$E_{\text{bound}} = E_C$, takes the form

$$E = E_C + \frac{\hbar^2}{2m_{\text{eff}}}|\mathbf{k} - \mathbf{k}_C|^2 \qquad \text{(VIII.4.02)}$$

However, the warning of pages 180 and 200 must be repeated, i.e., that (VII.2.23), (VII.4.04), and (VIII.4.02) are simplifying assumptions which, to take an example, are not fulfilled in the case of the cubic face-centered lattice at one band edge (the upper for negative exchange integral, the lower for positive exchange integral). Even if at the band edge $\mathbf{k} = \mathbf{k}_C$ Eq. (VIII.4.02) is fulfilled, it is not strictly correct to use the same m_{eff} higher up in the band.

If we compare the equation derived from (VIII.4.02)

$$|\mathbf{k} - \mathbf{k}_C| = \frac{\sqrt{2m_{\text{eff}}}}{\hbar}(E - E_C)^{1/2}$$

with (VII.1.02) and if we substitute $E_{\text{pot}} \to E_C$, $m \to m_{\text{eff}}$, and $\mathbf{k} \to \mathbf{k} - \mathbf{k}_C$, we can repeat exactly the computation of the density of states $D(E)$ carried out for the model of the potential well, and in analogy to (VIII.1.05) we obtain

$$2D(E)\,dE = V \cdot N_C \cdot \frac{2}{\sqrt{\pi}}\left(\frac{E - E_C}{\text{k}T}\right)^{1/2} d\left(\frac{E}{\text{k}T}\right) \qquad \text{(VIII.4.03)}$$

where the effective density of states N in the potential well is replaced by the effective density of states N_C in the conduction band.[1] Correspondingly, the definition of N_C is derived from that of N [Eq. (VIII.1.04)] by the substitution $m \to m_{\text{eff}}$:

$$N_C = 2 \cdot \left(\frac{2\pi m_{\text{eff}}\text{k}T}{h^2}\right)^{3/2} = 2.5 \cdot 10^{19}\left(\frac{m_{\text{eff}}}{m}\right)^{3/2} \cdot \left(\frac{T}{300°\text{K}}\right)^{3/2} \text{cm}^{-3} \qquad \text{(VIII.4.04)}$$

Using (VIII.4.03) in (VIII.4.01) and introducing the integration variable $\eta = (E - E_C)/\text{k}T$ as well as the concentration $n = N/V$, one obtains as the determining equation for E_F:

$$\frac{2}{\sqrt{\pi}}\int_{\eta=0}^{\eta=\infty} \frac{1}{e^{\eta - \frac{E_F - E_C}{\text{k}T}} + 1}\sqrt{\eta}\,d\eta = \frac{n}{N_C} \qquad \text{(VIII.4.05)}$$

Finally, by comparing (VIII.4.05) with the defining equation (A.II.1)

---

[1] It is apparent from (VIII.4.03) that the hypothesis (VIII.4.02) leads to zero state density $D(E)$ at the band edge $E = E_C$. Only in the linear case this does not apply, see Fig. VII.2.7.

of the function ζ discussed in Appendix II on page 384, one gets

$$E_F = E_C + \zeta\left(\frac{n}{N_C}\right) \tag{VIII.4.06}$$

Hence, on the basis of the assumption (VIII.4.02) *the conduction electrons behave like a free electron gas with a potential energy which equals the total energy E_C of the electrons at the lower edge of the conduction band. For the mass of the electrons, one has to use the effective mass*

$$m_{\text{eff}} = \frac{\hbar^2}{E''(k_C)} \tag{VII.6.23}$$

The foregoing treatment in which the fully occupied band below the conduction band is ignored and where the concentration n of the conduction electrons is considered temperature independent is permissible[1] only as long as

$$n \gg N_C$$

The number of conduction electrons in metals is about one per atom; hence $n \approx 10^{22}$ cm^{-3}. For N_C to reach this order of magnitude, one has to have

$$N_C = 2.5 \cdot 10^{19}\left(\frac{T}{300°\mathrm{K}}\right)^{3/2} \mathrm{cm}^{-3} \approx 10^{22}\ \mathrm{cm}^{-3}$$

$$\text{or} \quad T \approx 300°\mathrm{K} \cdot \left(\frac{10^{22}}{2.5 \cdot 10^{19}}\right)^{2/3} \approx 300°\mathrm{K} \cdot (4 \cdot 10^2)^{2/3} = 300°\mathrm{K} \cdot 55 = 16{,}500°\mathrm{K}$$

At such temperatures the metal is molten and therefore the case $n \leq N_C$ is of no interest. This means: *The electron gas in metals is always strongly degenerate.*

In many problems, therefore, the limiting value given by Eq. (A.II.3) can be used for ζ in (VIII.4.06)

$$\zeta = \mathrm{k}T \cdot \left(\frac{3}{4}\right)^{2/3} \pi^{1/3}\left(\frac{n}{N_C}\right)^{2/3} \tag{A.II.3}$$

From this, together with (VIII.4.04), one obtains for ζ the tempera-

---

[1] In this case, ζ is positive and the Fermi level $E_F = E_C + \zeta$ will lie within the conduction band in the temperature range in which we are interested. The band below the conduction band lies so far below the Fermi level E_F that it will remain fully occupied with great accuracy even when the slope of the Fermi distribution flattens with increasing temperature and the electrons move from states below E_F to states above E_F. The band below the conduction band lies, in the case $n \gg N_C$, too far below E_F to be noticeably affected by this process.

ture-independent[1] value

$$\zeta = \frac{1}{2}\left(\frac{3}{8\pi}\right)^{2/3} \frac{h^2}{m_{\text{eff}}} n^{2/3} \tag{VIII.4.07}$$

Using the equation $$E_F - E_C = \zeta = \frac{m_{\text{eff}}}{2} v_{\text{th}}^2 \tag{VIII.4.08}$$

one obtains for the velocity v_{th} of those electrons which are, energetically, at the Fermi level E_F

$$v_{\text{th}} = \left(\frac{3}{8\pi}\right)^{1/3} \frac{h}{m_{\text{eff}}} n^{1/3} = 7.71 \cdot 10^7 \text{ cm-sec}^{-1} \cdot \left(\frac{m}{m_{\text{eff}}}\right) \cdot \left(\frac{n}{10^{22} \text{ cm}^{-3}}\right)^{1/3} \tag{VIII.4.09}$$

in other words, *the order of magnitude of v_{th} is 10^8 cm-sec^{-1}*. Here the thermal velocity v_{th} has been computed as the velocity of the electrons at the Fermi level. This seems reasonable if one bears in mind that in a Fermi gas an increment of energy—be it from an external field or in the form of heat—changes only the energetic distribution of the electrons near the Fermi level while the electrons in the lower energy levels are not affected.

There are two other methods for estimating the velocity of the electrons at the Fermi level ($\approx$ center of the band), which predominantly determine the conduction process. First, by limitation to a one-dimensional case, one can derive directly from Eq. (VII.5.10) the equation

$$v = \frac{1}{\hbar} \cdot \frac{dE}{dk} > \frac{1}{\hbar} \frac{\Delta E}{\Delta k}$$

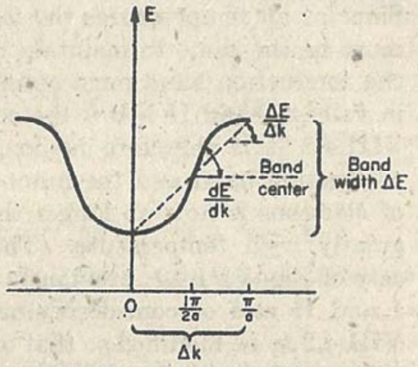


Fig. VIII.4.3. Computation of the electron velocity in the band center.

in which one can substitute for ΔE a band width of the order[2] of 10 ev = $1.6 \cdot 10^{-11}$ cm^2 g-sec^{-2} and for Δk a quantity of the order of $\pi/a \approx 1 \cdot 10^{+8}$ cm^{-1}. Thus one obtains the value

$$v \approx \frac{1}{1 \cdot 10^{-27} \text{ cm}^2 \text{ g-sec}^{-1}} \cdot \frac{1.6 \cdot 10^{-11} \text{ cm}^2 \text{ g-sec}^{-2}}{1 \cdot 10^{+8} \text{ cm}^{-1}} = 1.6 \cdot 10^8 \text{ cm-sec}^{-1}$$

which, according to Fig. VIII.4.3, is somewhat too low. The alterna-

---

[1] Owing to this temperature independence, the approximation (A.II.3) is *not* sufficient where the specific heat of the electron gas is concerned.

[2] See, e.g., F. Herman, *Phys. Rev.*, **88**: 1210 (1952), where a width of 22 ev is computed for the valence band of diamond.

tive method consists of using the approximation (VIII.4.02) in the equation $v = \frac{1}{\hbar}\frac{dE}{dk}$ with the result

$$v = \frac{\hbar}{m_{\text{eff}}} \cdot k$$

If we now use for m_{eff} the electron mass $9 \cdot 10^{-28}$ g and for k the value $\frac{1}{2}\pi/a \approx \frac{1}{2} \cdot 10^{8}$ cm^{-1} in the center of the band, we obtain

$$v = \frac{1 \cdot 10^{-27} \text{ cm}^2 \text{ g-sec}^{-1}}{9 \cdot 10^{-28} \text{ g}} \cdot \frac{1}{2} \cdot 10^{+8} \text{ cm}^{-1} \approx \frac{1}{2} 10^{+8} \text{ cm-sec}^{-1}$$

In other words, all three estimations result in the same order of magnitude.

c. The Band Model of the Insulator and of the Intrinsic Semiconductor[1] at Temperatures $T > 0$

It can be seen from Fig. VIII.4.2 that at temperatures > 0°K valence band II below conduction band III is no longer fully occupied.[2] Since at all temperatures the *total* number of electrons in the crystal must be the same to maintain neutrality, the number of electrons in the conduction band must equal the number of vacancies or "holes" in valence band II below the conduction band.[3] In the case of Fig. VIII.4.2 it is therefore no longer permissible to consider only the conduction band and the number of electrons therein. The number of electrons is now no longer the same at all temperatures but varies greatly with temperature. The simplified procedure which in the case of Fig. VIII.4.1, previously described, consisted of ignoring bands I and II and of considering only band III has in the case of Fig. VIII.4.2 to be modified so that only band I is ignored. The discussion must cover bands II and III and all the electrons which have to be accommodated in these two bands and whose number may again with fair accuracy be regarded as temperature independent.

The significance of the following discussion would be greatly reduced if explicit expressions were used for the density of states in the whole valence band II below the conduction band. Fortunately, an expedient relieves us from this necessity. This expedient consists of con-

---

[1] See pages 16 to 20 concerning this term.

[2] In principle, this was true in the case of Fig. VIII.4.1 just considered. However, there the number of vacancies or "holes" in valence band II is practically negligible as compared with the number of electrons in conduction band III.

[3] Strictly speaking, the holes in the electron occupancy of band I should also be considered. However, their number is again negligibly small.

sidering, instead of the distribution of *electrons* over the states of valence band II, the distribution of the *vacancies* or "holes" or "defect electrons."[1] If the probability of an electron level E being occupied by an electron is, according to (VII.10.01),

$$f(E) = \frac{1}{e^{\frac{E-E_F}{kT}} + 1} \qquad \text{(VII.10.01)}$$

then the probability that this energy level E is *not* occupied is

$$1 - f(E) = \frac{e^{\frac{E-E_F}{kT}} + 1 - 1}{e^{\frac{E-E_F}{kT}} + 1} = \frac{1}{1 + e^{\frac{E_F-E}{kT}}} \qquad \text{(VIII.4.10)}$$

On the other hand, for the computation of the density of states $D(E)$, one makes the *assumption*, in accordance with Eq. (VIII.4.02), that

$$E = E_V - \frac{\hbar^2}{2m_p} |\mathbf{k} - \mathbf{k}_V|^2 \qquad \text{(VIII.4.11)}$$

(m_p = absolute value of the effective mass of the electrons at the upper edge of the valence band[2]) from which it follows that

$$|\mathbf{k} - \mathbf{k}_V| = \frac{\sqrt{2m_p}}{\hbar} (E_V - E)^{1/2}$$

Comparison with (VIII.1.02) shows that we have to carry out the substitutions

$$E - E_{\text{pot}} \to E_V - E \qquad m \to m_p \qquad \text{and} \qquad \mathbf{k} \to \mathbf{k} - \mathbf{k}_V$$

and that hence we have now for the number of electron places in the energy interval $[E, E + dE]$ the expression[3]

$$2D(E)\, dE = V \cdot N_V \cdot \frac{2}{\sqrt{\pi}} \left(\frac{E_V - E}{kT}\right)^{1/2} d\left(\frac{E}{kT}\right) \qquad \text{(VIII.4.12)}$$

Here N_V is the effective density of states in the valence band

$$N_V = 2\left(\frac{2\pi m_p kT}{h^2}\right)^{3/2} = 2.5 \cdot 10^{19} \left(\frac{m_p}{m}\right)^{3/2} \left(\frac{T}{300°\text{K}}\right)^{3/2} \text{cm}^{-3} \qquad \text{(VIII.4.13)}$$

---

[1] This procedure is not limited to intrinsic semiconductors. The equations derived below are, therefore, of general validity for semiconductors—with the exception of (VIII.4.24), (VIII.4.25), and (VIII.4.26) where the relation $n = p$ (VIII.4.24), which defines the intrinsic semiconductor, enters decisively.

[2] The effective mass of the electrons at the upper edge of the valence band is negative, i.e., $-m_p$.

[3] Here the density of states $D(E)$ is of course zero at the upper band edge $E = E_V$. See also footnote 1, p. 298.

If we introduce the number P or the concentration $p = P/V$ of the vacancies or holes in the almost complete electron occupancy at the upper edge of the valence band, we obtain[1] for P

$$P = V \cdot p = 2 \int\limits_{E=-\infty}^{E=E_V} D(E) \cdot (1 - f(E))\, dE$$

$$= V N_V \frac{2}{\sqrt{\pi}} \int\limits_{E=-\infty}^{E=E_V} \left(\frac{E_V - E}{\mathrm{k}T}\right)^{1/2} \frac{1}{e^{\frac{E_F - E}{\mathrm{k}T}} + 1} \cdot d\left(\frac{E}{kT}\right)$$

or, with

$$\frac{E_V - E}{\mathrm{k}T} = \eta$$

as integration variable

$$\frac{2}{\sqrt{\pi}} \int\limits_{\eta=0}^{\eta=\infty} \frac{1}{e^{\eta - \frac{E_V - E_F}{\mathrm{k}T}} + 1} \cdot \sqrt{\eta} \cdot d\eta = \frac{p}{N_V} \qquad \text{(VIII.4.14)}$$

By comparison with the defining equation (A.II.1) of the function ζ, one can again derive

$$E_F = E_V - \zeta\left(\frac{p}{N_V}\right) \qquad \text{(VIII.4.15)}$$

With regard to the electron concentration n in the conduction band, we obtain in the present case of the insulator, as with a metal, in accordance with (VIII.4.06),

$$E_F = E_C + \zeta\left(\frac{n}{N_C}\right) \qquad \text{(VIII.4.16)}$$

where in the definition of the effective density of states N_C in the conduction band

$$N_C = 2\left(\frac{2\pi m_n \mathrm{k}T}{h^2}\right)^{3/2} = 2.5 \cdot 10^{19} \left(\frac{m_n}{m}\right)^{3/2} \left(\frac{T}{300^\circ\mathrm{K}}\right)^{3/2} \mathrm{cm}^{-3} \qquad \text{(VIII.4.17)}$$

we now call the effective mass of the conduction electrons m_n in order to achieve complete analogy with (VIII.4.13).

In Eqs. (VIII.4.15) and (VIII.4.16) E_F can be eliminated and one can see that the equation

$$\zeta\left(\frac{n}{N_C}\right) + \zeta\left(\frac{p}{N_V}\right) = E_V - E_C (<0) \qquad \text{(VIII.4.18)}$$

---

[1] The lower integration limit $E = -\infty$ is unimportant because according to (VIII.4.10) $1 - f(E)$ approaches zero very quickly as $E \to -\infty$ and no further important contributions to the integral arise.

must always hold, independent of the position E_F of the Fermi level. This equation will now be shown to represent the generalization of the law of mass action (see Eq. I.3.03) $np = n_i^2$ between electrons and holes, valid even in the case of degeneracy.

We can assume that at the edges of the conduction and of the valence band the effective masses m_n and m_p of the electrons will not differ in order of magnitude from the mass m of the free electron. Hence, as long as the temperature is not excessively low, N_C and N_V represent very high concentrations and we have[1]

$$p \ll N_V \quad \text{and} \quad n \ll N_C \tag{VIII.4.19}$$

Now the logarithmic limiting expression (A.II.2) for ζ can be used in (VIII.4.15), (VIII.4.16), and (VIII.4.18) and we obtain

$$E_F = E_V - \mathrm{k}T \ln \frac{p}{N_V} \tag{VIII.4.20}$$

$$E_F = E_C + \mathrm{k}T \ln \frac{n}{N_C} \tag{VIII.4.21}$$

and

$$\mathrm{k}T \ln \frac{n}{N_C} + \mathrm{k}T \ln \frac{p}{N_V} = -(E_C - E_V)$$

or

$$n \cdot p = N_C \cdot N_V \cdot \mathrm{e}^{-\frac{E_C - E_V}{\mathrm{k}T}} \tag{VIII.4.22}$$

This is—as just mentioned—the important law of mass action

$$n \cdot p = n_i^2 \tag{I.3.03}$$

between electrons and holes. By comparison with (VIII.4.22), one obtains for the "inversion density n_i" (see page 26)

$$\begin{aligned} n_i &= \sqrt{N_C N_V}\, \mathrm{e}^{-\frac{1}{2}\frac{E_C - E_V}{\mathrm{k}T}} \\ &= 2.5 \cdot 10^{19}\ \mathrm{cm}^{-3} \left(\frac{m_n}{m}\right)^{\frac{3}{4}} \left(\frac{m_p}{m}\right)^{\frac{3}{4}} \left(\frac{T}{300^\circ\ \mathrm{K}}\right)^{\frac{3}{2}} \cdot \mathrm{e}^{-\frac{1}{2}\frac{E_C - E_V}{\mathrm{k}T}} \end{aligned} \tag{VIII.4.23}$$

In the present case of the insulator we have already noted (page 302) that the number of all electrons in valence and conduction band taken

---

[1] If one bears in mind that in the first approximation $m_n \approx m_p$ and hence $N_C \approx N_V$, one can see that the later result VIII.4.25 confirms these assumptions for low temperatures as well. In other words, in the case of the intrinsic semiconductor with which we are dealing here, the concentrations n and p decrease much faster with temperature than N_C and N_V.

However, as in footnote 1, p. 303, it should be emphasized once again that Eqs. (VIII.4.20) to (VIII.4.23) do not hold only for intrinsic semiconductors but also for impurity semiconductors in which the assumption (VIII.4.19) of non-degeneracy is fulfilled as long as the doping is not excessive. Only the assumption (VIII.4.24) effects the limitation to the case of intrinsic conductors.

together must be temperature-independent and we have concluded from this that the concentration of the conduction electrons in the conduction band must be equal to the concentration of the holes in the valence band:

$$n = p \tag{VIII.4.24}$$

From the combination of (VIII.4.22) and (VIII. 4.24) follows

$$n = p = n_i = \sqrt{N_C N_V}\, e^{-\frac{1}{2}\frac{E_C - E_V}{kT}} \tag{VIII.4.25}$$

With the aid of these expressions the energetic position E_F of the Fermi level can be computed from (VIII.4.20) and (VIII.4.21):

$$E_F = \frac{1}{2}(E_C + E_V) + \frac{3}{4}\,kT \ln \frac{m_p}{m_n} \tag{VIII.4.26}$$

In this equation the definitions (VIII.4.13) and (VIII.4.17) for the state densities N_V and N_C have been used. Thus one can see that *in an insulator the Fermi level E_F lies practically halfway between the lower edge E_C of the conduction band and the upper edge E_V of the valence band. This result is independent of the temperature* T, apart from a small correction that may be required to allow for any differences in the effective masses m_n and m_p.

This result could have been obtained, even without calculation, from the symmetry of the distribution functions $f(E)$ and $1 - f(E)$ around the Fermi level E_F (see Fig. VII.10.1). Owing to this symmetry, the condition "concentration n of the electrons = concentration p of vacancies or holes" is just fulfilled if the Fermi level E_F lies halfway between E_C and E_V.

Thus the complete picture for the insulator at temperatures $T > 0$ encompasses an electron gas in the conduction band and a gas consisting of "holes" or "defect electrons" in the valence band. Stated more precisely, the assumptions (VIII.4.02) and (VIII.4.11) lead to the conclusion that the electron gas has an apparent potential energy E_C (lower edge of the conduction band) and the hole gas has an apparent potential energy E_V (upper edge of the valence band). Within these potential wells of depth E_C and E_V, respectively, the electron and hole gases, respectively, are "quasi-free."[1] Both gases have very low concentration and behave completely like Maxwell gases. Hence the thermal velocity v_{th} of the electrons can be computed from the equation[2]

---

[1] An influence of the lattice beyond these apparent potential energies E_C and E_V is expressed in the substitutions $m \rightarrow m_n$ and $m \rightarrow m_p$.

[2] Mean energy per degree of freedom = $\frac{1}{2}kT$ (principle of equipartition). See

$$\frac{m_{\text{eff}}}{2} v_{\text{th}}^2 = \frac{3}{2} \mathrm{k}T$$

to be $$v_{\text{th}} = \sqrt{\frac{3\mathrm{k}T}{m_{\text{eff}}}} = 1.168 \cdot 10^7 \,\mathrm{cm\,sec^{-1}} \cdot \sqrt{\frac{T}{300°\mathrm{K}}} \sqrt{\frac{m}{m_{\text{eff}}}} \quad \text{(VIII.4.27)}$$

Therefore, v_{th} in insulators (and semiconductors) is of the order of magnitude 10^{-7} cm-sec^{-1} (at 300°K).

§5. Fermi Statistics in Semiconductors

In §3 of Chap. I we have seen that at temperature $T = 0$ a semiconductor is an insulator because it does not contain any free charge carriers. These are created only if the temperature is raised. First, electrons are then supplied by donor impurities (n-type conduction). Secondly, electrons are pulled out of valence bands by temperature excitation and captured by acceptor impurities. The vacancies in the totality of valence electrons are then available as "defect electrons" for the transport of current (p-type or hole conduction). Finally, at high enough temperatures, electrons are raised directly from the valence band into the conduction band, and the current is then carried by equal numbers of excess electrons in the conduction band and of holes in the valence band.

We intend to calculate the position of the Fermi level E_F, first in an n-type semiconductor, then in a p-type semiconductor, and finally in a semiconductor with donors *and* acceptors. From the point of view of Fermi statistics, we are concerned (for example, in an n-type semiconductor) with the distribution of a certain total number of electrons among the donor levels and the levels of the conduction band. In Chap. II the same problems have been considered from the point of view of impurity reactions and the laws of mass action. Where we spoke before, for instance, of a "neutral donor $D^\times$," we shall now describe the same situation as an "electron at a donor level." At the end of §5, we shall return to the earlier point of view and derive an expression for the mass-action constant. This will form a transition to Chap. IX in which the law of mass action will be considered again, this time from the dynamic point of view. This will also be an opportunity to indicate the treatment of the inertia of impurity reactions, whereas in Chap. II and also here thermal equilibrium between the impurities and the electrons and holes is always assumed.

G. Joos, "Theoretical Physics," 2d ed., p. 587, Hafner Publishing Company, New York, 1950.

a. Semiconductors with Donors: *n*-Type Semiconductors

In analogy to the previous discussion of metals and insulators where we did not have to consider the electrons in the lower bands explicitly, we can now again ignore the lower bands, including the valence band, and need deal only with the conduction band and the donor levels below it (see Fig. VIII.5.1). For the density of states in the conduction band we use again the expression (VIII.4.03) based on the assumption (VIII.4.02) and obtain, as previously for metals and insulators,

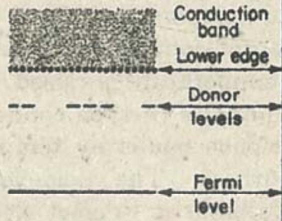


FIG. VIII.5.1. Position of the Fermi level E_F in the band model; *n*-type conductor; impurity saturation.

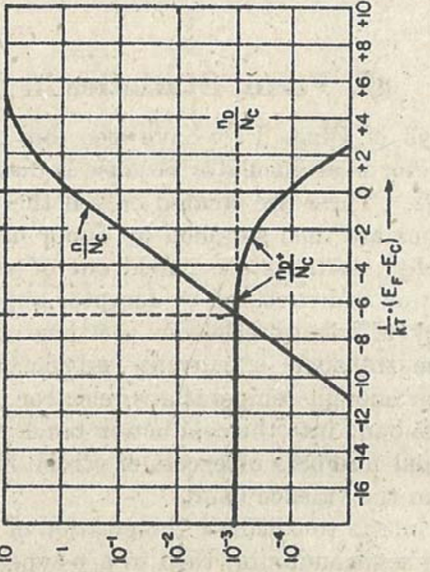


FIG. VIII.5.2. Graphic determination of the Fermi level E_F in *n*-type conductors; impurity saturation; $E_C - E_D = 2kT$; $n_D = 10^{-3}\ N_C$.

$$E_F = E_C + \zeta\left(\frac{n}{N_C}\right) \qquad \text{(VIII.4.06)}$$

i.e., a functional correlation $E_F(n)$ or $n(E_F)$, respectively, between the Fermi level E_F and the concentration n of the electrons in the conduction band. n as a function of E_F provides one of the two curves of Fig. VIII.5.2.

According to the principles of Fermi statistics, the probability with which a donor level E_D is occupied by an electron is[1]

---

[1] See Appendix III.

$$f(E_D) = \frac{1}{\frac{1}{2}\,e^{+\frac{1}{kT}(E_D - E_F)} + 1} \qquad \text{(VIII.5.01)}$$

Hence if we have in unit volume n_D donors at the energy level E_D, the fraction $n_D \cdot f(E_D)$ is occupied by an electron[1] and is therefore neutral:

$$n_{D^\times} = n_D \frac{1}{\frac{1}{2}\,e^{+\frac{1}{kT}(E_D - E_F)} + 1} \qquad \text{(VIII.5.02)}$$

Hence the concentration of the positively charged donors is

$$n_{D^+} = n_D - n_{D^\times} = n_D \left(1 - \frac{1}{\frac{1}{2}\,e^{+\frac{1}{kT}(E_D - E_F)} + 1}\right)$$

$$n_{D^+} = n_D \frac{1}{1 + 2\,e^{+\frac{1}{kT}(E_F - E_D)}} \qquad \text{(VIII.5.03)}$$

The functional dependence of this positive charge density n_{D^+} on E_F provides the second curve in Fig. VIII.5.2. To preserve the neutrality of the semiconductor as a whole, the concentration n of the negative electrons must be equal to the density n_{D^+} of the positive donors that remain. The resulting value E_F of the Fermi level, the *solution* of the equation

$$n(E_F) = n_{D^+}(E_F) \qquad \text{(VIII.5.04)}$$

is determined by the intersection of the two curves in Fig. VIII.5.2. With the assumed values for $E_C - E_D$ and n_D, a situation arises in which

$$n = n_{D^+} \approx n_D \qquad \text{(VIII.5.05)}$$

i.e., in which almost all donors are dissociated. In this state of "donor saturation,"[2] the Fermi level lies *below* the donor level ($E_F < E_D$), and just for this reason the majority of the donor levels are not occupied. For the $(E_C - E_D)$ values assumed in Figs. VIII.5.3 and VIII.5.4, on the other hand, we have an "impurity reserve"

$$n = n_{D^+} \ll n_D \qquad \text{(VIII.5.06)}$$

---

[1] If the Fermi level E_F is several kT below the donor level E_D, the number of electrons in the donor level is consequently

$$2n_D\,e^{+\frac{1}{kT}(E_F - E_D)}$$

This observation will be needed for footnote 1, p. 318.

[2] See Chap. II, §6, particularly p. 50.

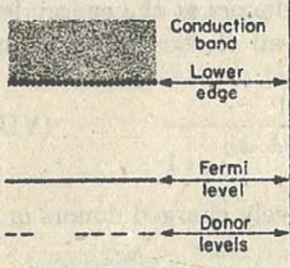


FIG. VIII.5.3. Position of the Fermi level E_F in the band model; n-type conductor; impurity reserve.

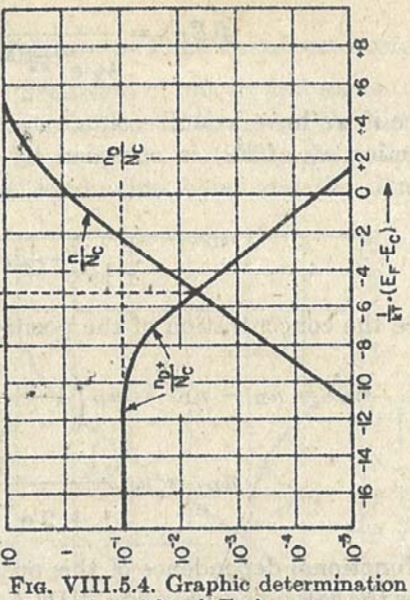


FIG. VIII.5.4. Graphic determination of the Fermi level E_F in n-type conductors; impurity reserve; $E_C - E_D = 8kT$; $n_D = 10^{-1}\ N_C$.

The Fermi level lies *above* the donor level ($E_F > E_D$), and therefore most of the donor levels are occupied.

b. Semiconductors with Acceptors: p-Type Semiconductors

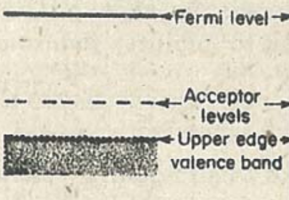
Here we have to investigate the distribution of the electrons among the valence band and the acceptor levels above it (see Fig. VIII.5.5). With the assumption (VIII.4.11) and the resulting density of states (VIII.4.12) in the valence band, we obtain the following functional correlation between the hole concentration p in the valence band and the Fermi level E_F

$$E_F = E_V - \zeta\left(\frac{p}{N_V}\right) \qquad \text{(VIII.4.15)}$$

With the aid of this equation, the curve $p(E_F)$ in Fig. VIII.5.6 can be drawn. Of the n_A acceptors in unit volume, the fraction $f(E_A)$ is occupied by an electron and therefore negatively charged:

$$n_{A^-} = n_A \frac{1}{2\,e^{+\frac{1}{kT}(E_A-E_F)} + 1} \qquad \text{(VIII.5.07)}$$

This supplies the curve n_A-(E_F) in Fig. VIII.5.6. To maintain



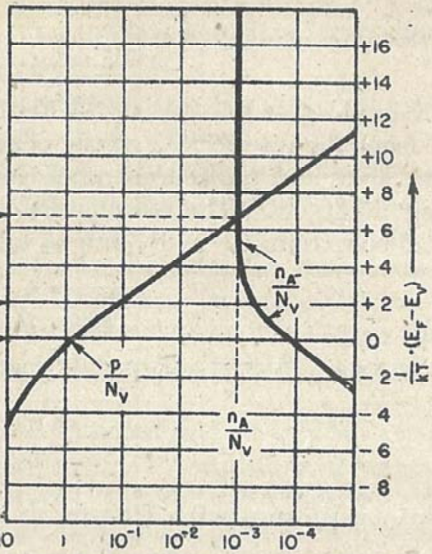
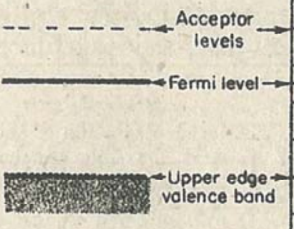


FIG. VIII.5.5. Position of the Fermi level E_F in the band model; p-type conductor; impurity saturation.

FIG. VIII.5.6. Graphic determination of the Fermi level E_F in p-type conductors; impurity saturation; $E_A - E_V = 2kT$; $n_A = 10^{-3} \cdot N_V$.



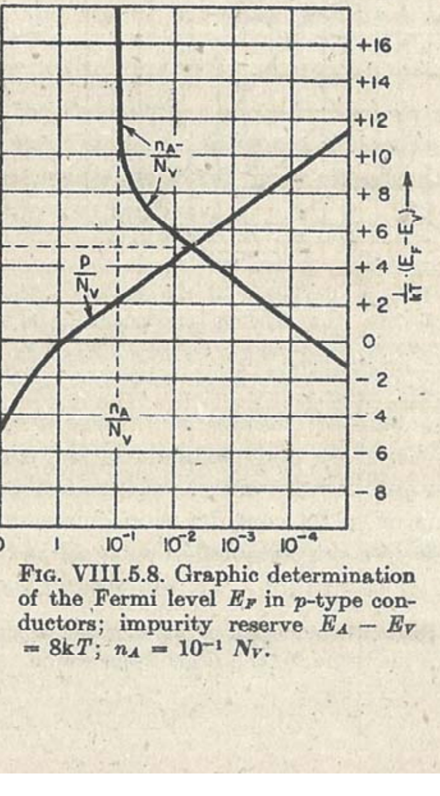


FIG. VIII.5.7. Position of the Fermi level E_F in the band model; p-type conductor; impurity reserve.

FIG. VIII.5.8. Graphic determination of the Fermi level E_F in p-type conductors; impurity reserve $E_A - E_V = 8kT$; $n_A = 10^{-1}\, N_V$.

neutrality

$$p(E_F) = n_{A^-}(E_F) \tag{VIII.5.08}$$

is again required so that the actual Fermi level in Fig. VIII.5.6 can be found from the intersection of the concentration curve $p(E_F)$ of the positive holes and the concentration curve $n_{A^-}(E_F)$ of the negatively charged acceptors. The assumptions made in Figs. VIII.5.5 and VIII.5.6 concerning $E_A - E_V$ and n_A lead to impurity *saturation*

$$\begin{aligned} n_{A^-} &\approx n_A \\ E_F &> E_A \end{aligned} \tag{VIII.5.09}$$

Figs. VIII.5.7 and VIII.5.8, however, lead to an impurity *reserve:*

$$\begin{aligned} n_{A^-} &\ll n_A \\ E_F &< E_A \end{aligned} \tag{VIII.5.10}$$

c Semiconductors with Donors *and* Acceptors

The poisoning of an n-type conductor by added acceptors and of a p-type conductor by added donors. If the donor concentration in an n-type conductor is increased (see Fig. VIII.5.9), the intersection of the n_{D^+} and of the n curve moves upward and to the right, i.e., the Fermi level E_F and particularly the electron concentration n assume higher and higher values. Hence the n-type conductivity of the semiconductor under consideration increases with the donor concentration n_D, at first proportionally with n_D as long as one does not leave the range of saturation. In the reserve range it increases[1] with $n_D^{1/2}$ [see

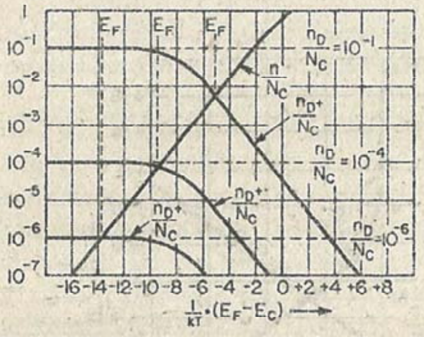


FIG. VIII.5.9. Variation of the Fermi level E_F and the electron concentration n with the donor concentration $n_D(E_C - E_D = 8\mathrm{k}T)$.

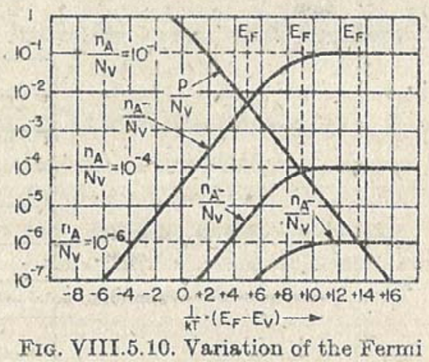


FIG. VIII.5.10. Variation of the Fermi level E_F and the hole concentration p with the acceptor concentration $n_A(E_A - E_V = 8\mathrm{k}T)$.

---

[1] A further attenuation of the increase occurs when the electron concentration exceeds the value N_C and degeneracy sets in.

Eqs. (II.6.15) and (II.6.17)]. Correspondingly, addition of acceptors to a p-type conductor increases the hole concentration and thereby the hole conduction (see Fig. VIII.5.10).

These results are not at all surprising. It is the donor content that makes a semiconductor into an n-type conductor; hence it seems obvious that an increase in the number of donors enhances the n-type conductivity. However, this argument might lead one to expect that addition of acceptors to an n-type conductor causes side by side with the n-type conduction additional p-type conduction, thus improving the total conductivity. In fact, however, the reverse effect takes place. The total conductivity decreases because the n-type conduction is poisoned by the added acceptors and additional p-type conduction is not produced.

This effect is easy to understand if one remembers that the acceptors have a tendency to capture electrons. By capturing ("trapping") of conduction electrons, the number of free carriers is reduced. Alternatively, one can say that the introduction of neutral acceptors produces additional deep electron levels without a corresponding increase of the total number of electrons which have to be accommodated. Hence it is certain that the Fermi level is lowered, with the result that the occupation of the states in the conduction band is reduced, i.e., that the concentration of conduction electrons is reduced.

Again, a complete understanding can be provided only by a quantitative treatment; therefore, in Figs. VIII.5.11 and VIII.5.12, and VIII.5.13 and VIII.5.14 the situations are shown which arise when acceptors are introduced into the impurity semiconductors of Figs. VIII.5.1 to VIII.5.4, the concentration of the acceptors in both cases being half an order of magnitude below that of the donor concentration. The condition for neutrality is then expressed by

$$n_{A^-} + n = n_{D^+} + p \qquad \text{(VIII.5.11)}$$

From Figs. VIII.5.12 and VIII.5.14 one can see that in both cases the hole concentration p is of no practical importance. The Fermi level E_F lies *so* high above the states of the valence band that these are all filled, with the exception of a negligible number (see Figs. VIII.5.11 and VIII.5.13). The same applies to the acceptors so that for all E_F values of interest the term n_{A^-} in the charge balance (VIII.5.11) is constant.

If the same applies for the donors, i.e., if there exists, even in the absence of acceptors, *saturation* of donors

$$n = n_{D^+} \approx n_D \quad \text{and} \quad E_F < E_D \qquad \text{(VIII.5.05)}$$

(Figs. VIII.5.1 and VIII.5.2, and VIII.5.11 and VIII.5.12), the poisoning effect of the acceptors is readily understood quantitatively. The addition of acceptors creates additional *deep* levels and thereby produces a lowering of the Fermi level and thus, if anything, enhances the donor saturation (VIII.5.05). In the charge balance (VIII.5.11) the two terms n_{A^-} and n_{D^+} are then practically independent of the position of the Fermi level E_F, and the term p becomes negligible. Hence simply

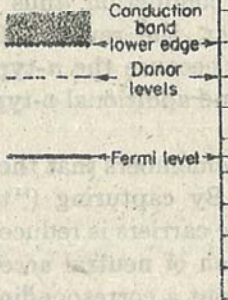


Fig. VIII.5.11. Position of the Fermi level E_F in the band model; semiconductor with donors and acceptors.

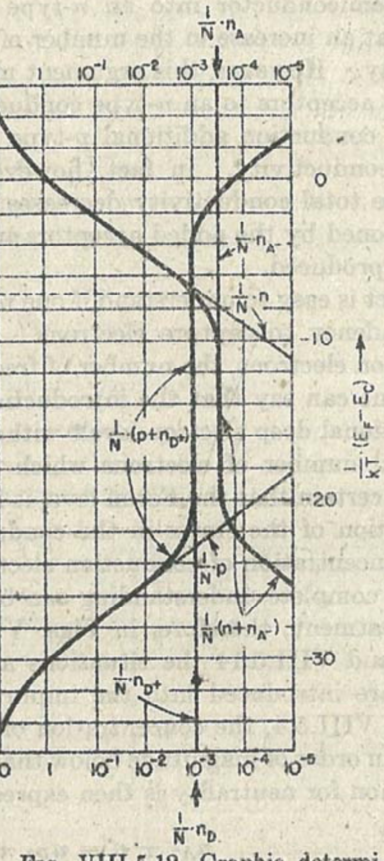


Fig. VIII.5.12. Graphic determination of the Fermi level E_F in a semiconductor with donors and acceptors: $E_C - E_D = 2kT$; $E_C - E_A = 22kT$; $E_C - E_V = 30kT$; $n_D = 10^{-3}\ N$; $n_A = 3 \cdot 10^{-4}\ N$; $N = N_C = N_V$.

$$n = n_{D^+} - n_{A^-}$$

$$n \approx n_D - n_A \qquad \text{(VIII.5.12)}$$

Figures VIII.5.13 and VIII.5.14 show the more complicated case where a donor *reserve* existed before the introduction of acceptors. Into the impurity semiconductor of Figs. VIII.5.3 and VIII.5.4, an acceptor concentration is introduced whose value is again one third of the donor concentration. A comparison of Figs. VIII.5.4 and VIII.5.14 shows that the electron concentration is reduced by about a factor 10 while in the preceding case the factor was only about $3/2$.

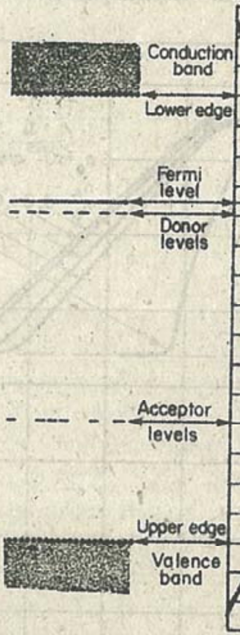


Fig. VIII.5.13. Position of the Fermi level E_F in the band model; semiconductor with donors and acceptors.

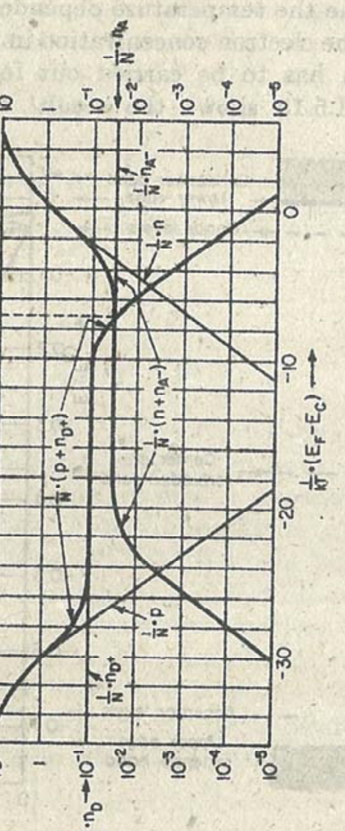


Fig. VIII.5.14. Graphic determination of the Fermi level E_F in a semiconductor with donors and acceptors: $E_C - E_D = 8kT$; $E_C - E_A = 22kT$; $E_C - E_V = 30kT$; $n_D = 10^{-1}N$; $n_A = 3 \cdot 10^{-2}N$; $N_V = N_C = N$.

The temperature T enters into the graphic determination of the Fermi level, e.g., in Fig. VIII.5.12, first, in the scale unit kT of abscissas, but also, secondly, in the scale unit of the ordinates

$$N = 2\left(\frac{2\pi m k T}{h^2}\right)^{3/2} .$$

To determine the temperature dependence of the Fermi level E_F and thereby of the electron concentration in the different levels, the graphic construction has to be carried out for a number of temperatures. Figure VIII.5.15 shows the result[1] for assumptions which apply approximately to the case of germanium. Essentially, the Fermi level E_F is lowered with increasing temperature. Hence, with increasing temperature the donors are all dissociated. At higher temperatures, the Fermi level lies in the center of the band, independent of the impurity concentrations n_D and n_A. This corresponds to the state

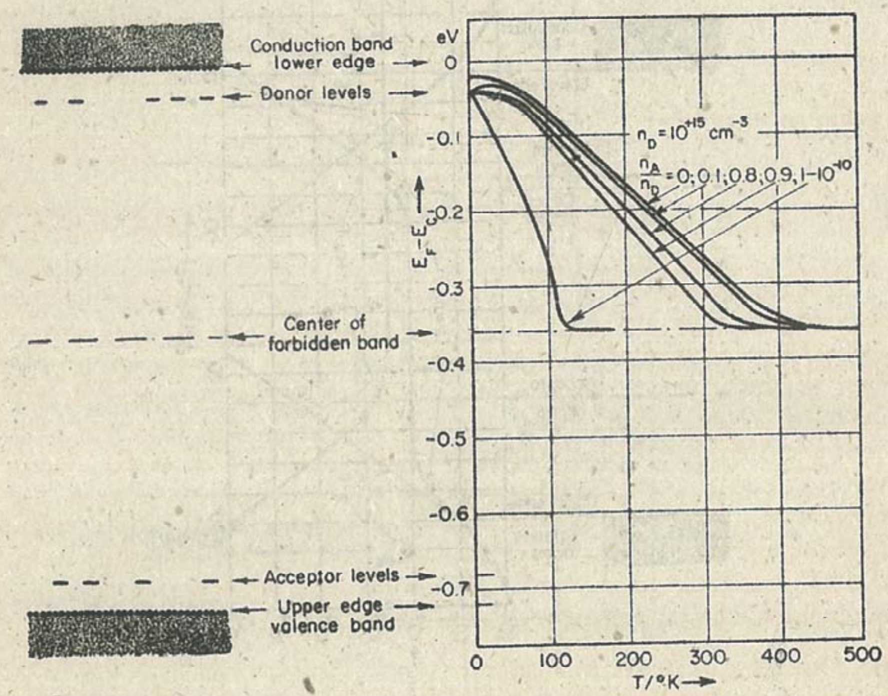


FIG. VIII.5.15. Temperature dependence of the Fermi level E_F. Germanium with 10^{15} donors per cm^3 and varying acceptor concentration.

---

[1] For similar investigations, see H. Mueser, *Z. Naturforsch.*, **5A**: 18 (1950). K. Seiler, *Z. Naturforsch.*, **5A**: 393 (1950). R. A. Hutner, E. S. Rittner, and F. K. DuPré, *Philips Research Repts.*, **5**: 188 (1950). W. Shockley, "Electrons and Holes in Semiconductors," pp. 465ff., D. Van Nostrand Company, Inc., Princeton, N.J., 1950.

of intrinsic conduction which will therefore be reached in every semiconductor at high enough temperatures, provided that the crystal does not melt first [see Eq. (VIII.4.25) and the discussion on page 25].

We now refer again to the fact that with known position of the Fermi level E_F the electron concentrations in the various levels are given by the equations (VIII.5.02), (VIII.5.07) and (VIII.4.06), (VIII.4.15). In the case of nondegeneracy

$$n \ll N_C \tag{VIII.5.13}$$

$$p \ll N_V \tag{VIII.5.14}$$

the last two equations can be replaced by (VIII.4.20) and (VIII.4.21). From these follow

$$\ln \frac{n}{N_C} = -\frac{1}{\mathrm{k}T}(E_C - E_F) \tag{VIII.5.15}$$

$$\ln \frac{p}{N_V} = -\frac{1}{\mathrm{k}T}(E_F - E_V) \tag{VIII.5.16}$$

Hence the distances $E_C - E_F$ and $E_F - E_V$ between Fermi level E_F and the band edges E_C and E_V, respectively, measured in units of $\mathrm{k}T$, are a logarithmic measure of the concentrations n and p, respectively. However, if the Fermi level does not lie in the forbidden zone but, for instance, in the valence band, then (VIII.5.16) would lead to a hole concentration $p \geqq N_V$ and the assumption (VIII.5.14) is no longer valid, i.e., degeneracy has set in. Therefore the equations (VIII.4.20), (VIII.4.21) and (VIII.5.15), (VIII.5.16), respectively, apply with reasonable accuracy only when the Fermi level is several $\mathrm{k}T$ away from the band edges. In this case the results can be translated into the language of the laws of mass action, which has certain advantages.

d. The Law of Mass Action

The computation of the concentration of the conduction electrons and the related problem of the saturation and of the reserve of one type of impurities have been treated in Chap. II, by using the reaction equations and the laws of mass action.

At that time it was noted that the application of the laws of mass action presupposes "sufficient" dilution of the electron gas. However, it could not be stated precisely what degree of dilution was sufficient, nor was it possible to give a reason for the given value of the mass-action constant.

From our new point of view the following can be said concerning these two questions. The dilution of the electron gas is sufficient for the law of mass action to be valid, provided that

$$n \ll N_C \tag{VIII.5.13}$$

because then we are in the linear part of the $n(E_F)$ curve of Figs. VIII.5.2, VIII.5.4, VIII.5.12, and VIII.5.14. Since in these figures n has been plotted on a logarithmic scale, this means the validity of the approximation (VII.4.21), which by solving for the concentration n of the conduction electrons results in[1]

$$n = N_C \, e^{+\frac{1}{kT}(E_F - E_C)} \qquad \text{(VIII.5.17)}$$

From this, together with Eqs. (VIII.5.02) and (VIII.5.03), we obtain the expression

$$\frac{n_{D^+} \cdot n}{n_{D^\times}} = n_D \frac{1}{1 + 2\,e^{+\frac{1}{kT}(E_F - E_D)}} \cdot N_C \, e^{+\frac{1}{kT}(E_F - E_C)} \, \frac{\frac{1}{2}\,e^{+\frac{1}{kT}(E_D - E_F)} + 1}{n_D}$$

$$= \frac{N_C}{2} \, e^{+\frac{1}{kT}(E_F - E_C)} \cdot e^{+\frac{1}{kT}(E_D - E_F)}$$

Hence

$$\frac{n_{D^+} \cdot n}{n_{D^\times}} = \frac{N_C}{2} \, e^{-\frac{1}{kT}E_{CD}} \qquad \text{(VIII.5.18)}$$

where

$$E_{CD} = E_C - E_D \qquad \text{(VIII.5.19)}$$

We can see that the Fermi level E_F which depends on the electron concentration n has been eliminated and that with a constant independent of concentration

$$K_D = \frac{N_C}{2} \, e^{-\frac{1}{kT}E_{CD}} \qquad \text{(VIII.5.20)}$$

the law of mass action

$$n_{D^+} \cdot n = K_D \cdot n_{D^\times} \qquad \text{(VIII.5.21)}$$

is valid.[2] The "internal work function E_{CD} of the donors" occurring in the expression (VIII.5.20) for the mass-action constant K_D is the

---

[1] If we compare (VIII.5.17) with footnote 1, p. 296, we see that the continuum of the available states in the conduction band behaves effectively like N_C states/cm^3, with a uniform energy level E_C. For this reason, Shockley chose the term "effective density of states" for N_C. (See Shockley, *op. cit.*, p. 240.) See also footnote 1, p. 285.

[2] In addition to this statistical proof, we have already mentioned a kinetic proof for the law of mass action on p. 46 which we shall discuss in more detail in Chap. IX. Finally, the law of mass action can be derived from the general equilibrium condition of footnote 1, p. 294.

$$(\text{Electrochemical potential})_{\text{phase I}} = (\text{electrochemical potential})_{\text{phase II}}$$

if we use in this equilibrium condition the expression for the chemical potential of a Maxwell gas as obtained from thermodynamics. However, without the aid of statistics, the constant of the law of mass action remains undefined because the entropy constant in thermodynamics is undefined. See W. Weizel, "Theoretische Physik," pp. 732, 733, and 1237, Springer-Verlag OHG, Berlin, 1949.

activation energy which has to be supplied to an electron to raise it from a donor level into the conduction band (see Fig. VIII.5.16). A corresponding law of mass action applies to the acceptors:

$$n_{A^-}\cdot p = K_A \cdot n_{A^\times} \qquad \text{(VIII.5.22)}$$

with

$$K_A = \frac{N_V}{2}\cdot e^{-\frac{E_{AV}}{kT}} \qquad \text{(VIII.5.23)}$$

and with

$$E_{AV} = E_A - E_V > 0 \qquad \text{(VIII.5.24)}$$

if the concentration p of the holes is sufficiently small:

$$p \ll N_V \qquad \text{(VIII.5.25)}$$

Using the example of the semiconductor with donors *and* acceptors, we want to indicate briefly the treatment of such a somewhat complex problem with the methods of reaction kinetics[1] instead of a purely statistical procedure.

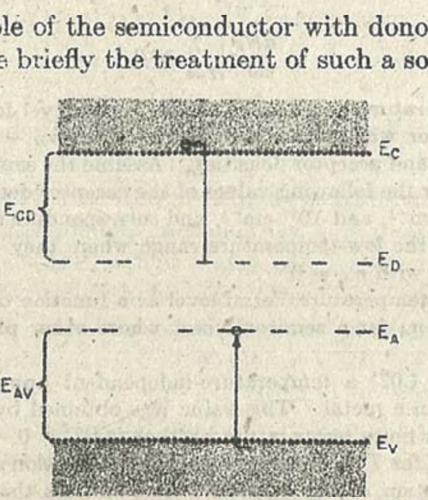


FIG. VIII.5.16. Activation energies of impurities.

In this case, we have the three laws of mass action:

$$n_{D^+}\cdot n = K_D n_{D^\times} \qquad \text{(VIII.5.21)}$$
$$n_{A^-}\cdot p = K_A n_{A^\times} \qquad \text{(VIII.5.22)}$$
$$n\cdot p = n_i^2 \qquad \text{(1.3.03)}$$

Furthermore we have two expressions for the impurities

$$n_{D^\times} + n_{D^+} = n_D \qquad \text{(VIII.5.26)}$$
$$n_{A^\times} + n_{A^-} = n_A \qquad \text{(VIII.5.27)}$$

and finally a neutrality condition

$$n_{A^-} + n = n_{D^+} + p \qquad \text{(VIII.5.11)}$$

---

[1] See W. Schottky, *Z. Elektrochem.*, **45**: 33 (1939).

These are six equations with the six unknowns n, p, $n_{D^\times}$, n_{D^+}, n_{A^-}, and $n_{A^\times}$ which are thus "determined," i.e., expressed in terms of the material constants K_D, K_A, n_i and the total impurity contents n_D and n_A.

§6. Problems

1. Plot the temperature dependence of the Fermi level for a semiconductor with the following properties: $E_C - E_V = 0.72$ ev, $N_C = \frac{1}{8}N$, $N_V = N$, $E_C - E_D = 0.04$ ev, $n_D - n_A = 10^{16}\ \mathrm{cm}^{-3}$, $n_A = 0$.

2. Figure VIII.5.15 shows that the Fermi level for $T = 0$ does not lie halfway between the donor level and the conduction band but at the donor level itself if the semiconductor contains acceptors as well as donors. Under what conditions, then, does the Fermi level rise initially with temperature, that is, under what conditions is

$$\left.\frac{dE_F}{dT}\right|_{T=0} > 0$$

3. Plot the temperature dependence of the Fermi level for a "partly compensated" semiconductor with a constant donor excess, $n_D - n_A$, of $10^{16}\ \mathrm{cm}^{-3}$, but with different donor and acceptor densities. Assume the same numerical values as in Prob. 1, except for the following values of the acceptor density: $n_A = 10^{15}\ \mathrm{cm}^{-3}$, $3 \cdot 10^{15}\ \mathrm{cm}^{-3}$, $10^{16}\ \mathrm{cm}^{-3}$, and $10^{17}\ \mathrm{cm}^{-3}$, and corresponding donor densities. Plot the curves only for the low-temperature range where they differ from the curve obtained in Prob. 1, with $n_A = 0$.

4. Plot the room-temperature Fermi level as a function of the "net" impurity concentration $n_D - n_A$ for a semiconductor whose other properties are given in Prob. 1.

5.* In Eq. (VIII.4.07) a temperature-independent approximation was given for the Fermi level in a metal. This value was obtained by replacing the actual Fermi function for a finite temperature by that for $T = 0$. The value obtained, therefore, holds only for $T = 0$ but is a good approximation otherwise. Calculate the first order deviation of the actual Fermi level from this approximation, that is, calculate the quantity

$$\Delta\zeta = T \cdot \left(\frac{d\zeta}{dT}\right)_{T=0}$$

6.* Determine the room-temperature Fermi level in an intrinsic semiconductor like indium antimonide which has a very narrow forbidden band and a very low effective mass for the electrons. Assume the following numerical values: $E_G = 0.18$ ev, $m_n = 0.03m$, $m_p = m$. *Note:* Make use of the graph of Fig. A.II.1 for the determination of the Fermi level.

CHAPTER IX

The Dynamic Approach to Impurity Equilibria and the Inertia of Impurity Reactions

In the investigation of any semiconducting material, one of the most important tasks is the determination of the number of free charge carriers, e.g., of the concentration of excess electrons in the conduction band. While the experimental methods for this purpose are characterized by the terms "Hall effect" and thermoelectric power, their counterparts in the realm of theory are "Fermi statistics" and "laws of mass action." We have discussed the statistical method in detail in Chap. VII, §10, and in the whole of Chap. VIII. The laws of mass action have been treated in Chap. II, §6, and on pages 317 to 319. It was shown that the statistical method is more comprehensive than the method of the laws of mass action because it is also applicable for large concentrations[1] $n \gg N_C$ and, furthermore, because it supplies the magnitude of the proportionality factors in the laws of mass action which remain undetermined in the thermodynamic treatment.[2] On the other hand, the laws of mass action have the advantage of great simplicity.

These remarks refer in the first place to the treatment of equilibrium states, but they are all the more valid when dynamic processes are under consideration, e.g., the slow rise and decay of luminescence in phosphors or the behavior of rectifying boundary layers at high frequencies.[3]

In the case of dynamic processes the statistical method, just because of its comprehensive and fundamental character, creates a number of very difficult problems.[4] Furthermore, it uses several concepts (life-

---

[1] N_C = effective density of states in the conduction band. See Eqs. (VIII.1.04) and (VIII.4.04).

[2] See, for instance, W. Weizel, "Lehrbuch der theoretischen Physik," pp. 732–733, Springer-Verlag OHG, Berlin, 1949.

[3] W. Schottky, *Z. Physik*, **132**: 261 (1952).

[4] W. Schottky, *Ann. Physik*, **6**: 193 (1949).

time, formation rate) which are introduced and defined more easily on the basis of the laws of mass action than within the scope of the statistical method itself. For this purpose, however, the laws of mass action have to be extended beyond the equilibrium form. This will be done below. In §1 we shall deal in a *general* way with the definition of the so-called mean lifetime τ, while the lifetimes of charged and neutral donors will be discussed specifically in §2 and §3. In §4, Eqs. (IX.4.02) and (IX.4.03) will give the previously mentioned extension of the simple law of mass action $n_{D^+} \cdot n = K_D \cdot n_{D^\times}$ for dynamic processes. In addition, the case of equilibrium will be reviewed once more from the newly gained point of view. Finally, in §5 a simple typical example of a nonstationary process will be considered and the time constant obtained for it will be discussed.

§1. Relaxation Time and Mean Lifetime for a Particular Type of Imperfection

An impurity, e.g., a donor D, which can react according to $D^\times \rightleftarrows D^+ + \ominus$ is by no means permanently in the charged state D^+; instead, after a shorter or longer period of time, it will associate with a free conduction electron and change into a neutral donor $D^\times$. Correspondingly, the state $D^\times$ is not permanent or final, but it too will, after a shorter or longer period of time, dissociate again by thermal excitation into $D^+ + \ominus$. In the first process, we have the "death" of a D^+ and the "birth" of a $D^\times$, in the second process the "death" of a $D^\times$ and the "birth" of a D^+; extending this nomenclature, we call the period of time for which the impurity is in the charged (neutral) state its "lifetime" in the charged (neutral) state. The processes just described are statistical events, and therefore the time intervals during which the donors of a given type are neutral or charged are by no means uniformly the same. Hence statements regarding, e.g., the temperature dependence of these times will primarily refer to mean values, i.e., mean lifetimes $\tau_{D^\times}$ or τ_{D^+}. Later on we shall have to discuss the fact that *different* mean lifetimes can be defined for a given particle. However, before we come to that we introduce a relaxation time τ_{rel} which is defined as follows:

We consider a group of N_S particles of a given type[1] S. The indi-

---

[1] For the remaining discussions of §1, type S may comprise, in addition to $D^\times$, D^+, $A^\times$, A^-, the minority carriers, i.e., the holes in an n-type conductor or the electrons in a p-type conductor. In that case the relaxation time τ_{rel} defined by (IX.1.01) is the "lifetime τ_p or τ_n" of the holes or the electrons. A detailed treatment of the lifetimes of minority carriers in connection with traps can be found in W. Shockley and W. T. Read Jr., *Phys., Rev.*, 87: 835 (1952).

vidual members of this group die one after the other in statistical irregularity. According to definition, we do not include in the group new particles of type S which were born after the beginning $t = t_0$ of the observation. Then the number N_S decreases steadily, and a decay or relaxation time τ_{rel} can be defined by the equation

$$dN_S = -\frac{1}{\tau_{\text{rel}}} N_S \, dt \qquad \text{(IX.1.01)}$$

In formulating this defining equation, one thinks in the first place of the case of sufficient dilution in which the fates of individual group members are independent of each other.[1] Then the probability that a given group member dies during the time from t to $t + dt$ is the same for all group members because it was postulated that they were all identical. Hence the number $|dN_S| = -dN_S$ of deaths during this time element $(t, t + dt)$ is proportional to the number $N_S(t)$ of the group members still alive at time t. Further, the expression (IX.1.01) refers to particles whose death is a purely random event which is quite independent of its previous history.[2] Then the time t cannot enter the number $-dN_S$ of deaths during the time $(t, t + dt)$ explicitly but only by way of the number $N_S(t)$ of the surviving group members at time t. This, however, presupposes that any physical parameters affecting the place and the environment of the events (such as the temperature) do not vary with time.[3] Hence τ_{rel} is a constant independent of time and concentration if the three assumptions of "sufficient dilution," "independence of the individual fate from previous history," and "constancy of environment in time" are satisfied. Integration of (IX.1.01) leads, for this case, to

$$N_S(t) = N_S(t_0)\, \mathrm{e}^{-\frac{t-t_0}{\tau_{\text{rel}}}} \qquad \text{(IX.1.02)}$$

Here the relaxation time defined by (IX.1.01) is also related to the initially mentioned mean lifetimes τ because the concentration of a group $N_S(t_0) = \nu(t_0)\, dt_0$ of particles born during the time interval $(t_0, t_0 + dt_0)$ decreases according to (IX.1.02), as expressed by the equation

$$N_S(t) = \nu(t_0)\, \mathrm{e}^{-\frac{t-t_0}{\tau_{\text{rel}}}} dt_0$$

Hence the number of group members with a lifetime between $\tau = t - t_0$ and $\tau + d\tau = t - t_0 + dt$ is, in accordance with Fig. IX.1.1,

$$|dN_S| = \frac{1}{\tau_{\text{rel}}} \nu(t_0)\, \mathrm{e}^{-\frac{t-t_0}{\tau_{\text{rel}}}} dt_0 \, dt = \nu(t_0)\, dt_0 \cdot \mathrm{e}^{-\frac{\tau}{\tau_{\text{rel}}}} \cdot d\left(\frac{\tau}{\tau_{\text{rel}}}\right)$$

---

[1] Opposite: Fermi gas!

[2] Opposite: A group of living beings where age affects the future fate decisively.

[3] Opposite: See end of §2, particularly footnote 1, p. 327.

The mean lifetime $\bar{\tau}$ can be computed by ascertaining the mean value only for the members of the group of particles characterized by identical birth date t_0

$$\bar{\tau} = \frac{\int\limits_{\tau=0}^{\tau=\infty} \tau \cdot |dN_S|}{\int\limits_{\tau=0}^{\tau=\infty} |dN_S|} = \frac{\nu(t_0)\, dt_0 \int\limits_{\tau=0}^{\tau=\infty} \tau\, e^{-\frac{\tau}{\tau_{\mathrm{rel}}}}\, d\left(\frac{\tau}{\tau_{\mathrm{rel}}}\right)}{\nu(t_0)\, dt_0 \int\limits_{\tau=0}^{\tau=\infty} e^{-\frac{\tau}{\tau_{\mathrm{rel}}}}\, d\left(\frac{\tau}{\tau_{\mathrm{rel}}}\right)}$$

If we integrate, we obtain

$$\bar{\tau} = \tau_{\mathrm{rel}}$$

Hence the relaxation time τ_{rel} defined by (IX.1.01) has also significance as a mean value provided that τ_{rel} is time-independent as in the foregoing case.

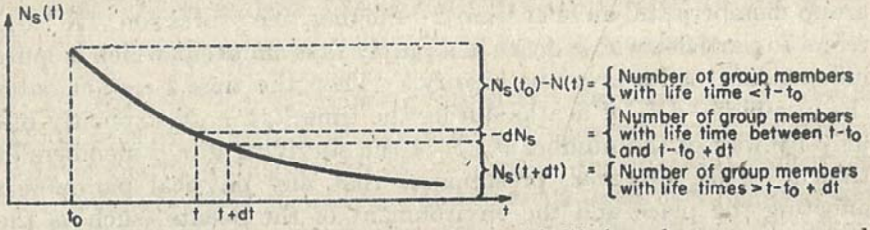


FIG. IX.1.1. The number of group members with lifetimes between $t - t_0$ and $t - t_0 + dt$ is

$$-dN_S = \frac{1}{\tau_{\mathrm{rel}}} \cdot N_S(t_0) \cdot e^{-\frac{t-t_0}{\tau_{\mathrm{rel}}}}\, dt = \frac{1}{\tau_{\mathrm{rel}}} \cdot \nu(t_0) \cdot e^{-\frac{t-t_0}{\tau_{\mathrm{rel}}}} \cdot dt_0 \cdot dt$$

Incidentally, the demonstrated identity of relaxation time τ_{rel} and mean lifetime of all particles *born* at the same time t_0 is by no means a matter of course. For instance, if we take the mean value $\bar{\bar{\tau}}$ of the lifetimes of all particles existing simultaneously at a given time t_1, we have to compute

$$\bar{\bar{\tau}} = \frac{\int\limits_{t_0=-\infty}^{t_0=t_1} \nu(t_0)\, dt_0 \int\limits_{\tau=t_1-t_0}^{\tau=\infty} \tau\, e^{-\frac{\tau}{\tau_{\mathrm{rel}}}}\, d\left(\frac{\tau}{\tau_{\mathrm{rel}}}\right)}{\int\limits_{t_0=-\infty}^{t_0=t_1} \nu(t_0)\, dt_0 \int\limits_{\tau=t_1-t_0}^{\tau=\infty} e^{-\frac{\tau}{\tau_{\mathrm{rel}}}}\, d\left(\frac{\tau}{\tau_{\mathrm{rel}}}\right)}$$

Assuming, by way of example, that the number of births $\nu(t_0)$ is constant in time we obtain for the thus defined mean lifetime $\bar{\bar{\tau}} = 2\tau_{\mathrm{rel}}$.

Assuming, in another example, that no particles were born at all in the time interval $t_0 = t_1 - 10\tau_{\mathrm{rel}}$ to $t_0 = t_1$, then when the mean values are ascertained at the time t_1 there exist only particles with lifetimes $\geqq 10\tau_{\mathrm{rel}}$. In this case, the mean value $\bar{\bar{\tau}}$ must become greater than $10\tau_{\mathrm{rel}}$.

Hence we have to be careful with the statement that the relaxation time τ_{rel} defined by (IX.1.01) is equal to the mean lifetime. For instance, the statement is correct if τ_{rel} is time-independent and if the mean value is taken for all simultaneously born particles of the type under consideration. If the mean value is taken for all particles simultaneously alive at a given moment t_1, the statement is not correct because in this total the long lifetimes are favored.

Theoretical deductions frequently refer to Eq. (IX.1.01), and therefore the relaxation time τ_{rel} of the particle type concerned which is defined by this equation is sometimes more important than any mean values of the statistically distributed lifetimes. It is now common practice to talk of τ_{rel} simply as the mean lifetime τ of the particles concerned although, as we have just shown, this identity depends on several assumptions which are often *not* fulfilled. The situation is similar to that concerning the mean collision time of the conduction electrons (see page 253). Here, too, in the continued development of the theory, the original character of a real mean value has given way to that of a relaxation or decay time of a group of particles, while the original name of "mean collision time" has been retained. In the following we, too, shall call the time value defined by Eq. (IX.1.01) simply τ_s = mean lifetime of the particles of type S. In §2 and §3, the particle types S will be represented by the charged donors D^+, and the neutral donors $D^\times$, so that in §2 we have $S = D^+$ and in §3, $S = D^\times$.

§2. Physical Statements about the Lifetime τ_{D^+}

The death of a charged donor D^+ occurs when it associates with a free conduction electron $\ominus$. The frequency of such recombination events will be proportional to the concentrations[1] n_{D^+} and $n_\ominus$, assuming low enough concentrations, i.e., sufficient "dilution" of the two reaction partners D^+ and $\ominus$; hence

$$\frac{\text{Number of association processes}}{(\text{Unit time}) \cdot (\text{unit volume})} = r_D n_\ominus n_{D^+} \qquad \text{(IX.2.01)}$$

---

[1] As in Chap. II we denote the electron concentration with $n_\ominus$, instead of simply with n as in the rest of this book.

One condition for such recombination events is that the two partners meet in space; then an electron $\ominus$ will be sufficiently close to a given donor D^+ more often if more electrons $\ominus$ are in thermal motion. Therefore the frequency of the association processes is proportional to the electron concentration $n_\ominus$. On the other hand, an electron $\ominus$, during its random thermal motion, will meet a charged donor D^+ more often if more such charged donors D^+ are present, i.e., if n_{D^+} is greater. This is the reason why the number of recombination processes is proportional to n_{D^+}.

The proportionality factor in (IX.2.01) is the so-called recombination coefficient r_D. It is frequently split up into the two factors of effective cross section σ_{D^+} and mean thermal velocity v_{th} of the electrons $\ominus$

$$r_D = \sigma_{D^+} \cdot v_{\text{th}} \qquad \text{(IX.2.02)}$$

where according to (VIII.4.27)

$$v_{\text{th}} = \sqrt{\frac{3kT}{m_{\text{eff}}}} = 1.168 \cdot 10^7 \frac{\text{cm}}{\text{sec}} \sqrt{\frac{T}{300°\text{K}}} \cdot \sqrt{\frac{m}{m_{\text{eff}}}} \qquad \text{(VIII.4.27)}$$

This splitting-up is based on the following consideration. One pictures the stationary donors D^+ as spheres of radius R while the electrons $\ominus$ are supposed to be without spatial extent. Now, if such a point electron passes the center of a donor sphere at a distance smaller than R, a collision occurs (see, however, below). In its zigzag flight with the velocity v_{th}, each one of the $n_\ominus$ electrons therefore sweeps out in unit time a broken cylindrical tube having the volume $\pi R^2 \cdot v_{\text{th}}$. This volume contains $n_{D^+} \cdot \pi R^2 v_{\text{th}}$ charged donors D^+, and a collision takes place with each one of them (however, see below). Therefore the number of collisions per unit time of *one* point electron $\ominus$ is

$$n_{D^+} \pi R^2 v_{\text{th}}$$

and that of $n_\ominus$ point electrons is

$$\pi R^2 v_{\text{th}} \cdot n_{D^+} n_\ominus = \sigma_{\text{coll}} v_{\text{th}} \cdot n_{D^+} \cdot n_\ominus$$

with $\sigma_{\text{coll}} = \pi R^2$ = collision cross section.

Each collision need not necessarily lead to a recombination, and sometimes only a fraction of the collisions is successful. One allows for this by introducing, in place of the cross section σ_{coll}, the

"Effective cross section σ_{D^+} for the recombination $D^+ + \ominus \rightarrow D^\times$"

and obtains

$$\frac{\text{Number of recombination processes}}{\text{(Unit time)} \cdot \text{(unit volume)}} = \sigma_{D^+} \cdot v_{\text{th}} \cdot n_{D^+} \cdot n_\ominus \qquad \text{(IX.2.03)}$$

Comparison of (IX.2.01) and (IX.2.03) results in Eq. (IX.2.02).

Here it must be emphasized that it is proper to assume the effective cross section σ_{D^+} for the recombination to be smaller than, or at most equal to, the collision cross section. However, there would be no justification at all for equating the collision cross section to the "geometric" cross section of the donors D^+ and assuming it to be $\approx 10^{-15}$ cm². Quantitative statements regarding σ_{coll} can be made only on the basis of a thorough and usually quite difficult analysis of the collision process, quite apart from the fact that the value $(3 \cdot 10^{-8}\ \text{cm})^2$ for the "geometric" cross section is also based on quite arbitrary assumptions.

Now we want to turn to the computation of the lifetime τ_{D^+} of the donors in the charged state. We consider a group of charged donors D^+ at time t whose concentration n_{D^+} decreases during the subsequent time interval dt by

$$-dn_{D^+} = r_D \cdot n_\ominus \cdot n_{D^+}\, dt \qquad \text{(IX.2.04)}$$

since each recombination process causes the death of a donor and because according to Eq. (IX.2.01) the number of recombination processes in unit time and unit volume is $r_D n_\ominus n_{D^+}$ and, finally, because, as agreed, charged donors D^+ created by dissociation of neutral donors $D^\times$ during the time interval $(t, t + dt)$ under consideration must not be counted when Eq. (IX.1.01) is used. Next we want to compare Eq. (IX.1.01) with Eq. (IX.2.04), and we obtain for the mean lifetime τ_{D^+} of the charged donors D^+

$$\frac{1}{\tau_{\text{rel}}} = \frac{1}{\tau_{D^+}} = r_D n_\ominus \qquad \text{(IX.2.05)}$$

The presence of the concentration $n_\ominus$ of the reaction partner $\ominus$ in the expression for τ_{D^+} requires some caution in the use of this lifetime. In dynamic processes, *all* concentrations will usually vary, including $n_\ominus$. In this case, τ_{D^+} is no longer a constant but will vary with time,[1] which must be remembered if integrations have to be carried out.

§3. Physical Statements about $\tau_{D^\times}$

This inconvenient though unavoidable difficulty does not arise in the case of the lifetime $\tau_{D^\times}$ of donors in the associated and, hence, neutral state. Whether a $D^\times$ dissociates or not can depend only on

---

[1] Here we have an example of the case mentioned in §1, where a "lifetime" is not constant in time because the "environment" varies with time. This is so because the environment comprises not only the dielectric constant of the fundamental lattice, the temperature of the crystal, etc., but also the concentration of any reaction partners.

the properties of the $D^\times$ and on the available thermal energy, i.e., on the temperature T, but not on any other concentration because no reaction partner is required. The concentration $n_{D^\times}$ itself, too, cannot play any role provided that the dilution is adequate. Therefore in dynamic processes $\tau_{D^\times}$ will not be time-dependent and can be treated as a constant because dynamic processes usually take place too fast for any change of temperature to occur.

The time-independence of $\tau_{D^\times}$ (at constant temperature) will be confirmed when, in §4, we shall derive the following expression for $1/\tau_{D^\times}$ (see page 330):

$$\frac{1}{\tau_{D^\times}} = r_D \cdot \frac{N_C}{2} \cdot \mathrm{e}^{-\frac{E_{CD}}{\mathrm{k}T}} = \sigma_{D^+} \cdot v_{\mathrm{th}} \cdot \frac{N_C}{2} \cdot \mathrm{e}^{-\frac{E_{CD}}{\mathrm{k}T}} \qquad \text{(IX.3.01)}$$

Here N_C is the effective density of states in the conduction band[1]

$$N_C = 2\left(\frac{2\pi m_{\mathrm{eff}}\mathrm{k}T}{h^2}\right)^{3/2} = 2.5 \cdot 10^{19}\ \mathrm{cm}^{-3}\left(\frac{m_{\mathrm{eff}}}{m}\right)^{3/2}\left(\frac{T}{300°\mathrm{K}}\right)^{3/2} \qquad \text{(VIII.4.04)}$$

One can see that only the temperature T can cause a time dependence of $\tau_{D^\times}$ because, apart from T, Eq. (IX.3.01) contains only atomic constants e, m, h, k, and the following factors which characterize the type of impurity:

Effective cross section σ_{D^+} and dissociation energy E_{CD}

The presence of the latter is not surprising. At a given temperature T, the more energy is required to remove the electron $\ominus$ from the donor core D^+, the less frequently will the thermal energy kT suffice to cause dissociation and, hence, the larger will be the lifetime $\tau_{D^\times}$.

Apart from this, the number of dissociations in unit time and unit volume is, of course, greater the more associated donors $D^\times$ are present because the probability of dissociation is equally great for each $D^\times$. Therefore we have

$$\frac{\text{Number of dissociation processes}}{(\text{Unit time}) \cdot (\text{unit volume})} = \frac{1}{\tau_{D^\times}}\, n_{D^\times} \qquad \text{(IX.3.02)}$$

By appropriate application of (IX.1.01) it follows that the proportionality factor is $1/\tau_{D^\times}$.

§4. The Generalization of the Law of Mass Action for Dynamic Processes

The concentration n_S of an impurity type S and its variation with time dn_S/dt are determined by the opposing actions of birth and death

---

[1] See Eqs. (VIII.1.04) and (VIII.4.04).

rate:

$$\frac{dn_S}{dt} = +\frac{\text{number of births}}{(\text{unit time}) \cdot (\text{unit volume})} - \frac{\text{number of deaths}}{(\text{unit time}) \cdot (\text{unit volume})} \qquad \text{(IX.4.01)}$$

For dissociated charged donors D^+, dissociations represent births and associations represent deaths. Equations (IX.3.02) and (IX.2.01) and application of (IX.4.01) to the impurity type $S = D^+$ lead, therefore, to

$$\frac{dn_{D^+}}{dt} = +\frac{1}{\tau_{D^\times}} n_{D^\times} - r_D n_\ominus n_{D^+} \qquad \text{(IX.4.02)}$$

In the case of impurity type $S = D^\times$, i.e., of associated neutral donors, we have the reverse situation where dissociations take the role of deaths and associations the role of births:

$$\frac{dn_{D^\times}}{dt} = +r_D n_\ominus n_{D^+} - \frac{1}{\tau_{D^\times}} n_{D^\times} \qquad \text{(IX.4.03)}$$

Addition yields

$$\frac{dn_{D^+}}{dt} + \frac{dn_{D^\times}}{dt} = 0$$

The total concentration n_D of all donors ($= n_{D^+} + n_{D^\times}$) remains constant in dynamic processes.[1]

Equations (IX.4.02) and (IX.4.03) are generalizations of the law of mass action which is valid only in the case of equilibrium

$$n_{D^+} \cdot n_\ominus = K_D n_{D^\times} \quad \text{(II.6.04) or (VIII.5.21)}$$

It is worth while to review the case of equilibrium from the newly gained point of view. In the case of equilibrium, the concentrations must be constant in time:

$$\frac{dn_{D^+}}{dt} = -\frac{dn_{D^\times}}{dt} = 0$$

Hence, in accordance with (IX.4.02) or (IX.4.03),

$$n_\ominus \cdot n_{D^+} = \frac{1}{r_D \tau_{D^\times}} n_{D^\times} \qquad \text{(IX.4.04)}$$

If we compare this form of the law of mass action with the form in which it is usually written, (II.6.04) or (VIII.5.21), we obtain

$$r_D \tau_{D^\times} K_D = 1 \qquad \text{(IX.4.05)}$$

Using the expression (VIII.5.20), derived from Fermi statistics, for the mass-action constant K_D and also Eq. (IX.2.02), we obtain the

---

[1] See also footnote 3, p. 47.

expression previously discussed in §3

$$\frac{1}{\tau_{D_\times}} = r_D \frac{N_C}{2} e^{-\frac{E_{CD}}{kT}} = \sigma_{D^+} \cdot v_{th} \cdot \frac{N_C}{2} \cdot e^{-\frac{E_{CD}}{kT}} \qquad \text{(IX.3.01)}$$

which we shall now compare once more with the expression for $1/\tau_{D^+}$ from §2:

$$\frac{1}{\tau_{D^+}} = r_D n_\ominus = \sigma_D v_{th} n_\ominus \qquad \text{(IX.2.05)}$$

Both equations have general validity, i.e., not only for the case of equilibrium.[1]

For the special case of equilibrium we obtain from (IX.4.04) by multiplication with r_D

$$r_D n_\ominus \cdot n_{D^+} = \frac{1}{\tau_{D^\times}} n_{D^\times}$$

and further with the aid of (IX.2.05)

$$\frac{1}{\tau_{D^+}} \cdot n_{D^+} = \frac{1}{\tau_{D^\times}} \cdot n_{D^\times} \qquad \text{(only valid in the case of equilibrium)} \qquad \text{(IX.4.06)}$$

and

$$\frac{\tau_{D^\times}}{\tau_{D^+}} = \frac{n_{D^\times}}{n_{D^+}} \qquad \text{(only valid in the case of equilibrium)} \qquad \text{(IX.4.07)}$$

In the form of (IX.4.06) the condition of equilibrium is again apparent because it says:

Number of deaths of
D^+ ($\equiv$ number of births of $D^\times$) = number of deaths of $D^\times$
hence
Number of births = number of deaths for $D^\times$

It can be seen from (IX.4.07) that for the case of *reserve* ($n_{D^+} \ll n_{D^\times}$) we have

$$\tau_{D_+} \ll \tau_{D^\times} \qquad \text{(IX.4.08)}$$

The equilibrium in this case of few D^+ is produced by the fact that, while these few D^+ die relatively quickly, an equivalent birth rate is achieved by the dissociation of many $D^\times$ at great time intervals $\tau_{D^\times}$.

In the reverse case of *saturation* ($n_{D^+} \gg n_{D^\times}$) we have

$$\tau_{D^+} \gg \tau_{D^\times} \qquad \text{(IX.4.09)}$$

---

[1] Equation (IX.4.05) used for the derivation of (IX.3.01) was obtained by considering the case of equilibrium; however, (IX.4.05) is a relationship between concentration-independent constants so that its limitation to the case of equilibrium is not even possible.

In conclusion, it may be mentioned that a change from a state of stronger association to a state of weaker association ($n_{D^\times}$ decreasing) is obtainable not only by enhanced dissociation. It can also happen that the association falls far below the value required for the preservation of the instantaneous state so that an unchanged or only slightly reduced dissociation is no longer fully compensated. This remark will be of importance at the end of §5.

§5. The Determining Time Constant in Nonstationary Processes

Owing to the finite lifetimes $\tau_{D^\times}$ and τ_{D^+}, the concentrations $n_{D^\times}$ and n_{D^+} cannot, after external interference, adjust themselves instantaneously to their new equilibrium values. The resulting inertia of impurity reactions can be of interest, for instance, for the high-frequency behavior of rectifier boundary layers.[1] The relevant time constant in this case is obtained by integration of the differential equations (IX.4.02) and (IX.4.03).

We carry out the integration in a particularly simple typical case (see Fig. IX.5.1): The electron concentration $n_\ominus$, after having varied in an irregular fashion for some time so that an equilibrium could not be established, retains, beginning at time $t = 0$, a certain value $n_\ominus$ (see Fig. IX.5.1):

$$n_\ominus = \text{const} \qquad \text{for } t > 0 \tag{IX.5.01}$$

Then, in the course of time, an impurity equilibrium will be established which is determined by the following two conditions:

1. (Death rate of the $D^+ \equiv$) birth rate of the $D^\times$ = death rate of the $D^\times$

$$\frac{1}{\tau_{D^+}} n_{D^+} = \frac{1}{\tau_{D^\times}} n_{D^\times} \tag{IX.5.02}$$

2. Total number of donors = const

$$n_{D^+} + n_{D^\times} = n_D \tag{IX.5.03}$$

From the foregoing result the equilibrium concentrations which are reached asymptotically for $t \to \infty$:

$$n_{D^\times}(\infty) = \frac{\tau_{D^\times}}{\tau_{D^+} + \tau_{D^\times}} n_D \tag{IX.5.04}$$

$$n_{D^+}(\infty) = \frac{\tau_{D^+}}{\tau_{D^+} + \tau_{D^\times}} n_D \tag{IX.5.05}$$

---

[1] See W. Schottky, *Z. Physik*, **132:** 261 (1952).

During the process of adjustment the concentrations $n_{D^\times}$ and n_{D^+} deviate from these equilibrium values by slowly decreasing amounts $\Delta(t)$, one toward higher values, the other toward lower values, so that at any time t the sum of all donors according to (IX.5.03) is equal to n_D:

$$n_{D^\times}(t) = n_{D^\times}(\infty) + \Delta(t) \qquad \text{(IX.5.06)}$$

$$n_{D^+}(t) = n_{D^+}(\infty) - \Delta(t) \qquad \text{(IX.5.07)}$$

With the agreed premises, the differential equations (IX.4.02) and (IX.4.03) which govern the process of adjustment are simplified because (IX.5.01) leads to a time-independent[1] $r_D n_\ominus = 1/\tau_{D^+}$ while $1/\tau_{D^\times}$ is time-independent anyway:

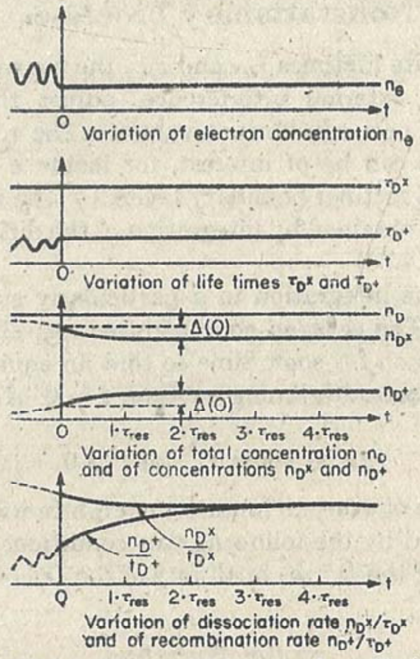


FIG. IX.5.1. The establishment of equilibrium.

$$\frac{dn_{D^\times}}{dt} = -\frac{1}{\tau_{D^\times}} \cdot n_{D^\times}(t) + \frac{1}{\tau_{D^+}} \cdot n_{D^+}(t) \qquad \text{(IX.5.08)}$$

with $\tau_{D^\times} = \text{const}_1$ and $\tau_{D^+} = \text{const}_2$

$$\frac{dn_{D^+}}{dt} = -\frac{1}{\tau_{D^+}} \cdot n_{D^+}(t) + \frac{1}{\tau_{D^\times}} \cdot n_{D^\times}(t) \qquad \text{(IX.5.09)}$$

If we now insert (IX.5.06) and (IX.5.07) and bear in mind (IX.5.02),

---

[1] This eliminates the objections to the use of τ_{D^+} mentioned at the end of §2.

we obtain for the deviation $\Delta(t)$ from (IX.5.08) as well as from (IX.5.09) the same differential equation:

$$\dot{\Delta}(t) = -\left(\frac{1}{\tau_{D^\times}} + \frac{1}{\tau_{D^+}}\right) \cdot \Delta(t) = -\frac{1}{\tau_{\mathrm{res}}}\,\Delta(t) \qquad \text{(IX.5.10)}$$

Here τ_{res} is a resultant time constant which according to

$$\frac{1}{\tau_{\mathrm{res}}} = \frac{1}{\tau_{D^\times}} + \frac{1}{\tau_{D^+}} \qquad \text{(IX.5.11)}$$

is

$$\tau_{\mathrm{res}} = \frac{\tau_{D^+} \cdot \tau_{D^\times}}{\tau_{D^+} + \tau_{D^\times}} \qquad \text{(IX.5.12)}$$

Integration of (IX.5.10) gives

$$\Delta(t) = \Delta(0) \cdot e^{-\frac{t}{\tau_{\mathrm{res}}}} \qquad \text{(IX.5.13)}$$

and thus we obtain from (IX.5.06) and (IX.5.07) and using (IX.5.04) and (IX.5.05) finally

$$n_{D^\times}(t) = \frac{\tau_{D^\times}}{\tau_{D^+} + \tau_{D^\times}}\, n_D + \Delta(0) \cdot e^{-\frac{t}{\tau_{\mathrm{res}}}} \qquad \text{(IX.5.14)}$$

$$n_{D^+}(t) = \frac{\tau_{D^+}}{\tau_{D^+} + \tau_{D^\times}}\, n_D - \Delta(0) \cdot e^{-\frac{t}{\tau_{\mathrm{res}}}} \qquad \text{(IX.5.15)}$$

Hence the relevant time constant τ_{res} is given by Eq. (IX.5.11). If, for instance, we have reserve during the whole process ($n_{D^+} \ll n_{D^\times}$), we have according to (IX.5.02) $\tau_{D^+} \ll \tau_{D^\times}$ and according to (IX.5.12)

$$\tau_{\mathrm{res}} \approx \tau_{D^+} \ll \tau_{D^\times} \qquad \text{(IX.5.16)}$$
$$(\text{Reserve } n_{D^\times} \gg n_D)$$

Inversely in the case of saturation

$$\tau_{\mathrm{res}} \approx \tau_{D^\times} \ll \tau_{D^+} \qquad \text{(IX.5.17)}$$
$$(\text{Saturation } n_{D^\times} \ll n_{D^+})$$

The resultant time constant is, therefore, always practically equal to the *smaller* one of the two time constants τ_{D^+} and $\tau_{D^\times}$.

This result may look surprising at first, but a physical understanding is obtained if we compute the number of dissociations and associations in unit time and unit volume. From (IX.3.02), (IX.5.06), and (IX.5.13) and from (IX.2.01) and (IX.2.05), (IX.5.07) and (IX.5.13), we obtain

$$\frac{\text{Number of dissociations}}{(\text{Unit time}) \cdot (\text{unit volume})} = \frac{n_{D^\times}(t)}{\tau_{D^\times}} = \frac{n_{D^\times}(\infty)}{\tau_{D^\times}} + \frac{\Delta(0)}{\tau_{D^\times}}\, e^{-\frac{t}{\tau_{\mathrm{res}}}} \qquad \text{(IX.5.18)}$$

$$\frac{\text{Number of associations}}{(\text{Unit time}) \cdot (\text{unit volume})} = \frac{n_{D^+}(t)}{\tau_{D^+}} = \frac{n_{D^+}(\infty)}{\tau_{D^+}} - \frac{\Delta(0)}{\tau_{D^+}} e^{-\frac{t}{\tau_{\text{res}}}} \tag{IX.5.19}$$

According to (IX.5.02), the first time-independent terms $n_{D^\times}(\infty)/\tau_{D^\times}$ and $n_{D^+}(\infty)/\tau_{D^+}$ represent in each case the final equilibrium rates. The final state of equilibrium is reached by the presence and the combined effect of the exponential terms in both equations, the determining exponential term being the one with the greater amplitude. Since the numerators of the amplitudes are equal, only the denominators $\tau_{D^\times}$ and τ_{D^+} are of importance. Therefore the term with the *smaller* time constant has the larger amplitude. This makes it also plausible that this shorter lifetime dominates the course of the whole process (see Fig. IX.5.1).

Finally a word about the orders of magnitude of $\tau_{D^\times}$ and τ_{D^+} with which we have to reckon. Both terms contain the recombination coefficient r_D or the effective cross section σ_{D^+}, respectively. Using $\sigma_{D^+} \approx (3 \cdot 10^{-8}\text{ cm})^2$ (IX.2.02) and (VIII.4.27) would lead to

$$r_D = 10^{+7}\text{ cm-sec}^{-1} \cdot 10^{-15}\text{ cm}^2 = 10^{-8}\text{ cm}^3\text{ sec}^{-1}$$

With $n_\ominus = 10^{16}\text{ cm}^{-3}$, we would then obtain for $\tau_{D^+} = 1/r_D n_\ominus$

$$\tau_{D^+} = \frac{1}{10^{-8}\text{ cm}^3\text{ sec}^{-1} \cdot 10^{16}\text{ cm}^{-3}} = 10^{-8}\text{ sec}$$

Assuming, for example, a dissociation energy $E_{CD} = 0.2$ ev for the donors, we obtain

$$\tau_{D^\times} = \frac{2}{r_D N_C e^{-\frac{E_{CD}}{kT}}} = \frac{2}{10^{-8}\text{ cm}^3\text{ sec}^{-1} \cdot 2.5 \cdot 10^{19}\text{ cm}^{-3}} e^{+8}$$

$$= \frac{2}{2.5 \cdot 10^{11}\text{ sec}^{-1}} 3 \cdot 10^3 \approx 2 \cdot 10^{-8}\text{ sec}$$

However, it has been mentioned on page 327 that the assumption $\sigma_{D^+} = 10^{-15}\text{ cm}^2$ is completely arbitrary. Other authors estimate values for the effective cross section which are several orders of magnitude smaller and lead to a corresponding increase of the lifetimes[1] $\tau_{D^\times}$ and τ_{D^+}.

§6. Problems (Recombination)

The recombination of injected electron-hole pairs is of great importance in semiconductor physics and technology. This recombination may take place by the electron dropping into the valence band across the forbidden band, releasing the energy difference in form of a photon or a phonon. In germanium,

---

[1] See W. Schottky, *Z. Physik*, **132**: 261 (1952), particularly p. 276.

this process would lead to a carrier lifetime of roughly 1 sec. It cannot, therefore, account for the much smaller observed lifetimes of the order of 10^{-7} to 10^{-3} sec. The recombination is greatly enhanced, however, if the electron first falls into an impurity level within the forbidden band and from there into the valence band.

We assume that the number of electrons falling into the impurity level is proportional to the number of electrons present in the conduction band and to the probability that the impurity levels are empty. Furthermore, the number of electrons returning to the conduction band is assumed to be proportional to the number of states in that band and to the probability that the impurity levels are occupied. If the latter probability is f_T, we therefore have

$$\frac{dn}{dt} = -\frac{n(1 - f_T)}{\tau_n} + \frac{N_C f_T}{\tau_n'} \tag{IX.6.01}$$

where τ_n and τ_n' are—as yet undetermined—constants with the dimensions of a lifetime. These constants are related to the number of the impurity centers and to their capture cross section.

An equation analogous to (IX.6.01) can be derived for the change of the hole density, introducing two additional constants, τ_p and τ_p':

$$\frac{dp}{dt} = -\frac{pf_T}{\tau_p} + \frac{N_V(1 - f_T)}{\tau_p'} \tag{XI.6.02}$$

1.* Derive a relationship between τ_n and τ_n' and between τ_p and τ_p', utilizing the fact that in thermal equilibrium no net changes take place. Assuming that the recombination center density is small, so that no carriers can be stored in the recombination centers, eliminate f_T and show that the net recombination rate is given by an expression of the form

$$\frac{dn}{dt} = \frac{dp}{dt} = -\frac{np - n_i^2}{(n + n^*)\tau_p + (p + p^*)\tau_n} \tag{IX.6.03}$$

What is the meaning of the quantities n^* and p^*?

2. Define from Eqs. (IX.1.01) and (IX.6.03) a lifetime for the excess carriers. Give the limiting values for low and for high injection levels. Under what condition do they coincide?

3. How does the low-level lifetime vary with the position of the recombination level inside the forbidden band, with the simplifying assumption that τ_n and τ_p do not vary with this position? How does this affect the recombination activity of impurity levels close to the band edges, which determine the conductivity, in comparison with levels deep inside the forbidden band which have little influence upon the conductivity?

4. What is the hole lifetime in an n-type germanium crystal with a resistivity of 10 ohm-cm if one assumes a recombination level 0.22 ev above the valence band and the values $\tau_n = 100$ μsec, $\tau_p = 10$ μsec? How does this lifetime depend on temperature if the τ's are temperature-independent?

5. Answer the questions of Prob. 4 for a p-type germanium crystal with a resistivity of 0.1 ohm-cm.

6.* At what conductivity has the low-level lifetime in a semiconductor its maximum, if one assumes that the τ's do not depend on the concentration of conductivity-determining impurities? Give numerical values for germanium, assuming $\tau_n:\tau_p = 10:1$, $p^* = 300n_i$, $n^* = \frac{1}{300}n_i$.

CHAPTER X

Boundary Layers in Semiconductors and the Metal-Semiconductor Contact

§1. The Electrostatic Macropotential and the Energy of an Electron in a Solid

When the potential inside a solid is mentioned, we think nowadays automatically of the periodic potential U in Fig. I.2.2. We have become so accustomed to the atomistic point of view in this connection that we base our thoughts more or less automatically on the microscopic scale and have lost the more primitive viewpoint of a macroscopic continuum.

As long as the conditions in the solid under consideration are macroscopically homogeneous, this atomistic treatment is satisfactory. However, the present chapter is concerned with phenomena at the surfaces and in the boundary layers of semiconductors, and here it is just the deviation from the macroscopic homogeneity, existing inside the solid, which is characteristic. Therefore we have to supplement the results previously obtained, by means of the atomistic treatment, with macroscopic considerations. For this purpose we shall, first of all, consider the energy of an electron within a solid quite simply from this macroscopic viewpoint of a continuum.

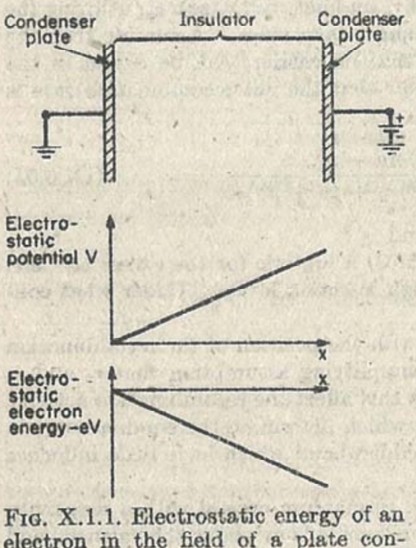


FIG. X.1.1. Electrostatic energy of an electron in the field of a plate condenser.

Consider an insulator as the dielectric between the plates of a charged plate condenser (Fig. X.1.1). Then, by way of example, we have high potential values on the right side of this insulator and low values on the left. Hence, because of its negative charge, an electron has a low electrostatic energy on the right side and a high one on the left. This statement at first contradicts the position-*in*dependent[1] energy values in the level schemes of Figs. I.3.6 and I.3.8 to I.3.11. The contradiction is caused by the fact that the representation in Chap. I does not concern the total energy of an electron although it was specifically a plot of the sum of kinetic and potential energy.[2] Even so, it would have been incorrect to call these energy values total energies, because only the microscopic aspect of the problem was considered, while macroscopic forces such as the condenser field in the preceding example were neglected.

To describe the previous situation, we could also say that in the level diagrams of Figs. I.3.6 and I.3.8 to I.3.11, only the electron energy E within the lattice was plotted, this energy being the result of the binding forces exerted by the crystal lattice on the electron.[3] To obtain the total energy $\mathbf{E}$ of an electron, an electric energy term $-eV$ has to be added to the lattice energy E, this additional energy being due to a macropotential V:

$$\mathbf{E} = E - eV \qquad \text{(X.1.01)}$$

The sources of the macropotential V are macroscopic surface or space charges, double layers, etc.[4] The following relationships show the kinetic and potential character,[5] respectively, of the lattice energy

---

[1] See p. 21 concerning the independence of position of the energy levels of the undisturbed lattice in contrast to the dependence on position of donor and acceptor levels.

[2] See footnote 3, p. 7.

[3] These binding forces are sometimes called chemical binding forces, and correspondingly the lattice energy is called chemical binding energy. These terms are used to emphasize that the chemical characteristics of the solid under consideration affect this component of the energy in particular.

[4] It appears at least doubtful whether in all possible cases such a subdivision of the total energy into electrostatic energy and lattice energy is unique. C. Herring and M. H. Nichols, in *Revs. Mod. Phys.*, **21**: 185–270 (1949), define the electrostatic macropotential by ascertaining the mean value of the electrostatic micropotential. W. Schottky, in *Physik regelmäss. Ber.*, **3**: 17 (1935), particularly in footnote 1, page 19, has introduced a so-called *Leerraumpotential* to define the electrostatic macropotential accurately. We believe that in the boundary layers of semiconductors, to be discussed later, there can be no doubt how one has to formulate the electrostatic energy $-eV$ of the semiconductor electrons and how great their lattice energy E is.

[5] In quantum mechanics, too, a subdivision of the energy according to

$$E = E_{\text{kin}} + E_{\text{pot}}$$

is possible. Here, however, E_{kin} and E_{pot} have to be considered as quantum-

E and of the electrostatic energy portion $-eV$:

$$\boldsymbol{E} = E_{\text{kin}} + \boldsymbol{E}_{\text{pot}} \qquad \text{(X.1.02)}$$

$$= \underbrace{E_{\text{kin}} + \overbrace{E_{\text{pot}}}^{}}_{}\overbrace{{} - eV}^{} \qquad \text{(X.1.03)}$$

$$\boldsymbol{E} = E - eV \qquad \text{(X.1.04)}$$

The lattice energy E has, therefore, partly kinetic[1] and partly potential character,[2] while the electrostatic portion of the energy is solely potential.

It need not worry us that the total energy of an electron is known only apart from an additive constant, i.e., that the zero level of the electron has to be fixed arbitrarily. It is only important, in considering two different solids, such as a metal and a semiconductor, that the total energy of the electrons in both solids be referred to the same zero level. Naturally, nothing can be said about the sign of the electrostatic energy produced by the macropotential. However, the contributions of the lattice energy in the various bands are negative because they have the nature of binding energies. Thus we obtain, by way of example, the energy diagram of Fig. X.1.2[3] for a metal in

---

mechanical mean values, because the operators of the kinetic and of the potential energy are not, as a rule, interchangeable with the Hamilton operator $(E_{\text{kin}} + E_{\text{pot}})_{\text{op}}$. Hence, a ψ function cannot be the eigenfunction of all three operators simultaneously. Usually, one considers a *stationary* solution of the Schrödinger equation, i.e., an eigenfunction of the Hamilton operator $(E_{\text{kin}} + E_{\text{pot}})_{\text{op}}$. Then the electron represented by such a ψ function has only a sharply defined lattice energy E. Kinetic energy E_{kin} and potential energy E_{pot} can be given only as quantum-mechanical mean values.

[1] This applies also to the lowest and the highest states of each band, though in these the ψ function has the character of a standing wave, and one may therefore be tempted, without justification, to assume the kinetic energy to be zero at these band edges. See also p. 206, footnote 1.

[2] At first sight it may seem strange that a chemical binding energy is not purely potential in character but contains also a kinetic portion. Yet, in the case of the single atom, we do not object to equating the binding energy of an electron to the atom core to the ionization energy of the electron concerned. However, if we fix the zero point in the usual way (the separated stationary electron having the energy zero), the binding energy will be equal to the energy of the electron in the appropriate atomic state and therefore has, by way of example, in the ground state of the H atom a potential portion $-Z^2e^2/n^2a_0$ and a kinetic portion $+\frac{1}{2}Z^2e^2/n^2a_0$, Z = atomic number = 1 for the H atom, n = principal quantum number = 1 for the ground state, a_0 = radius of the first Bohr orbit. [See also H. A. Bethe in H. Geiger and K. Scheel, "Handbuch der Physik," 2d ed., vol. XXV, part 1, p. 287, Eq. (3.29), Springer-Verlag OHG, Berlin, 1933.] Thus in this case, too, the chemical binding energy contains a considerable kinetic portion.

[3] In the figures of this chapter, the shading indicates only the presence of a

which, because of the extremely high conductivity, the same macropotential V prevails throughout.

On the other hand, in a semiconductor the electrostatic macropotential V need not by any means have the same value everywhere. For instance, the electrostatic energy can decrease or increase in layers several microns in thickness near the surface of a semiconductor.[1] In fact, this is usually the case, as will be discussed later in §6 and §7. At the moment, we are mainly concerned to prove that such a bending of the potential curve carries along the whole remaining energy diagram. This is so because the chemical binding forces, being deter-

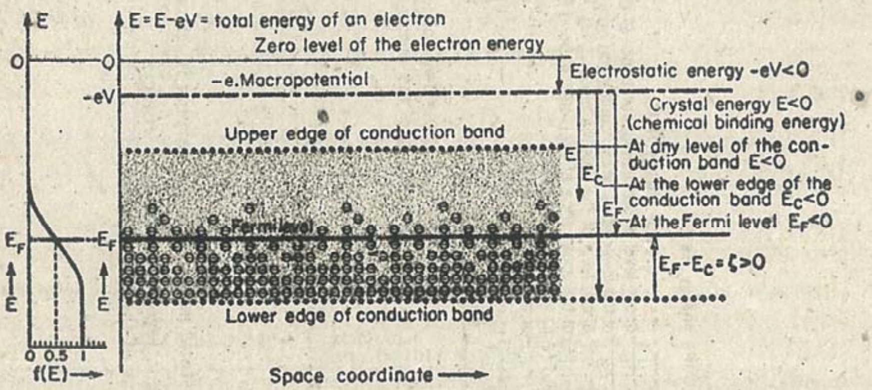


FIG. X.1.2. Band model of a metal.

$\mathbf{E}$ = total energy of an electron, counted from zero level of the electron energy,
E = crystal energy of an electron, counted from the macropotential,
$-eV$ = electrostatic energy of an electron.

---

continuum of states in which electrons can be accommodated, not the density of actual occupancy, in contrast to the figures in Chap. VIII.

An upper edge of the conduction band has been drawn in Fig. X.1.2 mainly for didactic reasons. The use of the Wigner-Seitz cellular method (see p. 200) has shown that in solids, such as Na, K, etc., the conduction band and the bands above it always overlap so that one cannot speak any longer of an upper edge of the conduction band. Therefore, an upper edge of the conduction band is omitted in the subsequent figures.

[1] Of course in such a case the additive composition of electrostatic energy $-eV$ and of lattice energy E can possess only the character of an approximation. The expression

$$\text{Total potential} = \text{periodic lattice potential } U + \text{macropotential } V$$

is now no longer entirely periodic. Hence, we have discarded in principle the basis for the computation of a particular lattice energy. However, if the local variation of the macropotential V is sufficiently small, the method described represents a good approximation. A more detailed discussion of this problem will be found in B. Kockel, *Z. Naturforsch.*, 7A: 10–16 (1952).

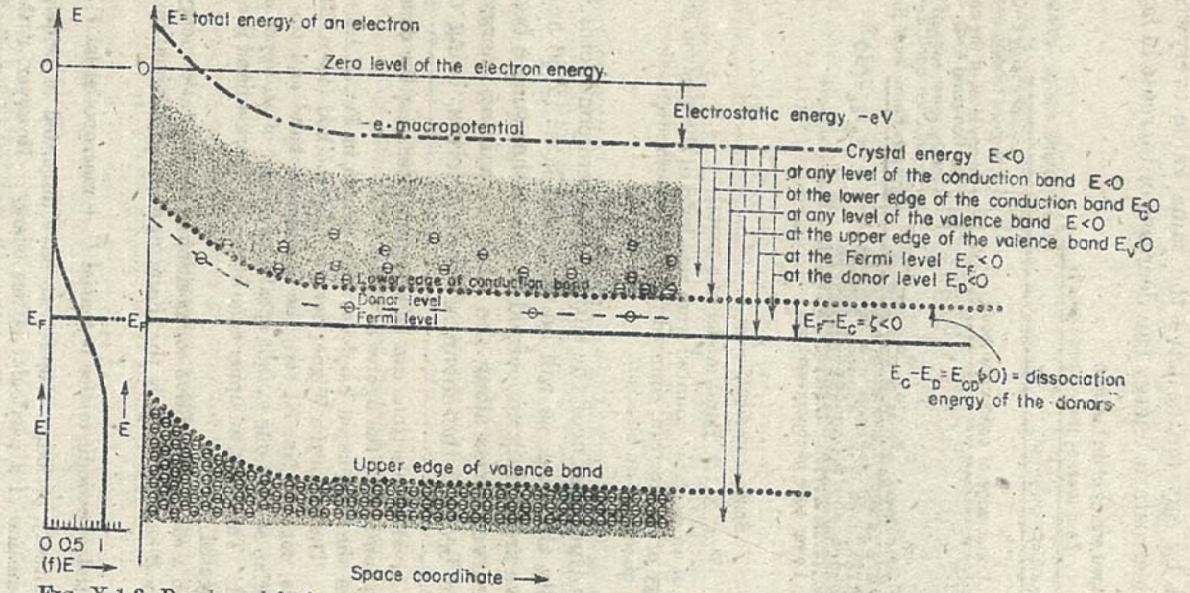


FIG. X.1.3. Band model of an impurity semiconductor with electrostatic macropotential varying in space.

E = total energy of an electron, referred to universal zero level of the electron energy.
E = crystal energy of an electron, referred to local value of the macropotential.
$-eV$ = electrostatic energy of an electron.

mined by the crystal lattice itself, have the same effect throughout the semiconductor; thus the binding portion E of each individual lattice level, for instance, E_C for the lower edge of the conduction band, has the same value throughout the semiconductor. The same applies to the upper edge of the valence band and for the impurity levels. Thus we obtain, for instance, Fig. X.1.3.

Finally, no binding forces whatever act in vacuum:

$$E = 0 \tag{X.1.05}$$

Apart from the electrostatic energy $-eV(x)$ as potential energy and the kinetic energy $1/2m|\mathbf{p}|^2$, no other energy term need be considered

$$\boldsymbol{E} = \frac{1}{2m}|\mathbf{p}|^2 - eV(x) \tag{X.1.06}$$

Comparison with (VIII.1.01)[1] shows that at point x the electron gas can be considered as being in a potential well of depth $-eV(x)$, and from (VIII.1.08) we can derive for the Fermi level

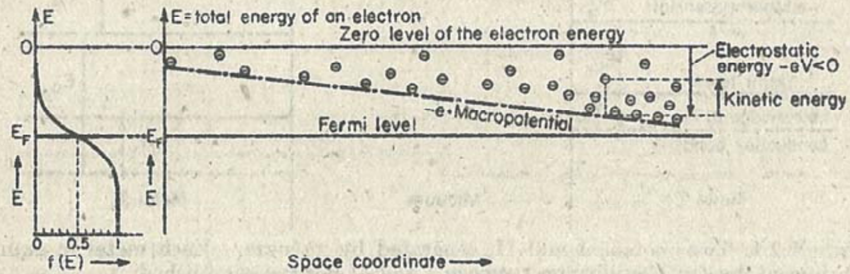


FIG. X.1.4. Energy diagram in vacuum. Linear variation of the electrostatic potential. The concentration of the electron gas is assumed to be so small that the potential curve does not noticeably deviate from a straight line.

$$\boldsymbol{E}_F = \zeta\left(\frac{n(x)}{N}\right) - eV(x) \tag{X.1.07}$$

Figure X.1.4 shows an $(\boldsymbol{E},x)$ representation corresponding to Figs. X.1.2 and X.1.3 for the case of a uniform potential rise, or a constant electric field,[2] in the vacuum region under consideration.

---

[1] In Chap. VIII, §1, the potential well is considered as the simplest model of a solid. Therefore E_{pot} in this case has the character of a chemical binding energy and E is used rather than $\boldsymbol{E}$. If we wished to include a macropotential V in this simple model, the potential energy of the electrons outside the well would not be zero but would be $-eV$. Correspondingly, the potential energy within the well would not be E_{pot} but $E_{pot} - eV(x)$. However, the consideration of a macropotential $V(x)$ is introduced only in the present Chap. X.

[2] Figures X.2.4 and X.3.1 show an experimental arrangement in which a linear potential variation is achieved at thermal equilibrium.

§2. Thermal Equilibrium between Two Metals. The Galvani Voltage

In Figs. X.1.2 to X.1.4 the Fermi level is drawn horizontal and the same occupancy function is used in all parts of the volume (hence in all layers in the plane problem under consideration). This is only a special application of the law concerning the thermal equilibrium between different phases which was derived on pp. 288 to 292. The application of this law acquires special significance if we now turn to the consideration of the thermal equilibrium between different solids, for example, between two metals which are in close contact. To begin with, Fig. X.2.1 shows the two metals separated. Thermal equilibrium exists within each metal, so that the Fermi level is horizontal within each individual metal. We assume that thermal equilibrium has not yet been established between the two metals, and hence the Fermi levels are of different height in the two metals.[1] If the two metals are now brought into intimate contact (Fig. X.2.2), thermal equilibrium is established between them and the Fermi levels must adjust themselves to the same height.

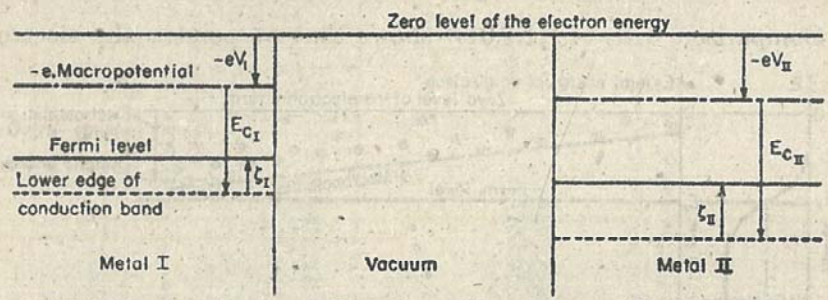


FIG. X.2.1. Two metals, I and II, separated by vacuum. Each metal in equilibrium in itself. Equilibrium between I and II not yet established.

The question arises how this is achieved. The distances E_{C_I} and $E_{C_{II}}$ between the edges of the conduction bands (drawn as dotted lines) and the dot-dash lines representing the electrostatic energy are chemical binding energies of an electron in the lowest levels of the two conduction bands and are thus fixed by the respective crystal properties. The heights ζ_I and ζ_{II} of the Fermi levels, drawn as heavy continuous lines, above the lower edges of the conduction bands, drawn as dotted lines, are fixed by the electron concentrations n_I and n_{II} [see Eq.

---

[1] As a *non*equilibrium state this state is naturally largely dependent on the previous history. Hence the relative position of the Fermi levels in the two metals is essentially arbitrary.

(VIII.4.07)]. None of these values is changed by making contact between the two metals. Hence, only the electrostatic energies are left for the adjustment of the Fermi levels, i.e., the distances $-eV_{\mathrm{I}}$ and $-eV_{\mathrm{II}}$, respectively, between the common zero level of the electron energy and the dot-dash lines. Therefore, the macropotentials V_{I} and V_{II} in the two metals I and II must assume a definite difference relative to each other so that the Fermi level may be at the same height in both metals. The question arises as to what actual physical process brings this about, in other words, what happens when contact is made between the two metals, i.e., during the transition from Fig. X.2.1 to X.2.2. A certain number of electrons from the uppermost atomic layer of one metal passes into the uppermost atomic layer of the other metal. Thus the surface of one metal is charged positive, the surface of the other metal is charged negative, and the resulting double layer produces a jump between the two macropotentials, i.e., the two dot-dash lines are displaced relative to each other. This process continues until the resulting potential gradient suffices to establish the Fermi levels at the same height, i.e., until thermal equilibrium is reached. The potential difference between the two internal values of the macropotential, which is characteristic for the two metals, is called the Galvani voltage.

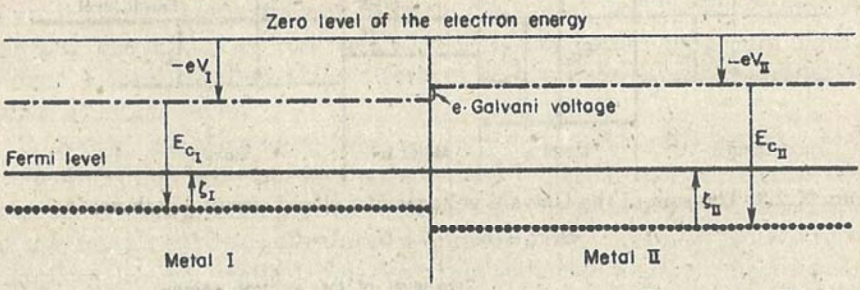


FIG. X.2.2. Two metals, I and II, in intimate contact. Thermal equilibrium. Difference $V_{\mathrm{II}} - V_{\mathrm{I}}$ between the two macropotentials = Galvani voltage.

If three or more metals at the same temperature are combined into a closed circuit and if no current flows through this circuit, i.e., if all these metals are in thermal equilibrium (see Fig. X.2.3), the Fermi level, drawn in heavy print, must be at the same height in all metals. The values of the electrostatic energy $-eV_{\mathrm{I}}$, $-eV_{\mathrm{II}}$, $-eV_{\mathrm{III}}$ are not fixed, but the heights $|E_{C\mathrm{I}}| - \zeta_{\mathrm{I}}$, $|E_{C\mathrm{II}}| - \zeta_{\mathrm{II}}$, $|E_{C\mathrm{III}}| - \zeta_{\mathrm{III}}$ of the electrostatic energy, in dot-dash lines, above the Fermi level, in heavy print, are determined by the lattice properties. Therefore, after moving through the circuit from left to right, we return in metal I to the same level of the dot-dash electrostatic energy and we can see that the sum of all Galvani voltages $G_{\text{II I}} + G_{\text{III II}} + G_{\text{I III}}$ in traversing a conducting loop must be exactly zero. We shall use this result below (see page 361).

Finally, it is known that hot metal surfaces emit electrons into the vacuum. We have to assume such an electron emission even for metals at normal temperature although it will be orders of magnitude

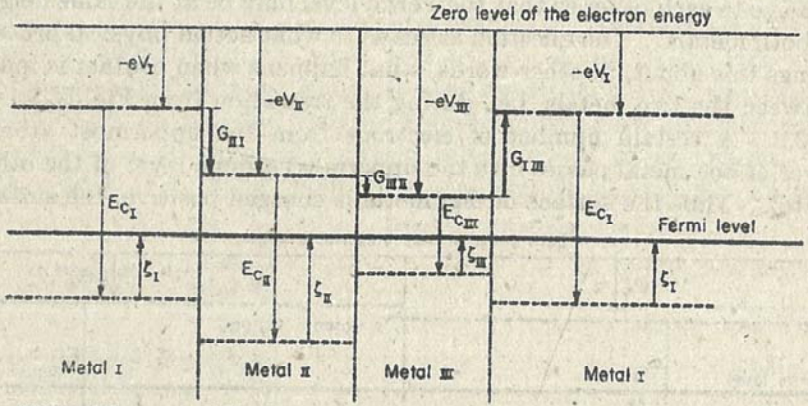


FIG. X.2.3. The sum of the Galvani voltages of a closed circuit equals zero:

$$G_{\text{II I}} + G_{\text{III II}} + G_{\text{I III}} = 0$$

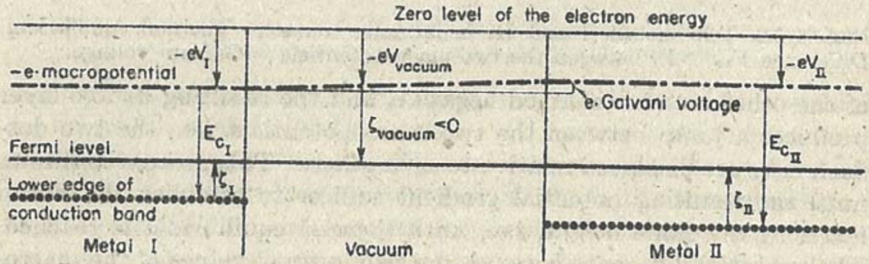


FIG. X.2.4. Two metals, I and II, separated by vacuum. Thermal equilibrium between I and II. No atomic double layers at the surfaces. Galvani voltage = difference of the internal values of the macropotential.

smaller than for hot surfaces. Therefore, even between the spatially separated metals of Fig. X.2.1, electrons will be exchanged, and this exchange will be stronger in one direction than in the other until thermal equilibrium is established, i.e., until the correct potential difference between the macropotentials V_I and V_{II} has been established and the Fermi levels have thus been brought to the same height (Fig. X.2.4).[1] Here the positively or negatively charged metal surfaces are

---

[1] In the case of hot surfaces this will occur relatively soon, but with cold surfaces it will take an extremely long time.

not in close proximity but cause an electric field right across the vacuum, corresponding to a linear potential gradient through the vacuum in Fig. X.2.4.

§3. Surface Double Layers. The Volta Potential Difference (Contact Potential)

The situation represented in Fig. X.2.4 by no means corresponds to reality, because at the boundary between each metal and the vacuum we have to assume a spontaneous double layer which can be due to many causes.

1. Within the metal the electrons are in vigorous thermal random motion. The electrons moving toward the surface of the metal, even if their energy is insufficient to leave the metal, will move somewhat beyond the lattice of the fixed positive ions before they turn back. Hence a thin negative skin is formed above a thin positive skin, in brief, a double layer.

2. At the metal-vacuum boundary a lattice exists only on one side, while the other side borders on essentially empty space. Hence the ions of the uppermost atomic layer are, by comparison with the bulk of the metal, subject only from one side to the "proper" forces of a complete lattice, while from the other side the near vacuum exerts practically no, i.e., "wrong," forces. Therefore the ions of the uppermost atomic layers take up displaced positions relative to the configuration in the bulk of the metal. Since quasi-neutrality prevails in the bulk of the metal, charges will result at the surface of the metal. However, neutrality must obtain in the total system, and therefore a double layer results.

3. A further contribution to the spontaneous double layer can be supplied, apart from the ion displacements discussed under 2, by ion deformations, i.e., polarizations of the ion cores which do not occur in the uppermost atomic layers in the same way as in the bulk of the metal because of the "wrong" forces of the vacuum.

4. Finally, we must also recall the well-known phenomena of monatomic impurity layers on metal surfaces. We may mention, by way of example, the problem of surface layers of thorium or cesium or other elements on tungsten which has been investigated in detail by Langmuir and his school. Here the adsorbed foreign atoms are pulled apart to form dipoles, which explains their adhesion as well as their double-layer effect.

Hence, in place of the continuous change of the electrostatic potential at the metal surfaces in Fig. X.2.4, we must assume a potential jump of unknown magnitude arising from spontaneous double layers (Fig. X.3.1). As a result the difference of the surface potentials is no longer identical with the difference of the macropotentials within the two metals as in Fig. X.2.4. Therefore we need a new name for this difference of the surface potentials which is usually different from the Galvani voltage. We call this potential difference Volta potential difference or contact potential.[1]

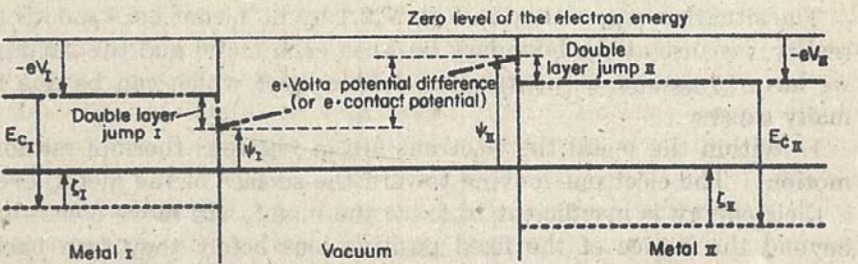


FIG. X.3.1. Two metals, I and II, with atomic double layers at the surfaces. Difference of the external surface potentials = Volta potential difference or contact potential. The figure indicates the validity of the statement: Volta potential difference (or contact potential) = difference of the vacuum work functions.

In contrast to the Galvani voltage, the contact potential can be determined directly by experiment. Here we mention only the case of the vacuum tube where the contact potential between cathode and grid has to be considered in the determination of the effective grid potential.

§4. The Work Function and the Photoelectric Activation Energy of Metals

Whereas the Fermi gas of electrons in metals is degenerate because of the extremely high electron concentration ($\sim 10^{22}$ cm^{-3}), the electron concentrations in vacuum are so low that the Maxwell-Boltzmann special case of Fermi statistics obtains. This, because of $n \ll N$, simplifies Eq. (X.1.07) in accordance with Eq. (A.II.2) to

$$E_F = \zeta - eV \approx kT \ln \frac{n}{N} - eV \qquad \text{(X.4.01)}$$

We apply this equation to the situation in the vacuum immediately

---

[1] We doubt whether the name contact potential is a happy choice. Experience has shown that the word "contact" in this name distracts attention from the fact that we deal with a potential difference between two free surfaces which are *not* in contact with each other but are separated by a large distance. The fact that the two solids are in intimate contact somewhere else only ensures thermal equilibrium between them and could in principle be replaced by waiting for a sufficiently long time if the solids were separated.

in front of the surface of metal I, where ζ_I is negative. We replace this by introducing the positive work function

$$\Psi_I = -\zeta_I = -(E_F + eV_I) = -eV_I - E_F$$

(see Fig. X.3.1) and obtain

$$n_{\text{vacuum}} = N\,\mathrm{e}^{-\frac{\Psi_I}{kT}} \tag{X.4.02}$$

Here we have assumed thermal equilibrium. Hence the number of electrons leaving the metal surface in unit time exactly equals the number of electrons returning to the metal surface in unit time. A unidirectional thermal current away from the metal surface is just compensated by a unidirectional thermal current in the opposite direction.

Of much greater importance experimentally are two deviations from thermal equilibrium: (1) that in which the electrode facing the metal surface under consideration has a much lower temperature than this surface and hence practically does not emit at all and (2) that in which the electric field between the two metals can be altered arbitrarily with a battery. In contrast to the case of thermal equilibrium we can now arrange, by applying the proper polarity, that fewer electrons return from the vacuum into the metal than leave the metal surface. In other words, an emission current can be drawn from the hot metal surface. Now, the larger the value of the suitably directed electric field is made, the smaller will be the number of electrons returning to the metal surface and the larger will be the emission current until beyond a certain field strength all electrons which have left the metal are carried away into the vacuum and no longer return to the metal surface. By further increase of the field, the emission current cannot, for the time being, be increased any further, i.e., we have saturation current.

Although this state of "complete stationary transport" and the state of thermal equilibrium differ tremendously and must be kept apart conceptually, a more detailed discussion[1] shows that the two cases are intimately related inasmuch as the saturation current, in the case of complete stationary transport, is equal to the unidirectional thermal current in the case of thermal equilibrium:

$$i_{\text{sat}} = \text{unidirectional thermal current density}$$

$$= e \cdot \int_{v_x=0}^{+\infty} \int_{v_y=-\infty}^{+\infty} \int_{v_z=-\infty}^{+\infty} v_x \cdot n(v_x v_y v_z) \cdot dv_x\, dv_y\, dv_z \tag{X.4.03}$$

---

[1] W. Schottky, Physik der Glühelektroden, in Wien and Harms, "Handbuch der Experimentalphysik," vol. 13, part 2, particularly pp. 31–42, Akademische Verlagsgesellschaft, Leipzig, 1928.

Since $n_{\text{vacuum}} \ll N$, we must here use the Boltzmann distribution for the distribution function $n(v_x v_y v_z)$. The former is derived from Eq. (VIII.1.12), in the form here required, by the following transformations:

$$N(E)\,dE = N \cdot \frac{2}{\sqrt{\pi}} \left(\frac{E - E_{\text{pot}}}{\text{k}T}\right)^{1/2} \mathrm{e}^{-\frac{E-E_{\text{pot}}}{\text{k}T}}\, d\left(\frac{E - E_{\text{pot}}}{\text{k}T}\right)$$

$$= N \cdot 2\pi^{-1/2} \cdot \left(\frac{m}{2\text{k}T}\right)^{+1/2} v\, \mathrm{e}^{-\frac{mv^2}{2\text{k}T}}\, \frac{m}{2\text{k}T} \cdot 2v\, dv$$

$$N(E)\,dE = N \cdot \pi^{-3/2} \left(\frac{m}{2\text{k}T}\right)^{+3/2} \mathrm{e}^{-\frac{mv^2}{2\text{k}T}} \cdot 4\pi v^2\, dv$$

With a transformation of the variables and simultaneous transition from the total numbers N to the densities n, we finally get

$$n(v_x v_y v_z)\, dv_x\, dv_y\, dv_z = n \cdot \pi^{-3/2} \left(\frac{m}{2\text{k}T}\right)^{+3/2} \cdot \mathrm{e}^{-\frac{m(v_x^2+v_y^2+v_z^2)}{2\text{k}T}} \cdot dv_x\, dv_y\, dv_z \tag{X.4.04}$$

Using this in (X.4.03) we obtain

$$i_{\text{sat}} = e \cdot n_{\text{vacuum}_{\text{I}}} \cdot \pi^{-3/2} \left(\frac{2\text{k}T}{m}\right)^{1/2} \cdot \int\limits_{u_x=0}^{\infty} \mathrm{e}^{-u_x^2} u_x\, du_x \cdot \int\limits_{u_y=-\infty}^{+\infty} \mathrm{e}^{-u_y^2}\, du_y \cdot \int\limits_{u_z=-\infty}^{+\infty} \mathrm{e}^{-u^2} du_z$$

$$i_{\text{sat}} = e \cdot n_{\text{vacuum}_{\text{I}}} \cdot \pi^{-3/2} \left(\frac{2\text{k}T}{m}\right)^{1/2} \cdot \frac{1}{2} \cdot \sqrt{\pi} \cdot \sqrt{\pi}$$

and with (X.4.02)

$$i_{\text{sat}} = \frac{1}{2} e \left(\frac{2\text{k}T}{\pi m}\right)^{1/2} \cdot N \cdot \mathrm{e}^{-\frac{\Psi_{\text{I}}}{\text{k}T}} \tag{X.4.05}$$

Using (VIII.1.04), we finally obtain

$$i_{\text{sat}} = \frac{4\pi e m \text{k}^2}{h^3} T^2\, \mathrm{e}^{-\frac{\Psi_{\text{I}}}{\text{k}T}} = AT^2\, \mathrm{e}^{-\frac{\Psi_{\text{I}}}{\text{k}T}} = 120\, \frac{\text{amp}}{\text{cm}^2} \left(\frac{T}{\text{degree K}}\right)^2 \mathrm{e}^{-\frac{\Psi_{\text{I}}}{\text{k}T}} \tag{X.4.06}$$

This is the well-known Richardson law for the saturation current density from a hot surface.

The decisive term Ψ in the exponent of the Richardson law (X.4.06) is called the thermal metal–vacuum work function. We see from Fig. X.3.1 that in such a representation this thermal metal–vacuum work function appears as the distance between the Fermi level and the electrostatic potential immediately in front of the metal surface. From Fig. X.3.1 we can hence derive the important law:

The difference of the thermal work functions of two metals equals the relative Volta potential difference of these two metals.

Finally, for later application, we mention that according to Fig. X.3.1 we have[1]

$$\Psi = E_{C_{\text{metal}}} - \zeta_{\text{metal}} \pm (\text{double layer})_{\text{metal vacuum}} \qquad (\text{X}.4.07)$$

Means other than thermal excitation can be used to liberate electrons from the binding forces of the solid and to push them into the vacuum. In particular we must mention here the photoelectric process, where the incident photon has to impart to the electron an increase of energy which is at least equal to the difference between the energy of an electron bound in the solid and the energy of a stationary electron in the vacuum in front of the metal surface. Since the crystal states above the Fermi level are very sparsely occupied by comparison with the states below the Fermi level, the described process will occur with great frequency only if the energy $\hbar\omega$ of the light quantum is sufficient[2] to lift an electron from the Fermi level to the electrostatic surface potential. In the limiting case $T \rightarrow 0$ this "red limit"[3] of the photoeffect becomes quite sharp, and this minimum energy required, in the limiting case $T \rightarrow 0$, for the liberation of an electron by impact with a light quantum can simply be called *the* photoelectric activation energy. In metals this energy is equal to the thermal work function because it is equal to the difference between the electrostatic surface potential and the Fermi level.

The fact that this equality happens to obtain for metals and the fact that this subject has in the past usually been discussed only in relation to metals have led to a replacement of the somewhat abstract thermodynamic definition of the thermal work function as the decisive term in the exponent of a Richardson law, by the undoubtedly simpler definition of the photoelectric activation energy. However, we shall see in the following paragraph that this is not permissible in the case of semiconductors.

---

[1] On this occasion we must remember that the several terms in Eq. (X.4.07), and hence also the work function Ψ itself, cannot a priori be considered independent of temperature. As long as the temperature dependence of Ψ is not known exactly, (X.4.06) does not give definite information about the temperature dependence of the saturation current. See W. Schottky and H. Rothe, Physik der Glühelektroden, in Wien and Harms, *op. cit.*, vol. 13, part 2, particularly Chap. 6.

[2] ω = angular frequency $= 2\pi f$; $\hbar = (1/2\pi)h = 1.054 \cdot 10^{-27}$ cm^2 g-sec^{-1}.

[3] The condition $\omega > \omega_{\min}$ establishes a low-frequency, or long wavelength (i.e., "red" light), limit to the photoelectrically active spectrum.

§5. The Work Function and the Photoelectric Activation Energy of Semiconductors

If we now consider, instead of two metals, a metal and a semiconductor at a great distance from each other, we obtain Fig. X.5.1 when equilibrium is established.[1]

This means that an equilibrium concentration n of the electrons exists also in front of semiconductor surfaces.. Using the separation Ψ_{sem} of the Fermi level from the electrostatic potential we obtain, just as before for the metal, the equation

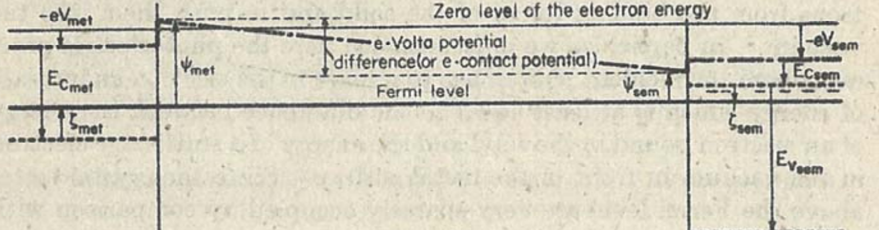


FIG. X.5.1. Metal and semiconductor widely separated by vacuum. Again the statement is valid: Volta potential difference (or contact potential) = difference of the vacuum work functions.

$$n = N\,\mathrm{e}^{-\frac{\Psi_{sem}}{kT}} \qquad (X.5.01)$$

corresponding to Eq. (X.4.02) whereupon we can then carry out the transition to the Richardson equation, just as in §4:

$$i_{sat} = \frac{1}{2}\,e\left(\frac{2kT}{\pi m}\right)^{1/2} N\cdot\mathrm{e}^{-\frac{\Psi_{sem}}{kT}} = \frac{4\pi e m k^2}{h^3}\,T^2\,\mathrm{e}^{-\frac{\Psi_{sem}}{kT}}$$
$$= 120\,\frac{\mathrm{amp}}{\mathrm{cm}^2}\cdot\left(\frac{T}{\mathrm{degree\ K}}\right)^2 \mathrm{e}^{-\frac{\Psi_{sem}}{kT}} \qquad (X.5.02)$$

From this it can be seen that in the semiconductor, too, the distance Ψ_{sem} between electrostatic surface potential and Fermi level yields the decisive term in the exponent of the Richardson law, i.e., the thermal work function.

---

[1] This may be accomplished by electron emission from both sides into the vacuum and then requires a relatively long time. As an alternative, we can imagine the metal and semiconductor to be bent in ring shape. At the surfaces under consideration they are far apart, but somewhere else they are in direct contact with other parts of their surface. In this way, the equilibrium can be established *quickly*.

In analogy to (X.4.07) one can derive from Fig. X.5.1

$$\Psi_{\text{sem}} = |E_{C_{\text{sem}}}| + |\zeta_{\text{sem}}| \pm (\text{double layer})_{\text{sem vac}} \qquad (\text{X}.5.03)$$

In contrast to metals there is now no longer any relationship between this thermal work function and any significant photoelectric activation energy which is related to some threshold wavelength of the photoelectric effect. In metals the Fermi level is, in the limiting case $T \to 0$, the highest occupied electron level and therefore the most favorable starting point for an electron which is to be released photoelectrically. In the semiconductor, however, the Fermi level, except in the case of degeneracy, lies in the forbidden zone where, apart from the discrete impurity levels, no levels exist which can accommodate electrons. Hence, in the limiting case $T \to 0$, the photoelectric emission of electrons can originate only from an impurity-level or from the fully occupied valence band; for $T > 0$, in addition, from the Maxwell tail in the conduction band. This shows that only in a few well-defined special cases[1] the Fermi level in the semiconductor can be the source of photoemission of an electron. Hence, thermionic work function Ψ_{sem} and photoelectric activation energy are not identical in the semiconductor (see also Fig. X.5.2).

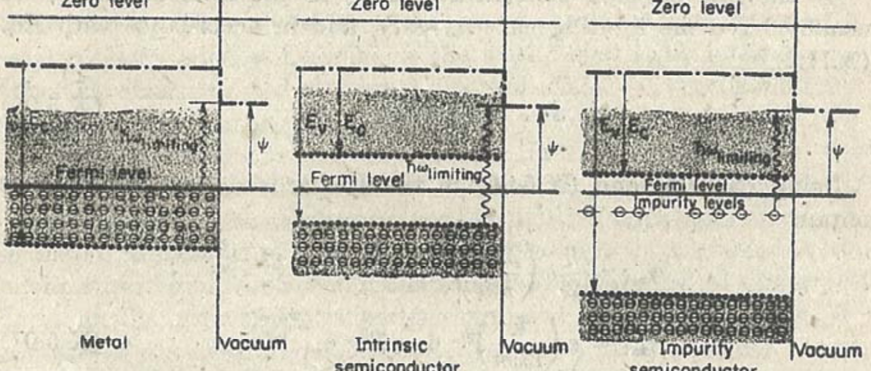


FIG. X.5.2. Photoelectric activation energy $\hbar\omega_{\text{lim}}$ and thermionic work function ψ. Limiting case of $T \to 0$.

---

[1] In a semiconductor with a large number of impurities, degeneracy can set in; then the Fermi level lies in the lower part of the conduction band and is therefore occupied by electrons. Another special case is the transition from "saturation" to "reserve," where the Fermi level coincides with the impurity level.

In both these cases, there exist electrons with even higher energy, and therefore, in these special cases, too, the thermionic work function Ψ_{sem}, referred to the Fermi level, does not appear as the photoelectric activation energy which is related to a long wavelength threshold.

In order to retain this identity which is so useful for the understanding of the physical meaning, we might be tempted to change the definition of the work function of semiconductors. For instance, we might choose for it the separation of the electrostatic surface potential and the lower edge of the conduction band:

$$\Psi^*_{\mathrm{sem}} = |E_{C_{\mathrm{sem}}}| \pm (\text{double layer})_{\mathrm{sem\,vac}} \tag{X.5.04}$$

Comparison with (X.5.03) yields

$$\Psi_{\mathrm{sem}} = \Psi^*_{\mathrm{sem}} + |\zeta_{\mathrm{sem}}| \tag{X.5.05}$$

At the low electron concentrations n_S in the semiconductor, we usually have the limiting case $n_S \ll N$, and in accordance with Eq. (A.II.2)

$$\zeta_{\mathrm{sem}} = \mathrm{k}T \ln \frac{n_S}{N} < 0 \tag{X.5.06}$$

Using (X.5.05) and (X.5.06) in the Richardson law (X.5.02), we obtain

$$i_{\mathrm{sat}} = \frac{1}{2} e \left(\frac{2\mathrm{k}T}{\pi m}\right)^{\frac{1}{2}} \cdot N \,\mathrm{e}^{-\frac{\Psi_{\mathrm{sem}}^*}{\mathrm{k}T}} \cdot \mathrm{e}^{-\ln\frac{N}{n_S}}$$

$$i_{\mathrm{sat}} = e \left(\frac{\mathrm{k}}{2\pi m}\right)^{\frac{1}{2}} \cdot n_S \cdot T^{\frac{1}{2}} \cdot \mathrm{e}^{-\frac{\Psi_{\mathrm{sem}}^*}{\mathrm{k}T}} \tag{X.5.07}$$

The term Ψ^*_{sem} occurring in this form of the emission law can, as previously mentioned, be plausibly defined as the work required to raise an electron from the lower edge of the conduction band to a position at rest in front of the surface. This procedure has the disadvantage that, in contrast to the numerical coefficient $4\pi e m \mathrm{k}^2/h^3 = 120$ amps/cm² of form (X.5.02) of the Richardson law, the numerical coefficient $e(\mathrm{k}/2\pi m)^{1/2} \cdot n_S$ of Eq. (X.5.07) is no longer a universal quantity because of n_S and will also be strongly temperature dependent in many semiconductors, again because of n_S. Furthermore, the correlation between work function and contact potential is lost. Finally, this whole transformation of the Richardson law is limited to cases where the electron gas in the semiconductor is not degenerate. However, in very highly conducting samples of Si and Ge, degeneracy is quite likely.

For all these reasons, the form (X.5.02) of the Richardson law has been retained in the literature for semiconductors, together with the corresponding definition of the semiconductor work function as referred to the Fermi level.

It must be emphasized that in semiconductors the work function Ψ_{sem} is definitely not temperature independent because of the temperature dependence of the Fermi level. This requires even more care in the interpretation of an experimental [$\ln i_{\text{sat}}$ versus $1/T$] plot than in the case of metals.

§6. Semiconductor Boundary Layers. The Metal-Semiconductor Contact

Figure X.5.1 was based specifically on the case of metal and semiconductor surfaces which are far apart. The purpose of this assumption was to keep the electric field due to the contact potential sufficiently small. Strong fields would cause special effects even with metal surfaces, such as lowering of the work function by image forces (Schottky emission) and also by the tunnel effect. In semiconductor surfaces we have to reckon, in addition, with the fact that charges must be located at the ends of the field lines which are due to the Volta potential difference. At the surface of a metal, considerable surface charges can be accommodated by slightly increasing or decreasing the large metallic electron density in the uppermost atomic layer. On the other hand, the accommodation of an equal amount of surface charge on the semiconductor surface requires[1] very large increases or decreases of the electron density in a layer up to 1,000 to 10,000 atoms deep because the available electron density is smaller by several orders of magnitude. We therefore deal no longer with an actual *surface* charge; instead, the field lines penetrate partly up to 10^{-5} or 10^{-4} cm into the semiconductor and end successively at the charges of a *space* charge distributed over distances up to 10^{-4} cm (see Fig. X.6.1).

In a metal, the sudden ending of all field lines at a surface charge causes a break in the potential curve, whereas in a semiconductor the gradual dissipation of the field in a spatially extended space charge causes a curving of the potential variation over some distance. If the semiconductor had the dielectric constant unity, like vacuum, the absence of surface charge would cause the potential variation to traverse the semiconductor surface with constant slope. In the general case $\varepsilon > 1$, it is not the field strength $\mathbf{E} = -dV/dx$ that has to be continuous at the surface but the dielectric displacement $\mathbf{D} = -\varepsilon\, dV/dx$. This means that the curved potential variation in the semiconductor starts at the surface with a slope that is smaller by the factor $1/\varepsilon$ than that with which it ends at the vacuum side of the surface. In any

---

[1] See, however, the later discussion of surface states.

case we can see that, when semiconductor and metal come close together, a space charge will be established in a boundary layer of the semiconductor, resulting in a curvature in the plot of the electrostatic electron energy $-eV$. The strength of the chemical binding, however, does not change because it is determined by the crystal structure. Therefore, the dotted line representing the lower edge E_C of the conduction band must have the same curvature (Fig. X.6.2), while the Fermi level E_F, indicated by a heavy continuous line, will remain horizontal as long as thermal equilibrium obtains. In the boundary layer, therefore, the distance $E_C - E_F$ changes. According

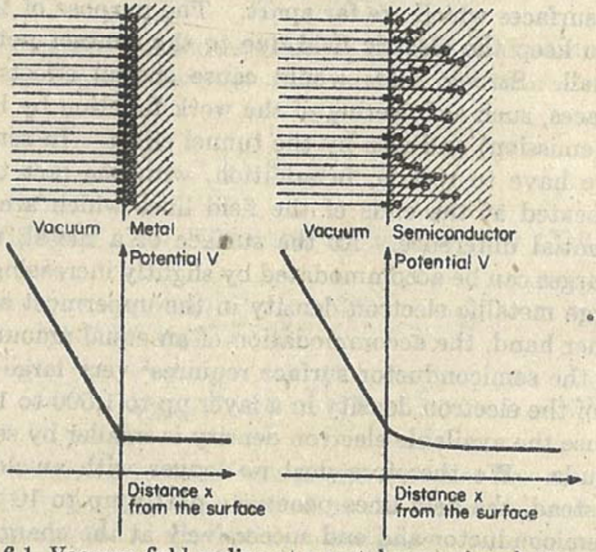


FIG. X.6.1. Vacuum field ending at a metal or semiconductor surface.

to page 317 and Eq. (VIII.5.15) this distance is a logarithmic measure for the electron concentration n:

$$E_C - E_F \approx \mathrm{k}T \ln \frac{N_C}{n} > 0 \qquad \text{for } n \ll N_C \qquad \text{(X.6.01)}$$

Thus the electron concentration n varies in the boundary layer and, since it deviates from its neutral value n_S, a space charge is created. Atomic double layers on the metal and on the semiconductor which, for greater clarity, were neglected in the potential representations of Fig. X.6.1, have been taken into consideration in Fig. X.6.2.

When the metal and semiconductor surfaces (see Fig. X.6.2) come infinitely close together, the dot-dash line indicating the electrostatic energy term $-eV$ has, in the end, to be raised so much that the whole

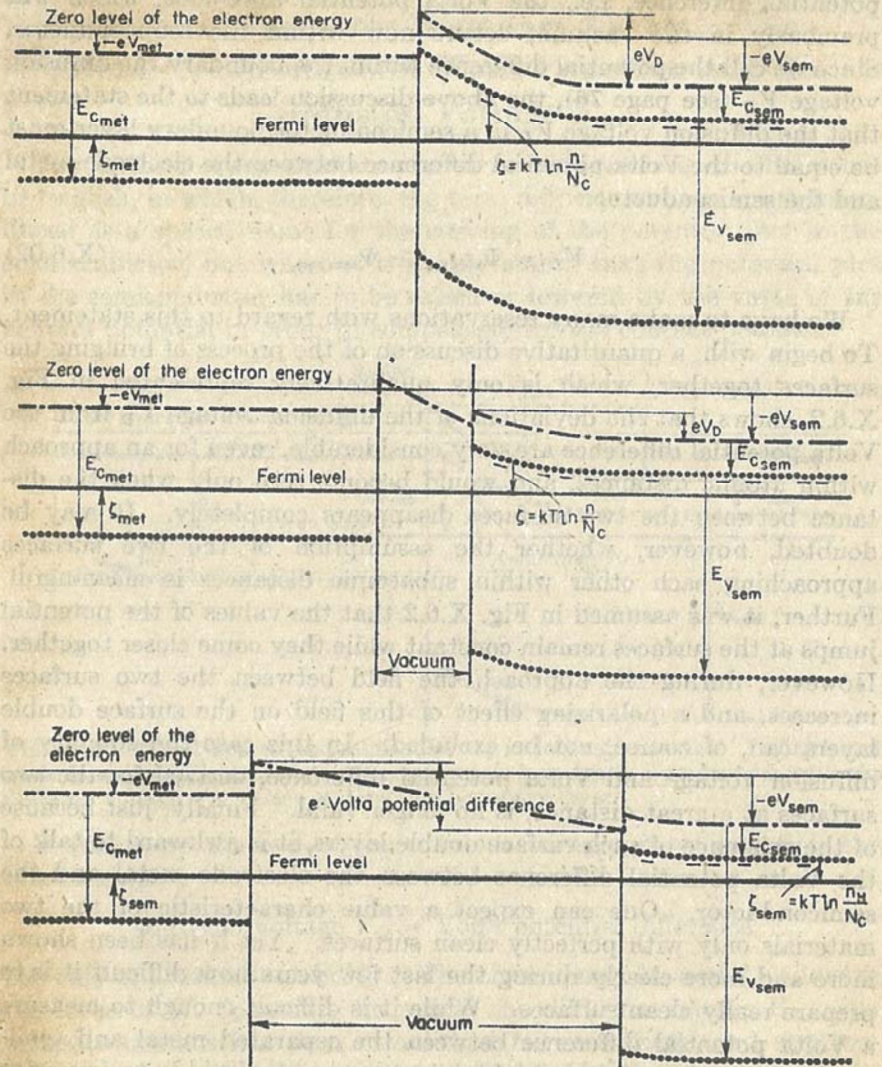


Fig. X.6.2. A semiconductor surface approaching a metal surface. Gradual distortion of the bands in the semiconductor by an amount equal to the diffusion voltage V_D. Hence the limiting condition of continuous transition of dielectric displacement $\varepsilon\mathbf{E}$ at the semiconductor surface is fulfilled; a space-charge boundary layer is created.

potential difference, i.e., the Volta potential difference, which was previously in the vacuum, occurs now within the semiconductor. Since we call the potential difference within the boundary the diffusion voltage V_D (see page 76), the above discussion leads to the statement that the diffusion voltage V_D in a semiconductor boundary layer must be equal to the Volta potential difference between the electrode metal and the semiconductor:

$$V_D = \Psi_{\text{met vac}} - \Psi_{\text{sem vac}} \tag{X.6.02}$$

We have to make many reservations with regard to this statement. To begin with, a quantitative discussion of the process of bringing the surfaces together, which is only qualitatively represented in Fig. X.6.2, shows that the deviations of the diffusion voltage V_D from the Volta potential difference are very considerable,[1] even for an approach within atomic distances, and would become zero only when the distance between the two surfaces disappears completely. It may be doubted, however, whether the assumption of the two surfaces approaching each other within subatomic distances is meaningful. Further, it was assumed in Fig. X.6.2 that the values of the potential jumps at the surfaces remain constant while they come closer together. However, during the approach the field between the two surfaces increases, and a polarizing effect of this field on the surface double layers can, of course, not be excluded. In this case the equality of diffusion voltage and Volta potential difference, derived for the two surfaces at a great distance, is no longer valid. Finally, just because of the existence of such surface double layers, it is awkward to talk of the Volta potential difference between the electrode metal and the semiconductor. One can expect a value characteristic of the two materials only with perfectly clean surfaces. Yet it has been shown more and more clearly during the last few years how difficult it is to prepare really clean surfaces. While it is difficult enough to measure a Volta potential difference between the separated metal and semiconductor surfaces without obtaining wrong results due to impurity layers, it is an even greater problem to avoid all contamination during the process of making contact. In fact, experiments to prove a relationship between diffusion voltage and Volta potential difference have been only partially successful.[2]

---

[1] This is particularly true in a semiconductor with very many impurities where even a small rise in potential, and the small deviations from neutrality linked with it, releases strong space charges.

[2] See, for instance, S. Poganski, *Z. Physik*, **134**: 469 (1953), particularly Fig. 3.

§7. Semiconductor Boundary Layers. The Metal-Semiconductor Work Function and the Diffusion Voltage V_D

The method of representation used so far dominates the literature in English, in which, therefore, the term diffusion voltage is not introduced as a special name for the curving of the potential plot in the semiconductor, but where it is simply stated that the potential plot in the semiconductor has to be raised or lowered by the value of the contact potential in order to compensate for the contact potential.

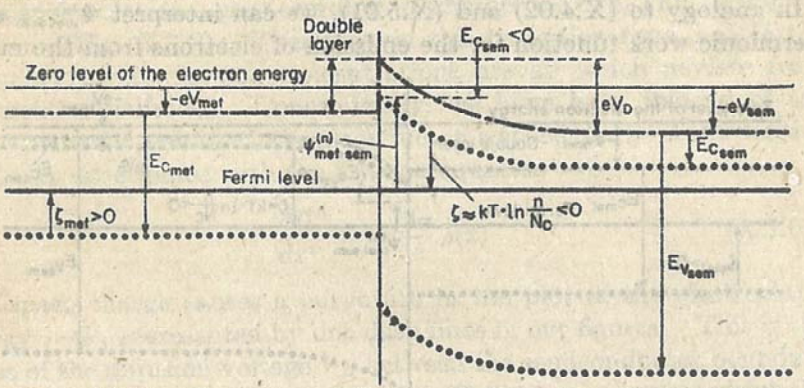


FIG. X.7.1. Metal and n-type semiconductor in intimate contact. Definition of work function $\psi^{(n)}_{met\text{-}sem}$. Note that this work function is *not* identical with the diffusion voltage V_D. Case of depletion boundary layer.

However, at the end of the previous paragraph we saw that the relationship

$$\text{Diffusion voltage } V_D = \text{Volta potential difference}$$

has only approximate validity. Therefore it may be useful, following the ideas of W. Schottky,[1] to regard the diffusion voltage V_D, after intimate contact has been established between metal and semiconductor, as an independent new term which is identical neither with the Galvani voltage nor with the Volta potential difference (≡ contact potential), nor with the vacuum work function, nor with the photoelectric activation energies. The question now arises whether, in the description of the intimate contact between metal and semiconductor, it is at all useful to start from the case of widely separated solids. If,

[1] In the publications of Schottky, these matters are treated rather briefly. The author is in the happy position of being able to refer to uncompleted and therefore unpublished manuscripts of Schottky.

instead, we depart directly from the case of intimate contact between metal and semiconductor, we obtain, by way of example, Fig. X.7.1 or X.7.2. In thermal equilibrium an electron concentration n_B is established[1] in the semiconductor at the boundary with the metal. Applying Eq. (X.6.01) to the semiconductor boundary, we obtain in the case of degeneracy

$$n_B = N_C \, e^{-\frac{\Psi^{(n)}_{\text{met sem}}}{kT}} \qquad \text{(X.7.01)}$$

Here $\Psi^{(n)}_{\text{met sem}} = (E_C - E_F)_{\text{bound}}$ is the distance, measured at the boundary, of the conduction band from the Fermi level, drawn as a heavy continuous line.

In analogy to (X.4.02) and (X.5.01), we can interpret $\Psi^{(n)}_{\text{met sem}}$ as a thermionic work function for the emission of electrons from the metal into the semiconductor. The main justification for such an interpretation is the fact that this term is independent of the impurity content of the semiconductor and of the position of the Fermi level which is, so to speak, accidental. Thus we have, according to Fig. X.7.1,

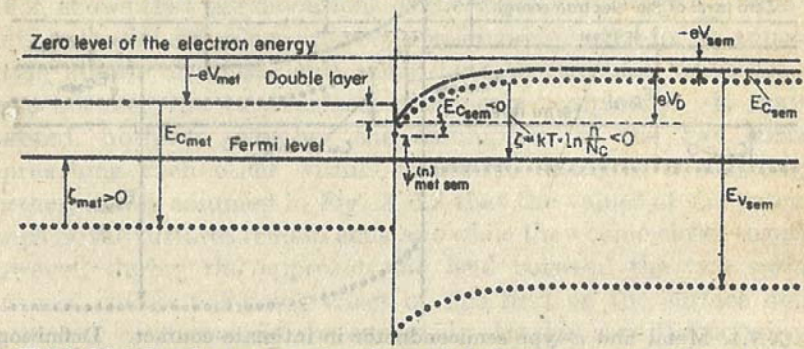


FIG. X.7.2. Metal and n-type semiconductor in intimate contact. Case of accumulation boundary layer.

$$\Psi^{(n)}_{\text{met sem}} = |E_{C_{\text{met}}}| - \zeta_{\text{met}} + (\text{double layer})_{\text{met sem}} - |E_{C_{\text{sem}}}| \qquad \text{(X.7.02)}$$

Compared with the thermionic work function for the emission of electrons into vacuum [see Eq. (X.4.07)], we have only one new term, the binding energy $E_{C_{\text{sem}}}$, i.e., a characteristic of the undisturbed semiconductor lattice.

Therefore the chemical binding of the electrons in the metal lattice ($E_{C_{\text{met}}}$), the number of the electrons therein (ζ_{met}), the surface conditions at the boundary metal semiconductor (double layer), and the chemical binding of the electrons in the semiconductor ($E_{C_{\text{sem}}}$) determine a work

---

[1] The subscript B indicates the semiconductor boundary.

function $\Psi^{(n)}_{\mathrm{met\,sem}}$ for the emission of electrons from the metal into the semiconductor and thereby, at least in thermal equilibrium, a boundary concentration n_B of the electrons quite independent of the impurity content of the semiconductor and of the position of the Fermi level in the semiconductor. Conversely, the electron concentration deep within the semiconductor is decisively influenced by the impurity content of the semiconductor because this impurity content enters, here, into the quasi-neutrality condition which, in turn, determines a value n_S of the electron concentration. Therefore a boundary layer must be formed in which a gradual transition takes place from the boundary value n_B to the semiconductor value[1] n_S. We have a depletion layer or an accumulation layer according to whether $n_B < n_S$ (Fig. X.7.1) or $n_B > n_S$ (Fig. X.7.2). Therefore, in this transition zone, i.e., in the boundary layer, electron concentrations prevail which deviate from the neutral value n_S. Consequently, we have here, instead of the quasi-neutrality *within* the semiconductor, a space charge of the density $\rho(x)$. In accordance with the Poisson equation

$$\frac{d^2V}{dx^2} = -\frac{4\pi}{\varepsilon}\,\rho(x) \tag{X.7.03}$$

this space charge causes a curvature in the plot of the electrostatic energy $-eV$, represented by dot-dash lines in our figures. This is the cause of the diffusion voltage V_D between the semiconductor boundary and the inside of the semiconductor. The name is explained by the fact that the field current generated by the diffusion voltage V_D balances out the diffusion current generated by the concentration gradient from n_S to n_B so that the zero current of thermal equilibrium is established. This has already been discussed on pages 76ff., together with the difficulties involved in visualizing the existence of an electrostatic potential difference within a conductor with zero current.

We want to discuss these difficulties once more from the newly gained point of view and note, to begin with, that the objections arise from an assumption that there is a law to the effect that the potential in a conductor is constant in the absence of a current. This law is incorrect if we mean by the potential the electrostatic potential represented by dot-dash lines in our figures. All the same, the law contains a core of truth, because in the thermodynamics of electric systems it is proved that this law is correct if we mean by the potential the electro*chemical* potential.[2] The electrochemical potential is the sum of electrostatic

---

[1] The subscript S indicates the value for the semiconductor.

[2] See, for instance, W. Schottky and H. Rothe, Physik der Glühelektroden, in Wien and Harms, *op. cit.*, vol. 13, part 2, chap. III, in particular Eq. (5), p. 18.

energy per particle and chemical potential ($\equiv$ free energy per particle). The electrostatic energy per electron is $-eV$. The other term of the sum can be computed, with the usual assumptions of solid-state physics, as chemical potential of a Fermi gas which is not free but has a potential energy equal to the total energy $E_C < 0$ of the lower edge of the conduction band (see pages 300 and 306). The chemical potential of a *free* Fermi gas with the concentration n is equal[1] to $\zeta(n)$. According to pages 300 and 306, the conduction electrons can be treated, under certain assumptions, not as a free gas, but as a gas in a potential well of depth E_C. Their chemical potential is, therefore, expressed by $E_C + \zeta(n)$ and their electrochemical potential by $-eV + E_C + \zeta(n)$. The law of the constancy of the electrochemical potential in a conductor in thermal equilibrium, i.e., with zero current, requires therefore that the sum $-eV + E_C + \zeta$ be constant. This, however, is just the distance of the Fermi level (in heavy type) from the zero level of the electron energy. Thus we have again a proof, from the thermodynamic point of view, that the Fermi level must be horizontal throughout all conductors and parts of conductors, provided that thermal equilibrium obtains.

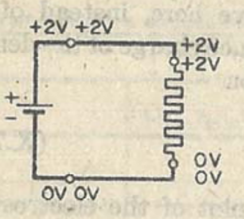


FIG. X.7.3

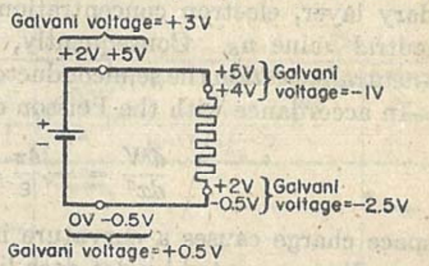


FIG. X.7.4

Incidentally, it is this electrochemical potential which we use, without ever worrying about the thermodynamics of electric systems, in the usual considerations concerning potential distributions in electric circuits in everyday work. To take a simple example, in the circuit of Fig. X.7.3 we would use the indicated potential values without troubling about the fact that the different parts of the conductors are made of different metals. The point is that we deal here with values of the

---

[1] See L. Brillouin, "Die Quantenstatistik," Springer-Verlag OHG, Berlin, 1931, Eq. (32a) on p. 141 in connection with Eq. (28) on p. 139. A very simple and elegant proof, unfortunately based on incorrect arguments, is given in H. Fröhlich, "Elektronentheorie der Metalle," p. 64, Springer-Verlag OHG, Berlin, 1936, and by A. Sommerfeld in Geiger and Scheel, *op. cit.*, vol. 24, part 2, p. 342. However, the derivation in the latter can be corrected by elementary mathematical means.

electrochemical potential, whereas indicating the electrostatic macropotentials might lead to Fig. X.7.4. Since the Galvani potential jumps summed over the complete circuit add up exactly to zero (see page 344 and Fig. X.2.3), we obtain by this procedure the same value of 2 volts for the terminal voltage as previously when we used the electrochemical potential.[1]

In other words, the initial difficulty in understanding the existence of electrostatic potential differences within a semiconductor with zero current may be due to a confusion between terminal voltage and a difference of electro*static* potentials, whereas in reality the terminal voltage which disappears in the case of zero current is equal to the difference of the electro*chemical* potentials. The latter are, in fact, commonly used in everyday work, probably quite intuitively.

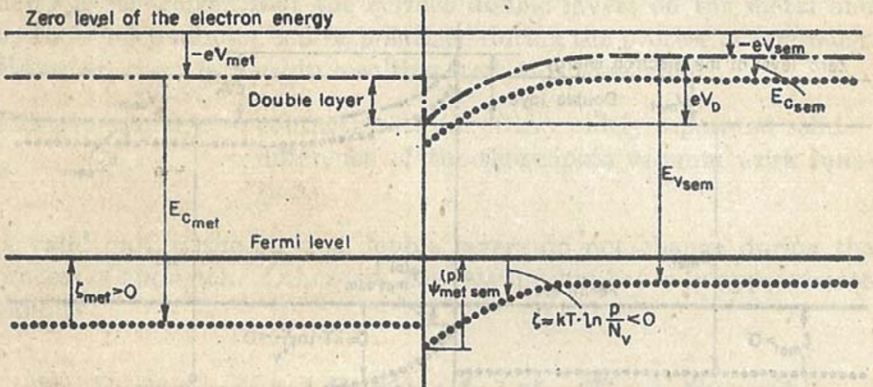


FIG. X.7.5. Metal and *p*-type semiconductor in intimate contact. Case of depletion boundary layer.

Figures X.7.5 and X.7.6 show a depletion and an accumulation layer in a *p*-type semiconductor. Here a work function $\Psi^{(p)}_{met\,sem}$ for the emission of holes from the metal into the semiconductor is shown. At first sight, this term may appear somewhat strange. But we have shown in Chap. III that one can always describe the totality of electrons in a band also as a totality of holes (see pages 65 to 66) and that, correspondingly, one can interpret the transition of a number of valence electrons into a neighboring metal electrode as the escape of a number of holes from the metal into the valence band. We also recall

---

[1] Here it may be recalled that, just because of the existence of Volta potential differences and the resulting vacuum fields, one has to take care in electrostatic measurements of terminal potentials that the two poles or knife edges or plates of the electrostatic instrument be made of the same material. This precaution is, of course, superfluous if the terminal voltage is measured galvanometrically.

that the distance $E_F - E_V$ of the Fermi level E_F from the upper edge E_V of the valence band usually equals $-\zeta(p/N_V)$ [see Eq. (VIII.4.15)] and, in the Maxwell-Boltzmann case, becomes equal to $kT \ln N_V/p$ [see Eq. (VIII.5.16)]. Accordingly, this distance is a logarithmic measure of the concentration p of the holes (see page 317). Using this relationship for the semiconductor boundary, we obtain for the boundary density p_B of the holes

$$p_B = N_V \, e^{-\frac{\Psi^{(p)}_{\text{met sem}}}{kT}} \tag{X.7.04}$$

The analogy to Eqs. (X.4.02), (X.5.01), and (X.7.01) shows that we are justified in introducing a work function $\Psi^{(p)}_{\text{met sem}}$ for holes.

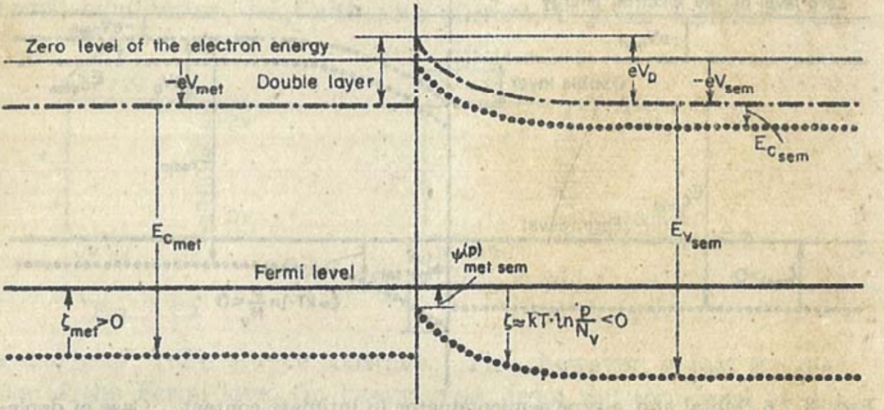


FIG. X.7.6. Metal and p-type semiconductor in intimate contact. Case of accumulation boundary layer.

Since we have defined $(E_C - E_F)_{\text{bound}} = \Psi^{(n)}_{\text{met sem}}$ and

$$(E_F - E_V)_{\text{bound}} = \Psi^{(p)}_{\text{met sem}}$$

we obtain by addition the expression

$$\Psi^{(n)}_{\text{met sem}} + \Psi^{(p)}_{\text{met sem}} = E_C - E_V \tag{X.7.05}$$

The sum of electron and hole work functions in a metal-semiconductor combination is independent of the metal and of any surface double layers and is equal to the width $E_C - E_V$ of the forbidden gap in the semiconductor.

We can summarize as follows: In §6 and §7 we have compared two different descriptions for the situation prevailing at the boundary of a metal and of a semiconductor. The second treatment, originated by Schottky, considers at once the intimate contact between metal and semiconductor. It starts with the concept of the thermionic work function $\Psi^{(n)}_{\text{met sem}}$ for the emission of electrons from the metal into the semiconductor, then points out the difference in electron concentrations n within the semiconductor (n_S) and at the boundary (n_B) and thus arrives at the decisive diffusion voltage V_D. The semiconductor density n_S is determined by the impurity content and the neutrality condition, the boundary density n_B, on the other hand, by the work function $\Psi^{(n)}_{\text{met sem}}$.

The first treatment, used mainly in the literature in English, describes the intimate contact of semiconductor and metal as the limiting case of the two solids separated by finite distances a, assuming that the solids are in thermal equilibrium even in the state of spatial separation. This treatment is unobjectionable as long as the possibility is recognized that the surface double layers on the metal and on the semiconductor can be polarized during the process of approach. However, the relationship resulting from this treatment:

Diffusion voltage = contact potential of the widely separated solids
= difference of the thermionic vacuum work functions

is valid only if the surface double layers do not change during the process of approach. Otherwise, the relationship has only approximate validity.

§8. Experimental Results for the Work Function of Semiconductors and the Metal-Semiconductor Contact

In the course of the last few years, it has been finally realized that in commercial selenium rectifiers and in point detectors we do not deal with semiconductor-metal contacts but with disguised p-n rectifiers.[1] Prior to that, however, many experiments were carried out to obtain evidence for the behavior of semiconductor-metal contacts in accordance with the Schottky theory. In the first place we have to mention H. Schweickert[2] who produced contacts between selenium and a number of different metals and who found a connection between the blocking resistance and the work function of the metal with which the contact is made.

---

[1] S. Poganski, *Z. Physik*, **134**: 469 (1953). A. Hoffmann and F. Rose, *Z. Physik*, **136**: 152 (1953). R. Thedieck, *Physik. Verhandl.*, **3**: 31 (1952); **3**: 212 (1952); *Z. angew. Phys.*, **5**: 165 (1953). L. B. Valdes, *Proc. IRE*, **40**: 445 (1952).

[2] H. Schweickert, *Verhandl. deut. physik. Ges.*, **3**: 99 (1939). The results of Schweickert are also reproduced in W. Schottky, *Z. tech. Physik*, **21**: 322 (1940).

From Eqs. (IV.4.23) and (IV.4.24) for the characteristics, one obtains for the zero resistance[1]

$$R_0 = \frac{\mathbf{V}}{e\mu_n n_B \mathbf{E}_B} \quad \text{with} \quad \mathbf{V} = \frac{kT}{e} = 25.9 \cdot \left(\frac{T}{300°\mathrm{K}}\right) \cdot 10^{-3} \text{ volt}$$

and with the aid of (X.7.01)

$$R_0 = \frac{\mathbf{V}}{e\mu_n N_C \mathbf{E}_B} e^{+\frac{\Psi^{(n)}_{\text{met sem}}}{kT}} \tag{X.8.01}$$

We have already noted on page 358 that the vacuum work function $\Psi^{(n)}_{\text{met vac}}$ and the semiconductor work function $\Psi^{(n)}_{\text{met sem}}$ of a metal contain a number of identical terms. If we make contact between the same semiconductor, selenium, and a variety of metals we would therefore expect that the $\Psi^{(n)}_{\text{met sem}}$ vary in the same way as the $\Psi^{(n)}_{\text{met vac}}$ and that therefore, in view of (X.8.01), we would find that the resistance R_0 varies strongly with the vacuum work function of the metal with which contact is made. The measurements of Schweickert which are reproduced in Fig. X.8.1 show, in fact, such an effect. At first it may

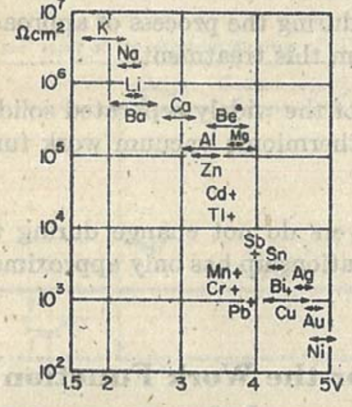


FIG. X.8.1. Work function and resistance of selenium rectifiers [from H. Schweickert, *Verhandl. deut. physik. Ges.* **3**, 99 (1939)].

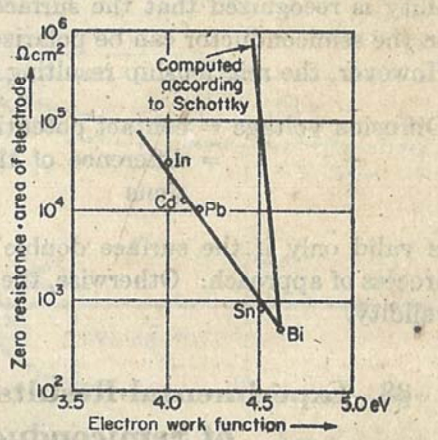


FIG. X.8.2. Selenium rectifier without barrier layer. Dependence of the zero resistance upon the electron work function of the metal of the blocking electrode. [From S. Poganski, *Z. Physik*, **134**, 476 (1953)].

---

[1] H. Schweickert plots in his diagram the maximum blocking resistances of the semiconductor-metal contacts under consideration. However, the maximum blocking resistance is outside the scope of the boundary-layer theory. Therefore, we use the differential resistance at zero bias voltage, the so-called zero resistance, to demonstrate the relationships to be expected from the boundary-layer theory.

seem surprising that with increasing work function the resistance R_0 decreases rather than increases. The explanation is the fact that selenium is a p-type conductor and that instead of (X.8.01) one has to use the expression

$$R_0 = \frac{\mathbf{V}}{e\mu_p N_V \mathbf{E}_B} \, \mathrm{e}^{+\frac{\Psi^{(p)}_{\mathrm{met\,sem}}}{\mathrm{k}T}} \qquad \text{(X.8.02)}$$

From this one obtains with the aid of (X.7.05)

$$R_0 = \frac{\mathbf{V}}{e\mu_p N_V \mathbf{E}_B} \, \mathrm{e}^{+\frac{E_C - E_V}{\mathrm{k}T}} \, \mathrm{e}^{-\frac{\Psi^{(n)}_{\mathrm{met\,sem}}}{\mathrm{k}T}} \sim \mathrm{e}^{-\frac{\Psi^{(n)}_{\mathrm{met\,sem}}}{\mathrm{k}T}} \qquad \text{(X.8.03)}$$

and thus the decrease of the contact resistance with the vacuum work function of the contact-forming metal which has, in fact, been observed by Schweickert.

Other authors[1] obtain only in part results similar to those of Schweickert. It has become more and more apparent that it is extremely difficult to obtain a real metal-semiconductor contact without intermediate layers of reaction products and to clean the surfaces sufficiently before making the contact. S. Poganski[2] has probably carried out the neatest comparison between measurements on metal-semiconductor contacts and the boundary-layer theory, with very good qualitative agreement. Quantitatively, however, the variation of blocking characteristics with the work function of the electrode material is much too small (see Fig. X.8.2).

In the course of the quoted investigations, a large number of Volta potential differences and work functions have been determined. The methods used will be described below, partly because they are of considerable intrinsic interest, but also because they represent excellent material for getting familiar with the use of the terms introduced in §1 to §6, such as photoelectric activation energy, Galvani voltage, etc.

a. Methods for Measuring Volta Potential Difference and Work Function

The experimental arrangement (see Fig. X.8.3) shows two parallel plane surfaces of the two solids to be investigated, facing each other. These are not in thermal equilibrium, but a current of electrons flows across the vacuum from surface I to surface II. The energy for this

---

[1] Unpublished work by Brattain and Shive at Bell Telephone Laboratories, 1940; see J. Bardeen, *Phys. Rev.*, **71:** 2 (1947). A. V. Joffe, *J. Phys. U.S.S.R.*, **10:** 49 (1946). W. E. Meyerhof, *Phys. Rev.*, **71:** 727 (1947). S. Benzer, *J. Appl. Phys.*, **20:** 804 (1949).

[2] S. Poganski, *Z. Physik*, **134:** 469 (1953).

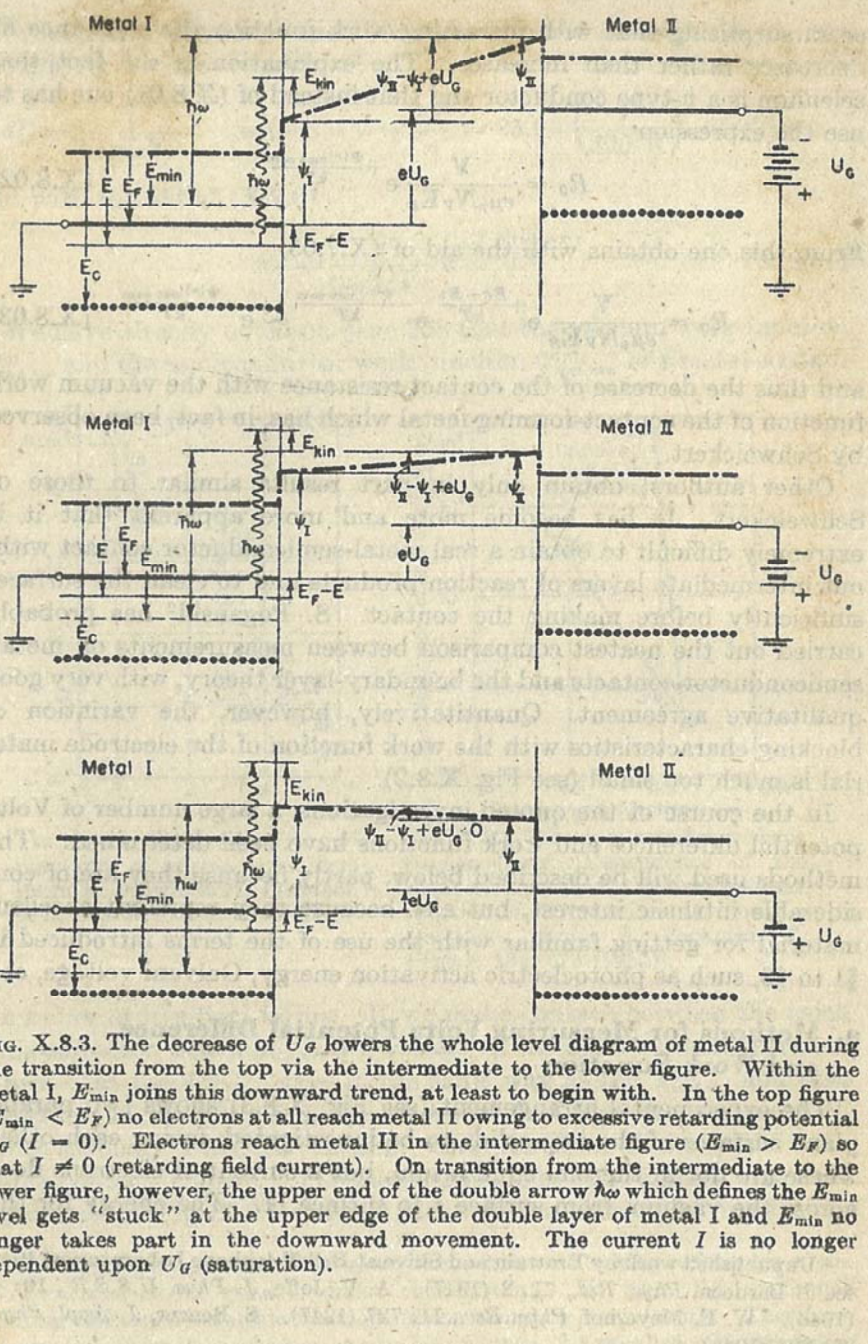


FIG. X.8.3. The decrease of U_G lowers the whole level diagram of metal II during the transition from the top via the intermediate to the lower figure. Within the metal I, E_{min} joins this downward trend, at least to begin with. In the top figure ($E_{min} < E_F$) no electrons at all reach metal II owing to excessive retarding potential U_G ($I = 0$). Electrons reach metal II in the intermediate figure ($E_{min} > E_F$) so that $I \neq 0$ (retarding field current). On transition from the intermediate to the lower figure, however, the upper end of the double arrow $\hbar\omega$ which defines the E_{min} level gets "stuck" at the upper edge of the double layer of metal I and E_{min} no longer takes part in the downward movement. The current I is no longer dependent upon U_G (saturation).

current flow is supplied either by heating or by illumination of solid I. The polarity of the battery in the diagram is such that it provides a retarding potential so that the current can be partially or completely cut off.[1]

As a first case we consider the *photoelectric* emission of electrons. In Fig. X.8.3 it is assumed that both solids I and II are metals. As a simplification it is also assumed that both metals are at temperature $T = 0$.

We now consider an electron on the lattice energy level E. Let a photon transfer its total energy[2] $\hbar\omega$ to this electron. If the electron obtains a velocity component in the direction of the surface, it can leave the metal I, provided that $\hbar\omega$ is large enough, i.e., $\hbar\omega > \Psi_I + (E_F - E)$. In this process it will first lose the energy $E_F - E$ up to the Fermi level and then the work function Ψ_I so that in vacuum, right in front of the surface I, it has the kinetic energy

$$E_{\text{kin}} = \hbar\omega - \Psi_I - (E_F - E) \tag{X.8.04}$$

The electron has to overcome the electrostatic potential difference $(1/e)(\Psi_{II} - \Psi_I) + U_G$ with this kinetic energy, to reach the metal II and thus to contribute to the current J from I and II.

This requires at least[3]

$$E_{\text{kin}} = \hbar\omega - \Psi_I - (E_F - E) \geqq \Psi_{II} - \Psi_I + eU_G \tag{X.8.05}$$

Hence only the electrons in the levels above

$$E_{\text{min}} = E_F - [\hbar\omega - \Psi_{II} - eU_G] \tag{X.8.06}$$

contribute to the current J from I to II.

If the retarding potential U_G is too large, the term in parenthesis in Eq. (X.8.06) becomes negative and E_{min} is above the Fermi level E_F. However, at $T = 0$ the levels above E_F are not occupied by electrons so that the energy $\hbar\omega$ of the photons is not large enough to raise a single electron from I to II and the current J is zero (see Fig. X.8.3). The limit $U_G^{(0)}$, where an electron transition from I to II is just possible in the most favorable case, is reached when the term in parenthesis in (X.8.06) becomes zero:

$$U_G^{(0)} = \frac{1}{e}(\hbar\omega - \Psi_{II}) \tag{X.8.07}$$

---

[1] In the actual performance of such measurements, cylindrical or spherically symmetrical arrangements are preferable because of the reduced edge distortions.

[2] ω = angular frequency $= 2\pi f$; $\hbar = \frac{1}{2\pi} h = 1.054 \cdot 10^{-27}$ cm^2 g sec^{-1}.

[3] A part of E_{kin} is likely to be associated with lateral velocity components which are of no use in overcoming the electrostatic potential difference.

It is surprising that this limit depends only on the work function Ψ_{II} of the anode, while the work function Ψ_I of the cathode from which the electrons are released has dropped out. With decreasing retarding potential $U_G < U_G^{(0)}$, more and more lattice energy levels $E < E_F$ can contribute to the current J and therefore J increases (see Fig. X.8.4).

However, this increase is not unlimited. When the retarding potential U_G has become so small that no rise of the electrostatic potential has to be overcome in the vacuum (see Fig. X.8.3, bottom), then condition (X.8.05) has to be replaced by the condition

$$E_{\text{kin}} = \hbar\omega - \Psi_I - (E_F - E) \geqq 0 \qquad \text{(X.8.08)}$$

According to (X.8.08) the levels from

$$E = E_F - (\hbar\omega - \Psi_I) \qquad \text{to} \qquad E = E_F \qquad \text{(X.8.09)}$$

now contribute to the current J, independent of the magnitude of the retarding potential U_G. Thus the current J is saturated at a retarding potential U_G for which we obtain from (X.8.08) and (X.8.05)

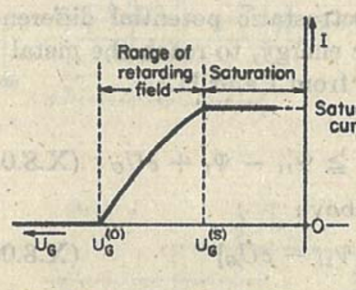


FIG. X.8.4. Plot of I as a function of U_G in photoelectric emission (diagrammatic).

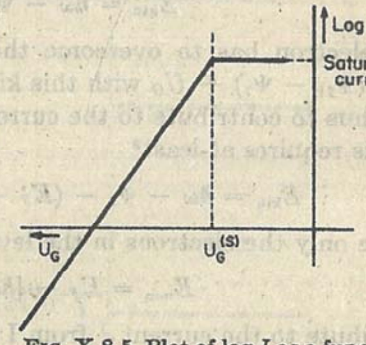


FIG. X.8.5. Plot of log I as a function of U_G in thermionic emission (diagrammatic).

$$U_G^{(s)} = \frac{1}{e}(\Psi_I - \Psi_{II}) \qquad \text{(X.8.10)}$$

(see Fig. X.8.4). Therefore, plotting a current-voltage characteristic, with an arrangement according to Fig. X.8.3, provides information concerning the work function Ψ_{II} of the anode through measurement of the limiting voltage $U_G^{(0)}$ according to (X.8.07) and concerning the Volta potential difference $(1/e)(\Psi_I - \Psi_{II})$ between cathode I and anode II through measurement of the saturation voltage $U_G^{(s)}$ according to (X.8.10). Hence $U_G^{(s)}$ and $U_G^{(0)}$, between them, also provide information concerning the work function Ψ_I of the cathode.

Nothing is changed in the foregoing considerations if the anode II is not a metal but a semiconductor. But the situation is quite different when a metal is replaced by a semiconductor in the cathode I. As is emphasized in Fig. X.5.2, we now have to bear in mind that the Fermi level E_F of a semiconductor is not occupied by electrons. Further, the situation for $T = 0°\mathrm{K}$ differs now in some respects *basically* from that for $T > 0°\mathrm{K}$. We have to refer the interested reader to specialized papers,[1] also with regard to the conclusions that can be drawn from the shape of the current rise between $U_G^{(0)}$ and $U_G^{(s)}$ in Fig. X.8.4 concerning the distribution $N(E)\,dE$ of the electron levels E below the Fermi level E_F.

We can now deal more briefly with the *thermionic* emission current J from I to II. Here we need modify only the computation of the emission current density of §4 in so far as we have to integrate over the x component v_x of the velocity, not from 0 to ∞, but from the value[2] $\sqrt{(2/m)(\Psi_{II} - \Psi_I + eU_G)}$, required to overcome the electrostatic potential difference $(1/e)(\Psi_{II} - \Psi_I) + U_G$, to ∞. Thus we obtain an expression for the retarding-field current

$$i = AT^2\, e^{-\frac{1}{kT}(eU_G + \Psi_{II})} \tag{X.8.11}$$

Of course, the retarding-field current cannot exceed the saturation value (X.4.05), and so we obtain again the Volta potential difference

$$U_G^{(s)} = \frac{1}{e}(\Psi_I - \Psi_{II}) \tag{X.8.12}$$

as saturation value of the retarding potential U_G (see Fig. X.8.5).

Thus, for thermionic emission of the current J from I to II, the work function Ψ_{II} of the anode can be determined[3] by measuring the retarding-field current and the variation of i/AT^2 with temperature at a constant retarding voltage U_G, while the transition into the saturation range indicates the Volta potential difference.

Condition (X.8.10) and (X.8.12) for the onset of saturation mean that the plot of the electrostatic potential in vacuum is horizontal (see Fig. X.8.6). However, this disappearance of the vacuum field between the two surfaces I and II also necessitates the disappearance of charges

---

[1] L. Apker, E. Taft, and J. Dickey, *Phys. Rev.*, **73:** 46 (1948); **74:** 1462 (1948); **76:** 270 (1949). E. Taft and L. Apker, *Phys. Rev.*, **75:** 344 (1949). See also J. A. Becker, *Elec. Eng.*, **68:** 937 (1949).

[2] This is the immediate result of (X.8.05) with $E_{\mathrm{kin}} = (m/2)v_x^2$.

[3] See, for instance, S. Sano, *Elec. J. Tokyo*, **5:** 75 (1941); and H. Benda, *Frequenz*, **7:** 226–232 (1953).

on the two surfaces. The condition "surface charge zero" can be tested experimentally, for example, by moving the two surfaces relative to each other. If the surfaces are not charged, no current can flow through a wire connecting the two solids I and II. These considerations indicate another method[1] for measuring the Volta potential difference. One varies the voltage U_G between the two surfaces until no current flows through a connecting wire when the surfaces are moved relative to each other (see Fig. X.8.7).

This method has also been used by W. E. Meyerhof[2] when he tried to measure the Volta potential difference between strongly n-type and strongly p-type silicon. We mention this particular example because it shows again clearly that two solids can, from the viewpoint of semiconductor physics, be entirely different materials although they differ only by impurities in the order of 10^{-3} and are therefore both "silicon" in ordinary chemical language. In strongly n-type silicon, the Fermi level is close to the lower edge E_C of the conduction band; in strongly p-type silicon, on the other hand, close to the upper edge E_V of the valence band. Hence, according to Fig. X.8.8, a Volta potential difference of almost 1.2 volts must be generated between the two surfaces, corresponding to the width of the forbidden zone in silicon. It must be mentioned that in Fig. X.8.8 the surface double layers on the n-type and on the p-type silicon are assumed to produce jumps of the same magnitude. This is an improbable assumption in the presence of accidental impurities in the surface layers, and hence it is not surprising that the expected potential difference of 1.2 volts has not been measured without special precautions in the cleaning process. The cleaning methods that were used led Meyerhof to a value of only 0.3 volt and W. H. Brattain and W. Shockley[3] to a value of only 0.6 volt.

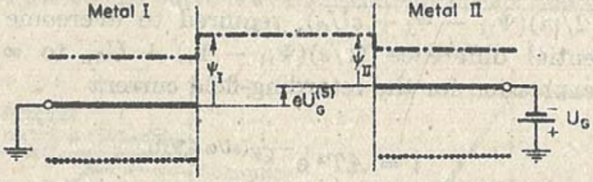


FIG. X.8.6. Saturation occurs when the electrostatic potential in vacuum is a straight horizontal line

$$U_G^{(s)} = \frac{1}{e}(\psi_I - \psi_{II})$$

---

[1] This is the original method of A. Volta, *Ann. chim. et phys.*, 40: 225 (1801).

[2] W. E. Meyerhof, *Phys. Rev.*, 71: 727 (1947).

[3] W. H. Brattain and W. Shockley, *Phys. Rev.*, 72: 345 (1947).

potential difference $= \frac{1}{e}(\psi_{II} - \psi_{I}) + U_G$ independent of distance a.

	(1)	(2)	(3)
distance a:	great	decreasing	small
vacuum field $\frac{\frac{1}{e}(\psi_{II} - \psi_I) + U_G}{a}$:	small	increasing	large
surface charges:	small	increasing	large

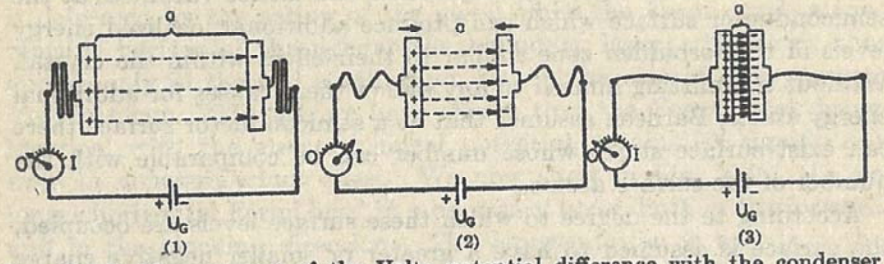


FIG. X.8.7. Measurement of the Volta potential difference with the condenser method. With suitable choice of U_G, namely $U_G = \frac{1}{e}(\psi_I - \psi_{II})$ = Volta potential difference$_{I\ II}$, the vacuum field vanishes (see Fig. X.8.6). In this case the surface charges vanish too for all values of the distance a. Hence on displacing the plates no current will flow in the connecting wire.

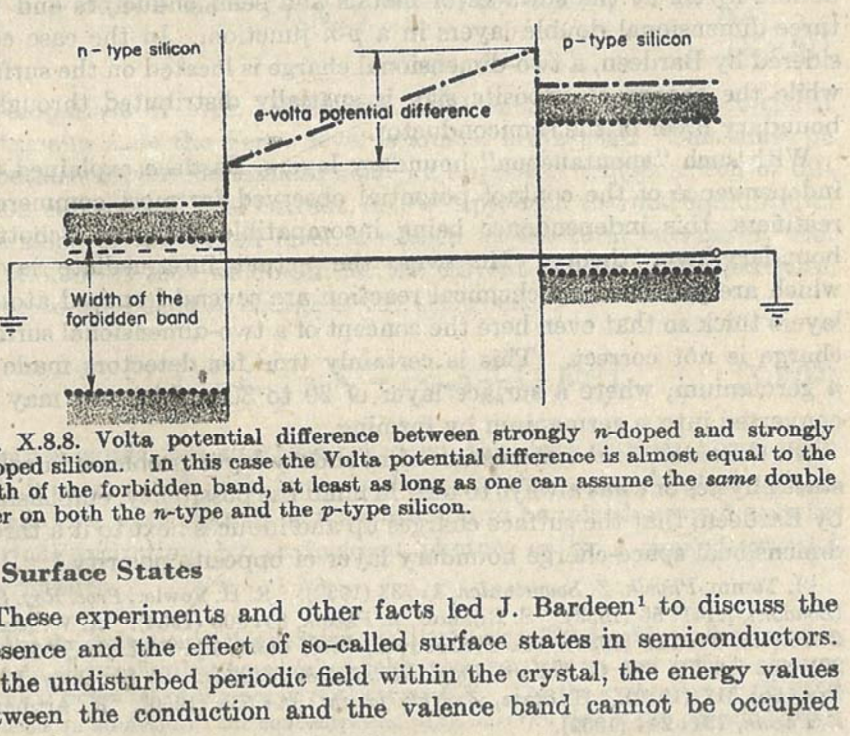


FIG. X.8.8. Volta potential difference between strongly n-doped and strongly p-doped silicon. In this case the Volta potential difference is almost equal to the width of the forbidden band, at least as long as one can assume the *same* double layer on both the n-type and the p-type silicon.

b. Surface States

These experiments and other facts led J. Bardeen[1] to discuss the presence and the effect of so-called surface states in semiconductors. In the undisturbed periodic field within the crystal, the energy values between the conduction and the valence band cannot be occupied

---

[1] J. Bardeen, *Phys. Rev.*, 71: 717 (1947).

under stationary conditions. However, at the surface of the crystal, the periodic potential is radically disturbed because the continuation into the vacuum is not periodic. I. Tamm[1] and a number of other workers after him have discussed the problem as to whether these disturbances create additional energy states in the forbidden zone. We also have to reckon with impurity atoms and lattice vacancies at the semiconductor surface which can produce additional localized energy levels in the forbidden zone similar to their effect within the crystal. Without committing himself to any one of these causes for additional energy states, Bardeen assumes that at a semiconductor surface there can exist surface states whose number can be comparable with the number of the surface atoms.

According to the degree to which these surface levels are occupied, the surface is assumed to have a greater or smaller negative charge which, in turn, generates a positive space-charge layer in the semiconductor. In this way boundary layers are produced, even in free surfaces of semiconductors without a contact, and the whole represents something halfway between the previously mentioned two-dimensional double layers at the surfaces of metals and semiconductors and the three-dimensional double layers in a p-n junction. In the case considered by Bardeen, a two-dimensional charge is located on the surface while the charge of opposite sign is spatially distributed through a boundary layer of the semiconductor.

With such "spontaneous" boundary layers, Bardeen explained the independence of the contact potential observed for most commercial rectifiers, this independence being incompatible with the Schottky boundary-layer theory. However, the actual intermediate layers which are produced by chemical reaction are several hundred atomic layers thick so that even here the concept of a two-dimensional surface charge is not correct. This is certainly true for detectors made of n germanium, where a surface layer of 20 to 30 μ thickness may be converted into p germanium by forming.

All the same, in the discussion of boundary-layer problems in solid-state physics one has always to bear in mind the possibility, emphasized by Bardeen, that the surface charges up and induces next to it a three-dimensional space-charge boundary layer of opposite polarity.[2]

[1] I. Tamm, *Physik. Z. Sowjetunion*, **1**: 733 (1932). R. H. Fowler, *Proc. Roy. Soc. (London)*, **A141**: 56 (1933). S. Rijanow, *Z. Physik*, **89**: 806 (1934). A. W. Maue, *Z. Physik*, **94**: 717 (1935). E. T. Goodwin, *Proc. Cambridge Phil. Soc.*, **35**: 205, 221, 232 (1939). W. G. Pollard, *Phys. Rev.*, **56**: 324 (1939). W. Shockley, *Phys. Rev.*, **56**: 317 (1939). H. Statz, *Z. Naturforsch*,. **5A**: 534 (1950). K. Artmann, *Z. Physik*, **131**: 244 (1952).

[2] See W. H. Brattain and J. Bardeen, *Bell System Tech. J.*, **32**: 1 (1953).

§9. The Electrochemical Potentials $E_F^{(n)}$ and $E_F^{(p)}$ of Electrons and Holes

In §6 and §7 we have shown how in the boundary layer of a semiconductor-metal contact the whole band diagram of the semiconductor is deformed by the action of the metal while the Fermi level, alone, remains horizontal throughout the boundary layer. However, this applies only in thermal equilibrium, i.e., in the case of zero current. Yet, in Chap. VIII, §3, we have shown that the Fermi level, being identical with the electrochemical potential, retains its significance even in nonequilibrium cases. We now show in Fig. X.9.1 the no longer horizontal Fermi level in a boundary layer, both in the forward and in the blocking direction. The representation is based on the same assumptions as in Fig. IV.5.1. From the plot of the concentration $n(x)$ in this figure, we obtain the plot of $E_F^{(n)}(x)$ with the aid of the defining equation[1]

$$E_F^{(n)} = -eV(x) + E_C + \zeta\left(\frac{n(x)}{N_C}\right) \approx -eV(x) + E_C + \mathrm{k}T \ln \frac{n(x)}{N_C} \tag{X.9.01}$$

Comparison of Figs. IV.5.1 and X.9.1 shows that in the so-called Boltzmann zone the Fermi level is almost horizontal. This must be so because in the Boltzmann zone we approach compensation of diffusion current and field current, i.e., we approach thermal equilibrium. The slope of the Fermi level is related to the total current by Eq. (VIII.3.05) so that we obtain for the current density of the particles, owing to the negative charge of the electrons,

$$\left(\frac{1}{-e}\, i_{\mathrm{tot}}\right) = s_n = -\mu_n n \frac{d}{dx}\left(\frac{1}{e} E_F^{(n)}\right) \tag{X.9.02}$$

In other words, the electrons roll down the slope of their Fermi level $E_F^{(n)}$, in agreement with their tendency in the energy-band diagrams to fall downward of their own accord and to be raised upward only by thermal excitation, by an incident photon, or by a similar external interference.

---

[1] For the derivation of this defining equation, see pp. 359 ff. We shall soon introduce an electrochemical potential of holes, and therefore we now call the electrochemical potential of the negative electrons $E_F^{(n)}$. For the evaluation of (X.9.01) we need, in addition to the concentration $n(x)$, also the plot of the potential $V(x)$. See E. Spenke, *Z. Physik*, **126**: 67 (1949).

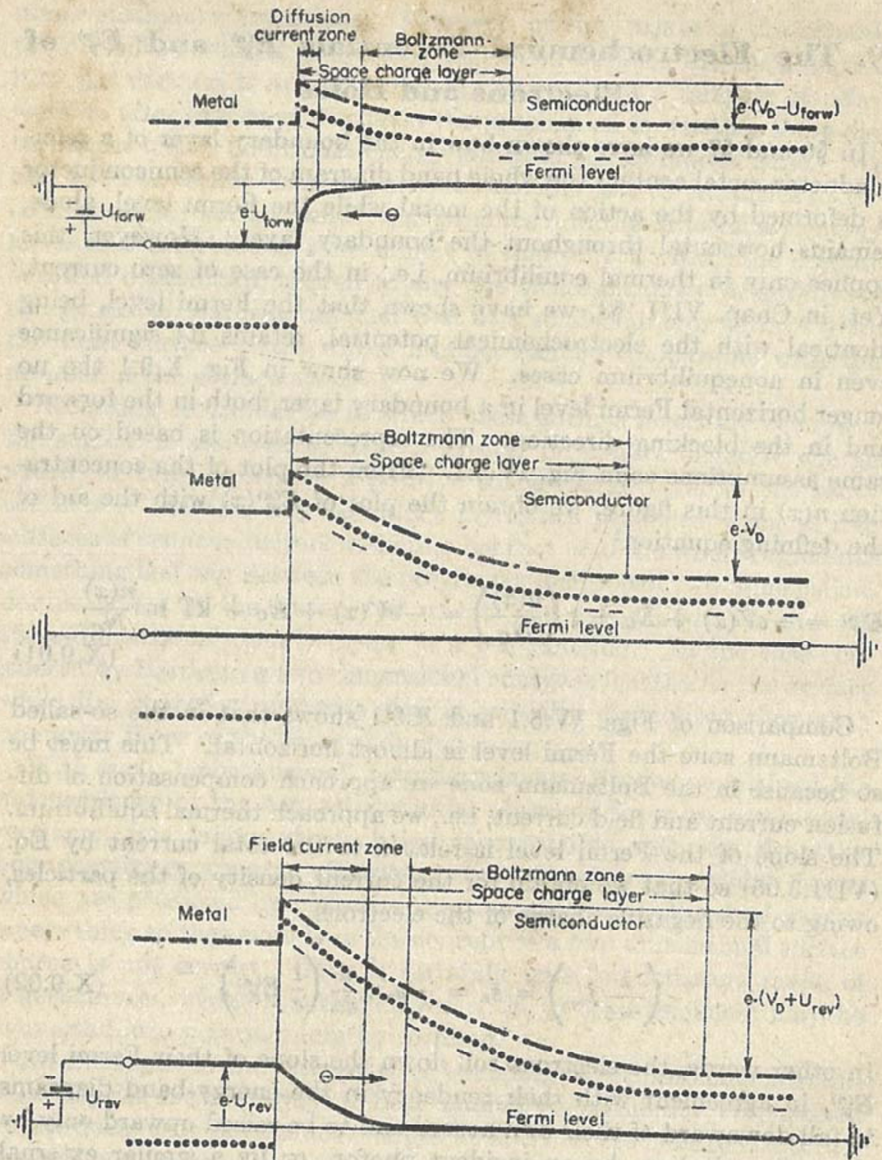


FIG. X.9.1. Plot of the Fermi level in the depletion boundary layer of an n-type semiconductor. Top: case of current flow, direction of current: ← ⊖. Middle: case of no current flow. Bottom: case of blocking, direction of current: ⊖ →.

Exactly the opposite applies to the holes. Let us consider, for example, the recombination of a hole $\oplus$ with a negative acceptor A^-, which is accompanied by a loss of energy[1] and therefore takes place spontaneously. If we, further, remember that this is represented in the energy-band diagram by the rise of a hole from the valence band into an acceptor level above it, we can see that the holes, like air bubbles in water, have an urge to rise upward in the energy-band diagram and that only external effects, such as thermal or optical excitation, can push them down. If we now wish to define the electrochemical potential or the Fermi level $\boldsymbol{E}_F^{(p)}$ of holes, we have to expect that the holes climb up the slope of this Fermi level and that, accordingly, the current density of the particles is given by the equation

$$s_p = +\mu_p p \frac{d}{dx}\left(\frac{1}{e}\boldsymbol{E}_F^{(p)}\right) \tag{X.9.03}$$

If this is, in fact, to be correct, we have to define

$$\boldsymbol{E}_F^{(p)} = E_V - eV(x) - \zeta\left(\frac{p(x)}{N_V}\right) \approx E_V - eV(x) - \mathrm{k}T \ln \frac{p(x)}{N_V} \tag{X.9.04}$$

because then we obtain

$$\frac{d\boldsymbol{E}_F^{(p)}}{dx} = +e(-V'(x)) - \frac{d\zeta(p)}{dp}p'(x) \approx +e(-V'(x)) - \mathrm{k}T\frac{1}{p}p'(x)$$

or

$$+\mu_p \cdot p \cdot \frac{d}{dx}\left(\frac{1}{e}\boldsymbol{E}_F^{(p)}\right) = +\mu_p p\boldsymbol{E}(x) + \frac{\mu_p}{e}\frac{d\zeta(p)}{d\ln p}(-p'(x))$$
$$\approx \mu_p p\boldsymbol{E}(x) + \mu_p\frac{\mathrm{k}T}{e}\cdot(-p'(x))$$

and with the aid of the Nernst-Townsend-Einstein relation (VIII.3.10) and its generalization (VIII.3.08) we obtain in fact

$$+\mu_p p \frac{d}{dx}\left(\frac{1}{e}\boldsymbol{E}_F^{(p)}\right) = s_{\text{field}} + s_{\text{diff}} = s_p$$

The signs have been chosen accordingly in the diagrams of Fig. X.9.2, in which the depletion boundary layer of an n-type semiconductor for the cases of forward, back, and zero current is shown. We need not go into further details, but we want to mention that the choice of signs for the definition of the Fermi level $\boldsymbol{E}_F^{(p)}$ of holes has the additional advantage that in thermal equilibrium we obtain

$$\boldsymbol{E}_F^{(p)} \equiv \boldsymbol{E}_F^{(n)} \equiv \boldsymbol{E}_F \tag{X.9.05}$$

---

[1] In this process the hole loses energy while the lattice (for example) gains energy.

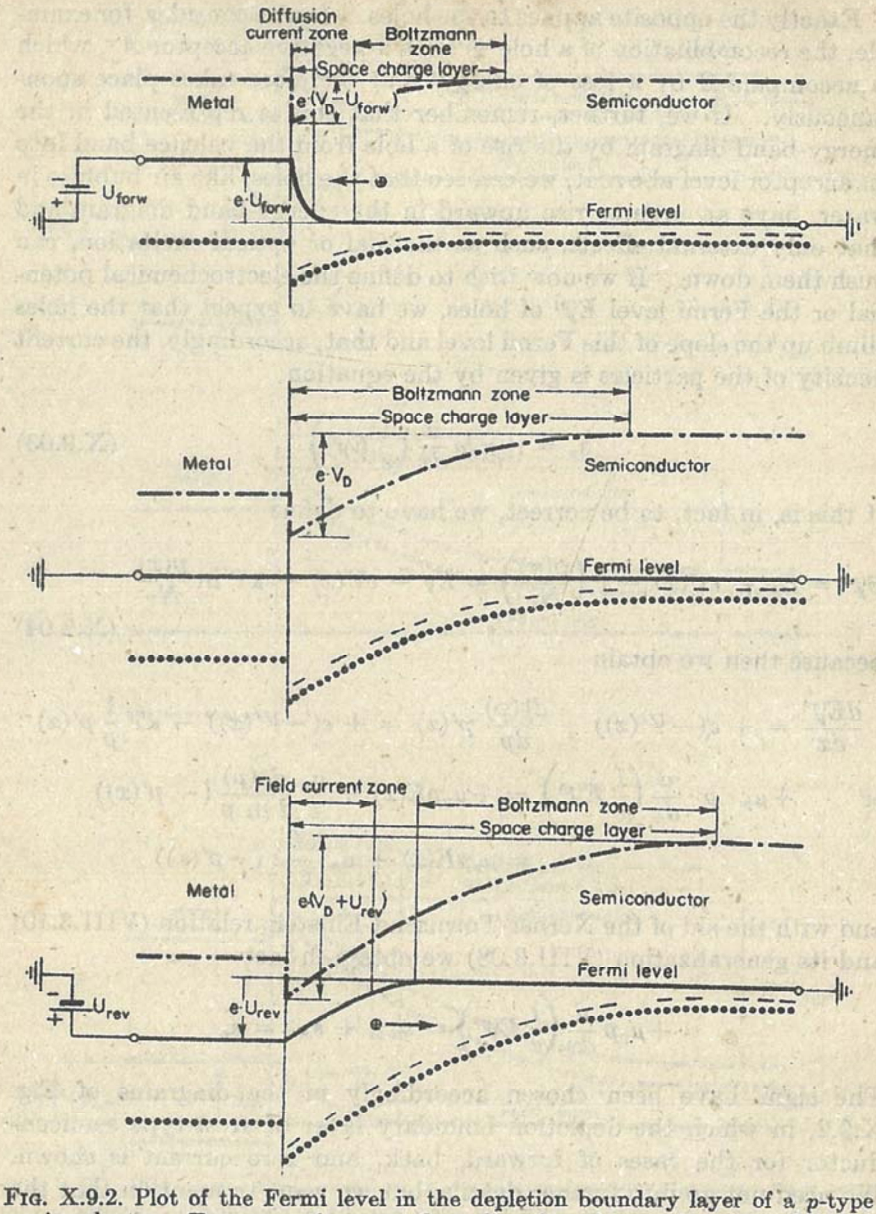


FIG. X.9.2. Plot of the Fermi level in the depletion boundary layer of a p-type semiconductor. Top: case of current flow, direction of current: ← ⊕. Middle: case of no current flow. Bottom: case of blocking, direction of current: ⊕ →.

For thermal equilibrium, we can use in the defining equation (X.9.04) the generalized law of mass action (VIII.4.18), and we can write

$$\boldsymbol{E}_F^{(p)} = E_V - eV(x) - \zeta\left(\frac{p}{N_V}\right) = E_V - eV(x) - \left(E_V - E_C - \zeta\left(\frac{n}{N_C}\right)\right)$$

Hence
$$\boldsymbol{E}_F^{(p)} = E_C - eV(x) + \zeta\left(\frac{n}{N_C}\right)$$

and with (X.9.01)
$$\boldsymbol{E}_F^{(p)} = \boldsymbol{E}_F^{(n)}$$

Thus by the chosen definition we have achieved that in thermal equilibrium the electrochemical potentials $\boldsymbol{E}_F^{(n)}$ and $\boldsymbol{E}_F^{(p)}$ of electrons and holes coincide in the common horizontal Fermi level $\boldsymbol{E}_F$. If current flows, i.e., in nonequilibrium, the two Fermi levels $\boldsymbol{E}_F^{(n)}$ and $\boldsymbol{E}_F^{(p)}$ separate. They also are no longer horizontal; instead their slopes represent the forces acting on the two current densities of particles s_n and s_p, which can also be computed from (X.9.02) and (X.9.03). This means that the electrons "voluntarily" fall down along their Fermi level $\boldsymbol{E}_F^{(n)}$ while the holes "voluntarily" climb up their Fermi level $\boldsymbol{E}_F^{(p)}$.

Using these principles, we now draw in Fig. X.9.3 band diagrams of a *p-n* junction for forward current, reverse current, and thermal equilibrium. We use as a basis the special case of low recombination according to Shockley. Then we have still approximately Boltzmann equilibrium (see page 100) in the junction zone, and therefore the Fermi levels in this zone have to be drawn approximately horizontal. From this we obtain for the separation of $\boldsymbol{E}_F^{(n)}$ and $\boldsymbol{E}_F^{(p)}$ in the junction zone the values eU_{forw} and eU_{rev}. Hence we get for the concentrations at the beginning of the diffusion tails the values

$$n_p\,\mathrm{e}^{+\frac{e}{kT}U_{\text{forw}}}\ \left(\text{or } n_p\,\mathrm{e}^{-\frac{e}{kT}U_{\text{rev}}}\right) \quad \text{and} \quad p_n\,\mathrm{e}^{+\frac{e}{kT}U_{\text{forw}}}\ \left(\text{or } p_n\,\mathrm{e}^{-\frac{e}{kT}U_{\text{rev}}}\right)$$

This is so because the distance of the Fermi level $\boldsymbol{E}_F^{(n)}$ from the lower edge $-eV + E_C$ of the conduction band, i.e., the term $(-eV + E_C) - \boldsymbol{E}_F^{(n)}$, is, according to Eq. (X.9.01), a logarithmic measure of the electron concentration n in the Maxwell-Boltzmann case. Corresponding considerations apply to the hole concentration according to Eq. (X.9.04).

On page 100, these results were explained by the identity of the logarithmically plotted concentration curve with the potential curve $V(x)$ in the case of exact or approximate Boltzmann equilibrium.

Finally, we have to mention that the case

$$\boldsymbol{E}_F^{(n)} > \boldsymbol{E}_F^{(p)}$$

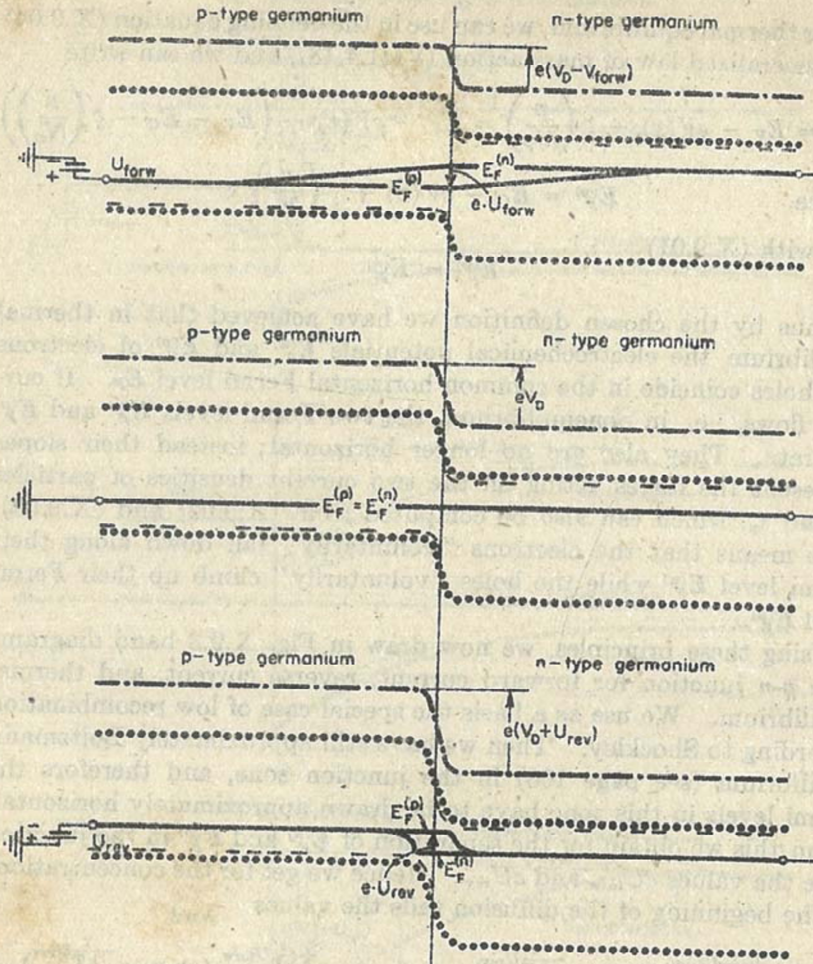


FIG. X.9.3. Plot of the Fermi levels $E_F^{(n)}$ and $E_F^{(p)}$ in a p-n junction. Top: case of current flow: ⊕ → ← ⊖. Middle: case of no current flow. Bottom: case of blocking, direction of current: ← ⊕ ⊖ →.

leads, according to the defining equations (X.9.01) and (X.9.04), to

$$E_C - E_V > \mathrm{k}T \ln \frac{N_C N_V}{np}$$

and with the aid of (VIII.4.23) to

$$np > n_i^2$$

and thus, according to page 26, signifies that recombination outweighs dissociation. According to Fig. X.9.3, top, the junction zone of a *p-n* junction is such a region if it is polarized in the forward direction (see page 94).

Correspondingly, $E_F^{(n)} < E_F^{(p)}$ signifies a region where dissociation predominates. An example, according to Fig. X.9.3, bottom, is the junction zone of a *p-n* junction which is polarized in the reverse direction.

§10. Problems

1. A parallel-plate condenser of 1 cm^2 area has two plates made of different metals. The condenser plates are connected by a wire. Calculate the current in the wire if the two metals have a work function difference of 1 volt, and if the two plates are vibrating against each other at a frequency of 100 cycles/sec such that their separation varies between d_1 and d_2. Assume $d_1 = 0.01$ cm, and calculate the current as a function of d_2. For what value of d_2 does the current attain one-half of its maximum possible value? *Note:* The current will, in general, not be sinusoidal. Use the average of the absolute value $\overline{|i|}$ as a measure.

2. The two plates in a vibrating condenser such as that of Prob. 1 are connected through a battery with a fixed voltage of 1.5 volts. The current is found to be 1.2 ma. If the battery is reversed, a current of 0.7 ma and of opposite phase is obtained. What is the work function difference of the two metals?

3. According to Richardson's law the plot of log (i_S/T^2) versus $1/T$ should be a straight line. The slope of the straight line is determined by the work function while the intercept with the ordinate axis should be the same for all materials.

a. Show that the plot is still a straight line if the work function varies linearly with temperature. What are now the values of the slope of the line and of the intercept?

b. A plot of log (i_S/T^2) vs. $1/T$ for a certain metal is a straight line with a slope corresponding to 6.50 ev while the intercept is 6,500 $\text{amp/cm}^2/°\text{C}^2$ rather than 120 $\text{amp/cm}^2/°\text{C}^2$. What is the actual work function at 800°C if Richardson's law is assumed to be correct?

4. An electron which is held a short distance from a metal surface induces positive charge on this surface which, in turn, attracts the electron (the so-called "image force"). Calculate the image force and the potential energy of the electron if the potential energy at infinite distance is normalized to zero. At what distance from the metal surface is the image force equal to the force exerted by an electric field of 10^4 volt-cm^{-1} on the electron?

5.* Calculate the total potential energy of an electron under the influence of both the image force and an electric field of a polarity such as to pull electrons out of the metal. As a reference potential energy, choose the potential energy right at the metal surface in the absence of an image force. How much is the work function lowered by an electric field of $5 \cdot 10^4$ volt cm^{-1}?

6. How are the results of Prob. 5 modified if the electron is not in vacuum but in a semi-conductor with a dielectric constant $\varepsilon = 16$?

APPENDIX I

The Integral $\int\limits_{x=-\frac{G}{2}a}^{x=+\frac{G}{2}a} U(x)\, e^{jkx}\, dx$

In the integral over the fundamental domain

$$I_k = \int\limits_{x=-\frac{G}{2}a}^{x=+\frac{G}{2}a} U(x)\, e^{jkx}\, dx \tag{A.I.1}$$

let $U(x)$ be a lattice-periodic function

$$U(x - na) \equiv U(x) \tag{A.I.2}$$

Let the wave number k have the following values, compatible with the requirement of periodicity in the fundamental domain [see Eq. (VII.2.05)]:

$$k = \frac{2\pi}{a} \cdot \frac{n}{G} \qquad n = 0, \pm 1, \pm 2, \ldots, \pm \frac{G}{2} \tag{A.I.3}$$

We then maintain that

$$I_k = 0 \qquad \text{except for } k = 0 \tag{A.I.4}$$

To prove this we carry out the integration, step by step, for one lattice cell at a time:

$$I_k = \sum_{n=-\frac{G}{2}}^{n=+\frac{G}{2}-1} \int\limits_{x=na}^{x=(n+1)a} U(x)\, e^{jkx}\, dx \tag{A.I.5}$$

Now we split off in the integrand the phase rotation e^{jkna} which is already present at the beginning of the nth cell and leave under the integral sign only the advance of the phase rotation within the cell $e^{jk(x-na)}$. With the use of (A.I.2) we obtain

$$I_k = \sum_{n=-\frac{G}{2}}^{n=+\frac{G}{2}-1} e^{jkna} \int\limits_{x-na=0}^{x-na=a} U(x-na)\, e^{jk(x-na)}\, dx \tag{A.I.6}$$

Introducing the integration variable $x - na = \xi$, we see that the integral is now independent of the summation index and can therefore be placed in front of the sum sign:

$$\begin{aligned} I_k &= \int\limits_{\xi=0}^{\xi=a} U(\xi)\, e^{jk\xi}\, d\xi \cdot \sum_{n=-\frac{G}{2}}^{n=+\frac{G}{2}-1} e^{jkna} \\ &= \int\limits_{\xi=0}^{\xi=a} U(\xi)\, e^{jk\xi}\, d\xi \cdot e^{-jka\frac{G}{2}} \cdot \sum_{n+\frac{G}{2}=0}^{n+\frac{G}{2}=G-1} e^{jka\left(n+\frac{G}{2}\right)} \end{aligned} \tag{A.I.7}$$

The sum is now a simple geometric progression

$$1 + q + q^2 \cdots + q^l = \frac{(1 - q^{l+1})}{(1 - q)}$$

and by carrying out the summation with $q = e^{jkna}$ and $l = G - 1$, i.e., $l + 1 = G$, we obtain for the wanted integral

$$I_k = \int\limits_{\xi=0}^{\xi=a} U(\xi)\, e^{jk\xi}\, d\xi \cdot e^{-jka\cdot\frac{G}{2}} \cdot \frac{1 - e^{jka\cdot G}}{1 - e^{jka}} \tag{A.I.8}$$

According to (A.I.3) we have $ka \cdot G = 2\pi n$. Therefore, the last numerator in (A.I.8) becomes $1 - 1 = 0$ and the inegral I_k vanishes, as was stated. The argument fails only when $k = 0$ because then each term of the geometric progression in (A.I.7) will be one to begin with and the summation formula $(1 - q^{l+1})/(1 - q)$ will be unusable because numerator and denominator vanish simultaneously. The result of the summation will then simply be G, and the integral will be

$$I_0 = G \cdot \int\limits_{\xi=0}^{\xi=a} U(\xi)\, d\xi \tag{A.I.9}$$

while we have

$$I_k = 0 \quad \text{for } k \neq 0 \tag{A.I.10}$$

Thus the preceding statement is proved in its entirety.

In conclusion, we mention that integrals of the type under consideration occur in the computation of coefficients of a Fourier expansion with the fundamental domain $G \cdot a$ as period. In this context, the proof just given means that in such a Fourier expansion the G Fourier coefficients

$$C_{-\frac{G}{2}};\ C_{-\frac{G}{2}+1};\ \cdots\ C_{-1}\quad C_{+1};\ \cdots\ C_{+\frac{G}{2}-1};\ C_{+\frac{G}{2}}$$

vanish if the represented function has, in reality, not the period $G \cdot a$, but the much smaller period a. This is so because in the Fourier expansion of such a function the period a is the greatest wavelength which occurs. However, the coefficients just mentioned are associated with the wavelengths $2\pi/|k|$, i.e., according to Eq. (A.I.3)

$$\frac{G\cdot a}{\frac{G}{2}};\ \frac{G\cdot a}{\frac{G}{2}-1};\ \cdots\ \frac{G\cdot a}{1}\qquad \frac{G\cdot a}{1};\ \cdots\ \frac{G\cdot a}{\frac{G}{2}-1};\ \frac{G\cdot a}{\frac{G}{2}}$$

or

$$2a;\ \frac{1}{1-\frac{2}{G}}\,2a;\ \cdots\ G\cdot a\qquad G\cdot a;\ \cdots\ \frac{1}{1-\frac{2}{G}}\,2a;\quad 2a$$

which are all greater than a.

APPENDIX II

The Function $\zeta(n/N)$

We define a function $\zeta(n/N)$ by the equation

$$\frac{2}{\sqrt{\pi}} \int_{\eta=0}^{\eta=\infty} \frac{1}{e^{\eta-\frac{\zeta}{kT}}+1} \sqrt{\eta}\, d\eta = \frac{n}{N} \qquad \text{(A.II.1)}$$

We want to find out how the quantity ζ depends on n/N.

Since the definition (A.II.1) is solved for the independent variable n/N, we determine, first, the inverse function n/N as a function of ζ. Here we shall discuss in detail only the two limiting cases "ζ strongly negative" and "ζ strongly positive."

1. $\zeta \to -\infty$. In this case the exponential term in the denominator of the integrand is very large compared with 1. Hence we can write as a first approximation

$$\frac{2}{\sqrt{\pi}} \int_{\eta=0}^{\eta=\infty} e^{-\eta+\frac{\zeta}{kT}} \sqrt{\eta}\, d\eta \approx \frac{n}{N}$$

$$e^{+\frac{\zeta}{kT}} \cdot \frac{2}{\sqrt{\pi}} \int_{\eta=0}^{\eta=\infty} e^{-\eta} \sqrt{\eta}\, d\eta = e^{\frac{\zeta}{kT}} \cdot 1 \approx \frac{n}{N}$$

Since we have assumed the term ζ to be strongly negative to obtain this approximation, we have

$$\frac{n}{N} \ll 1 \qquad \text{or} \qquad n \ll N$$

Solving the approximation $e^{\frac{\zeta}{kT}} \approx n/N$ derived for ζ; we obtain the wanted function $\zeta(n/N)$ for the limiting case $n \ll N$:

$$\zeta \approx kT \ln \frac{n}{N} \qquad \text{(A.II.2)}$$

2. $\zeta \to +\infty$. In this case the exponential term in the denominator of the integrand is very small compared with 1 as long as $\eta < \zeta/kT$.

As soon as $\eta > \zeta/\mathrm{k}T$, the exponential term very soon outweighs the term 1 and greatly reduces the value of the integrand. We, therefore, obtain an approximation if in the integration range $0 < \eta < \zeta/\mathrm{k}T$ we write 1 for the denominator and in the integration range $\zeta/\mathrm{k}T < \eta < \infty$ the value ∞. Thus we obtain the approximation

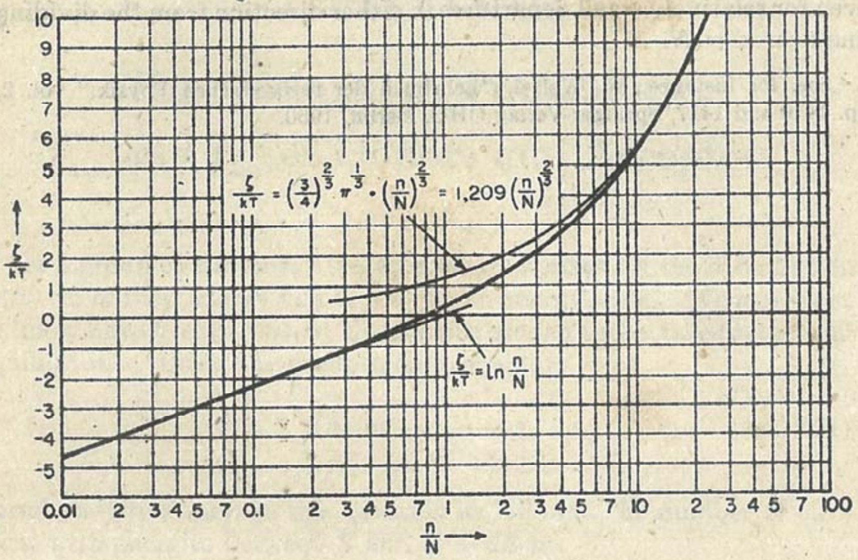


FIG. A.II.1. The function $\frac{\zeta}{\mathrm{k}T}$.

$$\frac{2}{\sqrt{\pi}} \int_{\eta=0}^{\eta=\frac{\zeta}{\mathrm{k}T}} \sqrt{\eta}\, d\eta \approx \frac{n}{N}$$

$$\frac{2}{\sqrt{\pi}} \cdot \frac{2}{3} \cdot \left(\frac{\zeta}{\mathrm{k}T}\right)^{3/2} \approx \frac{n}{N}$$

To obtain this approximation we assumed the term ζ to be strongly positive. Hence we now have

$$\frac{n}{N} \gg 1 \qquad \text{or} \qquad n \gg N$$

If we solve $\frac{2}{\sqrt{\pi}} \cdot \frac{2}{3} \left(\frac{\zeta}{\mathrm{k}T}\right)^{3/2} \approx \frac{n}{N}$ for ζ, we obtain the wanted function $\zeta(n/N)$ for the limiting case $n \gg N$:

$$\zeta \approx \mathrm{k}T \left(\frac{3}{4}\right)^{2/3} \pi^{1/3} \left(\frac{n}{N}\right)^{2/3} \qquad \text{(A.II.3)}$$

More accurate evaluations of the integral (A.II.1) which go beyond these two limiting cases use expansions of power series.[1] Figure A.II.1 shows ζ/kT as a function of n/N. We can see that the asymptotic variations (A.II.2) and (A.II.3) are valid with great accuracy even for relatively small departures in either direction from the dividing line near $n = N$.

---

[1] See, for instance, W. Weizel, "Lehrbuch der theoretischen Physik," vol. 2, pp. 1416 and 1417, Springer-Verlag OHG, Berlin, 1950.

APPENDIX III

The Occupation Probabilities f_{don} and f_{acc} of Donors and Acceptors[1]

Let a group of stationary states be characterized by the distribution $D(E)\,dE$ of their energy values E along an energy scale. We now want to know how N electrons are distributed among these states in thermal equilibrium. Using the occupation probability

$$f(E) = \frac{1}{e^{\frac{E-E_F}{kT}} + 1} \qquad \text{(A.III.01)}$$

Fermi statistics answers this question as follows: The number of electrons with energies between E and $E + dE$ is

$$N(E)\,dE = 2D(E) \cdot f(E) \cdot dE \qquad \text{(A.III.02)}$$

If now the problem has to be solved how N electrons have to be distributed among the states of the conduction band and among N_D donors, the following seems to be the simple answer: The number

$$N - N_{D^\times} = \int_{E=E_C}^{\infty} 2D(E)f(E)\,dE = 2\int_{E=E_C}^{\infty} \frac{D(E)}{e^{\frac{E-E_F}{kT}} + 1}\,dE \qquad \text{(A.III.03)}$$

goes into the conduction band and the number

$$N_{D^\times} = N_D \cdot f(E_D) = N_D \frac{1}{e^{\frac{E_D-E_F}{kT}} + 1} \qquad \text{(A.III.04)}$$

goes into the N_D donor levels $E = E_D$.

---

[1] See on this subject, N. F. Mott and R. W. Gurney, "Electronic Processes in Ionic Crystals," pp. 157ff., Clarendon Press, Oxford, 1948. W. Shockley, "Electrons and Holes in Semiconductors," p. 248, prob. 1, and p. 475, prob. 2, D. Van Nostrand Company, Inc., Princeton, N.J., 1950. P. T. Landsberg, *Proc. Phys. Soc.* (*London*), **65A:** 604 (1952). E. A. Guggenheim, *Proc. Phys. Soc.* (*London*), **66A:** 121 (1953). P. T. Landsberg, *Proc. Phys. Soc.* (*London*), **66A:** 662 (1953). J. H. Crawford and D. K. Holmes, *Proc. Phys. Soc.* (*London*), **67A:** 294 (1954).

In this solution the N_D donors are considered as N_D available places, but this involves an incorrect assumption. An electron can be attached with *two* different spin directions to an empty donor, in the same way as it can be attached to a hydrogen nucleus. At first sight one might think that the mistake could be eliminated by replacing N_D with $2N_D$. But this procedure ignores the fact that when an electron has been attached to a positive donor core the latter has become electrically neutral so that a second electron with opposite spin no longer encounters a potential trough and thus cannot be attached, for electrostatic reasons. In other words, *before* an electron has been attached the donor core offers *two* places to an electron, but *after* the electron has been attached it offers only *one* place, which is already occupied by the attached electron. This means that N_D donors offer neither N_D available places nor twice that number $2N_D$. Rather, the number of available places depends on the momentary degree of occupation, one might say it changes during the process of occupation. Thus it is apparent that the distribution of N electrons among the conduction band and the N_D donor levels is not the type of problem in which a *fixed* number of available places have to be filled with N electrons.[1] Hence, to solve this problem we have to reconsider the usual arguments, which in normal problems lead to the occupation probability

$$f(E) = \frac{1}{e^{\frac{E-E_F}{kT}} + 1} \qquad \text{(A.III.01)}$$

---

[1] The "normal" type of problem (occupation of N_P places P with the energy value E_P) still has the same solution

$$N_P \frac{1}{e^{\frac{E_P-E_F}{kT}} + 1}$$

If E_F lies several kT below E_P, this expression takes on the simpler form

$$N_P e^{+\frac{E_F-E_P}{kT}}$$

This remark is relevant for the footnote 1, p. 318.

The occupation of the conduction band alone is covered by the *normal* type of problem, because the quantum states of the conduction band result from a one-electron problem with fixed potential which has been chosen in such a way that the electrostatic effects of the many electrons which have later to be accommodated in these quantum states have been taken into account (self-consistent potential). Therefore, on pages 285 and 299, the possibility of spin reversal in the states of the conduction band could be allowed for simply by a factor 2 in front of the number of these states.

However, we shall have to refer to the literature[1] for the wave-mechanical basis of the counting rules ("each place can be occupied by only 0 or 1 electron," "the electrons cannot be distinguished from each other").

We subdivide the energy scale into intervals 1, 2, . . . i which are grouped around the values $E_1, E_2, \ldots E_i$ and contain $Z_1, Z_2, \ldots Z_i$ available places. The occupation of the ith interval with N_i electrons can be realized in

$$\frac{Z_i!}{N_i!(Z_i - N_i)!} \tag{A.III.05}$$

different ways. This can be derived as follows: One numbers the Z_i available places of the considered ith interval from 1 to Z and arranges them in this sequence (see Fig. A.III.1, top). Then, starting from the left, one fills the N_i first places and leaves the remaining $(Z_i - N_i)$ places unoccupied. The occupation of the ith interval with N_i electrons can be realized in a different way if one arranges the Z_i available places in a different sequence, i.e., in a different permutation, and if one again fills the N_i places on the left while the $(Z_i - N_i)$ places to the right remain empty (see Fig. A.III.1, second line). This procedure can be repeated $Z_i!$ times, but it leads only to a novel possibility if during the permutation at least one unoccupied place is exchanged for a filled one. Permutation of the N_i filled places at the left among themselves (Fig. A.III.1: transition from second to third line) and

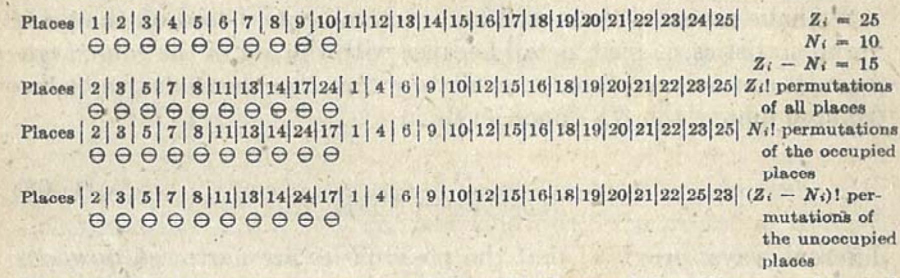


FIG. A.III.1. Different occupations of $Z_i = 25$ places of the ith energy interval with $N_i = 10$ electrons.

$$\frac{Z_i!}{N_i!(Z_i - N_i)!} \text{ different combinations are possible.}$$

---

[1] See, for instance, L. Nordheim in Müller-Pouillet, "Lehrbuch der Physik," vol. IV, part 4, p. 251, Vieweg-Verlag, Brunswick, Germany, 1933; or R. C. Tolman, "The Principles of Statistical Mechanics," pp. 364ff., Oxford University Press, 1938; or W. Weizel, "Lehrbuch der theoretischen Physik," vol. II, pp. 1193 and 1044ff., Springer-Verlag OHG, Berlin, 1950.

permutation of the $(Z_i - N_i)$ unoccupied places at the right (Fig. A.III.1: transition from third to fourth line) do *not* lead to a novel possibility. Therefore the number of possibilities for the occupation of the ith interval with N_i electrons is not $Z!$, but only

$$\frac{Z_i!}{N_i!(Z_i - N_i)!} \tag{A.III.06}$$

Occupation of the whole energy range with $N_1, N_2, \ldots N_i$ electrons can therefore be realized in

$$\prod_{i=1}^{\infty} \frac{Z_i!}{N_i!(Z_i - N_i)!} \tag{A.III.07}$$

different ways.

We have repeated these well-known arguments, based on the usual Fermi statistics, in such detail because with the aid of the same arguments it can now be stated immediately that $N_{D^\times}$ electrons can be distributed among the N_D donor levels in

$$\frac{N_D!}{N_{D^\times}!(N_D - N_{D^\times})!} \tag{A.III.08}$$

different ways, provided that the possibilities are narrowed down by allowing, for instance, only electrons with clockwise spin. In this case the problem is identical with the one just discussed in detail, namely, how N_i electrons can be distributed among the Z_i available places of the ith interval.

However, in the occupation of the donors it is by no means only the electrons with clockwise spin that are allowed. If the spin of only one donor electron, e.g., at the donor farthest to the left, is reversed

$$\frac{N_D!}{N_{D^\times}!(N_D - N_{D^\times})!}$$

new possibilities are created, corresponding to a factor $2 = 2^1$ in front of the number of possibilities. In the same way the spin reversal of all $N_{D^\times}$ accommodated electrons produces a factor 2 for each electron, hence a total factor of $2^{N_D{}^\times}$, so that finally

$$2^{N_D{}^\times} \frac{N_D!}{N_{D^\times}!(N_D - N_{D^\times})!} \tag{A.III.09}$$

is the number of possibilities in which the $N_D{}^\times$ electrons can be accommodated in N_D donor levels.

With the aid of (A.III.09) and (A.III.07) we thus obtain for the distribution, characterized by $N_{D^\times}$, N_1, N_2, . . . N_i of

$$N_{D^\times} + \sum_{i=1}^{\infty} N_i = N \tag{A.III.10}$$

electrons among N_D donor levels and the conduction band the following number of possibilities

$$W = 2^{N_{D^\times}} \frac{N_D!}{N_{D^\times}!(N_D - N_{D^\times})!} \prod_{i=1}^{\infty} \frac{Z_i!}{N_i!(Z_i - N_i)!} \tag{A.III.11}$$

In equilibrium the distribution with the largest number of realizable possibilities will be established. The occupation numbers $N_{D^\times}$, N_1, N_2, . . . N_i have to be varied until W or

$$\ln W = \begin{cases} N_{D^\times} \ln 2 + \ln N_D! - \ln N_{D^\times}! - \ln (N_D - N_{D^\times})! \\ + \sum\limits_{i=1}^{\infty} \{\ln Z_i! - \ln N_i! - \ln (Z_i - N_i)!\} \end{cases} \tag{A.III.12}$$

is a maximum. But it must be remembered that, in addition to the side condition (A.III.10), another side condition must be fulfilled, namely, that the total energy

$$U = N_{D^\times}E_D + \sum_{i=1}^{\infty} N_iE_i \tag{A.III.13}$$

must be conserved. We take the side conditions (A.III.10) and (A.III.13) into account, using two Lagrange factors α and β in the conventional way, by differentiating, not $\ln W$, but

$$\begin{aligned} \ln W + \alpha\Big(N - N_{D^\times} - \sum_{i=1}^{\infty} N_i\Big) &+ \beta\Big(U - N_{D^\times}E_D - \sum_{i=1}^{\infty} N_iE_i\Big) \\ &= \text{const} + N_{D^\times} \ln 2 - \ln N_{D^\times}! - \ln (N_D - N_{D^\times})! \\ &\quad - \sum_{i=1}^{\infty} \{\ln N_i! + \ln (Z_i - N_i)!\} + \alpha N + \beta U - \alpha N_{D^\times} \\ &\quad - \beta N_{D^\times}E_D - \sum_{i=1}^{\infty} \{\alpha N_i + \beta N_iE_i\} \end{aligned} \tag{A.III.14}$$

for the variables N_1, N_2, . . . , N_i, . . . , $N_{D^\times}$, α, and β and by equating the derivations with zero. In this procedure we use the Stirling formula

$$\ln A! \approx A \ln A - A \tag{A.III.15}$$

which leads to the differentiation rule

$$\frac{d}{dA} \ln A! \approx 1 \times \ln A + A \frac{1}{A} - 1 = \ln A \qquad \text{(A.III.16)}$$

By differentiation for each of the variables $N_1, N_2, \ldots, N_i, \ldots$ we obtain a system of equations

$$- \ln N_i + \ln (Z_i - N_i) - (\alpha + \beta E_i) = 0 \qquad i = 1, 2, \ldots \qquad \text{(A.III.17)}$$

Differentiating (A.III.14) for $N_{D^\times}$ we obtain

$$+ \ln 2 - \ln N_{D^\times} + \ln (N_D - N_{D^\times}) - (\alpha + \beta E_D) = 0 \qquad \text{(A.III.18)}$$

and differentiating for α and β we finally obtain again the side conditions

$$\sum_{i=1}^{\infty} N_i + N_{D^\times} = N \qquad \text{(A.III.10)}$$

$$\sum_{i=1}^{\infty} N_i E_i + N_{D^\times} E_D = U \qquad \text{(A.III.13)}$$

From (A.III.17) we obtain at once the occupation probability

$$\frac{N_i}{Z_i} = \frac{1}{e^{\alpha + \beta E_i} + 1} \qquad \text{(A.III.19)}$$

of the states in the conduction band and from (A.III.18), correspondingly, the occupation probability $N_{D^\times}/N_D$ of the donor levels

$$\frac{N_{D^\times}}{N_D} = \frac{1}{\frac{1}{2} e^{\alpha + \beta E_D} + 1} \qquad \text{(A.III.20)}$$

With these two equations the problem under discussion is already solved, because the occupation probability (A.III.19) of the states 1, 2, . . . in the conduction band must be identical with the Fermi occupation probability (A.III.01). From comparison of the two equations we obtain

$$\beta = \frac{1}{kT} \qquad \text{(A.III.21)}$$

and

$$\alpha = -\frac{E_F}{kT} \qquad \text{(A.III.22)}$$

Hence, using (A.III.20) for the occupation probability of the donor levels E_D, we obtain

$$f_{\mathrm{don}}(E_D) = \frac{N_{D^\times}}{N_D} = \frac{1}{\frac{1}{2}\, e^{\frac{1}{kT}(E_D - E_F)} + 1} \qquad \text{(A.III.23)}$$

instead of the original incorrect equation (A.III.04).

Finally, we want to consider briefly the case where, apart from the conduction band, N_A acceptor levels have to be filled, rather than N_D donor levels. By occupation with an electron a donor level changes from a D^+ to a $D^\times$, and hence the number of electrons accommodated in donor levels is $N_{D^\times}$. Acceptor levels, by occupation with an electron, are changed from $A^\times$ into A^-, and hence the number of electrons accommodated in acceptor levels is N_{A^-}. If we ignore the question of spin reversal, the number W of realization possibilities must, in the acceptor case, follow from Eq. (A.III.11) by replacing $N_{D^\times}$ with N_{A^-} and N_D with N_A. This results in the expression

$$\frac{N_A!}{N_{A^-}!(N_A - N_{A^-})!} \times \prod_{i=1}^{\infty} \frac{Z_i!}{N_i!(Z_i - N_i)!}$$

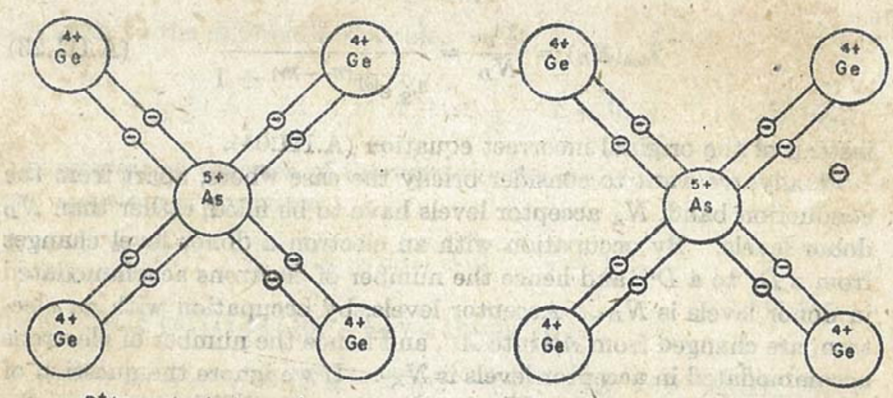
However, in determining the power of 2 in front of this expression, the preceding formal substitutions have to be supplemented by some new considerations. In the case of those donor levels in which an additional electron was accommodated, i.e., in the $D^\times$, the spin of an individual electron could be reversed (see Fig. A.III.2), whereas this did not apply to the empty donor levels D^+. The spin reversal in the $N_{D^\times}$ filled donor levels led in (A.III.11) to the factor $2^{N_{D^\times}}$. On the other hand, in the acceptor levels there is a single electron with reversible spin associated with the "empty" acceptors $A^\times$ and not with the filled levels A^- (see Fig. A.III.2). Hence we must now use the factor

$$2^{N_{A^\times}} = 2^{(N_A - N_{A^-})}$$

and we obtain for the number W of the realization possibilities

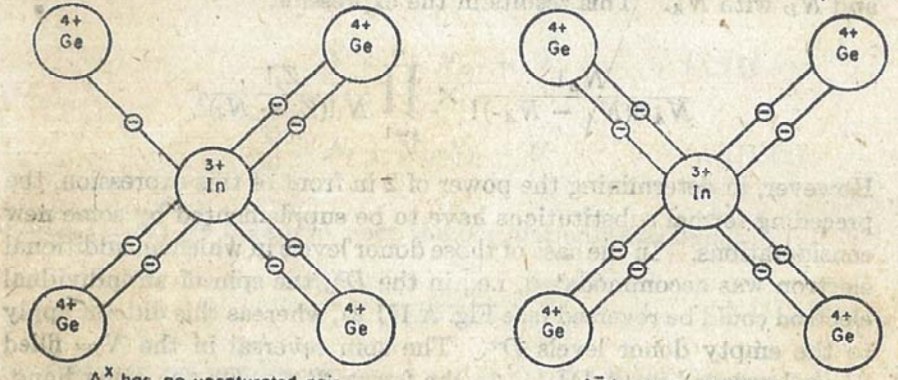
$$W = 2^{N_A - N_{A^-}} \frac{N_A!}{N_{A^-}!(N_A - N_{A^-})!} \prod_{i=1}^{\infty} \frac{Z_i!}{N_i!(Z_i - N_i)!} \qquad \text{(A.III.24)}$$

If we apply the same considerations that were used in connection with (A.III.11) to (A.III.24) we obtain, generally speaking, equations derived from the earlier ones by replacing $D^\times$ with A^- and D with A. The only modification is that the changed power of the factor 2 produces the term $(N_A - N_{A^-}) \ln 2$ in the equation corresponding to



D^+ has only electron pairs. No spin can be reversed.

$D^\times$ has a single 5th valence electron whose spin can be reversed.



$A^\times$ has an unsaturated pair bond, in which the spin of the single valence electron can be reversed.

A^- has only electron pairs. No spin can be reversed.

FIG. A.III.2. A reversible spin is available at $D^\times$ and at $A^\times$.

(A.III.12) and after differentiating for N_{A^-}, in the equation corresponding to (A.III.18), it leads to the term $-\ln 2$ instead of $+\ln 2$. Thus, in place of (A.III.23) we finally obtain

$$f_{\text{acc}}(E_A) = \frac{N_{A^-}}{N_A} = \frac{1}{2\,e^{\frac{1}{kT}(E_A - E_F)} + 1} \qquad \text{(A.III.25)}$$

Author Index

Subject Index