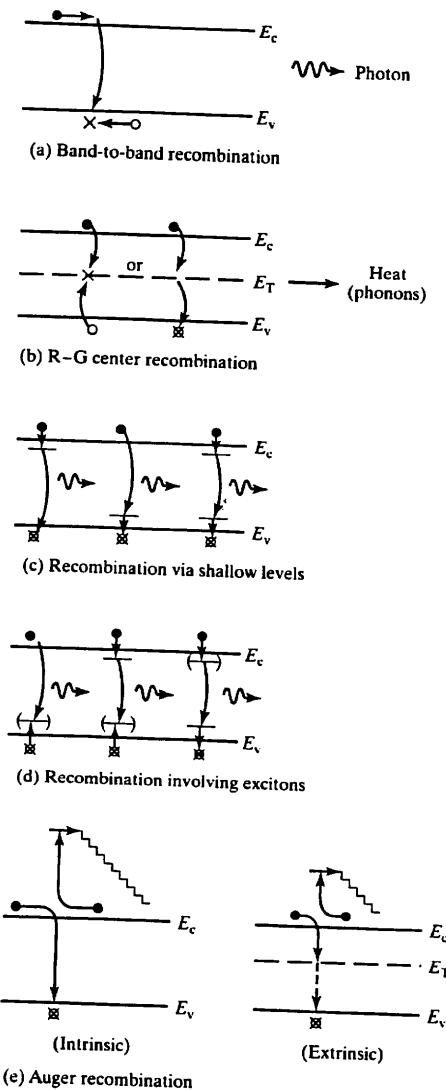


Figure 5.1 Energy-band visualizations of recombination processes.

P.02

formation of an exciton, the difference between the electron and hole energies of the coupled pair can be less than the band gap energy. The formation of an exciton is therefore viewed as introducing a temporary level into the band gap slightly above the valence band edge or slightly below the conduction band edge. In Fig. 5.1(d) these levels are enclosed by parentheses. As pictured in Fig. 5.1(d), recombination of the exciton components can give rise to subband-gap radiation. Recombination involving excitons is a very important mechanism at low temperatures and is the major light-producing mechanism in Light Emitting Diodes (LEDs) containing shallow-level isoelectronic centers.

- (5) *Auger Recombination.* The final recombination processes to be considered are the non-radiative Auger (pronounced Oh-jay) type processes. In an Auger process (see Fig. 5.1(e)), band-to-band recombination or trapping at a band gap center occurs simultaneously with the collision between two like carriers. The energy released by the recombination or trapping subprocess is transferred during the collision to the surviving carrier. Subsequently, this highly energetic carrier “thermalizes”—loses energy in small steps through collisions with the semiconductor lattice. The “staircases” in Fig. 5.1(e) represent the envisioned stepwise loss of energy. Because the number of carrier-carrier collisions increases with increased carrier concentration, Auger recombination likewise increases with carrier concentration, becoming very important at high carrier concentrations. Auger recombination must be considered, for example, in treating degenerately doped regions of a device structure and in the detailed operational modeling of concentrator-type solar cells, junction lasers, and LEDs.

Generation Processes

Any of the foregoing recombination processes can be reversed to generate carriers. Band-to-band generation, for example, is pictured in Fig. 5.2(a). Note that either thermal energy or light can provide the energy required for the band-to-band transition. If thermal energy is absorbed the process is alternatively referred to as direct thermal generation; if externally introduced light is absorbed the process is called photogeneration. The thermally assisted generation of carriers with R-G centers acting as intermediaries is envisioned in Fig. 5.2(b). Figure 5.2(c) pictures the photoemission of carriers from band gap centers, typically a rather improbable process. Finally, impact ionization, the inverse of Auger recombination, is shown in Fig. 5.2(d). In this process an electron-hole pair is produced as a result of the energy released when a highly energetic carrier collides with the crystal lattice. The generation of carriers through impact ionization routinely occurs in the high E-field regions of devices and is responsible for the avalanche breakdown in *pn* junctions.

5.1.2 Momentum Considerations

All of the various recombination-generation processes we have cited occur at all times in all semiconductors—they even occur under equilibrium conditions. The critical issue is not whether the processes are occurring, but rather the rates at which the various processes are occurring. Typically, one need be concerned only with the dominant process the process proceeding at the fastest rate. We have already pointed out that

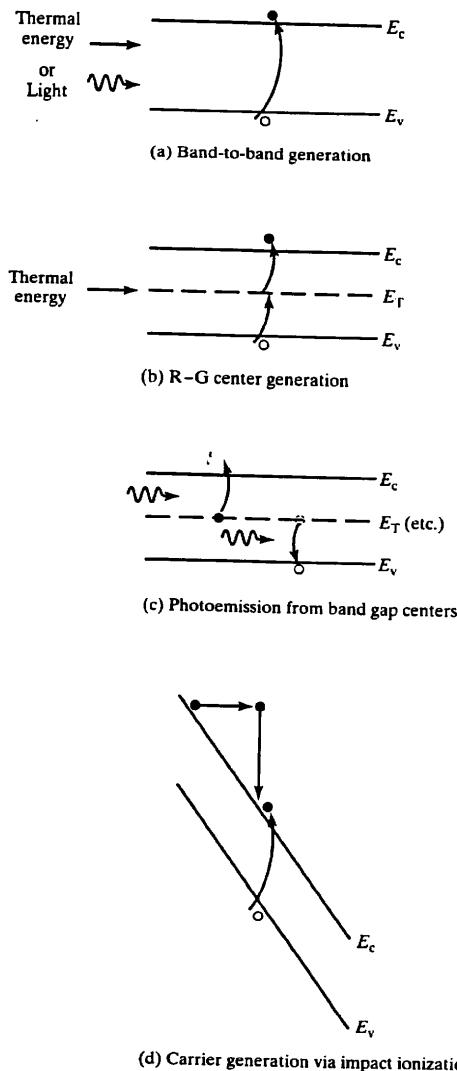


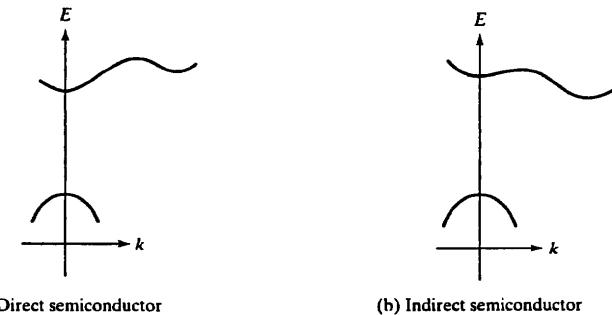
Figure 5.2 Energy-band visualizations of generation processes.

P.03
nazir
NOV-20-2004 17:08

improbable (occur at a much slower rate than competing processes). Thus, based on our survey, one might expect either band-to-band or R-G center recombination-generation to be dominant in low-field regions of a nondegenerately doped semiconductor maintained at room temperature. From the energy band explanation alone one might ever speculate that band-to-band recombination-generation would dominate under the cited "standard" conditions. Visualization of R-G processes using the energy band diagram, however, can be misleading. The E - x plot examines only changes in energy whereas crystal momentum in addition to energy is conserved in any R-G process. Changes in the crystal momentum must also be examined and, as it turns out, momentum-conservation requirements play an important role in setting the process rate.

Crystal-momentum-related aspects of R-G processes are conveniently discussed with the aid of E - k plots. In Subsection 3.3.2 we noted that semiconductors can be divided into two basic groups depending on the general form of the E - k plot. Direct semiconductors such as GaAs are characterized by E - k plots where the conduction band minimum and the valence band maximum both occur at $k = 0$. In indirect semiconductors such as Si and Ge, the conduction band minimum and the valence band maximum occur at different values of k . The two general plot forms are sketched in Fig. 5.3. To employ these plots in visualizing an R-G process, one also needs to know the nature of transitions associated with the absorption or emission of photons and phonons. Photons, being massless entities, carry very little momentum, and a photon-assisted transition is essentially vertical on an E - k plot. (For GaAs the $\Gamma \rightarrow X$ k -width of the E - k diagram is $2\pi/a$, where a is the GaAs lattice constant; $a = 5.65 \text{ \AA}$ at room temperature. By way of comparison, $k_{\text{photon}} = 2\pi/\lambda$. If $E_{\text{photon}} = E_G = 1.42 \text{ eV}$ at room temperature, $\lambda = 0.87 \mu\text{m}$. Clearly, $\lambda \gg a$ and $k_{\text{photon}} \ll 2\pi/a$.) Conversely, the thermal energy associated with lattice vibrations (phonons) is in the 10–50 meV range, whereas the effective phonon mass and associated momentum are comparatively large. Thus on an E - k plot a phonon-assisted transition is essentially horizontal.

Let us now re-examine the band-to-band recombination process. In a direct semiconductor where the k -values of electrons and holes are all bunched near $k = 0$, little change in momentum is required for the recombination process to proceed. The conservation of both energy and crystal momentum is readily met simply by the emission

Figure 5.3 General forms of E - k plots for direct and indirect semiconductors

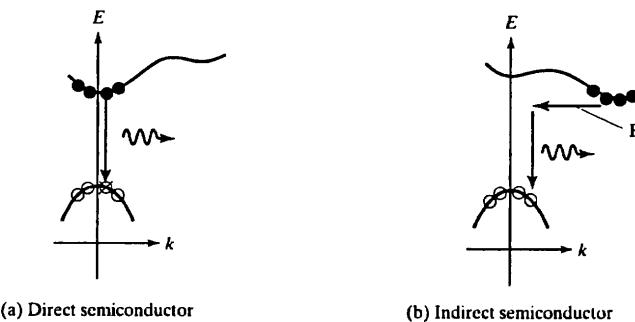


Figure 5.4 *E*-*k* plot visualizations of recombination in direct and indirect semiconductors.

of a photon (see Fig. 5.4(a)). In an indirect semiconductor, on the other hand, there is a large change in crystal momentum, associated with the recombination process. The emission of a photon will conserve energy but cannot simultaneously conserve momentum. Thus for band-to-band recombination to proceed in an indirect semiconductor a phonon must be emitted (or absorbed) coincident with the emission of a photon (see Fig. 5.4(b)).

The rather involved nature of the band-to-band process in indirect semiconductors understandably leads to a diminished recombination rate. Band-to-band recombination is in fact totally negligible compared to R-G center recombination in indirect semiconductors. Although band-to-band recombination proceeds at a much faster rate in direct semiconductors, the R-G center process can never be neglected and even dominates in many instances. Because of its central and often dominant role in the recombination-generation process, we herein concentrate on the R-G center mechanism, devoting the next two sections to its detailed analysis. The analytical procedures developed, it should be noted, are directly applicable to other R-G processes.

5.2 RECOMBINATION-GENERATION STATISTICS

5.2.1 Definition of Terms

R-G statistics is just the technical name given to the mathematical characterization of recombination-generation processes. Since all R-G processes act to change the carrier concentrations as a function of time, "mathematical characterization" simply means obtaining expressions for $\partial n/\partial t$ and $\partial p/\partial t$ due to the process under consideration. In what follows we concentrate on obtaining relationships for $\partial n/\partial t$ and $\partial p/\partial t$ due to recombination-generation via a single-level center. That is, the semiconductor under analysis is assumed to contain only one type of R-G center which introduces allowed states at an energy E_T into the central portion of the band gap. Actual semiconductors may contain a number of deep-level centers, but the process is typically dominated by a single center.

For the purposes of analysis let us define:

$\frac{\partial n}{\partial t} _{R-G}$... Time rate of change in the electron concentration due to both R-G center recombination and R-G center generation.
$\frac{\partial p}{\partial t} _{R-G}$... Time rate of change in the hole concentration due to R-G center recombination-generation.
n_T	... Number of R-G centers per cm^3 that are filled with electrons (equivalent to the previously employed N_T^- if the centers are acceptor-like or $N_T - N_T^+$ if the centers are donor-like).
p_T	... Number of empty R-G centers per cm^3 .
N_T	... Total number of R-G centers per cm^3 , $N_T = n_T + p_T$.

It should be emphasized that $\partial n/\partial t|_{R-G}$ and $\partial p/\partial t|_{R-G}$ are net rates, taking into account the effects of both recombination and generation. $\partial n/\partial t|_{R-G}$ will be negative if there is a net loss of electrons ($R > G$) or positive if there is a net gain of electrons ($G > R$). The designation " $|_{R-G}$ " indicates that the carrier concentrations are changing "due to recombination-generation via R-G centers." The "due to" designation is necessary because, in general, the time rate of change of the carrier concentrations can be affected by a number of processes, including non-R-G processes.

5.2.2 Generalized Rate Relationships

Consider the possible R-G center to energy band transitions shown in Fig. 5.5. The possible transitions, four in all, are (a) electron capture at an R-G center, (b) electron emission from an R-G center, (c) hole capture at an R-G center, and (d) hole emission from an R-G center. The latter two transitions may alternatively be thought of as

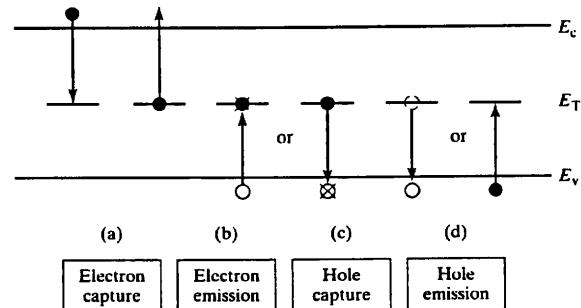


Figure 5.5 Possible electronic transitions between a single-level R-G center and the energy bands.

(c) an electron trapped at an R-G center falling into a vacant valence band state and (d) a valence band electron being excited to the R-G level. The former descriptions of valence band interactions (c) and (d) in terms of holes, however, are preferred and more convenient for our purposes. Since only transitions (a) and (b) affect the electron concentration, and only transitions (c) and (d) affect the hole concentration, one can obviously write

$$\frac{\partial n}{\partial t} \Big|_{R-G} = \frac{\partial n}{\partial t} \Big|_{(a)} + \frac{\partial n}{\partial t} \Big|_{(b)} \quad (5.1a)$$

$$\frac{\partial p}{\partial t} \Big|_{R-G} = \frac{\partial p}{\partial t} \Big|_{(c)} + \frac{\partial p}{\partial t} \Big|_{(d)} \quad (5.1b)$$

As is evident from Eqs. (5.1), fundamental process rates (a) to (d) must be expressed in terms of basic system variables to characterize mathematically the overall recombination-generation rates.

Examining fundamental process (a) in greater detail, we note that electron capture depends critically on the existence of electrons to be captured and the availability of empty R-G centers. If either the electron concentration or the empty-center concentration goes to zero, $\frac{\partial n}{\partial t}|_{(a)} \rightarrow 0$. Moreover, if the nonzero concentration of either electrons or empty R-G centers is conceptually doubled, the probability of capturing an electron doubles and $\frac{\partial n}{\partial t}|_{(a)}$ likewise doubles. In other words, process (a) is expected to proceed at a rate directly proportional to both n and p_T . Taking c_n to be the constant of proportionality, we can therefore write

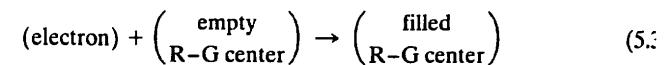
$$\frac{\partial n}{\partial t} \Big|_{(a)} = -c_n p_T n \quad (5.2)$$

where

c_n ... the electron capture coefficient (cm^3/sec)

is a positive-definite constant. Since c_n is taken to be a positive quantity, a minus sign is added to the right-hand side of Eq. (5.2) to account for the fact that electron capture acts to reduce the number of electrons in the conduction band. Note from a consideration of units that c_n must have the dimensions of $1/(\text{concentration-time})$ or cm^3/sec .

Later in this section we will present an alternative (and perhaps more satisfying) derivation of Eq. (5.2) based on collision theory. At this point it is nevertheless worthwhile to note that the same result can be obtained from "chemical reaction" type arguments. Specifically, each of the fundamental processes can be likened to a chemical reaction. In chemical terms process (a) becomes



According to the experimentally verified *rate law* from chemistry, an irreducible chemical reaction of the form $A + B \rightarrow C$ proceeds at a rate given by

$$\text{rate} = (\text{constant})[A][B] \quad (5.4)$$

where $[A]$ and $[B]$ are the concentrations of the reacting components. For process (a) the reacting components are electrons and empty R-G centers; the corresponding concentrations are n and p_T , respectively. Having arrived at the same result is actually not a surprise because the rate law from chemistry also follows from collision-theory considerations.

Turning to process (b) and invoking a parallel set of arguments, we anticipate a process rate which is directly proportional to the product of the filled R-G center concentration and the concentration of empty conduction band states, i.e.,

$$\frac{\partial n}{\partial t} \Big|_{(b)} = (\text{constant}) \left(\begin{array}{c} \text{concentration of empty} \\ \text{conduction band states} \end{array} \right) (n_T) \quad (5.5)$$

If the semiconductor is assumed to be nondegenerate, however, the vast majority of conduction band states will be empty at all times. For a nondegenerate semiconductor, then, the concentration of empty conduction band states will be essentially constant and may be incorporated into the process-rate proportionality constant. Thus

$$\frac{\partial n}{\partial t} \Big|_{(b)} = e_n n_T \quad (5.6)$$

where

e_n ... the electron emission coefficient ($1/\text{sec}$)

is again a positive-definite constant. Here the process rate is positive because electron emission always acts to increase the number of electrons in the conduction band.

Analogous arguments can be applied of course to processes (c) and (d). One readily deduces

$$\frac{\partial p}{\partial t} \Big|_{(c)} = -c_p n_T p \quad (5.7)$$

P.06
and, for a nondegenerate semiconductor,

$$\left. \frac{\partial p}{\partial t} \right|_{(d)} = e_p p_T \quad (5.8)$$

where c_p and e_p are the hole capture and emission coefficients, respectively.

Finally, substituting the fundamental process-rate expressions into Eqs. (5.1), we conclude that

$$r_N \equiv -\left. \frac{\partial n}{\partial t} \right|_{R-G} = c_n p_T n - e_n n_T \quad (5.9a)$$

$$r_P \equiv -\left. \frac{\partial p}{\partial t} \right|_{R-G} = c_p n_T p - e_p p_T \quad (5.9b)$$

In writing down Eqs. (5.9) we have also introduced a more compact net recombination-rate notation. The *net electron and hole recombination rates*, r_N and r_P , are of course positive if recombination is dominant and negative if generation is dominant. Although potentially confusing, this notation does find widespread usage and will be employed subsequently herein. Eqs. (5.9) themselves are very general relationships, applicable in almost any conceivable situation; nondegeneracy is the only limiting restriction. One typically encounters the direct use of these equations in more complex problems (e.g., transient analyses) and in the description of experiments designed to measure the capture and/or emission coefficients.

5.2.3 The Equilibrium Simplification

An intrinsically simpler form of Eqs. (5.9) can be established by invoking the *Principle of Detailed Balance*. Notably, the requirement of detailed balance under equilibrium conditions leads to an interrelationship between the capture and emission coefficients. The statement of the cited principle is as follows:

Principle of Detailed Balance. Under equilibrium conditions each fundamental process and its inverse must *self-balance* independent of any other process that may be occurring inside the material.

A corollary of the stated principle provides an excellent definition of the equilibrium state—namely, equilibrium is the special system state where each fundamental process and its inverse self-balance.

When applied to the R-G center interaction, detailed balance requires fundamental process (a) to self-balance with its inverse process (b), and fundamental process (c) to self-balance with its inverse process (d). Consequently,

$$\left. \begin{array}{l} r_N = 0 \\ r_P = 0 \end{array} \right\} \text{under equilibrium conditions} \quad (5.10a)$$

$$(5.10b)$$

The zero net recombination of carriers under equilibrium conditions forms the mathematical basis for interrelating the emission and capture coefficients. Substituting $r_N = r_P = 0$ into Eqs. (5.9), solving for the emission coefficients, and introducing the subscript "0" to emphasize that all quantities are to be evaluated under equilibrium conditions, one obtains

$$e_{n0} = \frac{c_{n0} p_{T0} n_0}{n_{T0}} = c_{n0} n_1 \quad (5.11a)$$

and

$$e_{p0} = \frac{c_{p0} n_{T0} p_0}{p_{T0}} = c_{p0} p_1 \quad (5.11b)$$

where

$$n_1 = \frac{p_{T0} n_0}{n_{T0}} \quad (5.12a)$$

$$p_1 = \frac{n_{T0} p_0}{p_{T0}} \quad (5.12b)$$

It is next assumed that the emission and capture coefficients all remain approximately equal to their equilibrium values under nonequilibrium conditions, i.e.,

$$e_n \approx e_{n0} = c_{n0} n_1 \approx c_n n_1 \quad (5.13a)$$

and

$$e_p \approx e_{p0} = c_{p0} p_1 \approx c_p p_1 \quad (5.13b)$$

Eliminating the emission coefficients in Eqs. (5.9) using Eqs. (5.13) then yields

$$r_N = -\left. \frac{\partial n}{\partial t} \right|_{R-G} = c_n (p_T n - n_T n_1) \quad (5.14a)$$

$$r_P = -\left. \frac{\partial p}{\partial t} \right|_{R-G} = c_p (n_T p - p_T p_1) \quad (5.14b)$$

P.07
Two comments are in order concerning the Eq. (5.14) results. First, like Eqs. (5.9), the Eq. (5.14) results enjoy wide-ranging applicability. There is of course the added assumption that the emission and capture coefficients (actually, the emission to capture coefficient ratios) remain fixed at their equilibrium values. It is difficult to assess the precise limitations imposed by this assumption, although in situations involving large deviations from equilibrium the validity of the equations is certainly open to question.

The second comment addresses the “simplified” nature of the results. At first glance it would appear that we have merely replaced two system parameters (e_n and e_p) with two new system parameters (n_1 and p_1). The emission coefficients are indeed system parameters that must be determined experimentally; n_1 and p_1 , however, are computable constants. To facilitate the computation of n_1 and p_1 , note that

$$n_1 = \frac{p_{T0} n_0}{n_{T0}} = \left(\frac{N_T - n_{T0}}{n_{T0}} \right) n_0 = \left(\frac{N_T}{n_{T0}} - 1 \right) n_0 \quad (5.15)$$

and for a nondegenerate semiconductor

$$n_0' = n_0 e^{(E_F - E_i)/kT} \quad (5.16)$$

(Same as 4.57a)

Moreover, referring to the n_T definition and Eq. (4.68), one finds

$$\frac{n_{T0}}{N_T} = \frac{1}{1 + e^{(E_T' - E_i)/kT}} \quad (5.17)$$

where $E_T' = E_T \pm kT \ln g_T$. The (+) is used if the R-G centers are acceptor-like and the (-) if the centers are donor-like. g_T is the degeneracy factor introduced in Chapter 4. Combining Eqs. (5.15) through (5.17) then gives

$$n_1 = n_0 e^{(E_T' - E_i)/kT} \quad (5.18a)$$

Likewise

$$p_1 = n_0 e^{(E_i - E_T')/kT} \quad (5.18b)$$

Assuming E_T and g_T are known, n_1 and p_1 are readily computed from Eqs. (5.18). For a quick approximate evaluation of these parameters, use can be made of the fact that $n_1 = n_0$ and $p_1 = p_0$ if the Fermi level is positioned such that $E_F = E_T'$. For example, we know that $n_0 = p_0 = n_i$ if $E_F = E_i$ and $n_0 > n_i$, $p_0 < n_i$ if E_F is positioned above midgap. Thus if E_T' is positioned near midgap we analogously conclude $n_1 \approx p_1 \approx n_i$ without referring to Eqs. (5.18). Likewise, if E_T' is positioned above midgap, $n_1 > n_i$ and $p_1 < n_i$. Also note that n_1 and p_1 obey the np product relation-

5.2.4 Steady-State Relationship

In the vast majority of device problems the analysis is performed assuming that the device is being operated under steady-state or quasi-steady-state¹ conditions. As we will see, the expressions for the net recombination rates take on a much more tractable form under the cited conditions.

Our first task will be to identify what goes on inside a semiconductor under steady-state conditions and to distinguish between the superficially similar equilibrium and steady states. In both the equilibrium and steady states the average values of all macroscopic observables within a system are constant with time—that is, dn/dt , dp/dt , dE/dt , dn_T/dt , etc. are all zero. Under equilibrium conditions the static situation is maintained by the self-balancing of each fundamental process and its inverse. Under steady-state conditions, on the other hand, the status quo is maintained by a *trade-off* between processes. This difference is nicely illustrated in Fig. 5.6, where the envisioned activity inside a small Δx section of a semiconductor is depicted under equilibrium and steady-state conditions. Please note from Fig. 5.6 that the steady-state net recombination rates are characteristically nonzero.

Although the net recombination rates do not vanish under steady-state conditions, there is nevertheless a readily established interrelationship between the net rates. Since n_T does not change with time, and assuming n_T can only change via the R-G center interaction, one can write

$$\frac{dn_T}{dt} = -\frac{\partial n}{\partial t}\Big|_{R-G} + \frac{\partial p}{\partial t}\Big|_{R-G} = r_N - r_P = 0 \quad (5.19)$$

or

$$r_N = r_P \quad \dots \text{under steady-state conditions} \quad (5.20)$$

The equal creation or annihilation of holes and electrons under steady-state conditions in turn fixes n_T for a given n and p . Specifically, equating the right-hand sides of Eqs. (5.14a) and (5.14b), remembering $p_T = N_T - n_T$, and solving for n_T , one obtains

$$n_T = \frac{c_n N_T n + c_p N_T p_1}{c_n(n + n_1) + c_p(p + p_1)} \quad (\text{steady-state}) \quad (5.21)$$

¹The quasi-steady-state or quasistatic assumption is invoked quite often in performing transient analyses where the rate of change of system variables such as n , p , E , etc. is slow compared to the rates of the dominant fundamental processes occurring inside the material. Under quasi-steady-state conditions, the instantaneous state of the system may be considered to be a progression of steady states, and steady-state relationships can be used to describe accurately the state of the system at any instant.

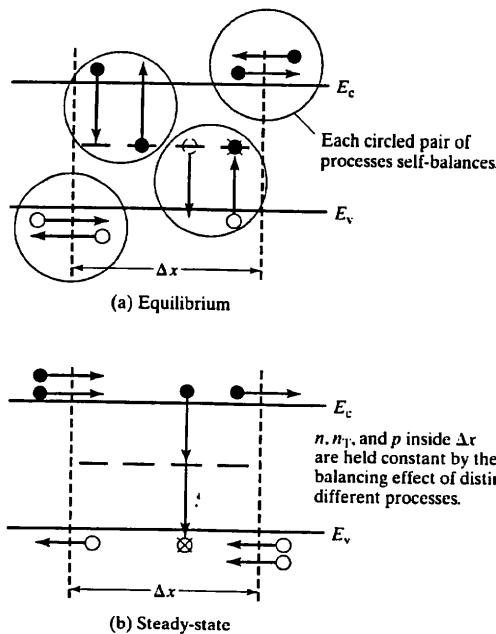


Figure 5.6 Conceptualization of activity inside a small Δx section of a semiconductor under equilibrium and steady-state conditions.

The Eq. (5.21) n_T expression can next be used to eliminate n_i (and p_T) in either Eq. (5.14a) or Eq. (5.14b). After a bit of manipulation which makes use of the fact that $n_i p_i = n_i^2$, we arrive at the result

$$R = r_N = r_P = \frac{np - n_i^2}{\frac{1}{c_p N_T} (n + n_i) + \frac{1}{c_n N_T} (p + p_i)} \quad (5.22)$$

where the symbol R has been introduced to identify the net steady-state recombination rate. Finally, $1/c_n N_T$ and $1/c_p N_T$ have units of time (seconds) and it is therefore reasonable to additionally introduce the time constants

$$\tau_n = \frac{1}{c_n N_T} \quad \dots \text{the electron minority carrier lifetime} \quad (5.23a)$$

and

$$\tau_p = \frac{1}{c_p N_T} \quad \dots \text{the hole minority carrier lifetime} \quad (5.23b)$$

which when substituted into Eq. (5.22) yields

$$R = \frac{np - n_i^2}{\tau_p(n + n_i) + \tau_n(p + p_i)} \quad (5.24)$$

Equation (5.24) is an extremely important result that is encountered again and again in the device literature. It should be emphasized that the R -expression applies to any steady-state situation and gives the net recombination rate for *both* electrons and holes. The τ 's, introduced as a mathematical expedient, are important material parameters and are to be interpreted as the average time an excess minority carrier will live in a sea of majority carriers. This interpretation follows from experiments where in *n*- or *p*-type semiconductor is weakly illuminated with carrier-generating light, the light is extinguished, and the subsequent decay back to equilibrium is monitored as a function of time. Under ideal conditions the time constant for the decay, equal to the average time required to eliminate the excess carriers created by the light, is just τ_n for a *p*-type semiconductor and τ_p for an *n*-type semiconductor. Note that the τ 's, simply referred to as the minority carrier lifetimes for identification purposes, vary inversely with the R-G center concentration, but are *explicitly independent* of the doping concentration. Although potentially computable constants, the τ 's are routinely deduced directly from experimental measurements. We will have more to say about the minority carrier lifetimes in Section 5.4.

5.2.5 Specialized Steady-State Relationships

The expression for the net steady-state recombination rate can be drastically simplified under certain conditions. In what follows we establish the most widely utilized recombination rate relationships—namely, the reduced steady-state expressions valid (1) under low-level injection conditions and (2) when the semiconductor is depleted of carriers.

Low Level Injection

The level of injection specifies the relative magnitude of changes in the carrier concentrations resulting from a perturbation. Low level injection is said to exist if the changes in the carrier concentrations are much *less* than the majority carrier concentration under equilibrium conditions. Conversely, high level injection exists if the changes are much *greater* than the equilibrium concentration of majority carriers. Mathematically, if n and p are the carrier concentrations under arbitrary conditions, n_0 and p_0 the equilibrium carrier concentrations, and $\Delta n = n - n_0$ and $\Delta p = p - p_0$ the changes in the carrier concentrations from their equilibrium values, then

low level injection implies

$$\begin{aligned}\Delta n, \Delta p \ll n_0 \quad (n \approx n_0) & \dots \text{in an } n\text{-type material} \\ \Delta n, \Delta p \ll p_0 \quad (p \approx p_0) & \dots \text{in a } p\text{-type material}\end{aligned}$$

Note that low level injection may also be viewed as a situation where the majority carrier concentration remains essentially unperturbed.

To achieve the desired R -expression simplification it is also necessary to make certain gross assumptions about the R-G center parameters. Specifically, we assume that the dominant R-G centers introduce an E_T' level fairly close to midgap so that $n_1 \approx p_1 \approx n_i$, and that the R-G center concentration is sufficiently low so that $\Delta n \approx \Delta p$. Moreover, τ_n and τ_p are taken to differ by no more than a few orders of magnitude. As it turns out, these are reasonable assumptions consistent with the experimentally observed properties of the dominant R-G centers in actual semiconductors.

Proceeding with the simplification, let us first substitute $n = n_0 + \Delta n$ and $p = p_0 + \Delta p$ into Eq. (5.24):

$$R = \frac{(n_0 + \Delta n)(p_0 + \Delta p) - n_i^2}{\tau_p(n_0 + \Delta n + n_i) + \tau_n(p_0 + \Delta p + p_i)} \quad (5.25a)$$

$$\approx \frac{n_0 p_0 + n_0 \Delta p + p_0 \Delta p + (\Delta p)^2 - n_i^2}{\tau_p(n_0 + \Delta p + n_i) + \tau_n(p_0 + \Delta p + p_i)} \quad (5.25b)$$

In the latter form of Eq. (5.25) we have set $\Delta n = \Delta p$ and $n_i = p_i = n_i$ in accordance with previously stated assumptions. Eq. (5.25b) is valid of course for either n - or p -type material. For illustrative purposes let us assume the semiconductor to be n -type. Examining the numerator on the right-hand side of Eq. (5.25b) we note that

$$\begin{aligned}n_0 p_0 &= n_i^2 \rightarrow \text{cancels } -n_i^2 \\ n_0 \Delta p &\gg p_0 \Delta p \quad (n_0 \gg p_0) \\ n_0 \Delta p &\gg \Delta p^2 \quad (n_0 \gg \Delta p)\end{aligned}$$

All but the $n_0 \Delta p$ term may be neglected.

In a like manner, examining the denominator, we note that

$$\begin{aligned}\tau_p(n_0 + \Delta p + n_i) &\approx \tau_p n_0 \quad (n_0 \gg \Delta p, n_0 \gg n_i) \\ \tau_p n_0 &\gg \tau_n(p_0 + \Delta p + p_i) \quad (n_0 \gg p_0 + \Delta p + p_i; \quad \tau_n \sim \tau_p)\end{aligned}$$

All but the $\tau_p n_0$ term may be neglected.

We therefore arrive at the drastically simplified result:

$$R = \frac{\Delta p}{\tau_p} \quad \dots \text{in an } n\text{-type material} \quad (5.26a)$$

$$R = \frac{\Delta n}{\tau_n} \quad \dots \text{in a } p\text{-type material} \quad (5.26b)$$

The recombination rate expressions obtained here are identical to the "standard case" expressions found in introductory texts and used extensively in device analyses. Eq. (5.25b), we should point out, is valid for any level of injection and, as is readily verified, simplifies to $R = \Delta p / (\tau_n + \tau_p)$ under high level injection conditions where $\Delta n = \Delta p \gg n_0$ or p_0 .

R-G Depletion Region

The simplified expression for the net recombination rate inside an R-G depletion region is another special-case result that is encountered quite often in device analyses. An R-G depletion region is formally defined to be a semiconductor volume where $n \ll n_i$ and (simultaneously) $p \ll p_i$. Since $np \ll n_1 p_1 = n_i^2$, a deficit of carriers always exists inside the envisioned depletion region.

Before continuing, it is important to carefully distinguish between an R-G depletion region and the "electrostatic" depletion region encountered in pn -junction analyses. In developing approximate expressions for the electrostatic variables (ρ, \mathcal{E}, V) in the pn -junction analysis, it is common practice to assume that the carrier concentrations are negligible compared to the net doping concentration over a width W about the metallurgical junction. This is the well-known depletion approximation and W is the width of the electrostatic depletion region. Unfortunately, the cited terminology can be somewhat misleading. The existence of an electrostatic depletion region merely requires n and p to be small compared to $|N_D - N_A|$; it does not imply the existence of a carrier deficit ($n < n_0; p < p_0$) within the region. When a pn -junction is zero biased an electrostatic depletion region exists inside the structure, but $n = n_0, p = p_0, np = n_i^2$ everywhere because equilibrium conditions prevail. Moreover, when the junction is forward biased there is actually an excess of carriers ($n > n_0, p > p_0$) in the electrostatic depletion region. A carrier deficit and associated R-G depletion region are created inside the electrostatic depletion region in a pn -junction only under reverse bias conditions. The width of the R-G depletion region is always smaller than W , but approaches W at large reverse biases.

With $n \ll n_i$ and $p \ll p_i$ in Eq. (5.24), we obtain by inspection,

$$R \approx -\frac{n_i^2}{\tau_p n_i + \tau_n p_i} \quad (5.27)$$

or

$$G = -R = \frac{n_i}{\tau_g} \quad \dots \text{in an R-G depletion region} \quad (5.28)$$

where

$$\tau_g = \tau_p(n_i/n_i) + \tau_n(p_i/n_i) \quad (5.29a)$$

$$= \tau_p + \tau_n \quad \dots \text{if } E_{T'} = E_i \quad (5.29b)$$

A negative R , of course, indicates that a net generation of carriers is taking place within the depletion region. Hence the introduction of the net generation rate symbol, $G = -R$, and the use of the subscript "g" to identify the *generation lifetime*, τ_g .

5.2.6 Physical View of Carrier Capture

We conclude the discussion of bulk recombination-generation statistics by presenting an alternative derivation of Eq. (5.2) based on a spatially oriented view of the capture process. This derivation has been included because it provides additional insight into the capture process while simultaneously introducing the very useful capture cross section concept.

The real-space visualization of electron capture at an R-G center is shown in Fig. 5.7(a). In this idealized view of the capture process, empty R-G centers are modeled as spheres randomly distributed about the semiconductor volume. Filled R-G centers are thought of as fixed dots and electrons as moving dots. The rather erratic path of the electron as it moves through the semiconductor is caused by collisions with vibrating lattice atoms and ionized impurity atoms. In the course of its travels an electron is considered to have been captured if it penetrates the sphere surrounding an empty R-G center site.

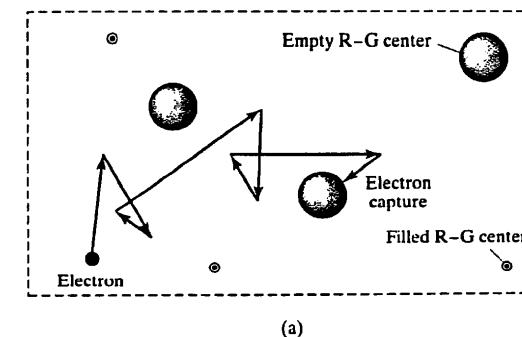
From the qualitative description of the capture process we expect the electron velocity to be a factor in setting the capture rate. Clearly, the greater the distance traveled by the electron per second, the greater the likelihood of electron capture within a given period of time. Now, it can be established in a relatively straightforward manner (see Problem 4.10) that the average kinetic energy of electrons in the conduction band of a nondegenerate semiconductor is $(3/2)kT$ under equilibrium conditions. Thus, the *thermal velocity* or average velocity under equilibrium conditions, v_{th} , is given by

$$\frac{1}{2} m^* v_{th}^2 = \frac{3}{2} kT \quad (5.30)$$

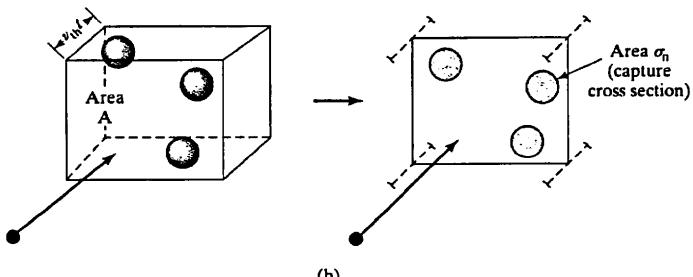
or

$$v_{th} = \sqrt{3kT/m^*} \approx 10^7 \text{ cm/sec at } 300 \text{ K}^{\dagger} \quad (5.31)$$

[†]Because of uncertainties in the effective mass to be employed, it is standard practice in R-G computations to use the free electron mass in calculating v_{th} . This applies to both electrons and holes.



(a)



(b)

Figure 5.7 (a) Real-space visualization of electron capture at an R-G center. (b) Construct employed in determining the capture rate. Left: volume through which an electron passes in a time t . Right: effective recombination plane.

Under nonequilibrium conditions there are of course added velocity components. However, the added velocity is typically small compared to v_{th} and, as a general rule, little error is introduced by assuming $v \approx v_{th}$ in modeling the capture process under arbitrary conditions.

We can now turn to the derivation proper. In a time t (assumed to be small) an electron will travel a distance $v_{th}t$ and will pass through a volume of material equal to $A v_{th}t$, where A is the cross-sectional area of the material normal to the electron's path. In this volume there will be p_T empty R-G centers per cm^3 , or a total number of $p_T A v_{th}t$ empty R-G centers. Since the R-G centers are assumed to be randomly distributed, the probability of the electron being captured in the volume can be determined by conceptually moving the centers of all R-G spheres to a single plane in the middle of the volume and noting the fraction of the plane blocked by the R-G centers

(see Fig. 5.7(b)). If the area of the plane blocked by a single R-G center is $\sigma_n = \pi r^2$, where r is the radius of the R-G spheres, the total area blocked by empty R-G centers will be $p_T A \sigma_n v_{th} t$. The fraction of the area giving rise to capture will be $p_T A \sigma_n v_{th} t / A$. The probability of electron capture in the volume is then $p_T \sigma_n v_{th} t$, and the capture rate (probability of capture per second) for a single electron is $p_T \sigma_n v_{th} t / t = p_T \sigma_n v_{th}$. Given n electrons per unit volume, the number of electrons/cm³ captured per second will be $n p_T \sigma_n v_{th}$, or

$$\frac{\partial n}{\partial t} \Big|_{(a)} = -\sigma_n v_{th} p_T n \quad (5.32)$$

Equations (5.2) and (5.32) are clearly equivalent if one identifies

$$c_n = \sigma_n v_{th} \quad (5.33a)$$

Analogously

$$c_p = \sigma_p v_{th} \quad (5.33b)$$

σ_n and σ_p , the electron and hole *capture cross sections*, are often used to gauge the relative effectiveness of R-G centers in capturing carriers. In fact, because of their “intuitive” appeal, the capture cross sections find a much greater utilization in the device literature than the more basic capture coefficients.

5.3 SURFACE RECOMBINATION-GENERATION

5.3.1 Introductory Comments

In many devices under certain conditions, surface recombination-generation can be as important as, or more important than, the “bulk” recombination-generation considered in the preceding section. Whereas bulk R-G takes place at centers spatially distributed throughout the volume of a semiconductor, surface recombination-generation refers to the creation/annihilation of carriers in the near vicinity of a semiconductor surface via the interaction with interfacial traps. Interfacial traps or surface states are functionally equivalent to R-G centers localized at the surface of a material. Unlike bulk R-G centers, however, interfacial traps are typically found to be continuously distributed in energy throughout the semiconductor band gap.

As pictured in Fig. 5.8, the same fundamental processes that occur in the semiconductor bulk also occur at the semiconductor surface. Electrons and holes can be captured at surface centers; electrons and holes can be emitted from surface centers. From the energy band description alone one might expect additional transitions to occur between surface centers at different energies. However, given realistic interfacial-trap densities, these seemingly plausible intercenter transitions are extremely unlikely

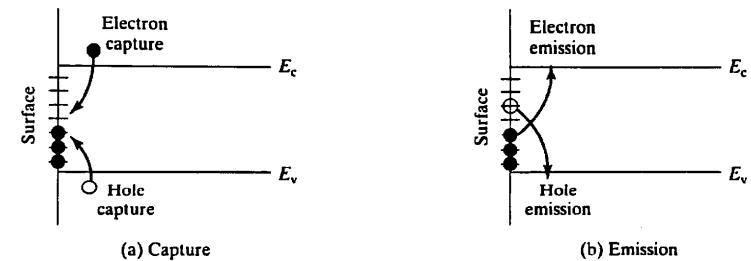


Figure 5.8 Recombination-generation at a semiconductor surface via transitions to and from interfacial traps. (a) Electron and hole capture leading to carrier recombination. (b) Electron and hole emission leading to carrier generation.

because of the spread-out or spatially isolated nature of the centers on the surface plane (see Fig. 5.9).

The very obvious physical similarity between surface and bulk recombination-generation leads to a parallel mathematical description of the processes. This will allow us to establish a number of surface relationships by direct inference from the corresponding bulk result. Nevertheless, there are two major differences:

- (1) Because surface states are arranged along a plane in space rather than spread out over a volume, the net recombination rates are logically expressed in terms of carriers removed from a given band per UNIT AREA-second.
- (2) Whereas a single level usually dominates bulk recombination-generation, the surface-center interaction routinely involves centers distributed in energy throughout the band gap. Hence, it is necessary to add up or integrate the single-level surface rates over the energy band gap.

5.3.2 General Rate Relationships (Single Level)

It is convenient to initially determine the net recombination rates associated with interface traps at a single energy, and to subsequently modify the results to account for

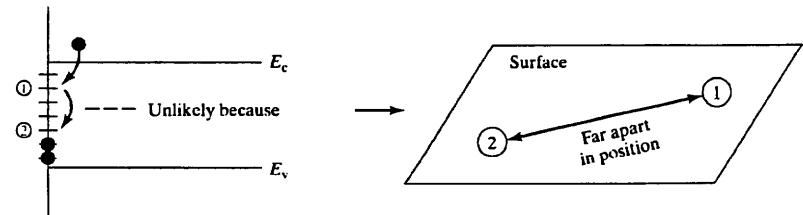


Figure 5.9 Visualization of an intercenter surface transition and a pictorial explanation of

P.12
the distributed nature of the states. To begin the analysis we therefore assume the band gap contains a single energy level, E_{IT} . Adding the subscript s to the corresponding bulk definitions, let us also define:

r_{Ns}	... Net electron recombination rate at surface centers (that is, the net change in the number of conduction band electrons/cm ² -sec due to electron capture and emission at the single level surface centers).
r_{Ps}	... Net hole recombination rate at the surface centers.
n_{Ts}	... Filled surface centers/cm ² at energy E_{IT} .
p_{Ts}	... Empty surface centers/cm ² at energy E_{IT} .
N_{Ts}	... Total number of surface states/cm ² ; $N_{Ts} = n_{Ts} + p_{Ts}$.
n_s	... Surface electron concentration (number per cm CUBED); $n_s = n _{surface}$.
p_s	... Surface hole concentration.
e_{ns}, e_{ps}	... Surface electron and hole emission coefficients (1/sec).
c_{ns}, c_{ps}	... Surface electron and hole capture coefficients (cm ³ /sec).

Given the one-to-one correspondence between physical processes and parametric quantities, we likewise expect a term-by-term correspondence in the expressions for the surface and bulk net recombination rates. For the bulk process we obtained the general relationships

$$r_N = c_n p_T n - e_n n_T$$

$$r_P = c_p n_T p - e_p p_T$$

Thus, for single level surface recombination-generation we conclude by analogy

$$\left[\begin{array}{l} r_{Ns} = c_{ns} p_{Ts} n_s - e_{ns} n_{Ts} \\ r_{Ps} = c_{ps} n_{Ts} p_s - e_{ps} p_{Ts} \end{array} \right] \quad (5.34a)$$

$$(5.34b)$$

Like their bulk counterparts, Eqs. (5.34) are very general relationships, with nondegeneracy being the only limiting restriction.

In the bulk analysis we next invoked detailed balance to obtain the simplified general relationships

$$r_N = c_n (p_T n - n_T n_i)$$

$$r_P = c_p (n_T p - p_T p_i)$$

Analogously, therefore

$$\left[\begin{array}{l} r_{Ns} = c_{ns} (p_{Ts} n_s - n_{Ts} n_{is}) \\ r_{Ps} = c_{ps} (n_{Ts} p_s - p_{Ts} p_{is}) \end{array} \right] \quad (5.35a)$$

where

$$e_{ns} = c_{ns} n_{is} \quad (5.36a)$$

$$e_{ps} = c_{ps} p_{is} \quad (5.36b)$$

and, taking the surface-center degeneracy factor to be unity,

$$n_{is} = n_i e^{(E_{II} - E_i)/kT} \quad (5.37a)$$

$$p_{is} = n_i e^{(E_i - E_{II})/kT} \quad (5.37b)$$

5.3.3 Steady-State Relationships

Single Level

Under steady-state conditions, as in the bulk,

$$r_{Ns} = r_{Ps} = R_s \quad (5.38)$$

if the filled-state population of interfacial traps at E_{IT} is assumed to change exclusively via thermal band-to-trap interactions. Equating the right-hand sides of Eqs. (5.35a) and (5.35b), one obtains

$$n_{Ts} = \frac{c_{ns} N_{Ts} n_s + c_{ps} N_{Ts} p_{is}}{c_{ns}(n_s + n_{is}) + c_{ps}(p_s + p_{is})} \quad \left(\begin{array}{l} \text{steady} \\ \text{state} \end{array} \right) \quad (5.39)$$

Equation (5.39) can then be used to eliminate n_{is} (and p_{is}) in either Eq. (5.35a) or Eq. (5.35b), yielding

$$\left[R_s = \frac{n_s p_s - n_i^2}{\frac{1}{c_{ps} N_{Ts}}(n_s + n_{is}) + \frac{1}{c_{ns} N_{Ts}}(p_s + p_{is})} \right] \quad (5.40)$$

Again, Eq. (5.40) is seen to be the direct surface analog of the corresponding bulk result [Eq. (5.22)]. Please note, however, that $1/c_{ns} N_{Ts}$ and $1/c_{ps} N_{Ts}$ are NOT time constants. In fact, $c_{ns} N_{Ts} = s_n$ and $c_{ps} N_{Ts} = s_p$ have units of a velocity, cm/sec, and are, respectively, the (single level) surface recombination velocities for electrons and holes. Conceptually, the recombination of excess carriers at a surface causes a flow of carriers toward the surface. Provided that low level injection conditions prevail and the surface bands are flat ($\mathcal{E}|_{surface} = 0$), the velocities at which the excess carriers flow into the surface will be s_n and s_p , respectively, in p - and n -type semiconductors containing a single

E_{IT} level. Because the single-level case is of little practical interest, the s_n and s_p velocities as defined above are unlikely to be encountered in the device literature. Nevertheless, it is commonplace to encounter the symbol s and an appropriately defined surface recombination (or generation) velocity in surface analyses. Functionally, the s in a surface R-G analysis replaces the τ in the corresponding bulk R-G analysis as the material constant characterizing the net carrier action.

Multi-Level

As already noted, surface centers are typically found to be continuously distributed in energy throughout the semiconductor band gap. The net recombination rates associated with the individual centers in the distribution must be added together to obtain the overall net recombination rate. A simple addition of rates is possible, we should interject, because the centers at different energies are noninteracting (i.e., as previously described, inter-center transitions are extremely unlikely). The task at hand is to appropriately modify the single-level result to obtain the net recombination rate associated with a continuous distribution of noninteracting surface centers. Although we specifically consider the steady-state case, a similar modification procedure can be readily applied to any of the foregoing single-level results.

Let $D_{IT}(E)$ be the density of interfacial traps (traps per $\text{cm}^2\text{-eV}$) at an arbitrarily chosen energy E ($E_v \leq E \leq E_c$). $D_{IT}(E)dE$ will then be the number of interfacial traps per cm^2 with energies between E and $E + dE$. Associating $D_{IT}(E)dE$ with N_{Ts} in the single-level relationship [$N_{Ts} \rightarrow D_{IT}(E)dE$ in Eq. (5.40)] and recognizing that these states provide an incremental contribution (dR_s) to the overall net recombination rate when there is a distribution of states, one deduces

$$dR_s = \frac{n_s p_s - n_i^2}{(n_s + n_{is})/c_{ps} + (p_s + p_{is})/c_{ns}} D_{IT}(E) dE \quad (5.41)$$

= net recombination rate associated with centers
between E and $E + dE$

Integrating over all band gap energies then yields

$$R_s = \int_{E_v}^{E_c} \frac{n_s p_s - n_i^2}{(n_s + n_{is})/c_{ps} + (p_s + p_{is})/c_{ns}} D_{IT}(E) dE \quad (5.42)$$

In utilizing the above relationship it must be remembered that all of the trap parameters can vary with energy. The anticipated variation of $D_{IT}(E)$ with energy is of course noted explicitly. With the integration variable E replacing E_{IT} in Eqs. (5.37), n_{is} and p_{is} are also seen to be functions of energy. In fact, n_{is} and p_{is} are exponential functions of energy. Even c_{ns} and c_{ps} can vary with the trap energy across the band gap. Like $D_{IT}(E)$, however, $c_{ns}(E)$ and $c_{ps}(E)$ must be determined from experimental measurements.

5.3.4 Specialized Steady-State Relationships

To conclude the discussion of surface recombination-generation we consider two special cases of practical interest that give rise to simplified R_s relationships. The special-case analyses, treating (1) low level injection when the bands are flat and (2) a depleted ($n_s \rightarrow 0, p_s \rightarrow 0$) surface, nicely illustrate simplification procedures and are the surface analog of the bulk analyses presented in Subsection 5.2.5.

Low Level Injection/Flat Band

We assume the semiconductor under analysis is n -type, the energy bands are flat at the surface ($\mathcal{E}|_{\text{surface}} = 0$, implying $n_{s0} = N_D$), and low level injection conditions prevail ($\Delta n_s = \Delta p_s \ll n_{s0}$). c_{ns} and c_{ps} are also taken to be comparable in magnitude. Under the stated conditions and introducing $n_s = n_{s0} + \Delta p_s$, $p_s = p_{s0} + \Delta p_s$, we find

$$n_s p_s - n_i^2 \approx n_{s0} \Delta p_s \quad (5.43)$$

and

$$(n_s + n_{is})/c_{ps} + (p_s + p_{is})/c_{ns} \approx (n_{s0} + n_{is})/c_{ps} + p_{is}/c_{ns} \quad (5.44)$$

Thus Eq. (5.42) simplifies to

$$R_s \approx \left[\int_{E_v}^{E_c} \frac{c_{ps} D_{IT}}{1 + \frac{n_{is}}{n_{s0}} + \frac{c_{ps} p_{is}}{c_{ns} n_{s0}}} dE \right] \Delta p_s \quad (5.45)$$

For a given set of trap parameters the integral in Eq. (5.45) is a system constant and must have the dimensions of a velocity. Logically taking this integral to be a surface recombination velocity, we can therefore write

$$R_s = s_p \Delta p_s \quad \dots n\text{-type material} \quad (5.46)$$

where

$$s_p = \int_{E_v}^{E_c} \frac{c_{ps} D_{IT}}{1 + \frac{n_{is}}{n_{s0}} + \frac{c_{ps} p_{is}}{c_{ns} n_{s0}}} dE \quad (5.47)$$

An analogous result is obtained for p -type material.

The simplification procedure can be carried one step further if more limiting assumptions are made concerning the trap parameters. Notably, D_{IT} and the capture coefficients are often assumed to be approximately constant (energy-independent) over the middle portion of the band gap. Indeed, a certain amount of experimental Si data tends to support this assumption^[4,5]. Let us pursue the implications of the assumption. Referring to Fig. 5.10(a), which provides a sketch of the denominator in the s_p integrand versus energy, note

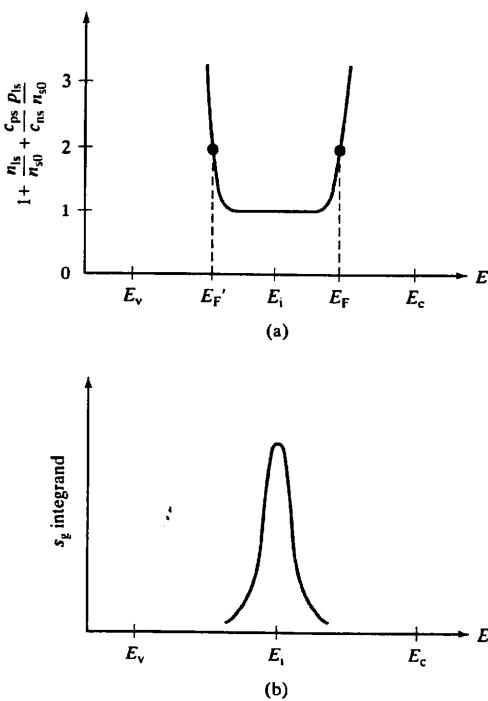


Figure 5.10 Observations related to special-case simplification of the R_s relationship. Sketches of (a) the denominator in the s_p integrand [Eq. (5.47)] and (b) the s_g integrand [Eq. (5.52b)] as a function of band gap energy.

$$1 + \frac{n_{ls}}{n_{s0}} + \frac{c_{ps} p_{ls}}{c_{ns} n_{s0}} = 1 + \frac{n_i}{N_D} \left[e^{(E-E_i)/kT} + \frac{c_{ps}}{c_{ns}} e^{(E_i-E)/kT} \right] \quad (5.48)$$

$$\approx 1 \quad \dots \text{if } E_F' \leq E \leq E_F \quad (5.49a)$$

$$\rightarrow \infty \quad \dots \text{for } E < E_F' \text{ and } E > E_F \quad (5.49b)$$

E_F' being the energy in the band gap where

$$\frac{n_i}{N_D} \frac{c_{ps}}{c_{ns}} e^{(E_i-E)/kT} = 1 \quad (5.50)$$

In other words, the denominator of the s_p integrand is approximately unity in the midgap region where D_{IT} and the capture coefficients are assumed to be approximately constant. Outside this range the denominator becomes large and the contribution to the overall integral is small. Thus, if D_{IT} and the capture coefficients are taken to be constant over the midgap range,

$$s_p \approx \int_{E_F'}^{E_F} c_{ps} D_{IT} dE \approx \underline{c_{ps} D_{IT}} (E_F - E_F') \quad (5.51)$$

evaluated at midgap

The Eq. (5.51) result is useful for estimating s_p . It also has the same general form as the single-level surface recombination velocity ($s_p = c_{ps} N_{Ts}$), thereby adding credibility to our use of the s_p symbol in the present context.

Before concluding, a comment is in order concerning the earlier $R_s = s_p \Delta p_s$ result. As is clearly evident from the analysis, this result is only valid under rather restrictive conditions. Nevertheless, in analyzing solar cells, photodetectors, and other photodevices, it is all but universally assumed that one can write $R_s = s \Delta p_s$ (or $R_s = s \Delta n_s$ for p -type material) **UNDER ARBITRARY CONDITIONS**, with the "surface recombination velocity" s being treated as a system constant. Admittedly, one can introduce a "generalized" surface recombination velocity, $s = R_s$ [Eq. (5.42)]/ Δp_s for n -type material and $s = R_s$ [Eq. (5.42)]/ Δn_s for p -type material. However, the s thus defined is not necessarily a system constant. The generalized s will vary with the level of injection and the amount of band bending. Moreover, under certain conditions s will even be a function of the perturbed carrier concentrations. Nonetheless, the somewhat questionable practice of treating s as a system constant (often a matter of expediency) persists, and it must be acknowledged.

Depleted Surface

If non-equilibrium conditions exist such that both $n_s \rightarrow 0$ and $p_s \rightarrow 0$ at the surface of a semiconductor, R_s by inspection reduces to

$$R_s = \int_{E_s}^{E_c} \frac{-n_i^2}{n_{ls}/c_{ps} + p_{ls}/c_{ns}} D_{IT}(E) dE \quad (5.52a)$$

$$= - \left[\int_{E_s}^{E_c} \frac{c_{ns} c_{ps} D_{IT} dE}{c_{ns} e^{(E-E_i)/kT} + c_{ps} e^{(E_i-E)/kT}} \right] n_i \quad (5.52b)$$

The integral in Eq. (5.52b) is a system constant with the dimensions of a velocity and is commonly referred to as the *surface generation velocity*, s_g . A negative R_s indicates, of course, that carriers are being generated at the surface. Taking the trap parameters in the Eq. (5.52b) integral to be reasonably well-behaved functions of energy, we find the s_g integrand to be a highly peaked function of E , maximizing at $E \approx E_i$ if

$c_{ns} = c_{ps}$ (see Fig. 5.10(b)). Given the highly peaked nature of the integrand, little error is introduced by (a) assuming the trap parameters are constant at their midgap values [$D_{IT}(E) = D_{IT}(E_i)$, etc.] and (b) setting the lower and upper integration limits to $-\infty$ and $+\infty$, respectively. A closed-form evaluation of the integral then becomes possible, giving

$$G_s \equiv -R_s = s_g n_i \quad (5.53)$$

$$s_g = \frac{\pi}{2} \sqrt{c_{ns} c_{ps}} k T D_{IT} \quad (5.54)$$

(Depleted surface)

where the trap parameters in the classic Eq. (5.54) result are to be evaluated at midgap.

5.4 SUPPLEMENTAL R-G INFORMATION

Collected in this section is a collage of practical R-G center information that is intended to enhance and supplement the preceding theoretical description of the R-G center process.

Multistep Nature of Carrier Capture

During our survey of R-G processes the recombination of carriers at R-G centers was noted to be typically non-radiative, implying that the energy lost in the recombination process gives rise to lattice vibrations or phonons. As noted subsequently, however, a single phonon can only carry away a small amount of energy. One is therefore faced with somewhat of a logical dilemma. If a large number of phonons must simultaneously collide with a carrier for capture to occur, the R-G center recombination process would be extremely unlikely! This dilemma is resolved by a more detailed view of carrier capture at R-G centers that recognizes the multistep nature of the process^[6-8]. In the cascade model^[7] an electron (or hole) is viewed to be first weakly bound in an excited-state orbit about the R-G center site. As the carrier moves about the R-G center it loses energy in small increments via collisions with the semiconductor lattice. With the sequential loss in energy to phonons, the carrier spirals in toward the R-G center and is ultimately "captured" or bound tightly to the center. In the multiphonon model^[8] the carrier is viewed as transferring most of its energy in an initial step that causes a violent lattice vibration in the vicinity of the R-G center. The vibration subsequently damps down to the amplitude of thermal vibrations after a small number of vibrational periods. During the damping the localized energy and momentum are carried away from the R-G center by lattice phonons. A cascade-like model appears to describe the initial portion of capture at ionized R-G centers, while a multiphonon-like model is considered to be applicable for the final portion of capture at ionized centers and for the entire capture process at neutral centers.

Manipulation of N_T

The carrier lifetimes (τ_n and τ_p) within a given material determine the response time of the R-G center interaction. The lifetimes in turn are inversely proportional to N_T , the

concentration of the dominant R-G center inside the material. Generally speaking, the dominant R-G center will be an unintentional impurity (or impurity-decorated defect) incorporated into the material during crystal growth or device processing. The resulting lifetimes are often variable and quite unpredictable. Thus procedures have evolved whereby N_T can be manipulated to optimize the characteristics of devices intended for τ -sensitive applications. The lifetimes can be controllably decreased by simply adding known amounts of an efficient R-G center to the semiconductor—e.g., by diffusing Au into Si. To increase the lifetimes, one or more "gettering" steps are included in the device fabrication procedure. "Gettering," or removal of R-G centers from the portion of the semiconductor containing active device junctions, can be accomplished, for example, by diffusing phosphorus into the back side of a Si wafer. During this high-temperature process, R-G centers move about the semiconductor and become trapped in the back-side layer away from the active front surface. For additional gettering information the reader is referred to the device fabrication literature^[9].

Selected Bulk Parametric Data

As we have indicated, the observed minority carrier lifetimes can vary dramatically with the quality of the starting semiconductor, the nature and number of device processing steps, and whether or not the R-G center concentration has been intentionally manipulated during device fabrication. Although a single universal value cannot be quoted, it is nevertheless worthwhile to indicate the general range of lifetimes to be expected under certain conditions in an extensively researched semiconductor such as silicon. Specifically, observed Si lifetimes can be grouped into three ranges as summarized in Table 5.1. The longest reported lifetimes, some exceeding 1 msec, are usually derived from device structures which are devoid of active *pn*-junctions and whose construction involves a minimum of high-temperature processing steps. On the other hand, *pn*-junction devices are routinely characterized by carrier lifetimes ranging from 1 to 100 μ sec. It should be noted that the τ_n and τ_p entries in Table 5.1, which are characteristic of junction devices, were simultaneously measured with a single test structure by varying the level of injection. (The vast majority of lifetime measurements reported in the literature give only τ_g , τ_n in *p*-type material, or τ_p in *n*-type material.) Finally, sub-microsecond lifetimes are readily achieved by design in Au-diffused Si structures. Gold introduces two levels into the Si band gap (see Fig. 4.14), but often one of the levels dominates the R-G center interaction. In such instances, as the Table 5.1 entry implies, the Au interaction may be acceptably modeled by single-level statistics.

Table 5.1 Observed Carrier Lifetimes in Si (300 K)

Lifetime Range	Result-Reference	Device Structure
10^{-4} – 10^{-2} sec	$\tau_g = 2$ msec [10]	Gettered Metal/SiO ₂ /Si (MOS) capacitors
10^{-6} – 10^{-4} sec	$\tau_n = 23.5$ μ sec [11] $\tau_p = 1.5$ μ sec [11]	<i>pn</i> -junction devices
10^{-8} – 10^{-6} sec	$\tau_n = 0.75$ μ sec [11] $\tau_p = 0.25$ μ sec [11]	Au-diffused <i>pn</i> -junction devices

Table 5.2 Measured Capture Cross Sections in Si (300 K)

R-G Center	Capture Cross Section [†]	Reference
Au	$\sigma_n \approx 1 \times 10^{-16} \text{ cm}^2$	
	$\sigma_p = 1 \times 10^{-13} \text{ cm}^2$... acceptor level [12]
	$\sigma_n = 6.3 \times 10^{-15} \text{ cm}^2$	
	$\sigma_p = 2.4 \times 10^{-15} \text{ cm}^2$... donor level [13]
Pt	$\sigma_n = 3.2 \times 10^{-14} \text{ cm}^2$... acceptor level
	$\sigma_p = 2.7 \times 10^{-12} \text{ cm}^2$	closest to midgap [14]
Zn [‡]	$\sigma_n = 2 \times 10^{-16} \text{ cm}^2$... $\text{Zn}^0 + e^- \rightarrow \text{Zn}^-$
	$\sigma_p = 6 \times 10^{-15} \text{ cm}^2$... $\text{Zn}^- + h^+ \rightarrow \text{Zn}^0$
	$\sigma_n = 5 \times 10^{-18} \text{ cm}^2$... $\text{Zn}^- + e^- \rightarrow \text{Zn}^{--}$ [15]
	$\sigma_p = 7 \times 10^{-14} \text{ cm}^2$... $\text{Zn}^{--} + h^+ \rightarrow \text{Zn}^-$

[†]A $v_{th} = 10^7 \text{ cm/sec}$ was assumed in converting capture coefficients to capture cross sections.

[‡]Observed σ_p capture cross sections were field-dependent. Quoted values are for $E = 10^4 \text{ V/cm}$.

Parametric data of a more fundamental nature, including R-G center concentrations, emission rate coefficients, and capture cross sections, can also be found in the device literature. A sampling of capture cross section results is presented in Table 5.2. Since Au in Si is the foremost example of an efficient R-G center, the Au in Si capture cross sections provide a standard of comparison for gauging the effectiveness of other centers. We should note that, with increasingly stringent material requirements in device manufacture, there arose a growing need for more detailed information about R-G centers. This in turn led to the development of routine measurement techniques, notably DLTS^[16] for determining fundamental R-G center parameters. Commercial DLTS (Deep Level Transient Spectroscopy) systems can be used to determine energy levels, emission coefficients, trap concentrations, and capture cross sections. The technique is primarily limited to the detection of purposely introduced centers in Si, but can readily detect the higher levels of unintentional R-G centers present in other materials.

Doping Dependence

The minority carrier lifetimes as defined by Eqs. (5.23) are explicitly independent of the acceptor and donor concentrations. However, Auger recombination, with $\tau_{\text{Auger}} \propto 1/(\text{carrier concentration})^2$, becomes the dominant recombination mechanism at high doping levels. As a result, the carrier lifetimes in Si exhibit a decrease with increased doping roughly as pictured in Fig. 5.11. In addition, for *n*-type Si, there is a theoretical prediction and supporting experimental evidence of a decrease in the minority carrier lifetime at doping levels below the onset of significant Auger recombination. This is believed to result from an increase in the defect density, and hence an added N_I , in direct proportion to the N_A doping concentration.^[17a]

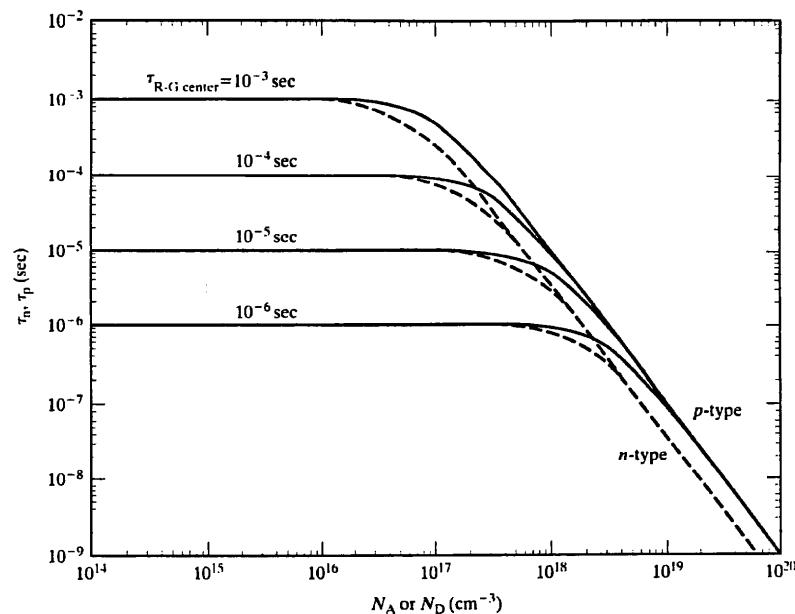


Figure 5.11 Effect of Auger recombination at high doping levels on the carrier lifetimes in Si at 300 K. Lifetimes computed employing the Auger recombination coefficients found in Reference [17b].

Selected Surface Parametric Data (Si/SiO₂)

The surface parametric data to be examined specifically applies to the Si/thermally grown SiO₂ interface. Because of its technological importance, the Si/SiO₂ interface has been the subject of an intensive experimental investigation. The available information on the interface state parameters characterizing the oxide-covered Si surface is quite extensive, far exceeding (and in greater detail than) that on all other semiconductor surfaces and interfaces combined. This is not to say that the Si/SiO₂ interface is thoroughly characterized. One complication stems from the fact that the surface state parameters are strongly process-dependent. This dependence is nicely illustrated by Fig. 5.12, reproduced from Ref. [18]. Whereas $D_{IT}(E)$ is at a minimum and approximately constant over the midgap region in an "optimally" processed structure, small variations in processing can give rise to interface-state densities which are larger by several orders of magnitude and which exhibit a decidedly different energy dependence. Surface state parameters also vary systematically with the Si surface orientation and are affected by ionizing radiation.

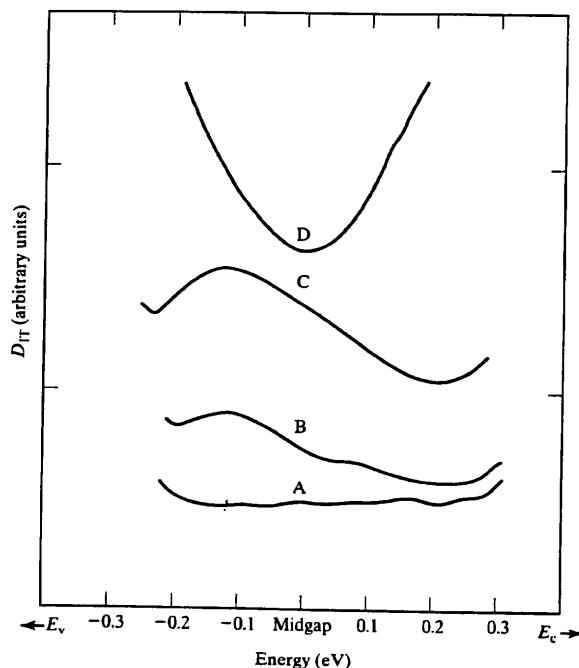


Figure 5.12 Observed variation of the Si/SiO₂ interfacial trap density distribution with processing steps immediately following thermal oxidation of the Si surface. The curves shown illustrate general trends. (A) Sample with near-optimum H₂ anneal. (B) Sample with nonoptimum H₂ anneal. (C) Unannealed sample pulled in dry O₂. (D) Unannealed sample pulled in N₂. (From Razouk and Deal^[18]. Reprinted by permission of the publisher, The Electrochemical Society, Inc.)

The parametric data presented in Fig. 5.13 and Fig. 5.14 are representative of an “optimally” processed, (100)-oriented, thermally oxidized Si surface. The plots are a superposition of the results reported by a number of investigators^[4, 5, 19–21, 23]. Given the independent fabrication of test structures, the difficulty of the measurements, and the use of different measurement techniques, the agreement between results is generally quite good.

In examining Fig. 5.13, please note that D_{IT} is approximately constant over the midgap region, with midgap values $\sim 10^{10}$ states/cm²·eV.[†] Most of the data sets

[†]Although $10^{10}/\text{cm}^2\cdot\text{eV}$ is a fairly representative value, midgap surface state densities as low as $2 \times 10^9/\text{cm}^2\cdot\text{eV}$ have been reported^[32].

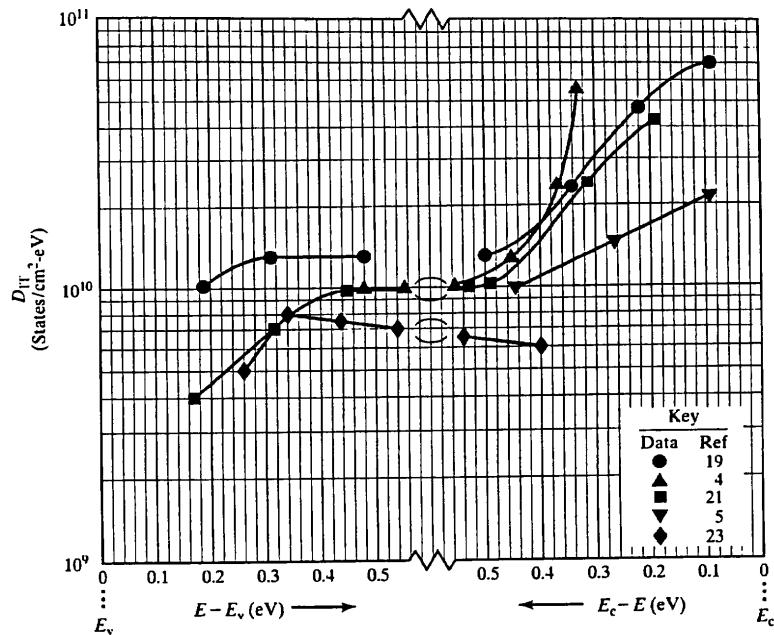


Figure 5.13 Measured Si/SiO₂ interfacial trap densities as a function of band gap energy^[4, 5, 19, 21, 23]. The data are representative of that derived from an optimally (or near-optimally) processed, (100)-oriented, thermally oxidized Si surface.

likewise show an increase in D_{IT} as one approaches the conduction band edge. Interestingly, a similar increase is not noted near the valence band edge. This, however, may be a function of the measurement techniques (DLTS and Charge Pumping) used to acquire the near- E_v Fig. 5.13 data. Other techniques applied to structures with higher surface state densities have consistently shown an upturn in D_{IT} near the valence band edge. Finally, note in Fig. 5.14 that available results for σ_{ns} and σ_{ps} ($c_{ns} = \sigma_{ns}v_{th}$, $c_{ps} = \sigma_{ps}v_{th}$) are primarily confined respectively to the upper and lower halves of the band gap. σ_{ns} is generally observed to be approximately constant or slowly varying near midgap, while falling off sharply as E approaches E_c . σ_{ps} results, on the other hand, have not exhibited a consistent trend. Measurement complications arise from the interplay between energy, temperature, and surface field dependencies. In addition, it is possible there may be more than one type of interface trap.

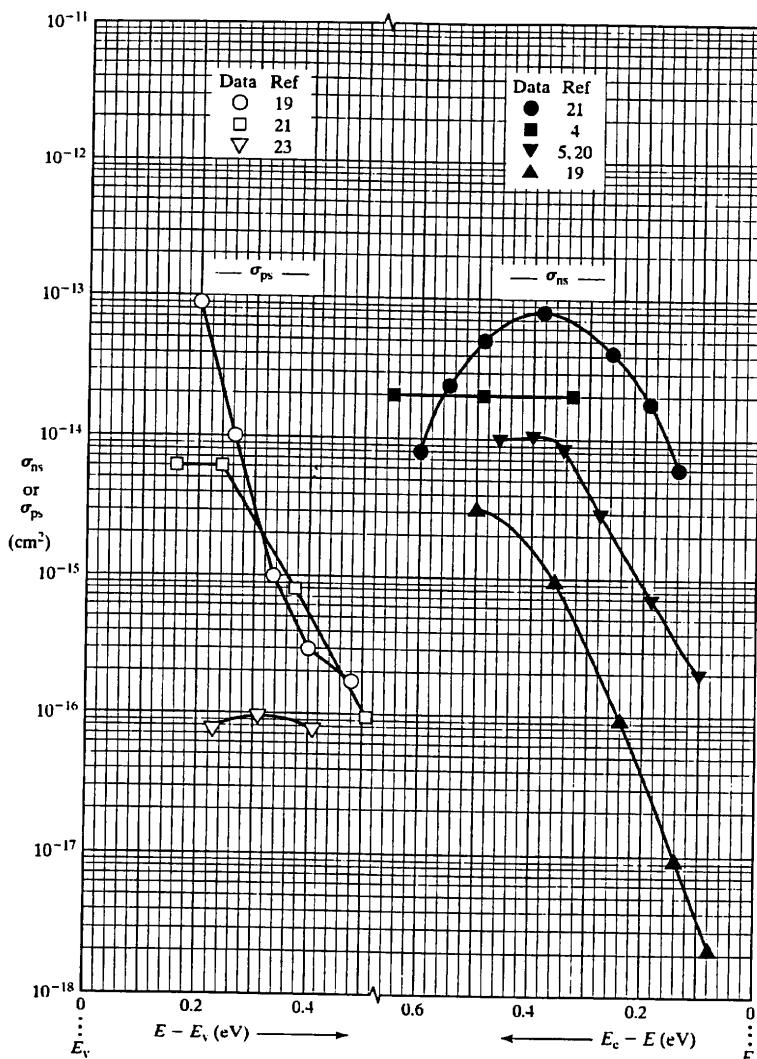


Figure 5.14 Measured electron and hole surface capture cross sections^[4,5,19,21,23] characteristic of an optimally processed, (100)-oriented, thermally oxidized Si surface.

REFERENCES

- [1] W. Shockley and W. T. Read, Jr., "Statistics of the Recombination of Holes and Electrons," *Phys. Rev.*, **87**, 835 (1952).
- [2] R. N. Hall, "Germanium Rectifier Characteristics," *Phys. Rev.*, **83**, 228 (1951).
- [3] R. N. Hall, "Electron-Hole Recombination in Germanium," *Phys. Rev.*, **87**, 387 (1952).
- [4] J. A. Cooper, Jr. and R. J. Schwartz, "Electrical Characteristics of the SiO₂-Si Interface Near Midgap and in Weak Inversion," *Solid-State Electronics*, **17**, 641 (1974).
- [5] H. Deuling, E. Klausmann, and A. Goetzberger, "Interface States in Si-SiO₂ Interfaces," *Solid-State Electronics*, **15**, 559 (1972).
- [6] H. J. Queisser, "Recombination at Deep Traps," *Solid-State Electronics*, **21**, 1495 (1980).
- [7] M. Lax, "Cascade Capture of Electrons in Solids," *Phys. Rev.*, **119**, 1502 (1960).
- [8] C. H. Henry and D. V. Lang, "Nonradiative Capture and Recombination by Multiphonon Emission in GaAs and GaP," *Phys. Rev. B*, **15**, 989 (Jan., 1977).
- [9] For an excellent discussion of gettering processes see J. D. Plummer, M. D. Deal, and P. B. Griffin, *Silicon VLSI Technology, Fundamentals, Practice, and Modeling*, Prentice Hall, Upper Saddle River NJ, 2000.
- [10] A. Rohatgi and P. Rai-Choudhury, "Process-Induced Effects on Carrier Lifetime and Defects in Float Zone Silicon," *J. Electrochem. Soc.*, **127**, 1136 (May, 1980).
- [11] W. Zimmerman, "Experimental Verification of the Shockley-Read-Hall Recombination Theory in Silicon," *Electronics Letts.*, **9**, 378 (Aug., 1973).
- [12] F. Richou, G. Pelous, and D. Lecrosnier, "Thermal Generation of Carriers in Gold-Doped Silicon," *J. Appl. Phys.*, **51**, 6252, (Dec., 1980).
- [13] J. M. Fairfield and B. V. Gokhale, "Gold as a Recombination Centre in Silicon," *Solid-State Electronics*, **8**, 685 (1965).
- [14] K. P. Lisiak and A. G. Milnes, "Platinum as a Lifetime-Control Deep Impurity in Silicon," *J. Appl. Phys.*, **46**, 5229 (Dec., 1975).
- [15] J. M. Herman III and C. T. Sah, "Thermal Capture of Electrons and Holes at Zinc Centers in Silicon," *Solid-State Electronics*, **16**, 1133 (1973).
- [16] D. V. Lang, "Deep-Level Transient Spectroscopy: A New Method to Characterize Traps in Semiconductors," *J. Appl. Phys.*, **45**, 3023 (July, 1974).
- [17a] R. J. Van Overstraeten and R. P. Mertens, "Heavy Doping Effects in Silicon," *Solid-State Electronics*, **30**, 1077 (1987).
- [17b] J. Dziewior and W. Schmid, "Auger Coefficients for Highly Doped and Highly Excited Silicon," *Appl. Phys. Lett.*, **31**, 346 (Sept., 1977).
- [18] R. R. Razouk and B. E. Deal, "Dependence of Interface State Density on Silicon Thermal Oxidation Process Variables," *J. Electrochem. Soc.*, **126**, 1573 (Sept., 1979).
- [19] T. J. Tredwell and C. R. Viswanathan, "Interface-State Parameter Determination by Deep-Level Transient Spectroscopy," *Appl. Phys. Lett.*, **36**, 462 (March 15, 1980).
- [20] T. Katsume, K. Kakimoto, and T. Ikoma, "Temperature and Energy Dependence of Capture Cross Sections at Surface States in Si Metal-Oxide-Semiconductor Diodes Measured by Deep Level Transient Spectroscopy," *J. Appl. Phys.*, **52**, 3504 (May, 1981).

- [21] W. D. Eades and R. M. Swanson, "Improvements in the Determination of Interface State Density Using Deep Level Transient Spectroscopy," *J. Appl. Phys.*, **56**, 1744 (Sept., 1984).
- [22] G. Declerck, R. van Overstraeten, and G. Broux, "Measurement of Low Densities of Surface States at the Si-SiO₂-Interface," *Solid-State Electronics*, **16**, 1451 (1973).
- [23] N. S. Saks and M. G. Ancona, "Determination of Interface Trap Capture Cross Sections Using Three-Level Charge Pumping," *IEEE Electron Device Lett.*, **11**, 339 (August, 1990).

SOURCE LISTING

- (1) Review articles on a variety of R-G topics were published in *Solid-State Electronics*, **21**, (1978). Relevant papers include:
- (a) N. F. Mott, "Recombination: A Survey," *Solid-State Electronics*, **21**, 1275 (1978).
 - (b) H. J. Queisser, "Recombination at Deep Traps," *Solid-State Electronics*, **21**, 1495 (1978).
- (2) Perhaps the most authoritative and complete treatment of single- and multi-level R-G center statistics can be found in:
- (a) C. T. Sah, "The Equivalent Circuit Model in Solid-State Electronics—Part I: The Single Energy Level Defect Centers," *Proc. IEEE*, **55**, 654 (May, 1967).
 - (b) C. T. Sah, "The Equivalent Circuit Model in Solid-State Electronics—Part II: The Multiple Energy Level Impurity Centers," *Proc. IEEE*, **55**, 672 (May, 1967).

PROBLEMS

GENERAL NOTE: Problems 5.5, 5.6, 5.7, and 5.9 require a working knowledge of quasi-Fermi levels. Quasi-Fermi levels are reviewed herein in Subsection 6.3.1.

5.1 Answer the following questions as concisely as possible.

- (a) Using the energy band diagram, indicate how one visualizes photogeneration, intrinsic Auger recombination, and recombination via SRH centers.
- (b) Prior to processing, a portion of a semiconductor sample contains $N_D = 10^{14}/\text{cm}^3$ donors and $N_T = 10^{11}/\text{cm}^3$ R-G centers. After processing (say in the fabrication of a device), the same portion of the semiconductor contains $N_D = 10^{16}/\text{cm}^3$ donors and $N_T = 10^{10}/\text{cm}^3$ R-G centers. Did the processing increase or decrease the minority carrier lifetime? Explain.
- (c) Briefly explain the difference between "equilibrium" and "steady-state."
- (d) Make a plot of the net recombination rate (R) versus position inside the depletion region of a *pn*-junction diode maintained under equilibrium conditions.

5.2 A semiconductor contains bulk traps that introduce an R-G level at $E_T' = E_i$. Steady-state conditions prevail.

- (a) If c_n is roughly the same order of magnitude as c_p , confirm that:
 - (i) $n_T \approx N_T \dots$ in an *n*-type semiconductor subject to low level injection.
 - (ii) $p_T \approx N_T \dots$ in a *p*-type semiconductor subject to low level injection.
- (b) The semiconductor is uniformly illuminated such that $\Delta n = \Delta p \gg n_0$ or p_0 .
 - (i) Determine n_T/N_T if $c_n = c_p$.
 - (ii) Determine n_T/N_T if $c_p \gg c_n$.

- (c) The semiconductor is depleted ($n \rightarrow 0, p \rightarrow 0$) and $e_n = e_p$. What fraction of the R-G centers will be filled for the specified situation?

- 5.3 Utilizing the general steady-state R -expression, confirm that band gap centers with E_i' near E_i make the best R-G centers. [Assume, for example, that the semiconductor is *n*-type, low level injection conditions prevail, and $\tau_n = \tau_p = \tau = \text{constant}$ independent of the trap energy. Next consider how R would vary with E_T' under the given conditions and conclude with a rough sketch of R versus E_T' across the band gap.]

- 5.4 (a) Read and summarize the paper by W. Zimmerman, *Electronics Letters*, **9**, 378 (August, 1973).

- (b) Starting with the R -expression established in the text [Eq. (5.24)], derive Zimmerman's equation for τ .

- 5.5 The energy band diagram for a reverse-biased Si *pn*-junction diode under steady-state conditions is pictured in Fig. P5.5.

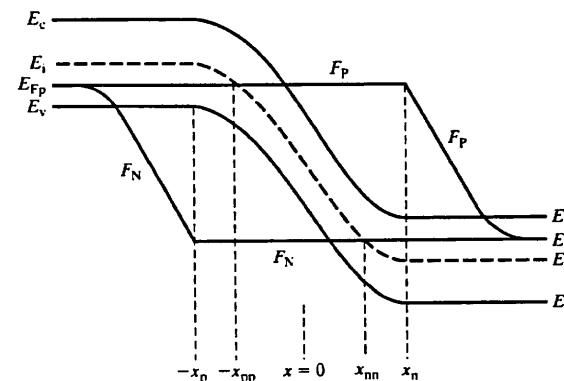


Figure P5.5

- (a) With the aid of the diagram and assuming single-level R-G center statistics, $\tau_n = \tau_p = \tau$, and $E_T' = E_i$, simplify the general steady-state net recombination rate expression to obtain the simplest possible relationship for R at (i) $x = 0$, (ii) $x = -x_{pp}$, (iii) $x = x_{nn}$, (iv) $x = -x_p$, and (v) $x = x_n$.

- (b) Sketch R versus x for x -values lying within the electrostatic depletion region ($-x_p \leq x \leq x_n$).

- (c) What was the purpose or point of this problem?

- 5.6 The energy band diagram for a forward-biased Si *pn*-junction diode maintained under steady-state conditions at room temperature ($T = 300$ K) is pictured in Fig. P5.6. Note that $E_{Fn} - E_i = E_i - E_{fp} = E_G/4$.

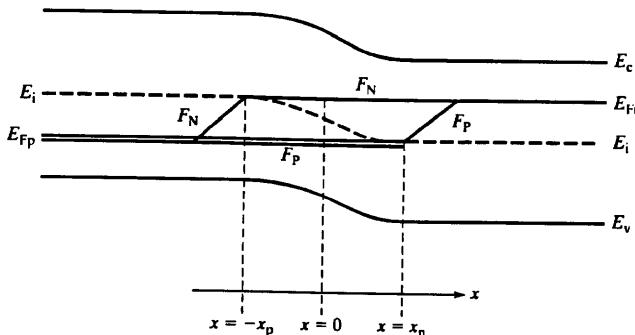


Figure P5.6

- (a) For the particular situation pictured in Fig. P5.6, and assuming single-level R-G center statistics, $\tau_n = \tau_p = \tau$, and $E_{T'} = E_i$, show that the steady-state net recombination rate inside the electrostatic depletion region ($-x_p \leq x \leq x_n$) can be simplified to

$$R \approx \frac{n_i}{\tau} \left[\frac{e^{E_i/4kT}}{e^{(F_N-E_i)/kT} + e^{(F_p-E_i)/kT}} \right]$$

- (b) Plot $R/(n_i/\tau)$ versus x for $-x_p \leq x \leq x_n$. Assume the E_i variation between $x = -x_p$ and $x = x_n$ is approximately linear in constructing the plot.
(c) What is the purpose or point of this problem?

- 5.7 The energy band diagram for the semiconductor component of an MOS device under steady-state conditions is pictured in Fig. P5.7.

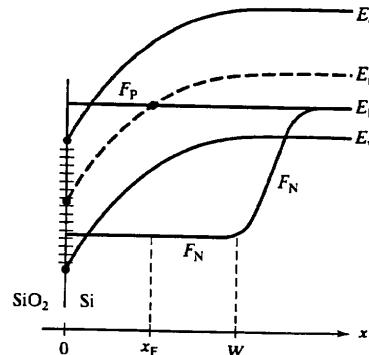


Figure P5.7

- (a) Assuming single-level R-G centers, $\tau_n = \tau_p = \tau$, and $E_{T'} = E_i$, make a sketch of the net recombination rate (R) versus x for x -values lying within the electrostatic depletion region ($0 \leq x \leq W$). Include the specific values of R at $x = 0$, $x = x_F$ and $x = W$ on your sketch.

- (b) Is the net surface recombination rate (R_s) for the situation pictured in Fig. P5.7 expected to be less than, approximately equal to, or greater than the R_s at a totally depleted ($n_s \rightarrow 0$, $p_s \rightarrow 0$) surface? Briefly explain how you arrived at your answer.

- 5.8 (a) For a totally depleted semiconductor surface ($n_s \rightarrow 0$, $p_s \rightarrow 0$) with c_{ns} , c_{ps} , and $D_{IT}(E)$ all approximately constant over the midgap region, confirm the text assertion that

$$G_s = -R_s = s_g n_i$$

where

$$s_g = (\pi/2) \sqrt{c_{ns} c_{ps}} k T D_{IT}$$

s_g is the surface generation velocity; all parameters in the s_g expression are evaluated at midgap.

- (b) Based on the data presented in Fig. 5.13 and Fig. 5.14, estimate the expected value of s_g for an optimally processed, (100)-oriented, thermally oxidized Si surface.

- 5.9 The surface of an n -bulk solar cell is subjected to intense illumination giving rise to high level injection ($\Delta n_s = \Delta p_s \gg n_{s0}$). The surface region of the device is characterized by the energy band diagram shown in Fig. P5.9.

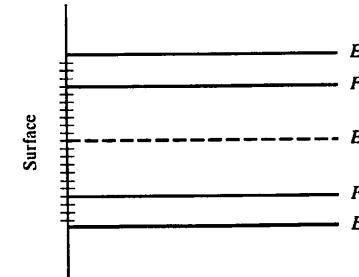


Figure P5.9

- (a) What fraction of the surface centers at $E = F_N$, $E = E_i$, and $E = F_p$ will be filled for the specified steady-state situation? (Let D_{Filled} be the number of filled surface centers per $\text{eV}\cdot\text{cm}^2$ at E . The fraction of filled surface centers is then D_{Filled}/D_{IT} .) Sketch D_{Filled}/D_{IT} versus E for $E_v \leq E \leq E_c$. Assume $c_{ns} = c_{ps}$ for all E in this part of the problem.

- (b) Defining $R_s = s^* \Delta p_s$, obtain an approximate closed-form expression for s^* involving only $\Delta E = F_N - F_p$ and the parameters c_{ns} , c_{ps} , and D_{IT} evaluated at midgap. Invoke and record all reasonable assumptions and simplifications needed to complete this problem.

5.10 In this problem we examine the statistics of thermal *band-to-band* recombination-generation.

- Using the energy band diagram, indicate the possible electronic transitions giving rise to (i) band-to-band recombination and (ii) band-to-band generation.
- Taking the semiconductor to be nondegenerate and paralleling the text derivation of Eqs. (5.14), establish the general relationship

$$r_b = c_b(n_p - n_i^2)$$

where r_b is the net band-to-band recombination rate and c_b is the band-to-band recombination coefficient (units of cm^3/sec). (Note from the nature of the band-to-band processes that the net electron and hole recombination rates are always equal; $r_{\text{electron}} = r_{\text{hole}} = r_b$. It also follows, of course, that there is no special relationship for steady-state conditions.)

- Show that the general r_b relationship reduces to

$$r_b = \frac{\Delta n}{\tau_b}$$

$$\tau_b \equiv \frac{1}{c_b(n_0 + p_0)}$$

under low level injection conditions where $\Delta n = \Delta p$.

- Given $c_b \approx 5 \times 10^{-15} \text{ cm}^3/\text{sec}$ in Si at 300 K, and assuming $\tau \leq 1 \text{ msec}$ due to recombination at R-G centers, does one have to worry about carrier recombination via the band-to-band process in nondegenerately doped Si at room temperature? Explain.