Electronic Semiconductors

Eberhard Spenke

Electronic Semiconductors

Eberhard Spenke

Translated with additions by
+ others
D. A. Jenny, H. Kroemer, E. G. Ramberg, and A. H. Sommer
RCA Laboratories, Radio Corporation of America
Princeton, New Jersey

McGraw-Hill Book Company, Inc.
New York Toronto London
1958

ELECTRONIC SEMICONDUCTORS

Library of Congress Catalog Card Number: 57-8020

The German edition of Dr. Spenke's book was published under the title *Elektronische Halbleiter* by Springer-Verlag, Berlin-Goettingen-Heidelberg.

THE MAPLE PRESS COMPANY, YORK, PA.

Translators' Note

This translation of the second German edition of Eberhard Spenke's "Elektronische Halbleiter" has been augmented by a section on junction capacitance (Chap. IV, §10), and a series of problems inserted at the end of every chapter. The problems not only serve to test the reader's familiarity with the contents of the preceding chapter, but challenge him to carry his studies beyond the limits of the text, eventually with the help of references to readily accessible sources. The more difficult problems have been marked with an asterisk.

An effort has been made to preserve Dr. Spenke's original approach as far as possible. However, American practice has been followed with respect to nomenclature and symbols wherever this has been possible without doing undue violence to the original text.

D. A. Jenny
H. Kroemer
E. G. Ramberg
A. H. Sommer

Foreword

This introduction to semiconductor theory is the result of a series of lectures on the fundamentals of semiconductor physics, particularly as related to rectifier and transistor problems, which my friend and coworker E. Spenke has given at physical meetings and to groups of younger colleagues in industry. The reception of these lectures demonstrated that a need existed in Germany for an able introduction to these fundamentals, even in the presence of the well-known book by Shockley. The derivation of the important relationships from precisely defined basic concepts and assumptions, in a manner which is clear even to the beginner, was recognized as a particular advantage of Spenke's representation. Spenke's audience is spared the unpleasant sensation of being suspended in mid-air.

The present arrangement of these lectures exhibits these advantages in enhanced measure. Care has been taken also that Chaps. I to V make good individual reading for the young technically trained person who has a special interest in the specific problem covered. The first three chapters deal with subjects of importance to semiconductor theory in general (and hence also to thermionics, luminescence, and photoconductivity):

I. The Conduction Mechanism in Electronic Semiconductors
II. The Nature, Models, and Reactions of Impurities and Imperfections
III. The Hole

Other chapters supply the fundamentals for two fields of application:

IV. The Mechanism of Crystal Rectifiers
V. The Physical Mechanism of Crystal Amplifiers (Transistors)

Spenke's own contributions to the basic publications for Chap. IV are well known.

Whereas the first part of this book constitutes reading matter which should provide a firm basis for the treatment of the semiconductor problems which are of greatest practical importance even to the beginner with some orientation in general physics, the second part is intended primarily for more advanced students who wish to have a thorough understanding of the fundamentals for the first part and to share in the clarification of the many unsolved problems. It covers the following chapters:

VI. Approximation Methods in the Quantum Mechanics of the Hydrogen Molecule
VII. The Band Model
VIII. Fermi Statistics of the Electrons in a Crystal
IX. The Dynamic Approach to Imperfection Equilibria and the Inertia of Impurity-Reactions
X. Boundary Layers in Semiconductors and Metal-Semiconductor Contacts

Unquestionably, good reviews of all these fields are available in books, particularly books published abroad; however, apart from Shockley's book, which treats transistor physics in particular, the reader is obliged to collect the fundamentals from a whole series of monographs, among which Bethe and Sommerfeld's classical article in Geiger and Scheel's "Handbuch der Physik" (vol. XXIV, part 2) may well take first place. However, this representation, in particular, is aimed primarily at the physics of electrons in a metal and provides an introduction which may be understood in detail only in combination with other articles from Geiger and Scheel's book. Furthermore, it lacks, of course, the new emphases and ideas which have colored semiconductor theory since 1931.

The thoroughness and realism of the author are a guarantee that, in this more advanced portion of the book, the existing limits to our knowledge are never obscured and that, at the same time, within these limits meaningful considerations and conclusions are derived and proved in reliable manner. The careful treatment of the acceleration of electrons in Chap. VII, §6, of the Zener effect in Chap. VII, §7 and §9, and of the statistics and kinetics of electrons as modified by electrostatic potentials in Chap. VIII are worthy of special mention. The last treatment also points out the significance of the Fermi potential for the current problems of diffusion theory and indicates its relationship with the general electrochemical potential of charged particles, whose gradient in the concentration and potential distribution replaces as driving force the field strength of the theory of homogeneous media. These concepts permit, in Chap. X, an unusually clear representation of the basis and development of the diffusion theory of metal-semiconductor boundary layers.

If I am to define the circle of readers for whom the reading of the second part may be useful, I am inclined to include, in addition to those who are active in this field in industrial research, all the younger scientists who may wish to have a part, both benefiting and contributing, in the solution of the preeminent semiconductor problems of our time. The manifold relationships of semiconductor physics to other fields demand the cooperation of theoretical and experimental physicists, of physical and inorganic chemists, of crystallographers and electrical engineers. I believe that all these specialists will find, in the second part of Spenke's book, representations which will extend their general knowledge of the semiconductor field in certain directions. Hence I trust that this treatment may have a stimulating effect on the advancement of theory and practice in this field as well, even though this cannot be the objective of a publication of this type.

It should be mentioned, finally, that Spenke's nomenclature follows closely international usage, so that the book should be easily readable for interested foreign workers, without the dreaded fruitless effort of a general transformation of symbols.

My best wishes speed the book on its way.

Pretzfeld, February, 1954

W. Schottky

Preface*

It was my privilege to take my first steps in the labyrinth of semiconductor physics under the guidance of Walter Schottky who, in the preceding foreword, has so kindly outlined the circle of readers to whom this introduction to the physics of electronic semiconductors may be of value. Thus, the book is not addressed to experienced semiconductor specialists, but to the "beginners" and the "more advanced." Hence it may be proper to enumerate some of the subjects and problems which are not treated in the present book even though they also belong to the fundamentals of electronic semiconductor physics. Among others, these are the problems in the forefront of present developments in the semiconductor field:

The attempts to pass beyond the current one-electron approximations by treating a solid really as a many-electron problem.

The theory of the mean free path with the difficult problem of energy and momentum transfer between conduction electrons and lattice.

Excitons and polarons, plasma interactions, and multiple collisions.

Conduction processes in an impurity conduction band, which may appear in highly doped semiconductors or at very low temperatures because, under these conditions, they are no longer overshadowed by the usual processes in the normal conduction band.

Semiconductor optics, the physics of phosphors, and the photoeffect.

Titanates, spinels, and ferrites.

Limiting frequencies in rectifiers and transistors.

The omission of all these interesting subjects naturally raises the question to what objectives and intentions of the author the treatment of these important problems has been sacrificed. The answer lies in the purely pedagogic purpose of the present book. Semiconductor electronics requires for its foundation primarily wave mechanics and statistics. However, crystallography, thermodynamics, and chemistry also have a share in it and, quite generally, "it is incredible what miserable quantities of thought and mathematics are needed to provide even the simplest tools for daily use in semiconductor physics" (from a conversation of W. Schottky with the author).

---

*To the first German edition.

The present Introduction aims to be of some help in this respect. Hence the author was not interested in the treatment of the abovementioned practical problems but in the most easily understood, and yet precise, treatment of fundamental concepts and fundamental equations. Customarily, these are passed over more quickly in proportion as the questions hidden under a simple external form are more involved. The objectives and methods employed may be clarified by an example.

A whole section is devoted to an equation as simple as Eq. (IX.1.01)

$$\frac{dN_S}{dt} = -\frac{1}{\tau_{\text{rel}}} N_S$$

In the usual presentations, this equation is introduced without further justification, with at most a note to the effect that τ_{rel} is the mean life of particles of type S. The statistical problems related to the concept of mean life are omitted altogether, and the concepts of mean life, mean life expectation, and relaxation time are commonly not distinguished. While in the present book such specialized problems are not treated in detail, the reader is at least made aware of the existence of the problem.

A whole series of kind helpers must be thanked for their valuable assistance. First of all, some associates at Pretzfeld are to be mentioned, namely, Dr. Arnulf Hoffmann, Dr. Adolf Herlet, and Hubert Patalong. Every newly written chapter was discussed with them in detail, and the author is indebted to all of them for innumerable suggestions for improvements in the text and the figures. Professor Helmut Volz and Dr. Hermann Haken (University of Erlangen) helped with a detailed "translation" of Heisenberg's original paper on the hole into the language of "ordinary" wave mechanics. Dr. Hermann Haken has also carried out a series of calculations on which §11 of Chap. VII is based. With the aid of Dr. Dieter Pfirsch (University of Frankfurt a. M.) a remark of Hintenberger regarding the effective mass of an electron in a crystal was confirmed quantitatively; the result is to be found in Chap. III, §2, and in Chap. VII, §6. For the author and subject index I am also indebted to Mr. Patalong, who also bore the chief burden of proofreading, along with Dr. Otfried Madelung, Dr. Bernhard Seraphin, and Claus Freitag. In the preparation of the manuscript, the untiring effort of Miss Aurelie Bathelt was invaluable. My thanks are extended once again to all these kind helpers. Finally, it must not remain unmentioned that I would not have arrived at a successful conclusion without the enduring patience and the continuous encouragement of my wife.

Pretzfeld, November, 1954

Eberhard Spenke

Contents

The Conduction Mechanism of Electronic Semiconductors and the Physics of Rectifiers and Transistors

Fundamentals of Semiconductor Physics

Appendixes

List of Symbols

a = lattice constant
a_0 = Bohr radius
$\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ = axis vectors in a translation lattice
A = cross-sectional area of n-p-n transistor
A = acceptor
$A^{\times}$ = neutral acceptor
A^{-} = negatively charged acceptor
A = integration constant
$b = \mu_n/\mu_p$ = ratio of electron to hole mobility
$\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ = axis vectors in reciprocal lattice
B = integration constant
C = conduction band
C_{stor} = storage capacity
C_{tr} = transition capacity
$\mathbf{c}$ = collector
D = donor
$D^{\times}$ = neutral donor
D^{+} = positively charged donor
D = diffusion constant
D_n = diffusion constant of electrons
D_p = diffusion constant of holes
$\mathbf{D}$ = dielectric displacement
$e = 1.6 \cdot 10^{-19}$ coulomb $= 4.803 \cdot 10^{-10}$ cm$^{3/2}$ g$^{1/2}$ sec^{-1} = elementary charge
$-e$ = charge of electron
$\mathrm{e} = 2.718 \ldots$ = Napierian base
$\mathbf{e}$ = emitter
E = crystal energy of an electron in a crystal (see Chap. X, §1)
E_A = electron energy at acceptor level
E_{at} = atomic eigenvalue
$E_{AV} = E_A - E_V$ = association (activation) energy of acceptors
E_C = electron energy at lower edge of conduction band
$E_{CD} = E_C - E_D$ = dissociation (activation) energy of donors

$E_{CV} = E_C - E_V$ = width of forbidden band
E_D = electron energy at donor level
E_F = electron energy at Fermi level
E_n = electron energy
E_p = hole energy
E_{pot} = potential energy
E_V = electron energy at upper edge of valence band
E = total energy of electron in a crystal (see Chap. X, §1)
$E_F^{(n)}$ = electrochemical potential of electrons
$E_F^{(p)}$ = electrochemical potential of holes
$\mathbf{E} = \{E_x, E_y, E_z\}$ = electric field strength
$\mathbf{E}_B$ = boundary field strength in a Schottky barrier layer
E_0 = field strength in filamentary transistor without injection
E_1 = incremental field strength due to injection
$f(E)$ = Fermi distribution function
f = frequency
$\mathbf{F} = \{F_x, F_y, F_z\}$ = external force
F = external force
g = rate of pair production
g = acceleration due to gravity
$\mathbf{g}$ = acceleration
$g_{11}, g_{12}, g_{21}, g_{22}$ = differential conductance values of *n-p-n* transistor

$$\left.\begin{array}{l} G, G_{ll}, G_{lr}, G_{rl}, G_{rr} \\ G_{ll_n}, G_{ll_p}, G_{lr_n}, G_{rl_n}, G_{rr_n}, G_{rr_p} \end{array}\right\} = \text{“conductances” of } n\text{-}p\text{-}n \text{ transistor}$$

G = number of cells along edge of fundamental domain
$G_{\mathrm{I\,II}}$ = Galvani voltage between solids I and II
h = Planck constant
$\hbar = \frac{1}{2\pi} h$ = Dirac constant
$\mathbf{h} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3$ = vector in reciprocal lattice
$\mathbf{H} = \{H_x, H_y, H_z\}$ = magnetic field strength
$\mathbf{i}$ = current density
i_b = base current increment
i_c = collector current increment
i_{c_n} = electron component of collector current increment
i_{diff} = density of diffusion current
i_e = emitter current increment
i_{e_n} = electron component of emitter current increment
i_{field} = density of field current
i_{forw} = current density in forward direction
i_n = density of current component carried by electrons
i_p = density of current component carried by holes

i_{rev} = current density in reverse direction
i_S = density of saturation current in rectifier theory
i_{sat} = density of saturation current in electron emission of solids
i_{tot} = density of total current
I = current
I_b = base current
I_c = collector current
I_{c_n} = electron component of collector current
I_{c_p} = hole component of collector current
I_e = emitter current
I_{e_n} = electron component of emitter current
I_{e_p} = hole component of emitter current
$j = \sqrt{-1}$ = imaginary unit
k = Boltzmann constant
k = wave number
$\mathbf{k} = \{k_x, k_y, k_z\}$ = wave number vector
K = mass-action constant
K_A = mass-action constant of acceptors
K_D = mass-action constant of donors
l = mean free path
l = thickness of barrier layer (Chap. IV, §4 and §5)
L_c = distance between collector and emitter in filamentary transistor
L_n = diffusion length of electrons
L_p = diffusion length of holes
m = electron mass
m_{eff} = effective mass
m_n = effective mass of electrons
m_p = effective mass of holes
M = mass
$\boldsymbol{M}$ = dipole moment
n = principal quantum number
n = electron concentration
n_A = total acceptor concentration
$n_{A^\times}$ = concentration of neutral acceptors
n_{A^-} = concentration of negative acceptors
n_D = total donor concentration
$n_{D^\times}$ = concentration of neutral donors
n_{D^+} = concentration of positive donors
n_S = neutrality concentration at the semiconductor end of Schottky barrier layer
n_i = intrinsic (inversion) density

n_n = electron concentration on *n* side of *p-n* junction
n_p = electron concentration on *p* side of *p-n* junction
n_B = concentration at metal boundary of Schottky barrier layer
$n_\ominus$ = electron concentration
$n_\oplus$ = hole concentration
n_0 = electron concentration in filamentary transistor *without* injection
n_1 = incremental electron density in filamentary transistor due to injection
N = band number
N = number of particles, number of particles of a selected group
N_C = effective density of states in conduction band
N_V = effective density of states in valence band
p = hole concentration
p_S = neutrality concentration of holes at Schottky barrier layer
p_n = hole concentration on *n* side of *p-n* junction
p_p = hole concentration on *p* side of *p-n* junction
p_B = concentration of holes at boundary of Schottky barrier layer
p_1 = incremental hole density in filamentary transistor due to injection
$\mathbf{p}$ = momentum
Q = cross section
r = recombination coefficient
r_b = differential base resistance
$r_c = \frac{L_c}{e\mu_n n_0 Q}$ = unmodulated collector resistance
r_e = differential emitter resistance
$r_{11}, r_{12}, r_{21}, r_{22}$ = differential resistance values of network or of transistor
R = radius
R_b = internal base resistance of point-contact transistor (see Figs. V.4.4 and V.4.5)
R_B = spreading resistance of rectifier
R_{diff} = differential resistance
R_K = external feedback resistance
R_L = load resistance
R_0 = zero-bias resistance of rectifier
s = density of particle current
S = distance between emitter and collector in point-contact transistor

t = time
t_{tr} = transit time
T = absolute temperature
$T_{relax} = \varepsilon/4\pi\sigma$ = dielectric relaxation time of solid
u_c = collector-voltage increment
u_e = emitter-voltage increment
u_L = voltage increment across load resistance R_L
U = terminal voltage
U_c = collector voltage
U_e = emitter voltage
U_{forw} = terminal voltage in forward direction
$U(\mathbf{r})$ = lattice potential
U_{rev} = terminal voltage in reverse direction
$U(x)$ = lattice potential (one-dimensional)
v = velocity
v_{drift} = drift velocity
v_{th} = thermal velocity
$\mathbf{v} = \{v_x, v_y, v_z\}$ = velocity
V = volume
V = valence band
V = electrostatic potential
V_D = diffusion voltage
V_{fund} = fundamental domain of crystal
$\mathbf{V} = kT/e$ = volt equivalent of temperature
W = width of p layer in n-p-n transistor
x_b, x_e, x_c = space coordinates in Fig. V.3.4
Z = atomic number
Z_{eff} = effective atomic number
$\alpha = \alpha_i\beta\gamma$ = current amplification factor
$\alpha_e = (1 + b)\beta\gamma$ = equivalent current amplification
α_i = inherent (true) current amplification
β = transport factor
γ = injection efficiency
γ_c = collection efficiency
γ_e = emitter injection efficiency
∂ = partial derivative
ε = dielectric constant
$\zeta(n/N)$ = special function (see p. 384)
Θ_n = Hall angle of electrons
Θ_p = Hall angle of holes
λ = wavelength
μ_n = electron mobility

μ_p = hole mobility
ρ = space-charge density
σ = conductivity
σ_{coll} = collision cross section
σ_{D^+} = effective cross section of donors
σ_i = intrinsic conductivity
τ = mean free time
τ = lifetime
$\tau_{D^\times}$ = lifetime of neutral donors
τ_{D^+} = lifetime of positively charged donors
τ_n = electron lifetime in p-type conductor
τ_p = hole lifetime in n-type conductor
τ_{rel} = relaxation time of assembly of particles, of current, etc.
$\psi = \psi(\mathbf{r}; \mathbf{k})$ = eigenfunction
$\psi = \psi(\mathbf{r}, t)$ = time-dependent wave function
ψ_{at} = atomic eigenfunction
Ψ = work function
Ψ_{met} = work function of metal
Ψ_{sem} = work function of semiconductor
$\Psi_{\text{met sem}}$ = work function from metal to semiconductor
$\Psi_{\text{met vac}}$ = work function from metal to vacuum
$\Psi^{(n)}_{\text{met sem}}$ = electron work function (see p. 358)
$\Psi^{(p)}_{\text{met sem}}$ = hole work function (see p. 362)
Ψ_{I} or Ψ_{II} = work function of metal I or II
$\omega = 2\pi f$ = angular frequency
$\ominus$ = conduction electron
$\oplus$ = hole
●( ) = substitutional imperfection (see Fig. II.4.2)
○ = interstitial lattice position (see Fig. II.4.2)
□ = vacancy (see Fig. II.4.2)
$\times$ = neutral
$\cdot$ = singly positively charged
$\cdot\cdot$ = doubly positively charged
$'$ = singly negatively charged
$''$ = doubly negatively charged

Electronic Semiconductors

The Conduction Mechanism of Electronic Semiconductors and the Physics of Rectifiers and Transistors

CHAPTER I

The Conduction Mechanism in Electronic Semiconductors

§1. Introduction

The semiconductors, among the crystalline solids, conduct an electrical current better than insulators but not so well as metals. Accordingly, their conductivity range at room temperature extends from 10^{-12} ohm^{-1} cm^{-1} to 10^{4} ohm^{-1} cm^{-1}. These limits are drawn somewhat arbitrarily, and it will be shown that there is no fundamental difference between insulators and semiconductors. It cannot be decided at the present stage of development whether there is a physically meaningful border line between metals and semiconductors.[1] The extreme sensitivity of the electrical resistivity to various factors, particularly the "chemical composition," is a characteristic property of semiconductors.

The significance of the chemical composition in semiconductor physics goes far beyond the previous usage of this term in chemistry. The chemical difference between two bodies means, to the unbiased reader, the difference, for instance, between cupric oxide (CuO) and cuprous oxide (Cu_2O). The semiconductor physicist, on the other hand, compares two samples by the magnitude of their electrical resistivity, whereby variations of a fraction of a per cent in the oxygen content can be of decisive significance. Minute deviations from the stoichiometric composition (of the order of 10^{-4}) may manifest themselves in the electrical resistivity by changes of several orders of mag-

---

[1] The negative temperature coefficient of the electrical resistivity, i.e., the fact that many semiconductors are better conductors at high temperatures than at low temperatures, is often used to define a semiconductor. The existence of substances such as niobium hydride and nitride, which are "low-temperature conductors," speaks against the validity of this definition, for they are superconductors with surprisingly high transition temperatures. On the other hand, they can hardly be called metals and must therefore be classified among the semiconductors, although they are not true "high-temperature conductors." A substantial number of substances fall in this group, such as CuS and the borides, carbides, and nitrides of zirconium, hafnium, titanium, vanadium, and tantalum.

nitude (possibly 10^4).[1] Similarly, a minute amount of a foreign substance, such as a chlorine content of the order of 10^{-4} in selenium (1 chlorine atom per 10^4 selenium atoms), can influence the electrical resistivity decisively. This is the reason why in semiconductor physics a difference between, for instance, selenium with an addition of 10^{-4} chlorine and 10^{-5} chlorine is called a "chemical difference," although both are essentially selenium and the only distinction lies in the minute amount of added foreign substance.

Other factors which can influence the resistivity appreciably shall be listed briefly; some of the meanings may be understood only by the reader who is already somewhat familiar with the semiconductor field: (1) The previous history of the material, particularly the thermal treatment in certain atmospheres. (2) The fine or coarse polycrystallinity or single crystallinity of the material, as well as possible predominant crystal directions and textures. (3) Possible deviations from the chemical homogeneity of the material, where "chemical" is to be understood in the foregoing sense as applied to semiconductor physics. This is particularly applicable to the neighborhood of crystal grain boundaries, where a certain lattice distortion can favor vacancies, interstitials, and precipitation of foreign atoms.[2] (4) Microscopic and colloidal precipitation of dissimilar phases (in the metallurgical sense), which can form metallic conductivity bridges or insulating layers. (5) The surface structure of the entire sample or the single crystallites can affect the resistance appreciably because of space-charge barriers with polarized characteristics. (6) The surrounding atmosphere can affect the properties, particularly at high temperatures. (7) The resistivity of a semiconductor is often highly temperature dependent. (8) The magnitude of the electrical field during measurements influences the semiconductor properties in many cases.

In view of the great number of effective parameters, it is not surprising that the field of semiconductor physics is characterized by a multitude of confusing phenomena and apparently contradictory observations. For many years one had to be satisfied with the observation and analysis of "order-of-magnitude effects" in the semiconductor field. Few physicists appreciate this type of research, which led to the bad reputation of semiconductor physics as the "physics of dirt effects." In spite of this reputation, an ever-increasing number of scientists have become interested in this field over the years. One

---

[1] The same sensitivity can be found in the optical absorption, so that the appearance of a substance may be changed drastically by, for instance, heating in the vapor of one of the components: for example, copper iodide, alkali halides.

[2] See Chap. II, The Nature, Models, and Reactions of Impurities and Imperfections.

reason was the application of the methods of wave mechanics to the giant crystal molecule in the late 1920s, which gave a great boost to solid-state physics and along with it to semiconductor physics. Furthermore, the continuously increasing technical importance of semiconductors as well as economic reasons demanded intensive research and development.[1] We shall name a few examples: thermistors, electrolytic capacitors, crystal rectifiers and detectors, transistors, oxide cathodes, overvoltage regulators and varistors of SiC, photovoltaic cells.

Even in technical problems apparently not related to electrical conduction, the principles of semiconductor physics are of decisive importance, as for instance: luminescent screens for cathode rays and X-rays, the photographic exposure and development process, the protection of metals against corrosion by various atmospheres, corrosion-resistant alloys.

Finally, after the Second World War, a prototype substance was found in the form of germanium, in which the complications do not seem to be insurmountable and it is possible to progress beyond order-of-magnitude relations toward a number of theoretically derived formulas which are accessible to quantitative verification by experiment. This is the reason why, in the following chapters, we shall always emphasize germanium and its properties rather than discuss the extensive research on other materials.

The entire field which might fall within the realm of this book has become so large that certain restrictions are unavoidable. We shall not deal with materials or mechanisms which involve primarily ionic conduction. Problems of cohesion and the magnetic semiconductors, the so-called ferrites, have to be left out. Nor shall we consider optical effects in detail; therefore we can eliminate the extensive subject of phosphors. But we shall treat the *conduction* mechanism of *electronic* semiconductors within the framework of the *band model* to which this first chapter is devoted.

§2. The Band Model

a. The One-electron Approximations of Solid-state Physics: The Atomistic Picture and the Band Model[2]

The formation of a crystal takes place when many atoms are arranged in a systematic manner to form a single giant molecule.

---

[1] However, one must not underestimate the driving force of scientific curiosity which does not tolerate the ever-increasing use of devices in practical applications without an understanding of the fundamental mechanisms involved.

[2] See also Chaps. VI and VII.

Hence the two models of solid-state theory, namely, the atomistic picture on one hand and the band model on the other, correspond to the two approximations of molecular physics which are the "method of atomic eigenfunctions" (Heitler-London) and the "collective treatment" or the "method of the molecular eigenfunctions" (Hund and Mulliken). The Heitler-London method uses the limiting case of widely spaced complete atoms as a starting point and considers the mutual interaction only as a perturbation. According to this concept, an electron belongs to a specific nucleus and its behavior is modified only by the presence of the surrounding atoms. In solid-state physics this approximation corresponds with the atomistic picture where the electrons are assigned to single lattice points—ions, atoms, or molecules (Fig. I.2.1). Their behavior differs from that in an isolated atom only inasmuch as it is modified by, for instance, a polarizable environment with a dielectric constant $\varepsilon \neq 1$. The atomistic picture is particularly useful in the treatment of problems where the electrons in the inner shells of the lattice atoms play an important part, such as in the emission and absorption of X-rays. The atomistic picture is indispensable for the theory of ferromagnetism and cohesion, and it is very valuable for the treatment of crystals with pronounced ionic character such as the alkali halides. Although it allows very plausible qualitative statements about the conductivity problem, the quantitative treatment must be left to the band model, which, as pointed out previously, is based on the approximation of Hund and Mulliken in molecular physics. The starting point in this case is the opposite extreme of the Heitler-London method, namely, very close-spaced nuclei. This limiting case does not allow the assignment of an electron to a single nucleus, for it is affected continuously by the force field of all the surrounding nuclei. In this method, the mutual interaction between the different electrons is treated as a small perturbation. Besides the obvious treatment by a perturbation calculation of higher order, it is possible to apply a first-order treatment by introducing screening factors for the field of the nuclei which act on the electron under consideration.

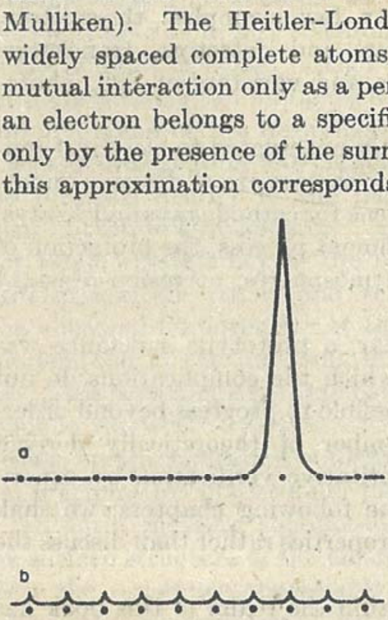


FIG. I.2.1. The two one-electron approximations of solid-state physics. The local probability for an electron in a crystal (schematically). (*a*) The atomistic picture. (*b*) The band picture.

The screening factors are a consequence of the charge of the surrounding electrons. In solid-state physics, this procedure is represented by the band model where a particular electron is affected by the field of *all* lattice points rather than just one (Fig. I.2.1). In the band model, the interaction between the electron under consideration and the other electrons is taken into account in the form of a modification of the nuclear potential caused by the charge of the surrounding electrons.

As in the case of the atomistic picture, this is essentially a one-electron approximation because the energy levels of a single electron are determined in a given fixed force field. The transition to the many-electron problem lies not in the introduction of an interaction between electrons, but simply in filling the energy levels of the one-electron problem with all the electrons which must be accommodated in the crystal. The occupancy of the energy levels follows the principles of the Fermi statistics, whereby a certain rough consideration for electron-electron interaction is given in the form of the Pauli exclusion principle.

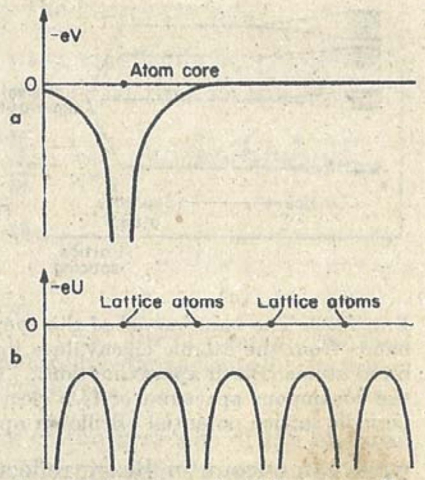


FIG. I.2.2. The potential energy distribution of an electron (*a*) in a single atom and (*b*) in the lattice of ordered atoms.

b. The Band Model

If one asks for the energy-level diagram which forms the basis for the band model, one recognizes that the name "band model" is in itself the answer to the question: The energy-level diagram of the stationary allowed energies shows a bandlike division into alternating "allowed" and "forbidden" energy ranges. This can be demonstrated satisfactorily with a purely conceptual approach without resorting to mathematical methods:

An electron in a crystal[1] is not in the force field of a single atom, but rather in the periodic[2] potential field of many uniformly distributed atoms (Fig. I.2.2). The discrete eigenvalues of the electron energy[3] in

---

[1] We speak of an "electron in a crystal" when we want to emphasize that an electron is not *free*, but is exposed to the strong lattice forces which lead to a more or less strong binding to the crystal.

[2] This is a hypothesis. See p. 163.

[3] As usual, this is the sum of kinetic energy and potential energy, the latter being

the single atom are multiply degenerate because of the spatial accumulation of many atoms in the crystal (Fig. I.2.3). Many different eigenfunctions are associated with one and the same atom eigenvalue; they are always grouped in the same manner around a different lattice point. Because of the exchange interaction, each atom eigenvalue is split into a multitude of quasi-continuous bandlike energy eigenvalues. The splitting becomes more pronounced as the atoms approach each other more closely, and therefore the exchange interaction increases.

Besides this method, quantitatively treated by Bloch[1] to obtain the band spectrum of a crystal, it is possible to start from the wave nature of the electron according to Brillouin.[1] In this case it is understandable that at certain wavelengths and propagation directions an electron wave can encounter Bragg reflections similar to those of X-rays. This means that for certain electron waves a reflected wave is formed because of interference with the lattice, which has the same wavelength and therefore, according to de Broglie, the same electron energy and hence is degenerate with the incident wave. The degeneracy is canceled by splitting of the common energy eigenvalue because of the interaction with the periodic lattice potential in space. This splitting becomes more pronounced as the amplitude of the lattice potential increases. The Bloch treatment demonstrates the splitting of the discrete atomic levels into the *allowed* bands, whereas the Brillouin

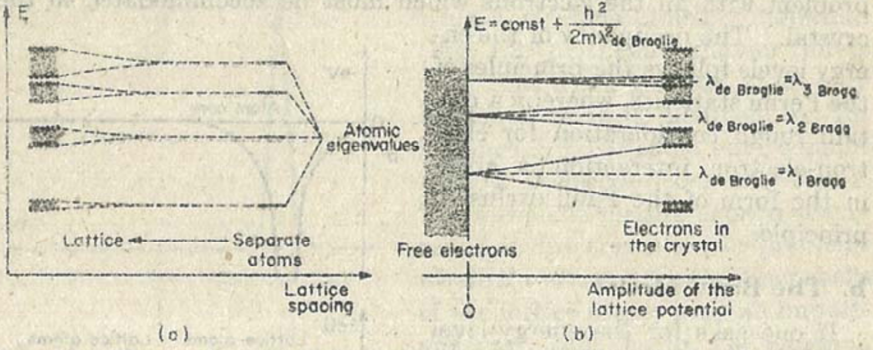


FIG. I.2.3. The band model of the energy terms. (*a*) The splitting of the *allowed* bands from the atomic eigenvalues upon spatial approximation of initially separated atoms (Bloch approximation). (*b*) The splitting of the *forbidden* bands from the continuous spectrum of free electrons upon increasing the amplitude of the periodic lattice potential (Brillouin approximation).

---

caused by the force field of the nucleus. We avoid the expression "total energy," because we shall see in Chap. X that in some solid state problems an electrostatic energy contribution must be considered, which stems from an electrostatic macropotential, see p. 338.

[1] See Chap. VII, §2 and §3.

treatment shows the splitting of an originally continuous energy spectrum of free electron waves into the *forbidden bands*.

If an electron has a sharply defined energy which corresponds to one of the described levels of the strictly periodic ideal lattice, it is represented by a similar eigenfunction as that of a free electron, hence with a plane wave:

$$\psi(x; k) = A\,e^{jkx} \qquad k = \frac{2\pi}{\lambda} = \text{wave number} \qquad (I.2.01)$$

The eigenfunction of an electron in a crystal differs from that of a free electron by a modulation of the wave amplitude due to the periodic lattice potential:

$$\begin{aligned} \psi(x; k) &= u(x; k) \cdot e^{jkx} \qquad (I.2.02) \\ u(x; k) &= \text{periodic with the lattice period} \end{aligned}$$

The absolute square $u(x; k)^2$ of the wave amplitude $u(x; k)$ gives the probability of the presence of the electron, so that in the case of a sharply defined energy the probability of the presence of the electron is the same for all lattice cells. If the sharply defined energy value of an electron lies in a *deep*[1] level (Fig. I.2.4), the probability of finding the electron near a lattice point[2] is appreciably higher than for locations between two lattice points, but these fluctuations are repeated from cell to cell throughout the entire lattice in a periodic manner, so that the electron cannot be assigned preferentially to a single lattice point. If the electron is in a *high*[3] level (Fig. I.2.5), the probability does not vary as much between lattice points and intermediate locations, and the electron is rather evenly "smeared out" throughout the entire lattice. Through formation of wave packets it is possible to attain a

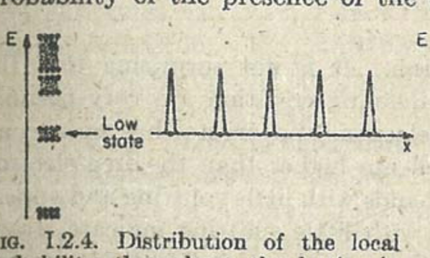


FIG. I.2.4. Distribution of the local probability throughout the lattice in low levels.

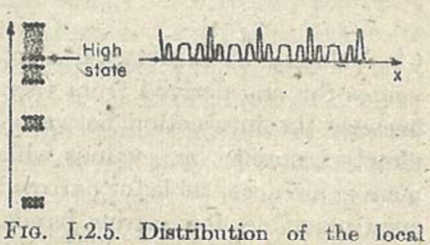


FIG. I.2.5. Distribution of the local probability throughout the lattice in high levels.

---

[1] Such a level belongs to a band which results from the splitting of one of the deep atom eigenvalues and therefore belongs to one of the *inner* shells.

[2] "Lattice point" refers to the position of a lattice atom, ion, or molecule in the crystal.

[3] This level corresponds with the outer shell of a lattice point.

more or less sharp localization of the electron, but only by combining several neighboring levels so that an accurately defined energy value has to be abandoned. Even so, an important property of these eigenfunctions of the band model is not sacrificed: All these functions represent an electron which moves unhindered and uniformly through the lattice, so that one is tempted to call it a free electron. The Brillouin picture points even more strongly to the existence of free electrons; for only at certain energy values is the electron prevented by Bragg reflection from propagating through the lattice, so that avoiding these forbidden energy values ought to result in a free electron. The fact that this is not entirely true is best demonstrated by studying the behavior of an electron in a crystal under the influence of an external force such as an electric field.[1]

For this purpose an effective mass m_{eff} is defined by the following equation:

$$\mathbf{F} = m_{\text{eff}} \cdot \frac{d\mathbf{v}}{dt} . \tag{I.2.03}$$

$\mathbf{v}$ = velocity of the crystal electron. It is not surprising that the values for m_{eff} derived from this defining equation are very peculiar because the interaction between electron and crystal has been completely ignored. m_{eff} values which are higher than the free electron mass m are acceptable for narrow bands with little splitting and appear even plausible, for narrow bands with little splitting indicate strong binding of the electrons to the lattice points, and it is entirely reasonable that this leads by way of a large effective mass to sluggish electrons. The dependence of the value of m_{eff} on the exact energetic location of the electron in its energy band is less understandable. The most surprising fact is at first that m_{eff} is negative in the upper part of a band.

How this comes about because of the neglected lattice forces in the defining equation (I.2.03) may be more easily understood from the following considerations. We start with the assumption that the velocity of the electron $\mathbf{v}$ and the external force $\mathbf{F}$ have the same direction, so that the external force $\mathbf{F}$ acts on the electron by increasing its energy. This raises the electron in the band diagram of Fig. I.2.3 so that it approaches finally the upper edge of the allowed band in which it was located at the beginning of the process. The more the electron approaches the band edge, the more the condition for Bragg reflection is fulfilled, and the lattice reflects more of the incident wave in the form of a reflected wave. At the upper band edge, the condition of

---

[1] See Chap. VII, §6.

equality of incident and reflected wave results in a standing electron wave, and the velocity **v** of the electron has become zero. During this entire process, the force **F** and the electron velocity **v** have had the same direction, so that the electron mass must be considered negative in order to explain zero velocity, if one ignores the actual cause, namely, the Bragg reflection of the lattice.

The foregoing may have aided the reader to gain a conceptual understanding for the negative values of the effective mass, and we can proceed to the deductions derived from this surprising result.

c. The Band Model and Conductivity Problems

In a solid, an electric current is carried by more than just one electron. Therefore, in order to treat conductivity problems, it is necessary to fill the band diagram of energy levels, which was developed in the previous section, with many electrons in a crystal, as was implied at the end of section *a*. The Pauli principle restricts the number of electrons which can occupy each single energy level to two electrons with opposite spin. The available electrons in a crystal can thus fill the energy diagram from the bottom up to a certain limit, above which the diagram remains empty according to the foregoing principles. The transition from the filled levels to the empty ones is very gradual at high temperatures and becomes more abrupt at lower temperatures until the limit of absolute zero is approached ($T \rightarrow 0$), where the transition is discontinuous. This transition is governed by the Fermi statistics, from which the expression "Fermi level" for the dividing line between filled and empty levels originates.

Below the Fermi level, one or more bands are completely filled with electrons. It can be shown that these bands do not contribute to the current; for the electrons in the lower part of the band have a positive effective mass and are accelerated in the direction of the external force, whereas the effective electron mass in the upper part of the band is negative, which results in an acceleration in the opposite direction. Their respective contributions to the current cancel exactly as the accurate calculation shows:[1] *A filled band does not contribute to the conductivity*.

The deep lying bands, corresponding to the inner electron shells of the free atoms, are the ones which are filled and do not contribute to the conductivity. From this standpoint, the objectionable result of negative effective masses in the upper part of a band leads to a reasonable result: The electrons of the inner shells do not play a part in the conduction process. One can consider them as bound electrons which,

[1] See pp. 262 to 263, 264, and 267.

for the consideration of conductivity problems, form an "atom core" together with the nucleus itself. The internal structure of this atom core is, in *this* connection, as uninteresting as the structure of the nucleus itself. The conduction of the current and similar transport phenomena depend entirely on the outer electrons of the lattice atoms.

This leads to a natural classification of the solids into insulators and metals (Fig. I.2.6). If the number of electrons to be placed in the energy diagram is such that a series of deep bands are just filled and no electrons are left over for the higher bands, the crystal is an insulator; for all the completely filled bands do not contribute to the conductivity. A conductor, namely a metal, therefore must have a partially filled uppermost band, for instance, a half-filled band.[1]

FIG. I.2.6. Band model of an insulator and a metal.

The behavior of the *conduction electrons* in such a partially filled *conduction band* of a metal is actually extremely complicated. The effective mass is not the same for all conduction electrons, but depends entirely on the particular location the electron occupies in the energy band at any one time. Besides, the effective mass is not isotropic, but is rather described by a tensor in any crystal system (even in symmetrical ones), so that force and acceleration of the electron are in general not in the same or exactly opposite directions but at an angle with each other. If the crystal system is not symmetrical, further complications set in. In the theoretical treatment of conductivity problems, these difficulties are removed through simplifying assumptions about the distribution of energy levels within the conduction band. This approach allows the treatment of the conduction electrons as if they formed a free Fermi gas with a potential energy which corresponds to the lower edge of the conduction band.[2]

The electrons in such a gas are never at rest. The kinetic theory of heat teaches that each electron is continuously moving in a zigzag path (Fig. I.2.7). The straight portions of such a zigzag path are called "free path l," and the time required to complete a free path is called "free time τ." The velocities of the electron during the free time, while on the free path, are distributed statistically around the "mean thermal velocity." At the end of the free path, the electron collides with a col-

---

[1] The actual application of these criteria, for instance, to the elements of the periodic table, meets considerable difficulties which are, at present, by no means solved. See Chap. VII, §11.

[2] See p. 299.

lision partner, which we shall now discuss in detail. The collision itself requires an extremely short time compared with the free time and can be considered a quasi-discontinuous change in the velocity of the electron. The collision partners are characteristically different from those in an ordinary gas of, for instance, H_2 molecules. In such a gas the H_2 molecules collide *with each other*, whereas in a Fermi gas the conduction electrons collide with *foreign* collision partners. These collision partners are of a peculiar nature. We saw in section *b* that in a lattice of ideal regularity an electron, because of its wave nature, does not in general[1] encounter any resistance and can therefore traverse paths of arbitrary length without any disturbance, scattering, or deflection. This ceases to be true as soon as the regularity of the lattice is disturbed in any way, which is unavoidable at finite temperatures because of the thermal motion of the lattice atoms.

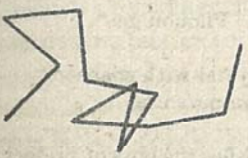


FIG. I.2.7. The path of an electron with the assumption of a uniform mean free path *without* external force.

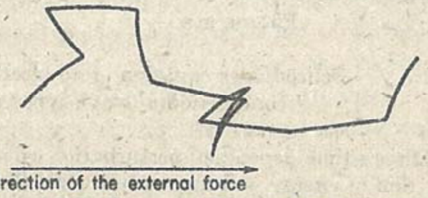


FIG. I.2.8. Same as Fig. I.2.7 but *with* external force.

At present, we can tentatively state that the conduction electrons are scattered by the thermal expansions and compressions of the lattice. In the theoretical treatment of the problem of interaction between the electron in the crystal and the self-oscillation of the lattice, these thermal density changes of the lattice are treated as a superposition of elastic and acoustical self-oscillations of the lattice.

However, this obvious picture of a "collision" requires a particle as a collision partner. According to the quantum laws, the interaction between electrons and the elastic and acoustical oscillations of a solid can be interpreted as collisions between electrons and "phonons." Most physicists, however, are more familiar with photons than with phonons, so that Fig. I.2.9 will be helpful as a schematic of notations where the concept of the phonon is compared with the corresponding well-known photon concept.

In this schematic, we see that the quantum expression $nh\nu$ for the energy of an oscillator leads to the Planck radiation equation, if the

---

[1] Unless its wavelength happens to correspond with a Bragg reflection.

oscillators represent the electromagnetic eigenoscillations of a cavity with perfectly reflecting walls.[1] For the phonon, the same quantum law results in the Einstein-Debye theory of the specific heat of solids,[2,3] if the oscillators represent the elastic self-oscillations of a solid. The next step on the "electromagnetic" side of the schematic is the interpretation of the cavity radiation as a photon gas.[4,5] The equivalent

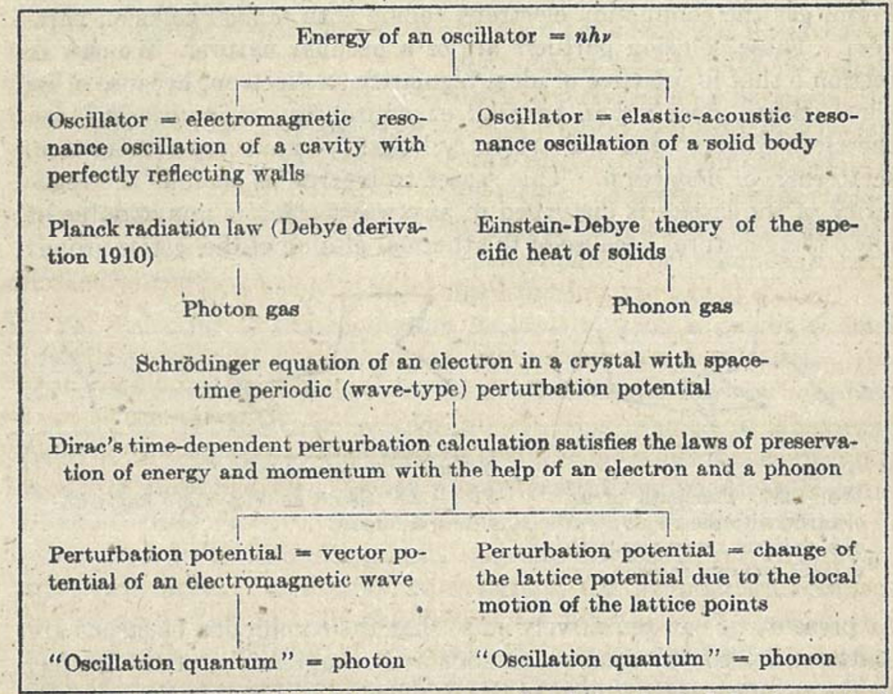


FIG. I.2.9. Schematic table of the chronological and the physical development of the "phonon" concept.

step on the "elastic" side of the schematic of notation interprets the thermal oscillations of the solid as a "phonon" gas[6] with Bose statistics.

If we calculate the interaction between an electron in the crystal and an electromagnetic wave,[7] we find that this electron is described by a Schrödinger equation with a periodic crystal potential of the lattice in

[1] P. Debye, *Ann. Physik*, **33**: 1427 (1910).

[2] A. Einstein, *Ann. Physik*, **22**: 180 (1907).

[3] P. Debye, *Ann. Physik*, **39**: 789 (1912).

[4] A. Einstein, *Ann. Physik*, **17**: 132 (1905).

[5] S. N. Bose, *Z. Physik*, **26**: 178 (1924).

[6] L. Nordheim, *Ann. Physik*, **9**: 607 (1931). A. H. Wilson, *Proc. Roy. Soc. (London)*, **133**: 458 (1931).

[7] See, for instance, Chap. VII, §8.

which the classically introduced (vector) potential of the light wave occurs as perturbation. The result can be interpreted as a collision between the electron in a crystal and a light quantum, although only the electron is introduced in the quantized form (Schrödinger equation), whereas the electromagnetic wave is used with its classical vector potential in the Schrödinger equation.

Similarly, the mathematical treatment of the interaction of an electron in a crystal and an elastic lattice wave leads also to a Schrödinger equation with a space periodic lattice potential and a space and time periodic perturbation potential of the lattice wave.[1] Although the expression for the lattice wave does not represent the quantized properties, the result can be interpreted as a collision between lattice electron and phonon. This pure particle interpretation is, in both cases, only possible on a historical basis, which is indicated by the first few lines of the schematic.

Besides these thermally determined phonons, a real lattice contains other deviations from strictly ideal periodicity, because of deviations from the ideal structure. For instance, atoms may be missing at certain lattice points (vacancies), or atoms can be misplaced in the wrong lattice points or be placed interstitially, or foreign atoms can be incorporated which are not part of the ideal lattice. All these atomic "imperfections" act as scattering centers or collision partners like the phonons.

Up to this point, we have treated the behavior of the Fermi gas of the conduction electrons in thermal equilibrium. Finally, we want to treat the Fermi gas under the influence of an externally applied electric field, so that a current is flowing. The previously straight paths between two collision processes are now bent into parabolas because of the continuous action of the external force field (Fig. I.2.8). All the electrons receive during each free path an additional velocity in the direction of the external force, so that the entire thermally agitated electron cloud is drifting slowly in this direction. Thus an electric current is created. As long as the drift velocities due to the external force are small compared to the mean thermal velocity, the drift velocity $\mathbf{v}$ is proportional[2] to the field strength $\mathbf{E}$:

$$\mathbf{v} = \mu_n \mathbf{E} \qquad (I.2.04)$$

The factor μ_n is called the *mobility* of the negative electrons.[3] Ohm's

---

[1] F. Bloch, *Z. Physik*, **52:** 555 (1929), particularly pp. 578ff.

[2] See, for instance pp. 244 to 246.

[3] In contrast to the corresponding mobility μ_p of the positive holes (defect electrons or missing electrons) which will be introduced later (see pp. 16 to 19 and 27).

law applies in this case

$$\mathbf{i} = \sigma \mathbf{E} \qquad (I.2.05)$$

where the conductivity is given by

$$\sigma = e\mu_n n \qquad (I.2.06)$$

n = concentration of the negative electrons.

An example of the opposite case is that of the cathode-to-anode current in a vacuum diode, where the electrons do not find collision partners because of the vacuum and therefore fall down the potential slope in one single free path, so that the motion of all electrons is essentially the same. There is obviously no linear, i.e., ohmic, relation between anode voltage and anode current; instead the laws of the thermal, space-charge limited, and saturation regions are valid.

§3. The Band Model of a Semiconductor

We have seen in §2, section c, that a completely filled band does not contribute to the conductivity and that a crystal with only filled bands is an insulator. The band model of an insulator shows that the uppermost filled band is followed by a forbidden band, which in turn is followed by an allowed but empty band. Such a distribution of the electrons in the band diagram is, strictly speaking, possible only at the temperature $T = 0°K$, even if exactly the right number of electrons are present just to fill completely the last occupied band. At any finite temperature $T > 0$, a fraction of the electrons from the last "filled" band are thermally excited or raised into the next higher "empty" band above the intermediate forbidden band (Fig. I.3.1). The electrons which are raised into the empty band represent, according to the principles of section c of §2, the gas[1] of the conduction electrons and

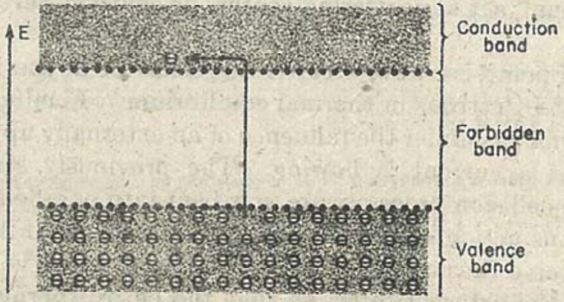


FIG. I.3.1. Thermal excitation of a valence electron into the conduction band.

---

[1] See pp. 299 and 306.

cause a certain amount of conductivity. Whether this effect is real and experimentally observable depends on one hand on the temperature and on the other on the width of the forbidden band. The smaller the width of the forbidden band, the smaller the energy for thermal excitation and therefore the lower the temperature at which the effect is observed. Wilson[1] has pointed out in his work that the thermal excitation of a fraction of the electrons from the band below the conduction band leaves this originally filled band only partially occupied, so that its contribution to the current is no longer zero. Therefore the electrons of this band carry a certain current which is added to the current of the conduction electrons. It is found that this current of the nearly filled band is of the same magnitude as if one assumed, in place of the holes in the filled band, positive electrons with positive effective mass but no other current carriers. The fact that this concept has proved valid in the interpretation of the sign of the Hall effect and the thermoelectric effect and in the rectification direction of metal semiconductor contacts has led to the introduction of the term "hole" (or defect-electron) conduction.

Many years after the work of Wilson, it was still uncertain whether the described conduction mechanism of a semiconductor is actually realized in one of the known semiconductors. It was shown only toward the end of the Second World War, almost simultaneously in Germany[2] and the United States,[3] that the current transport in germanium above 150°C is of the type just described. It is now necessary to supplement the somewhat abstract description of the "last completely filled band" and the "conduction band" with more concrete concepts, for which we elect germanium as the example. The germanium lattice has a diamond-type structure, where each atom is surrounded by four neighbors (Fig. I.3.2). The binding of these four neighbors consists of four electron-pair bonds as we know them for the homopolar binding of the hydrogen molecule. The bonds consist of the four valence electrons of the germanium atom under consideration together with one of the valence electrons from each of the four neighboring atoms to which the bond leads. We see that the diamond lattice points are occupied by Ge^{+4} ions and that the four valence electrons per atom are located in the pair bond between the Ge^{+4} ions (Fig. I.3.3). The localized particle representation is the atomistic parallel to the band model, in which the four valence electrons of the germanium just fill a band, which is therefore called the "valence

---

[1] A. H. Wilson, *Proc. Roy. Soc.* (*London*), **133:** 458 (1931).

[2] J. Stuke, Dissertation, Göttingen.

[3] K. Lark-Horovitz and V. A. Johnson, *Phys. Rev.*, **69:** 258 (1946).

band." The fact that the next higher band, the conduction band, is entirely empty to begin with corresponds in the atomistic picture accurately to the fact that all valence electrons are fixed in the pair bonds and therefore will not be available for the conduction process. In the case of thermal excitation, one or the other valence electron is separated from one of the pair bridges and becomes quasi-free to move through the lattice (Fig. I.3.4, upper left).[1] This corresponds in the band model to the thermal excitation of a valence electron from the upper edge of the valence band to the lower edge of the conduction band (Fig. I.3.4, upper right). Here we see again how the atomistic picture and the band model complement each other; that is, the band model teaches that the empty place among the valence electrons acts as a hole $(+m, +e)$ if one considers the wave nature of the electrons, which means that the hole can migrate through the unperturbed lattice without encountering any resistance and therefore contributes an essentially equal part to the conductivity as the freed electron $(+m, -e)$ in the conduction band. The migration without resistance of a hole in the band model corresponds to the motion of the valence electrons into neighboring empty places without energy expenditure in the atomistic picture (see Fig. I.3.4, left side).

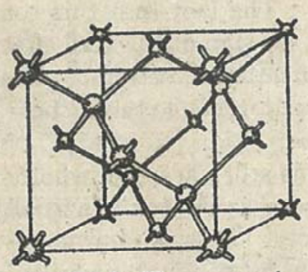


FIG. I.3.2. The germanium lattice.

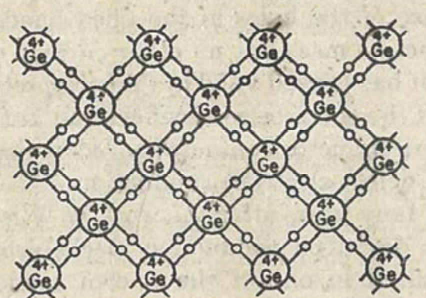


FIG. I.3.3. Two-dimensional representation of the germanium lattice.

This conduction mechanism is called intrinsic conduction in contrast to extrinsic or impurity conduction. The existence of the latter conduction type has, for a long time, overshadowed the intrinsic conduction; this is still the case in most germanium samples below 150°C. In the previous paragraph we became briefly acquainted with the

---

[1] The presentation of the migration of the conduction electron on the right-hand side of Fig. I.3.4 assigns to the abscissa the significance of a spatial coordinate within the solid, whereas no significance was attributed to the abscissa in Figs. I.2.6 and I.3.1. See also pp. 21 and 22 in connection with Fig. I.3.6.

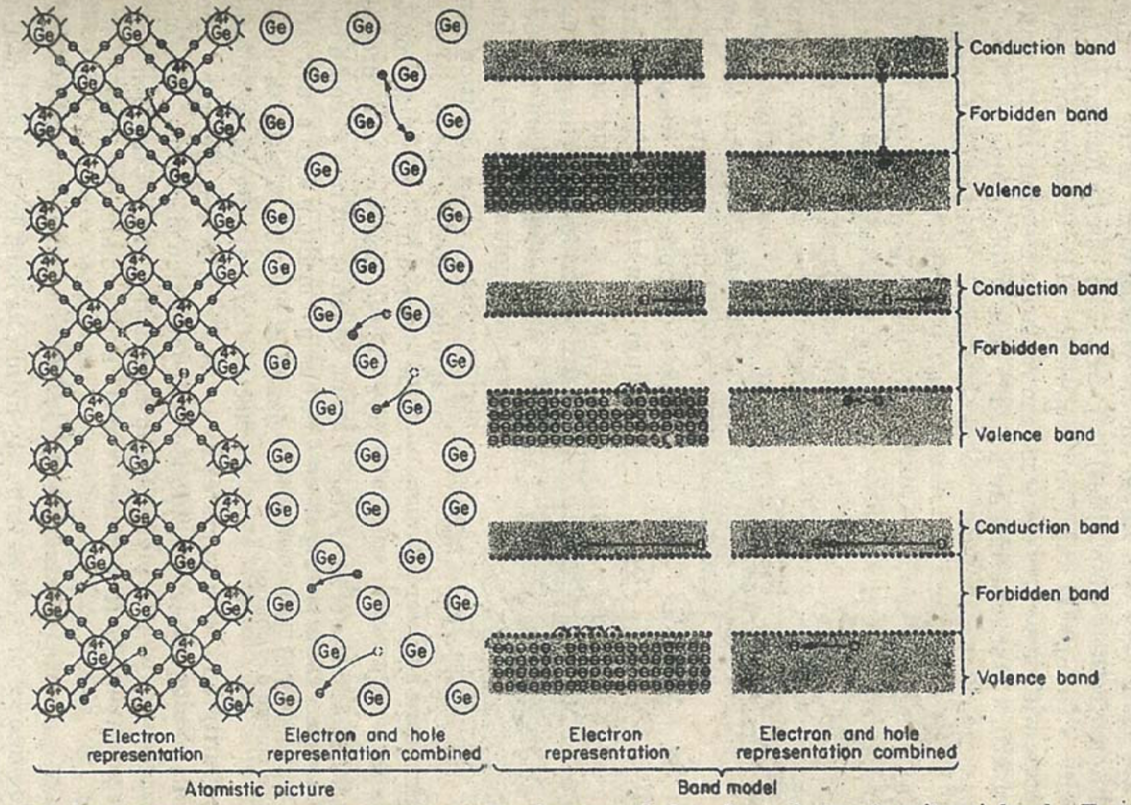


FIG. I.3.4. Formation of a conduction electron by the removal of a valence electron from its pair bond. Free migration of the conduction electron ⊖ and the hole ⊕.

imperfection as a collision partner of the conduction electrons. Now we shall treat an even more important function of these narrowly localized disorder points in the crystal lattices. The conditions in germanium are again very well suited for the visualization, as the substitutional imperfections (impurities) in germanium can be formed by the elements of Groups III and V in the periodic table, whereby an arsenic (As) atom with five outer electrons can replace a Ge atom in the lattice (Fig. I.3.5). More specifically an As^{5+} core is substituted in place of a Ge^{4+} core. Four of the five outer electrons of the As atom are used up in the four pair bonds for the binding to the four Ge neighbors, which compensates four positive charges of the As^{5+} core, so that there is only one positive charge left to bind the fifth outer electron. The field of this positive charge is weakened appreciably because the medium in which it acts is polarizable, with a dielectric constant $\varepsilon = 16$. Therefore the charge cloud of the fifth outer electron extends over a large volume around the impurity center. This condition can be approximated by the model of a single electron in the field of a positive charge, namely, a hydrogen atom, which is not in a vacuum but rather in a medium with a dielectric constant $\varepsilon = 16$. This fifth outer electron can be easily removed from the impurity and thus migrates away as a conduction electron: The impurity acts as a source or "donor" of conduction electrons. The entire process can be interpreted as the dissociation of a neutral donor $D^{\times}$ into a positively charged donor core D^{+} and a conduction electron $\ominus$:

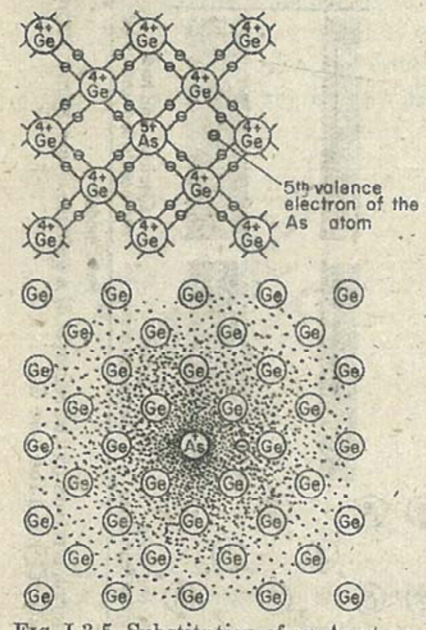


FIG. I.3.5. Substitution of an As atom for a Ge atom. Weak binding of the fifth valence electron in the form of an extended charge distribution.

$$D^{\times} \rightarrow D^{+} + \ominus \qquad \text{(I.3.01)}$$

The presence of such an impurity manifests itself in the band model through the lowering of one of the conduction-band levels of the unperturbed lattice into the forbidden band, so that the energy difference between the "impurity level" and the lower edge of the conduction band equals the dissociation energy of the donor. The donor is neutral if an electron is present in the impurity level; if the donor is ionized by, for instance, thermal excitation, the electron is raised from the impurity level to the conduction band. There is a fundamental difference concerning the character of the eigenfunction between the electrons in the impurity level and in the conduction band, besides the quantitative difference concerning the energy. The eigenfunctions of the conduction band are of the type of propagating waves, and the probability of finding the electron in a certain location is distributed evenly over all lattice cells, whereas the eigenfunction of the impurity level is concentrated around the donor in a hydrogen-like way. Even if the energy is sharply defined, the donor electron is strongly localized, whereas an electron in the conduction band can be localized only if the sharp energy fixation is sacrificed. In order to express this difference in the energy-level diagram, we shall introduce the abscissa in Fig. I.3.6 as a spatial coordinate within the semiconductor under consideration. The levels of the unperturbed lattice are thus drawn throughout the entire body, which signifies that the probability of finding an electron in a certain place is evenly distributed over all lattice cells. The impurity levels of the donor, however, are drawn as short dashes near the particular impurity in order to signify the localized character of an electron in this energy level.

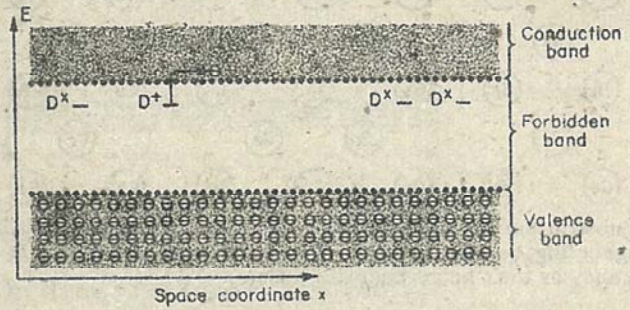


Fig. I.3.6. Band model with neutral and ionized donors.

So much for the substitutional incorporation of elements of Group V and the corresponding donors. As mentioned previously, the elements of Group III of the periodic table can also be incorporated in germanium. If a Ge atom is replaced by an In atom (Fig. I.3.7), only three valence electrons are introduced, so that one valence electron is missing in the four pair bonds to the four Ge neighbors. This vacancy can easily be filled by a valence electron from one of the neighboring bonds. The process of completing an incomplete pair bond from a neighboring complete pair bond by the motion of a valence electron allows the vacancy, or what we now call the hole, to move in the lattice.[1] The In atom remains in its original place and has now acquired an electron in addition to its original three valence electrons so that it is negatively charged. This is an example of an impurity which accepts an additional electron and is therefore called an "acceptor." This process includes the liberation of a hole and the charging of the impurity by one negative charge:

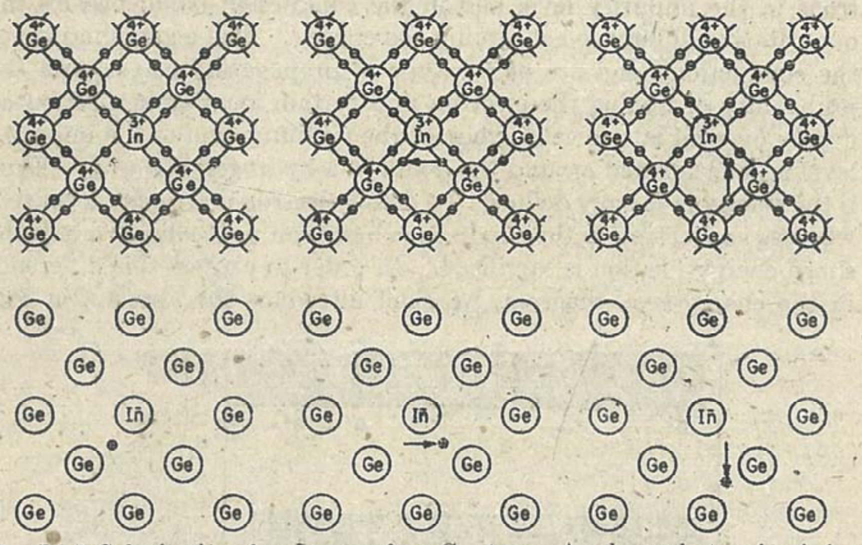


FIG. I.3.7. Substitution of an In atom for a Ge atom. A valence electron is missing so that neighboring valence electrons can move in (above). Thus the valence electron vacancy or the "hole" migrates (below).

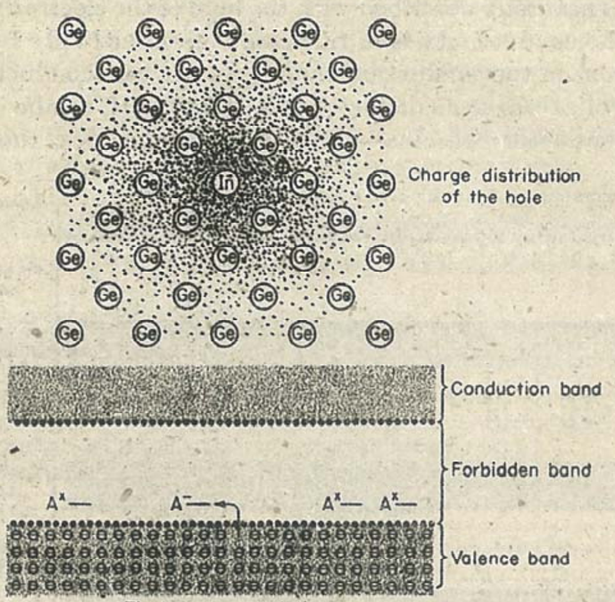
$$A^{\times} \rightleftarrows A^{-} + \oplus \qquad \text{(I.3.02)}$$

The acceptor can be represented by a "hydrogen" model, in the form of a positive "defect" electron in the field of a negative point charge, in a medium with the dielectric constant of germanium $\varepsilon = 16$ as in the case of the donor (Fig. I.3.8). Such an impurity introduces in

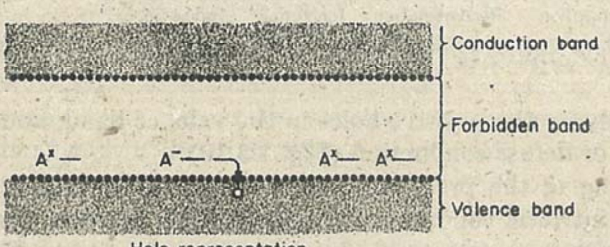
---

[1] This interpretation of the migration of a hole, namely, the successive motion of neighboring valence electrons into neighboring incomplete pair bonds in order to fill the holes, represents the atomistic or localized-particle standpoint. This should not lead one to ignore the wave nature of all electrons, including the valence electrons, which allows an unhindered migration of the valence electrons throughout the crystal on one hand and represents on the other hand the valence electrons as evenly "smeared out" over all pair bonds. The motion of one valence electron into the hole of an incomplete neighboring bond does not require any energy expenditure because it is not necessary to break a pair bond. An actual breaking of a pair bond occurs only if the valence electron is converted into a conduction electron, which obviously requires energy, rather than just changing place as a valence electron into another pair bond (see also pp. 18 to 20 and Fig. I.3.4).

the band model a localized impurity level above the valence band. Thermal excitation lifts an electron from the upper edge of the valence band to the impurity level of the neutral acceptor. Expressed in terms of holes, this is the ionization process $A^{\times} \rightarrow A^{-} + \oplus$, in which a hole is pushed down from the impurity level into the valence band.



Electron representation



Hole representation

FIG. I.3.8. The charge distribution of a hole around a substitutional In atom. A band model with neutral and ionized acceptors in the electron and the hole representation.

Electrons fall by themselves into lower levels, but energy must be supplied—for instance, by thermal or photon collisions—to raise an electron into a higher level. In contrast to this, holes spontaneously move upward in the band-level scheme like air bubbles in water, but energy is necessary to push them down into deeper levels.[1] Every

---

[1] See also pp. 375 and 377.

process in the band model can be expressed either in electron or hole language. The valence band may thus be regarded either as almost filled with electrons or as nearly devoid of holes. The choice between these two ways of expressing the same thing is a matter of convenience.[1]

A semiconductor which contains only donors as impurities, for instance, is naturally described with the help of the electron representation; for the conductivity is in this case determined only by the negative electrons in the conduction band. Such a semiconductor is called an *n*-type or excess semiconductor (Fig. I.3.9). If, on the other hand, the semiconductor contains only acceptor impurities, the current is carried only by the positive holes in the valence band, and one speaks of *p*-type or defect conduction (Fig. I.3.10).

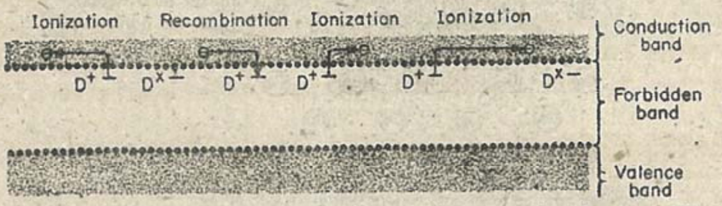


FIG. I.3.9. *n*-type conductor. Ionization and recombination equilibrium of donors: $D^{\times} \leftrightarrows D^{+} + \ominus$.

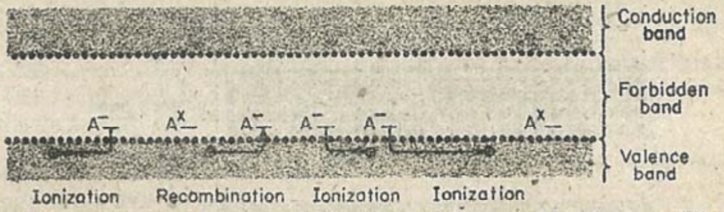


FIG. I.3.10. *p*-type conductor. Ionization and recombination equilibrium of acceptors: $A^{\times} \leftrightarrows A^{-} + \oplus$.

According to the preceding explanations, it appears as if the first described intrinsic conduction and the later introduced impurity conduction constitute two mutually exclusive opposites. This is, however, true only to a certain degree. We can understand this by considering a semiconductor with donors and analyzing its behavior with increasing temperature, starting out from a low enough temperature so that the thermal energy is barely sufficient to ionize an occasional donor. The conductivity is, in this case, only very small because of the low concentration of conduction electrons; however, one can increase the conductivity by raising the temperature, because a large number of not yet ionized donors is available (case of partial ionization). If the temperature is increased further, we approach the case

---

[1] See also pp. 65 to 66.

where all donors have given off their electrons to the conduction band. A further increase of temperature will not raise the conductivity as there are no neutral donors left (saturation case). A continuing temperature increase will, after a constant conductivity range, result in a further very steep conductivity increase caused by the mechanism of intrinsic conduction. This is the result of thermal excitation of electrons from the upper edge of the valence band to the lower edge of the conduction band. Since each such process produces an electron-hole pair, it is often called "pair formation." Impurity conduction and intrinsic conduction are thus not mutually exclusive, but can coexist. Impurity conduction predominates at low temperatures. The temperature at which the intrinsic conductivity overtakes the impurity conductivity depends only on the impurities. It is, of course, possible that in the case of very large forbidden bands and/or at very high impurity concentrations this temperature may be beyond the melting point of the crystal lattice.

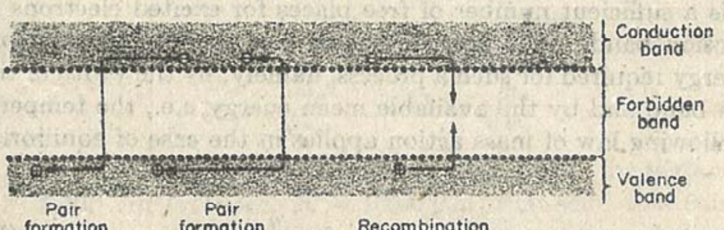


FIG. I.3.11. Intrinsic semiconductor. Pair formation and recombination equilibrium of the intrinsic semiconductor lattice: $0 \rightleftarrows \ominus + \oplus$.

We already mentioned the process of pair formation, namely, the simultaneous creation of an electron-hole pair by raising a valence electron from the upper edge of the valence band to the lower edge of the conduction band in the unperturbed lattice (Fig. I.3.11). There must be a complementary process to such a microscopic elementary process which occurs under equilibrium conditions as often as the elementary process. The recombination of an electron with a hole is the counterpart to the pair formation. Here a conduction electron in random thermal motion within the conduction band encounters a hole, namely, an electron vacancy in the otherwise full valence band, and falls back into the valence band. Such a process occurs more frequently the more frequently a conduction electron $\ominus$ meets a hole $\oplus$, in other words, the larger the number of holes. The number of recombinations per unit time and per unit volume is therefore proportional to the concentration of the positive holes p.[1] The fact that it is also

---

[1] Higher terms involving p^2, p^3, p^4, etc., can be neglected if the dilution is high enough.

proportional to the concentration of the negative electrons n follows from a similar consideration in which attention is fixed on a particular hole instead of an electron. Thus the same principles apply to the holes ⊕ as were previously applied to the electrons ⊖. We find that the number of recombinations per unit time and per unit volume is

$$r \cdot n \cdot p$$

where the factor r is called the recombination coefficient. The pair formation g, namely, the number of pair-formation processes per unit time and per unit volume, is, however, independent of the concentration. This means that whether a certain valence electron is raised from the valence band to the conduction band cannot depend on the concentration of electrons already present in the conduction band nor on the concentration of holes in the valence band—at least as long as there is a sufficient number of free places for excited electrons in the conduction band. The pair formation g is therefore determined by the energy required for such a process, namely, by the width of the forbidden band and by the available mean energy, i.e., the temperature. The following law of mass action applies in the case of equilibrium:

$$g = r \cdot n \cdot p$$

or[1]

$$p \cdot n = n_i^2 = \frac{g}{r} \qquad \text{(I.3.03)}$$

The so-called intrinsic density n_i is accordingly that concentration below which neither n nor p can fall without causing, at equilibrium, the other quantity p or n to rise above the concentration n_i. n_i could also be called the inversion density, because the transition from $n \gg n_i$, $p \ll n_i$ to $p \gg n_i$, $n \ll n_i$ is connected with the transition from n-type conductivity to p-type conductivity, which represents an inversion of the conductivity type of the semiconductor.

The name "intrinsic density" stems from the fact that in an intrinsic semiconductor the following neutrality requirement must be fulfilled:

$$n = p = n_i \qquad \text{(I.3.04)}$$

In order to render a semiconductor intrinsic at room temperature, it is necessary to reduce appreciably the generally present impurity concentration by purification[2] of the semiconductor. Hence it is plausible

---

[1] See also p. 305.

[2] The degree of extreme purity necessary for this purpose, in contrast to the meaning of the term purity in chemistry, can be demonstrated by the following numerical example. The intrinsic density n_i at room temperature is about 2.5×10^{13} cm^{-3}. In order to attain the intrinsic condition $n = p = n_i$, it is

that the condition of intrinsic conduction corresponds to the minimum conductivity attainable at a certain temperature in a particular semiconductor.

This is not entirely true, however. An extension of Eq. (I.2.06) gives for the conductivity

$$\sigma = e(\mu_n n + \mu_p p) \tag{I.3.05}$$

and with the law of mass action (I.3.03) this becomes

$$\sigma = e\left(\mu_n n + \mu_p \frac{n_i^2}{n}\right) \tag{I.3.06}$$

It is easy to see that the minimum of the conductivity is

$$\sigma = 2e\sqrt{\mu_n \mu_p}\, n_i \tag{I.3.07}$$

and that it is obtained for

$$n = n_i \sqrt{\frac{\mu_p}{\mu_n}} \qquad p = n_i \sqrt{\frac{\mu_n}{\mu_p}} \tag{I.3.08}$$

Since μ_n and μ_p are of the same order of magnitude (for instance,[1] in Ge $\mu_n = 3{,}600\ \text{cm}^2/\text{volt-sec}$, $\mu_p = 1{,}700\ \text{cm}^2/\text{volt-sec}$), this condition for minimum conductivity lies not too far from the condition for intrinsic conduction.

$$\sigma_i = e(\mu_n + \mu_p)n_i \tag{I.3.09}$$

$$n = n_i \qquad p = n_i \tag{I.3.04}$$

The described processes of pair formation and recombination can be represented by a reaction

$$0 \rightleftarrows \ominus + \oplus \tag{I.3.10}$$

in which the zero on the left side represents the fully periodic lattice which is unperturbed even with respect to the electron distribution. The law of mass action (I.3.03) which we formulated for this reaction equation (I.3.10) could similarly be established for the reaction equa-

---

necessary to reduce the impurity concentrations below $n \approx 2.5 \times 10^{13}\ \text{cm}^{-3}$, for each impurity contributes an electron or a hole. The concentration of the germanium atoms is $4.52 \times 10^{22}\ \text{cm}^{-3}$. Therefore there cannot be more than *one* impurity atom per $4.52 \times 10^{22}/2.5 \times 10^{13} = 1.8 \times 10^9$ germanium atoms. This signifies a purity in the chemical sense of between "9 and 10 nines"! On a straight line through the germanium lattice, one would encounter $\sqrt[3]{1.8 \times 10^9} = 1.2 \times 10^3$ germanium atoms between two impurity atoms. The numerical values are taken from E. M. Conwell, *Proc. IRE*, **40**: 1327 (1952).

[1] E. M. Conwell, *Proc. IRE*, **40**: 1330 (1952). P. P. Debye and E. M. Conwell, *Phys. Rev.*, **93**: 693 (1954).

tions (I.3.01) and (I.3.02) of the donor and the acceptor ionization, respectively; for the counterparts of donor and acceptor recombination exist here also, a fact which we have already taken into consideration in Figs. I.3.9 and I.3.10.

However, we shall leave this to the next chapter on impurity reactions and impurity equilibria, where we shall treat these conditions in more detail.

§4. Problems

1. Verify Eqs. (I.3.07) and (I.3.08). Give numerical values for the intrinsic conductivity and the minimal conductivity in germanium, assuming the values for n_i, μ_n, and μ_p given in §3. Give the corresponding resistivities.

2. Transform Eq. (I.3.06) into expressions for the electron and hole density as a function of the conductivity. For each conductivity value, two sets of electron and hole density values are possible. What are these values for a Ge crystal with a resistivity of 30 ohm-cm?

3. If μ_n and μ_p are not equal, the conductivity minimum corresponds to the type of conduction with the lower mobility, for $\mu_p < \mu_n$ to p-type conduction. For what hole density, then, will the conductivity again be equal to the intrinsic conductivity, and what would be the conductivity of a crystal whose electron concentration has that value? Give numerical values for germanium.

4.* (Cyclotron resonance.) The free carriers in a semiconductor move in circular orbits if the semiconductor is brought into a magnetic field. An electromagnetic wave passing through the semiconductor will be strongly absorbed if its frequency is equal to the rotational frequency of the carriers. Since the rotational frequency depends upon the mass of the carriers, one can in this way determine the effective mass of the electrons and holes.

Determine the mutual relationships between the following quantities: effective mass, rotational frequency, magnetic field strength, velocity of the carriers, radius of the orbit. Show that the rotational frequency does not depend on the velocity of the carrier. What does this mean for the sharpness of the resonance?

5.* In order to get a sharp cyclotron resonance, it is necessary that the mean free path l be long compared to the circumference of the cyclotron orbit. What magnetic field is necessary to satisfy this condition under the following two conditions: (1) $l = 10^{-5}$ cm, $T = 300°$ abs and (2) $l = 10^{-4}$ cm, $T = 10°$ abs, and an effective mass equal to the free mass in both cases?

Note Concerning Cyclotron Resonance. The effective masses, actually, are not constant but depend upon the direction within the crystal in which the electron (or hole) is accelerated. For our more illustrative purposes, we can neglect this directional dependence which goes far beyond the scope of this book. The interested reader is referred to Herman[1] or to Dresselhaus, Kip, and Kittel[2] for more details and references.

---

[1] F. Herman, *Proc. IRE.*, **43**: 1703 (1955).

[2] G. Dresselhaus, A. F. Kip, and C. Kittel, *Phys. Rev.*, **78**: 368 (1955).

CHAPTER II

The Nature, Models, and Reactions of Impurities and Imperfections

The band model, as described in the previous chapter, was formulated in the course of the development of the theory of metals and was taken over by semiconductor physics. The idea of the atomic lattice defect—the imperfection or impurity confined to a lattice point and its immediate vicinity—is, in contrast, typical of semiconductor physics, and its conception can, in a way, be considered as the beginning of modern semiconductor physics in general. It may be desirable to defer general considerations to the end of this chapter and to begin immediately with the discussion of specific concrete examples.

§1. Substitutional Impurities in Valence Crystals

Few types of imperfections are as easily understood with respect to their nature and characteristics as the substitutional impurities in the valence crystals of Group IV of the periodic table. In these lattices of the diamond type (Fig. II.1.1), such as the germanium lattice, each Ge atom is bound to four tetrahedrally arranged identical Ge neighbors by four electron-pair bonds. Any one germanium atom contributes four valence electrons, one to each pair bond, whereas the remaining electron of each pair bond, with opposite spin, is provided by the Ge neighbor to which the pair bond leads.

If one admixes or alloys a small amount of an element from Group III or V of the periodic table to a Ge sample, these atoms replace Ge atoms in their regular lattice positions, i.e., they are *substituted* for a Ge atom (Fig. II.1.2). For instance, an As atom introduces five valence electrons into the Ge lattice, but only four of these can be absorbed by the four pair bonds to the four Ge neighbors. The fifth valence electron is bound only very weakly to the As core if it remains at all attached to it.

The strength of this bond can be estimated, according to Bethe,[1] as follows:

The charge cloud of the fifth valence electron extends over a relatively large volume as a consequence of the weak binding and therefore encompasses many of the Ge neighbors of the arsenic core. This fact can be taken into consideration by the following approximation: The coulomb field of the positively charged arsenic core which carries a single positive charge, because of the absence of the fifth valence electron, is reduced by a factor $1/\varepsilon_{Ge}$, where $\varepsilon_{Ge} = 16.1$ is the macroscopic dielectric constant of Ge. Thus we have the well-known hydrogen problem in a medium with the dielectric constant ε_{Ge}. Formally,

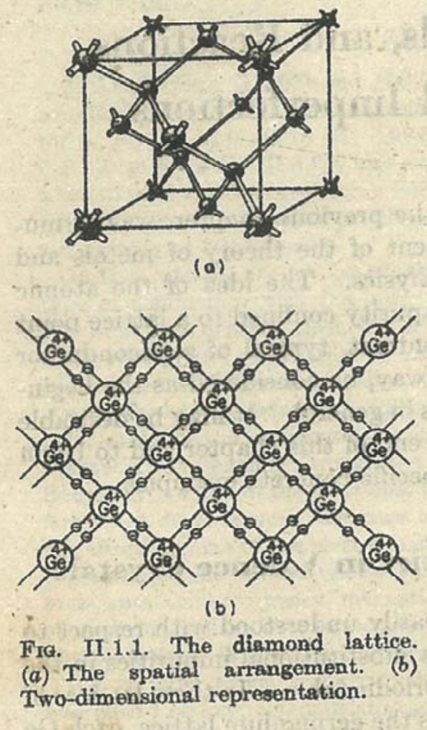


FIG. II.1.1. The diamond lattice. (a) The spatial arrangement. (b) Two-dimensional representation.

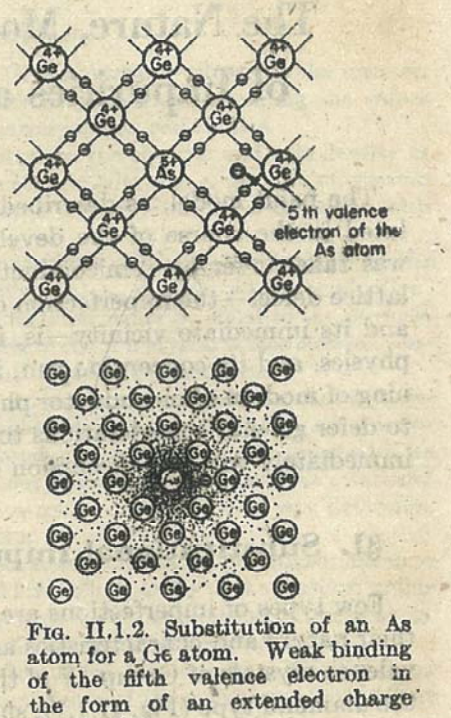


FIG. II.1.2. Substitution of an As atom for a Ge atom. Weak binding of the fifth valence electron in the form of an extended charge distribution.

---

[1] H. A. Bethe, R. L. Report No. 43-12, 1942. See also H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," pp. 65 and 66, McGraw-Hill Book Company, Inc., New York, 1948. The corresponding approximation for imperfections in ionic crystals can be found already in N. F. Mott and R. W. Gurney, "Electronic Processes in Ionic Crystals," pp. 80–86, Clarendon Press, Oxford, 1940 and 1948. Finally, the corresponding result for an electron in the field of a defect electron (hole), namely, for an "exciton," has been derived by G. H. Wannier, *Phys. Rev.*, **52**: 191 (1937).

this is equivalent to the vacuum problem with an effective atomic number $Z_{eff} = 1/\varepsilon_{Ge}$. The reduction of the coulomb field causes an increase of the first Bohr radius[1] $a_0 = \hbar^2/Z_{eff}me^2$ by the factor $1/Z_{eff} = \varepsilon_{Ge} = 16.1$, so that $a_0 = 16.1 \cdot 0.53 \cdot 10^{-8}$ cm $= 8.5 \cdot 10^{-8}$ cm. Further we obtain from the 1s eigenfunction $(1/\sqrt{a_0^3\pi})e^{-\frac{r}{a_0}}$ the radius $2a_0 = 17 \cdot 10^{-8}$ cm of a sphere which contains three quarters of the charge cloud of a 1s electron. The unit cell of the Ge lattice is a cube with a side dimension of $5.62 \cdot 10^{-8}$ cm containing eight Ge atoms, so that the previously mentioned spherical charge cloud includes $4(\pi/3)(17 \cdot 10^{-8}/5.62 \cdot 10^{-8})^3$ unit cells and $4\ (\pi/3)\ (3.02)^3 \cdot 8 \approx 925$ Ge atoms. This surprisingly large extension of the charge cloud in space justifies the calculation with the macroscopic dielectric constant. The ionization energy for the fifth valence electron can be calculated by multiplying the product of the ionization potential $2\pi^2me^4Z_{eff}^2/h^2$ of hydrogen with e by the factor $Z_{eff}^2 = 1/(\varepsilon_{Ge})^2 = 1/(16.1)^2 = 1/259.2$, which gives finally $13.59/259.2$ ev $= 0.052$ ev.

The binding of the fifth valence electron of a substitutional As atom to its As^+ core is very weak in comparison with that of the valence electron of hydrogen.

Hence even slight perturbations of thermal or other nature can remove the fifth valence electron, leaving the positively charged As^+ core. The substitutional impurity, just described, acts as a neutral electron "source" or "donor" $D^\times$ which is dissociated into a positive donor residue D^+ and a negative electron $\ominus$: $D^\times \rightarrow D^+ + \ominus$.

The picture is somewhat different if a Ge atom is replaced by an atom of Group III of the periodic table such as an indium atom (Fig. II.1.3), which introduces only three valence electrons into the Ge lattice. One of the four pair bonds leading from the lattice point of the indium atom to the four germanium neighbors is initially incomplete. This bond can be completed at the expense of another pair bond, if a valence electron of this other pair bond changes over to the originally incomplete pair bond of the substituted In atom.[2] The initially neutral indium atom becomes thus negatively charged, and the vacancy in a pair bond, the "hole" in the totality of the valence electrons, moves[3]

---

[1] See, for instance, H. Geiger and K. Scheel, "Handbuch der Physik," vol. XXIV, part 1, pp. 273 and 274, Springer-Verlag OHG, Berlin, 1933. The quadratic dependence of the ionization energy on the nuclear charge number Z can be explained as follows: A reduction in nuclear charge obviously causes a dilation of the charge cloud of the electron. Any mean orbital radius is therefore proportional to Z^{-1}. In order to obtain the ionization energy, one must introduce the value of the orbital radius into the denominator of the expression for the potential Ze/r. This is the reason for the proportionality of the ionization energy to Z^{+2}.

[2] See footnote 1, p. 22.

therefore in the force field of the negatively charged In atom. We recognize again a hydrogen problem, where, in contrast to the well-known vacuum case, the difference lies not only in the changed dielectric constant of the medium but also in an exchange of sign of the charges of nucleus and electron.

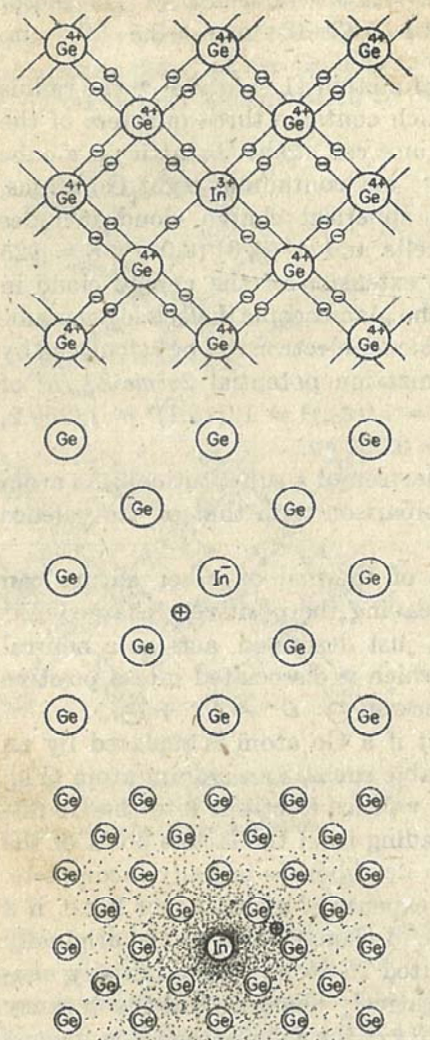


FIG. II.1.3. Substitutional impurity: replacing a Ge atom by an In atom.

Previously we found that the fifth valence electron of a substitutional As atom is bound only lightly to the As atom which acts as a donor. This result can now be applied to the case of the substitution of indium, and it can be said that a hole ⊕ is only weakly bound to an indium atom. This positive particle ⊕ is easily removed or given up, leaving a negatively charged indium atom or rather an ion In^-. The description of this process in the hole language is only a figure of speech, and the real process consists in the *acceptance* or capture of an electron in the force field of the In^-, which is the origin of the name *acceptor A* for such an impurity. Without regard to the origin of this name, we prefer, in this connection, the hole representation,[1] where we speak of the dissociation of a neutral acceptor $A^\times$ into a negative acceptor "residue" A^- and a positive hole ⊕

$$A^\times \rightarrow A^- + \oplus$$

We can summarize the preceding explanations in the following generalizations:

---

[1] See, however, p. 36, where, in ZnO, acceptors take electrons from the *conduction* band rather than from the *valence* band and thus reduce the initially present excess conductivity.

In the diamond-type valence lattice of the elements of Group IV of the periodic table, the chemical additions of elements from Groups III and V are built in *substitutionally*, which means that the foreign atoms occupy lattice points in the host lattice. This leads to donor formation from the elements of Group V, which give off negative electrons, and acceptor formation from the elements of Group III, which take up electrons or, in other words, give off holes.

In the following it is convenient to use an abbreviated symbolism for the various types of impurities and imperfections. In common with K. Hauffe, O. Stasiw, and J. Teltow (Fig. II.4.2), we adopt the system recently suggested by W. Schottky. The previously discussed substitutional impurities in valence lattices can be represented by

$$\begin{array}{lll} \mathrm{P}\bullet^{\cdot}\,(\mathrm{Ge}) & \mathrm{As}\bullet^{\cdot}\,(\mathrm{Ge}) & \mathrm{Sb}\bullet^{\cdot}\,(\mathrm{Ge}) \\ \mathrm{Al}\bullet'\,(\mathrm{Ge}) & \mathrm{Ga}\bullet'\,(\mathrm{Ge}) & \mathrm{In}\bullet'\,(\mathrm{Ge}) \end{array}$$

where, for instance, the first symbol $\mathrm{P}\bullet^{\cdot}\,(\mathrm{Ge})$ refers to the substitution of the substituent phosphor P for the substituted germanium (Ge) and $\bullet$ indicates the substitutional character of the impurity in contrast to the vacancies and interstitials to be described later. “·” and “′” indicate the effective state of charge[1] of the entire impurity, namely, the difference compared to the normally occupied lattice, so that “·” stands for a single positive charge, “··” a double positive charge, and “′” signifies a single negative charge, “″” a double negative charge, etc.

The foregoing symbols represent the substitutional impurities in the condition where they have given off their electron $\ominus$ or hole $\oplus$, respectively. If this is not the case and the electron $\ominus$, for instance, is located around the substituent P, on a hydrogenic orbit enlarged by ε_{Ge}, the impurity as a whole is neutral. This shall be indicated by the symbol “×”. In this associated condition, the foregoing impurities are represented by

$$\begin{array}{lll} \mathrm{P}\bullet^{\times}\,(\mathrm{Ge}) & \mathrm{As}\bullet^{\times}\,(\mathrm{Ge}) & \mathrm{Sb}\bullet^{\times}\,(\mathrm{Ge}) \\ \mathrm{Al}\bullet^{\times}\,(\mathrm{Ge}) & \mathrm{Ga}\bullet^{\times}\,(\mathrm{Ge}) & \mathrm{In}\bullet^{\times}\,(\mathrm{Ge}) \end{array}$$

The dissociation-association equilibria between a donor and an electron and between an acceptor and a hole, respectively, are given by the following reaction equations:

$$\mathrm{As}\bullet^{\times}\,(\mathrm{Ge}) \rightleftarrows \mathrm{As}\bullet^{\cdot}\,(\mathrm{Ge}) + \ominus$$
$$\mathrm{In}\bullet^{\times}\,(\mathrm{Ge}) \rightleftarrows \mathrm{In}\bullet'\,(\mathrm{Ge}) + \oplus$$

---

[1] Here the history of this actual state of charge is completely ignored. This remark will be of special importance in ionic crystals. See p. 40 particularly footnote [1].

§2. Substitutional Impurities in Ionic Crystals[1]

Substitutional impurities have also been investigated in ionic crystals, such as metal oxides, sulfides, selenides, and tellurides, by C. Wagner[2] and coworkers, K. Hauffe[3] and his students, and finally by E. J. W. Verwey, P. W. Haaymann, and F. C. Romeyn.[4] For instance, K. Hauffe and A. L. Vierk[5] replace Zn^{++} ions by Al^{+++} ions in a ZnO lattice, which consists of Zn^{++} and O^{--} ions, by adding Al_2O_3. In contrast to the unperturbed lattice, the substituted impurity is positively charged with a single charge and must therefore be represented with the Schottky symbol $\mathrm{Al}\,\bullet^{\cdot}\,(Zn)$. A negative electron can again be freed or captured in the coulomb field of this positive charge, so that this impurity acts as a donor:

$$\mathrm{Al}\,\bullet^{\times}\,(\mathrm{Zn}) \rightleftarrows \mathrm{Al}\,\bullet^{\cdot}\,(\mathrm{Zn}) + \ominus$$

Several complications enter in the calculation of the dissociation energy compared with the corresponding problem in valence crystals. One is the difference between the dielectric constant of the crystal for slowly varying electric fields and that for rapidly varying ones. An appreciable part of the dielectric screening of an electric field in an ionic crystal stems from the polarization of the crystal, because the positive and negative ions yield somewhat to the external field and are therefore displaced in space compared with their normal field-free position. This contribution of dielectric screening of an electric field

---

[1] We restrict ourselves to electronic conduction in ionic crystals, although the systematic substitution of ions of the host crystal was accomplished in crystals with predominantly ionic conduction earlier than in those with predominantly electronic conduction. In the case of the silver halides, we refer to E. Koch and C. Wagner, *Z. physik. Chem.*, **B38**: 295 (1937). O. Stasiw and J. Teltow, *Ann. Physik*, **5/40**: 181 (1941), and **6/1**: 261 (1947). O. Stasiw, *Ann. Physik*, **6/5**: 151 (1949). J. Teltow, *Ann. Physik*, **6/5**: 63, 71 (1949). J. Teltow, *Z. physik. Chem.*, **195**: 197, 213 (1950). The substitution of polyvalent ions in alkali halides was especially discussed by H. Pick, *Ann. Physik*, **5/35**: 73 (1939); and G. Heiland and H. Kelting, *Z. Physik*, **126**: 689 (1949).

[2] C. Wagner, *J. Chem. Phys.*, **18**: 62 (1950).

[3] K. Hauffe, *Ann. Physik*, **6/8**: 201 (1950). K. Hauffe and A. L. Vierk, *Z. physik. Chem.*, **196**: 160 (1950). K. Hauffe and J. Block, *Z. physik. Chem.*, **196**: 438 (1950). K. Hauffe and J. Block, *Z. physik. Chem.*, **198**: 232 (1951). K. Hauffe and H. Grunewald, *Z. physik. Chem.*, **198**: 248 (1951).

[4] E. J. W. Verwey, P. W. Haaymann, and F. C. Romeyn, *Chem. Weekblad*, **44**: 705 (1948). See also E. J. W. Verwey, P. W. Haaymann, F. C. Romeyn, and F. W. van Oosterhout, *Philips Research Repts.*, **5**: 173 (1950).

[5] K. Hauffe and A. L. Vierk, Über die elektrische Leitfähigkeit von Zinkoxyd mit Fremdoxydzusätzen, *Z. physik. Chem.*, **196**: 160 (1950).

disappears for rapidly changing electric fields, because the large and heavy ions are not capable of following the rapid field changes. The only remaining electrical screening is due to the polarization of the lattice ions themselves. NaCl, for instance, exhibits for static fields $\varepsilon = 5.62$, whereas for alternating fields above 5.10^{12} cycles, namely, for optical fields, $\varepsilon_{\text{opt}} = 2.25$. Based on the Franck-Condon principle, one can say that the separation of an electron from the charged perturbation center is such a sudden process that the ions must be considered as stationary or immobile. Therefore one introduces ε_{opt} instead of $\varepsilon_{\text{stat}}$ into the equation $E = -13.59/n^2\varepsilon^2$ ev. The resulting values are considerably higher than in the case of Ge ($\varepsilon = 16.1$) and Si ($\varepsilon = 11.2$), since $\varepsilon_{\text{opt}} = 2.25$ for NaCl. We find that the ground state with $n = 1$ lies below the continuum of the conduction band by about $13.59/1^2(2.25)^2$ ev ≈ 2.67 ev. This separation energy, which is enormous compared to that for valence crystals, has further consequences.

Besides the ground state we must now consider the excited states since the state with the quantum number 2 is still $13.59/2^2(2.25)^2$ ev $= 0.67$ ev below the conduction band. In the "dissociation" of a donor, for instance, an electron may first be raised by optical excitation from the ground state to the first excited state, for which only $(2.67 - 0.67)$ ev $= 2.00$ ev are necessary. The remaining step to the conduction band can, under certain circumstances, be effected thermally.

Thermal excitation for an energy step of 0.67 ev seems improbable, at least at room temperature. Here, another complication enters the picture. After the first optical excitation of the electron, we find that the ions in the vicinity of the impurity are not in equilibrium any more. They rearrange themselves in a slow and inertia-retarded process into a new equilibrium configuration of lower energy. The remaining excitation of the electron can now be effected by less than 0.67 ev, if it is thermal. In the case of the rapid single processes, one must consider the energies of the initial state and the end state without any ionic rearrangement, namely, for stationary ions, whereas the thermal-excitation energies are determined as differences between real equilibrium conditions arising in the statistics of collective systems with large numbers of participants.

Unfortunately, we must restrict ourselves to this short discussion and refer for further details to the well-known book by Mott and Gurney.[1] However, it appeared appropriate to indicate the reasons why the conditions in ionic crystals are fundamentally more complicated than in valence crystals.

---

[1] N. F. Mott and R. W. Gurney, "Electronic Processes in Ionic Crystals," Clarendon Press, Oxford, 1948. See, for instance, pp. 85, 115, and 160.

Up to now, we have mentioned only the formation of donors by introducing ions of higher valence. One can also build in lower-valence ions in order to form acceptors. ZnO is inherently an excess conductor, and we indicated previously in footnote [1] on page 32 that, in such a case, an acceptor A receives an electron from the conduction band with the liberation of energy and does not require energy to raise an electron from the valence band:

$$A^{\times} \rightarrow A^{-} - \ominus$$

The number of current carriers is thus reduced by 1, and the conductivity of the ZnO is lower as a consequence of the "trapping" of the previously freely migrating electron. Schottky speaks, in these cases, very descriptively of the "poisoning" of excess conduction through introduction of additional acceptors. We summarize once more as follows:

Addition of Al_2O_3 to ZnO produces[1] donors

$$\mathrm{Al}\,\bullet^{\times}\,(\mathrm{Zn}) \rightleftarrows \mathrm{Al}\,\bullet^{\cdot}\,(\mathrm{Zn}) + \ominus$$

Addition of Li_2O to ZnO produces acceptors

$$\mathrm{Li}\,\bullet^{\times}\,(\mathrm{Zn}) \rightleftarrows \mathrm{Li}\,\bullet^{\prime}\,(\mathrm{Zn}) - \ominus$$

Shortly before the experiments of K. Hauffe and A. L. Vierk,[2] E. J. Verwey, P. W. Haaymann, and F. C. Romeyn[3] obtained the following results with another host lattice, namely, NiO:

Addition of Cr_2O_3 to NiO produces donors

$$\mathrm{Cr}\,\bullet^{\times}\,(\mathrm{Ni}) \rightleftarrows \mathrm{Cr}\,\bullet^{\cdot}\,(\mathrm{Ni}) - \oplus$$

Addition of Li_2O to NiO produces acceptors

$$\mathrm{Li}\,\bullet^{\times}\,(\mathrm{Ni}) \rightleftarrows \mathrm{Li}\,\bullet^{\prime}\,(\mathrm{Ni}) + \oplus$$

The action on the conductivity of the p-type NiO is exactly opposite to that on the n-type ZnO, so that the introduction of low-valence ions increases the conductivity by the addition of acceptors. The introduction of higher-valence ions poisons, however, the p-type conduction of NiO by the addition of donors.

In order to understand the difference in behavior between ZnO and

---

[1] The oxygen balance is established through exchange with a surrounding gas phase (see p. 45).

[2] K. Hauffe and A. L. Vierk, Über die elektrische Leitfähigkeit von Zinkoxyd mit Fremdoxydzusätzen, *Z. physik. Chem.*, **196:** 169 (1950).

[3] E. J. W. Verwey, P. W. Haaymann, and F. C. Romeyn, *Chem. Weekblad*, **44:** 705 (1948). See also E. J. W. Verwey, P. W. Haaymann, F. C. Romeyn, and F. W. van Oosterhout, *Philips Research Repts*, **5:** 173 (1950).

NiO, we must know why ZnO is in itself an n-type conductor and why NiO is p-type. For this purpose we have to introduce a new concept, namely, that these two lattices contain atomic imperfections other than foreign impurities and that two different types of imperfections are effective in NiO and ZnO, respectively. Since we are not yet familiar with these types of imperfections, we shall first study them in valence crystals.

§3. Vacancies and Interstitials in Valence Crystals

If we ask about the possible types of *atomic* imperfections which can occur in a Ge crystal without any addition of foreign atoms, we find two possibilities for a deviation from the ideal lattice occupancy:

1. A Ge atom can be missing from a regularly occupied lattice point. A "lattice vacancy" is thus formed.
2. A lattice atom is in a so-called "interstitial position," which means that a Ge atom is placed in a suitable location between the atoms of an already complete diamond lattice. The diamond lattice, especially, offers a relatively large amount of space for these interstitials.

Like the substitutional impurities, the interstitials and the vacancies act as donors and acceptors, respectively. This is relatively easy to understand in the case of a Ge atom in an interstitial position.[1] The ionization energy of a Ge atom in vacuum is 8.13 ev. The Ge atom in an interstitial position acts roughly in the same manner as a Ge atom with a dielectric constant $\varepsilon_{\text{germanium}} = 16.1$. Its ionization energy is therefore reduced by the factor $1/(\varepsilon_{\text{germanium}})^2 = \frac{1}{259}$ compared with the vacuum value, yielding only 8.13/259 ev = 0.0314 ev. We see that a Ge atom in an interstitial position can give off an electron with an excitation energy of only 0.0314 ev, thus acquiring a positive charge:

$$\text{Ge}\,\bigcirc^{\times} \rightleftarrows \text{Ge}\,\bigcirc^{\cdot} + \ominus$$

The second ionization energy of germanium in vacuum is 16.0 ev. If the calculation with the macroscopic dielectric constant $\varepsilon = 16.1$ were applicable to the case of double ionization of the Ge atom in the interstitial position, we would find the still very small value of 16.0/259 ev = 0.0618 ev. It is, however, questionable whether this ε calculation is permissible, for the second electron to be removed from the Ge core moves in the force field of a doubly positive charge of the remaining core, so that its orbit radius must be considerably smaller

---

[1] Schottky symbol: Ge $\bigcirc^{\times}$ or Ge $\bigcirc^{\cdot}$, depending on the charge condition (see Fig. II.4.2).

than that of the first electron. James and Lark-Horowitz[1] favor the interpretation of the interstitial Ge atom as a doubly ionizable donor. The higher ionization energies beyond the second one are then assumed to be larger than the width of the forbidden band, which eliminates further imperfection levels in the forbidden band. These higher ionization levels would fall in the valence band and would hence be of no importance for the conduction processes.

The action of a lattice vacancy as an acceptor is more difficult to understand. One may give the following explanation: The Ge^{++++} ions are connected by the valence electron bonds in a perfect lattice, as we have described previously (Fig. II.1.1). The common point of the four valence bonds to the four nearest neighbors is largely screened from the outside by the Ge^{++++} under consideration, so that it is much less likely for a conduction electron to approach the Ge^{++++} ion than for a valence electron. In addition, there is a certain repulsion of the conduction electron by the valence electrons. The energy of a conduction electron, in order to approach the Ge^{++++} ion, must be appreciably higher than that of a valence electron, which manifests itself in the higher energy of the levels in the conduction band compared with those in the valence band, as indicated by the width of the forbidden band. If at one lattice point a Ge^{++++} including its valence electrons, namely, a whole Ge atom, is missing, there seems to be no reason for a conduction electron either to seek or to avoid this point in view of charge neutrality. However, there is an appreciable amount of empty space at this vacancy location. An electron approaching this point is not attracted by a positive Ge^{++++} ion, but it will move as far as possible from the surrounding valence electrons in view of their repulsion, so that the vacancy location becomes an energetically more favorable position for the electron than the normal level in the conduction band. Nevertheless, valence electrons are still favored energetically over conduction electrons in view of the absence of an attracting Ge^{++++} ion. Thus a localized electron energy level below the conduction band but above the valence band in the forbidden band is formed by the Ge vacancy, which becomes charged negatively if it is occupied by an electron. This means that the lattice vacancy acts as an acceptor, for an acceptor is an imperfection which provides electrons with an energy level below the conduction-band levels and becomes negatively charged in the energy-liberating process of capture.

Lark-Horowitz and his coworkers[2] produced vacancies and inter-

---

[1] H. M. James and K. Lark-Horowitz, *Z. physik. Chem.*, **198**: 107 (1951).

[2] K. Lark-Horowitz, in "Semiconducting Materials" (Reading Report), p. 47, Butterworths Scientific Publications, Ltd., London, 1951.

stitials in Ge and Si by bombardment with high-energy particles of various types, which led to a better understanding of the role of these imperfection types.[1]

The annealing effects found by Scaff and Theuerer[2] are not due to vacancies but rather a consequence of diffusion and precipitation processes of metallic impurities (mainly Cu).[3] If these effects are eliminated by working with extreme cleanliness and other precautions, p conductivity is observed which was previously overshadowed by the much larger Scaff-Theuerer effects. These acceptors are beyond doubt caused by thermally created imperfections.[4]

§4. Vacancies and Interstitials in Ionic Crystals

Vacancies acting as acceptors are easier to understand in ionic crystals than in valence crystals. Let us consider the heretofore practically important and intensively investigated cuprous oxide (Cu_2O). During the preparation of the samples to be investigated, an excess of oxygen was introduced by the proper choice of conditions. This oxygen excess should rather be called a defect in the stoichiometric metal content, for the deviation from stoichiometry is caused by vacancies in the lattice of the Cu^+ ions. There is no doubt that such a vacancy acts electrostatically like an additional negative charge; for one positive charge is missing at the vacated lattice point. The coulomb field of the additional negative charge can thus capture or give off a positive hole. The imperfection as a whole is neutral if the hole is captured; otherwise it is negatively charged. The vacancy in

---

[1] It remains to be seen whether the detailed interpretations of Lark-Horowitz will prove valid in the long run. James and Lark-Horowitz themselves point out in *Z. physik. Chem.*, **198:** 112 (1951), that their interpretation does not take into consideration an accumulation and cloud formation of vacancies and interstitials. Such accumulations at the periphery of the trajectories of fast particles appear more probable than a homogeneous distribution over the entire Ge lattice.

[2] J. H. Scaff and H. C. Theuerer, *Trans. Am. Inst. Mining Met. Engrs.*, **189:** 59 (1951).

[3] C. S. Fuller, H C. Theuerer, and W. van Roosbroeck, *Phys. Rev.*, **85:** 678 (1952). C. S. Fuller and J. D. Struthers, *Phys. Rev.*, **87:** 526 (1952). W. P. Slichter and E. D. Kolb, *Phys. Rev.*, **87:** 527 (1952). C. Goldberg, *Phys. Rev.*, **88:** 921 (1952). L. Esaki, *Phys. Rev.*, **89:** 1026 (1953). K. Seiler, D. Geist, K. Keller, and K. Blank, *Naturwiss.*, **40:** 56 (1953). F. van der Maesen, P. Penning, and A. van Wieringer, *Philips Research Repts.*, **8:** 241 (1953). G. Finn, *Phys. Rev.*, **91:** 754 (1953). F. van der Maesen and J. A. Brenkman, *Philips Research Repts.*, **9:** 225 (1954).

[4] R. A. Logan, *Phys. Rev.*, **91:** 757 (1953). S. Mayburg and L. Rotondi, *Phys. Rev.*, **91:** 1015 (1953).

the lattice of the positive Cu^+ ions acts as an acceptor according to the reaction equation with the Schottky symbol $\mathrm{Cu}\,\square'$.[1]

$$\mathrm{Cu}\,\square' + \oplus \rightleftarrows \mathrm{Cu}\,\square^{\times}$$

While the effective imperfections in $\mathrm{Cu_2O}$ are vacancies in the *cation* lattice, we recognize the well-known F centers, introduced by Pohl,[2] as vacancies in the anion lattice of the alkali halides (Fig. II.4.1). A vacancy in the lattice of the Cl^- ions in KCl acts as an attraction center for negative excess electrons:

$$\mathrm{Cl}\,\square^{\cdot} + \ominus \rightleftarrows \mathrm{Cl}\,\square^{\times}$$

The F center is equivalent to such a donor in the *associated* state (namely, the imperfection $\mathrm{Cl}\,\square^{\times}$). In order to obtain a model for this imperfection,[3] it is important to introduce the low infrared value of the dielectric constant, which lies between 1.74 and 3.80, into the hydrogen analog of Mott and Gurney.[4] The vacuum value of the Bohr radius of $0.53 \cdot 10^{-8}$ cm is thereby increased only very little to 0.92 to $2.01 \cdot 10^{-8}$ cm. Even if, as in §1 (page 31), we take again twice the Bohr radius to define a sphere which contains 76.2 per cent of the charge cloud of the electron, we arrive at only 1.84 to $4.02 \cdot 10^{-8}$ cm. The lattice constants of the alkali halides are, however, between 2.07 and $3.66 \cdot 10^{-8}$ cm. Although the hydrogen model still has a limited conceptual significance in the alkali halides, it is not nearly so meaningful as in Ge ($\varepsilon = 16.1$), Si ($\varepsilon = 11.9$), and also in $\mathrm{Cu_2O}$ ($\varepsilon = 12$); for the charge cloud of the outermost electron is concentrated in the immediate vicinity of the halogen vacancy, so that we are inclined to distribute the electron in a statistical manner among the six neighboring metal ions. The resulting model of a color center becomes thus a halogen vacancy with a neutral metal atom $\mathrm{K}^{\times}$ (Fig. II.4.1) as neighbor.[5]

[1] $\mathrm{Cu}\,\square'$ represents, by definition, an imperfection in the form of the absence of a Cu^+ ion as compared with the unperturbed lattice. Whether the neutral atom is removed first with the subsequent removal of a hole or the vacancy is formed by the removal of a Cu^+ ion is quite irrelevant. The symbol "′" describes the actual charge condition of the imperfection without regard to the previous history.

[2] R. W. Pohl, *Physik. Z.*, **39:** 36 (1938).

[3] W. Schottky, *Z. physik. Chem.*, **29/B:** 335 (1935), particularly p. 342; and *Wiss. Veröffentl. Siemens-Werken*, **14** (2): 1 (1935), particularly middle of p. 4. J. H. de Boer, *Trav. chim. Pays-Bas*, **56:** 301 (1937).

[4] Mott and Gurney, *op. cit.*, p. 12, Table 5.

[5] Even the experiments on the spin resonance of the F-center electron do not appear to lead to a decision between the two models. See in connection with this, A. F. Kip, C. Kittel, R. A. Levy, and A. M. Portis, *Phys. Rev.*, **91:** 1066 (1954); and D. L. Dexter, *Phys. Rev.*, **93:** 244 (1954).

The high dissociation energies of F centers, of the order of 2 to 3 ev, are a further consequence of the low effective values of the dielectric constant for fast processes. Thermal dissociation of the electron and the Cl vacancy is therefore for all practical purposes out of question. The raising of the electron to an excited state by light, with the subsequent thermal transition into the conduction band, is the process which led to the name of this type of imperfection (F center = color center).[1]

The action mechanism of ions in interstitial positions in the lattice is even more obvious. For instance, ZnO contains probably Zn^{++} ions in interstitial positions[2] which are plentiful in the wurtzite lattice[3] of ZnO. The positively charged Zn^{++} act of course as attraction centers for negative excess electrons. A zinc atom in an interstitial position acts therefore as a donor:

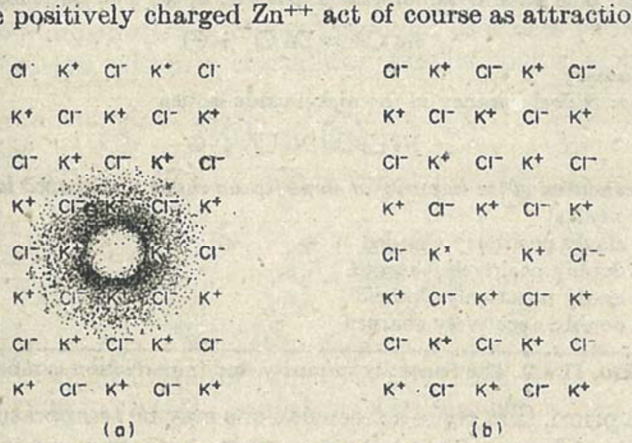


FIG. II.4.1. Model of a Pohl F center. (*a*) Charge cloud of the electron around the Cl^- vacancy. (*b*) Neutral K atom adjacent to the Cl^- vacancy. The charge cloud of the electron around the Cl^- vacancy is not very far-reaching. Thus representation *b* is possible.

$$\mathrm{Zn}\,\bigcirc^{\times} \rightleftarrows \mathrm{Zn}\,\bigcirc^{\cdot} + \ominus$$

We mention merely in passing that the halogen vacancies in the alkali halides can also capture two electrons (formation of F′ centers):

$$\mathrm{Cl}\,\square^{\cdot} + 2\ominus = \mathrm{Cl}\,\square^{\times} + \ominus = \text{F center} + \text{electron} = \mathrm{Cl}\,\square' = \text{F}'\text{ center}$$

---

[1] See Mott and Gurney, *op. cit.*, p. 135. See further the discussion on pp. 34 to 35.

[2] F. Stöckmann, *Z. Physik*, **127**: 563 (1950). Another type of imperfection in ZnO is probably the anion vacancy $\mathrm{O}\,\square^{\cdot}$ or $\mathrm{O}\,\square^{\cdot\cdot}$, which acts like $\mathrm{Zn}\bigcirc^{\times}$ as donor:

$$\mathrm{O}\,\square^{\times} \rightleftarrows \mathrm{O}\,\square^{\cdot} + \ominus$$

[3] The wurtzite lattice is closely related to the diamond lattice, for which the interstitial positions were shown in Fig. II.3.1.

Furthermore, we shall not enter into a discussion of the effects which occur when several imperfections form closely associated clusters.[1]

We turn now to the question of the laws which govern the concentration of the various impurity and imperfection types in a crystal. We assume, a priori, that these concentrations may be temperature dependent and that changes in temperature cause impurity and imperfection reactions, which lead to changes in the respective concentrations.

$\bullet\,(\;)$ = Substitutional Impurity

Example: Arsenic atom in a germanium position in the germanium lattice

$$\text{As}\,\bullet^{\times}\,(\text{Ge}) \rightleftarrows \text{As}\,\bullet^{\cdot}\,(\text{Ge}) + \ominus$$

$\bigcirc$ = *Interstitial Occupancy*

Example: Zinc atom in an interstitial position in the zinc oxide lattice

$$\text{Zn}\,\bigcirc^{\times} \rightleftarrows \text{Zn}\,\bigcirc^{\cdot} + \ominus$$

$\square$ = *Vacancy*

Example: Nickel vacancy in the nickel oxide lattice

$$\text{Ni}\,\square^{\times} \rightleftarrows \text{Ni}\,\square' + \oplus$$

Charge condition of the impurity or imperfection in the undisturbed lattice:

$\times$ = neutral
$\cdot$ = singly positively charged
$\cdot\cdot$ = doubly positively charged
$'$ = singly negatively charged
$''$ = doubly negatively charged

FIG. II.4.2. The Schottky impurity and imperfection symbols.

§5. Impurity and Imperfection Reactions

To permit dissociation or association processes during a chemical reaction, at least one of the reaction partners must be mobile, since spatial separations or combinations are involved. We are by now fully acquainted with the mobility of electrons $\ominus$ and holes $\oplus$, respectively, so that the impurity and imperfection reactions mentioned so far do not require a new concept. They constitute an analogy to the dissociation-association equilibrium of a donor

$$D^{\times} \rightleftarrows D^{+} + \ominus$$

or an acceptor

$$A^{\times} \rightleftarrows A^{-} + \oplus$$

---

[1] See, for instance, the comprehensive summary of H. Pick, *Naturwiss.*, **38**: 323 (1951), particularly pp. 328 and 329; and F. Seitz, *Phys. Rev.*, **83**: 134, (1951), particularly pp. 136 and 137.

The occurrence of diffusion, electrolysis, and in general the processes of chemical reactions in solids shows that the migration of matter is, after all, possible in solids which seem to be so compact and impenetrable. Although macroscopic imperfections such as cracks and large dislocations can play a role (e.g., diffusion at grain boundaries), it has become increasingly clear during the last twenty years that migration of matter can take place even within a single crystallite. This is possible only if the lattice atoms are not irremovably fixed in their prescribed lattice positions, so that deviations from the ideal lattice structure can occur. The presence of atomic disorder, i.e., the existence of atomic imperfections, allows migration processes through the transport of matter within a crystallite. Similar to the mobility of electrons ⊖ and holes ⊕ in the crystal lattice, a lattice vacancy Ge □, for instance, possesses a certain mobility in a Ge lattice which is, however, orders of magnitude smaller than the electronic mobilities $\mu_{\ominus}$ and $\mu_{\oplus}$. This mobility is particularly easy to comprehend in the case of a vacancy. If a neighboring atom oscillates momentarily with an unusually large amplitude of its thermal vibration, it can jump over into the new equilibrium position, provided that the amplitude is in the direction of the vacancy (Fig. II.5.1). Thus the Ge atom has moved the distance between two adjacent atoms to the left, whereas the vacancy has moved the same distance to the right. If this process is repeated, one can speak of a migration of the vacancy rather than an atom, for a different Ge atom is involved in the move in each case. The migration of a vacancy is, in effect, nothing else than the transport of the charge and mass of an atom through the crystal in the opposite direction.

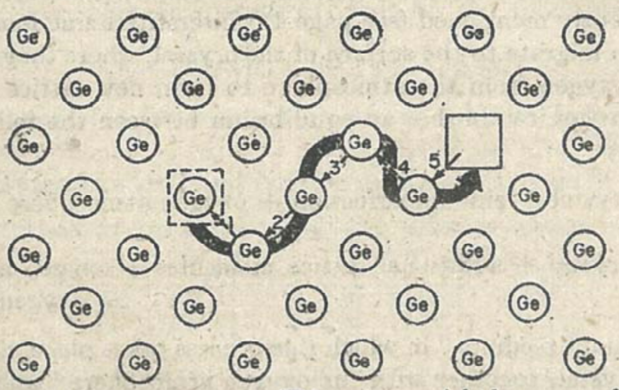


FIG. II.5.1. Migration of a vacancy Ge □ from left to right by means of the steps 1, 2, 3, 4, and 5 of the neighbor atoms from right to left.

A Ge atom which is located in an interstitial position can also migrate by jumping from one interstitial position to the next adjacent one, for there is plenty of space in the Ge lattice for this process. Furthermore, experimental observations leave no doubt that substitutional impurities and, in general, all types of imperfections have a certain mobility in the valence as well as in the ionic crystals.

Thus with the introduction of the Schottky symbols for the various imperfection and impurity types, the number of possible reactions to be investigated has increased enormously compared with the few just mentioned. Although the Schottky symbols and the introduction of reaction equations may initially have appeared to be trivial, they now become indispensable aids in visualizing the many phenomena.

We shall mention, as a first example, the reaction between zinc oxide and a surrounding oxygen atmosphere at high temperatures > 500°C. The previously mentioned (see page 41) interstitial zinc ions $\mathrm{Zn}\,\bigcirc^{\cdot}$ or $\mathrm{Zn}\,\bigcirc^{\times}$ can migrate to the surface of the crystal, where they recombine with the oxygen from the atmosphere to form new lattice molecules. Such a reaction establishes an equilibrium between the following two conditions:

Original crystal + zinc imperfection + oxygen atmosphere

$$\leftrightarrows$$

Original crystal + additional lattice molecules + oxygen atmosphere
− oxygen molecules

The actual "medium" in which this process takes place, namely, the original crystal, together with the oxygen atmosphere, can be left out on both sides of the reaction symbol $\rightleftarrows$. We obtain in a more quantitative form:

$$2\mathrm{Zn}\,\bigcirc^{\times} \rightleftarrows 2\mathrm{ZnO} - \mathrm{O}_2^{(\mathrm{gas})}$$

The neutral interstitials $\mathrm{Zn}\bigcirc^{\times}$ give off at least one electron $\ominus$ at these high temperatures

$$\mathrm{Zn}\,\bigcirc^{\times} \rightarrow \mathrm{Zn}\,\bigcirc^{\cdot} + \ominus$$

so that the final equation can be written

$$2\mathrm{Zn}\,\bigcirc^{\cdot} + 2\,\ominus + \mathrm{O}_2^{(\mathrm{gas})} \rightleftarrows 2\mathrm{ZnO} \qquad \text{(II.5.01)}$$

A further example is the reaction between an NiO crystal and a surrounding O_2 atmosphere. In analogy to the preceding ZnO example, crystal molecules are formed at the crystal surface and Ni is transported from the interior of the crystal to the surface by the migration of the imperfections in NiO. These imperfections, in contrast to the interstitial metal ions of ZnO, are negatively charged or neutral nickel vacancies[1] $\mathrm{Ni}\,\square''$ or $\mathrm{Ni}\,\square'$ or $\mathrm{Ni}\,\square^{\times}$. In order to effect a nickel transport from the interior to the surface, the nickel vacancies $\mathrm{Ni}\,\square^{\times}$ must migrate from the surface into the interior, so that the reaction equation has the following form:

$$\begin{gathered}\text{Original crystal} + \text{oxygen atmosphere}\\ \leftrightarrows\\ \text{Original crystal} + \text{additional lattice molecules} + \text{nickel vacancies}\\ \qquad\qquad + \text{oxygen atmosphere} - \text{oxygen molecules}\end{gathered}$$

or in the more quantitative form:

$$0 \rightleftarrows 2\mathrm{NiO} + 2\mathrm{Ni}\,\square^{\times} - \mathrm{O}_2^{(\mathrm{gas})}$$

The probable liberation of a hole $\oplus$ from the neutral nickel vacancy $\mathrm{Ni}\,\square^{\times}$ at these high temperatures leads to the following reaction equation:

$$\mathrm{O}_2^{(\mathrm{gas})} \rightleftarrows 2\mathrm{NiO} + 2\mathrm{Ni}\,\square' + 2\oplus$$

Further examples for reaction equations are the following four, which describe the introduction of ions of higher and lower valence into the two host lattices ZnO and NiO, as discussed on page 36:

$$\begin{gathered}\mathrm{Al_2O_3} \rightleftarrows 2\mathrm{Al}\,\bullet^{\cdot}\,(\mathrm{Zn}) + 2\ominus + 2\mathrm{ZnO} + \tfrac{1}{2}\mathrm{O}_2^{(\mathrm{gas})}\\ \mathrm{Li_2O} \rightleftarrows 2\mathrm{Li}\,\bullet'\,(\mathrm{Zn}) - 2\ominus + 2\mathrm{ZnO} - \tfrac{1}{2}\mathrm{O}_2^{(\mathrm{gas})}\\ \mathrm{Li_2O} \rightleftarrows 2\mathrm{Li}\,\bullet'\,(\mathrm{Ni}) + 2\oplus + 2\mathrm{NiO} - \tfrac{1}{2}\mathrm{O}_2^{(\mathrm{gas})}\\ \mathrm{Cr_2O_3} \rightleftarrows 2\mathrm{Cr}\,\bullet^{\cdot}\,(\mathrm{Ni}) - 2\oplus + 2\mathrm{NiO} + \tfrac{1}{2}\mathrm{O}_2^{(\mathrm{gas})}\end{gathered}$$

These equations describe the conductivity-increasing or decreasing effect of the foreign ions and the oxygen exchange with the surrounding atmosphere. The conductivity can be calculated as a function of the partial oxygen pressure of the surrounding atmosphere, as will be shown on page 51 with the help of simple examples.

§6. Calculation of Equilibrium Concentrations with the Aid of the Laws of Mass Action

The significance of these reaction equations, some of which we have encountered previously, goes far beyond a means of general orientation.

---

[1] The energy necessary for the creation of a certain type of imperfection determines whether a lattice contains predominantly interstitials or vacancies or both, namely, cations or anions. Concerning the calculation of such activation energies, see W. Jost, *Trans. Faraday Soc.*, **34** II: 860 (1938). W. Jost, "Diffusion und chemische Reaktion in festen Stoffen," T. Steinkopf, Leipzig, 1937. N. V. Mott and M. J. Littleton, *Trans. Faraday Soc.*, **34**: 485 (1938). E. S. Rittner, R. A. Hutner, and F. K. Du Pré, *J. chem. Phys.*, **17**, 198, 204 (1949); **18**: 379 (1950). P. Brauer, *Z. Naturforsch.*, **7a**: 372 (1952).

They are rather the basis for the computation of equilibrium concentrations of the reaction partners. We shall demonstrate this with the help of one of the simplest reactions, namely, the dissociation and association of a donor:

$$D^\times \rightleftarrows D^+ + \ominus$$

The concentrations of the participating partners in such a reaction follow the so-called law of mass action[1]

$$K_D \cdot n_{D^\times} = n_{D^+} \cdot n_\ominus$$

This law is best explained with the help of the following reasoning.[2] The number of dissociation processes per unit time and volume is proportional to the concentration $n_{D^\times}$ and inversely proportional to the mean lifetime $\tau_{D^\times}$ of the undissociated donors:

$$\frac{\text{Dissociation processes}}{\text{cm}^3\ \text{sec}} = \frac{1}{\tau_{D^\times}} \cdot n_{D^\times} \tag{II.6.01}$$

The number of association processes per unit time and volume is proportional to the concentrations of the two reaction partners, for association can take place only when two reaction partners D^+ and $\ominus$ meet in space, which occurs more often the more representatives of the two types are present:[3]

$$\frac{\text{Association processes}}{\text{cm}^3\ \text{sec}} = r_D n_{D^+} \cdot n_\ominus \tag{II.6.02}$$

r_D is the recombination coefficient for the recombination D^+ and $\ominus$. The number of dissociation and association processes must be the same at equilibrium:

$$\frac{1}{\tau_{D^\times}} n_{D^\times} = r_D n_{D^+} \cdot n_\ominus \tag{II.6.03}$$

The law of mass action is thus

$$\overset{7}{K}_D n_{D^\times} = n_{D^+} \cdot n_\ominus \tag{II.6.04}$$

---

[1] In the following, we shall encounter the concentrations of many other "particles" such as positively charged sodium vacancies Na □˙. We choose the symbol n for these values with the Schottky symbol as subscript, so that the above example is described by $n_{\text{Na}\,\square^\cdot}$. Correspondingly, we shall write $n_\oplus$ and $n_\ominus$ for the concentrations of electrons and holes instead of n and p in this chapter as well as in Chap. IX.

[2] The kinetic foundation of the law of mass action, as previously sketched, is treated in more detail on pp. 329 to 330. For the foundation on a statistical basis, see pp. 317 to 320.

[3] The fact that no further sum terms are added, such as $n_{D^+}^2 n_\ominus^1$, $n_{D^+}^3 n_\ominus^1$, . . . , $n_{D^+}^1 n_\ominus^2$, $n_{D^+}^1 n_\ominus^3$, . . . , etc., is based on the assumption of "sufficient" dilution of the reaction partners to which we restrict this discussion.

with the mass-action constant

$$K_D = \frac{1}{\tau_D \tau_{D^\times}} \tag{II.6.05}$$

If the temperature of the crystal is high enough, a thermal collision often has enough energy to dissociate an associated donor. In this case, the lifetime $\tau_{D^\times}$ is small and the mass-action constant $K_D = 1/\tau_D \tau_{D^\times}$ large. Statistics teaches in detail[1] that

$$K_D = \tfrac{1}{2} N_C \, e^{-\frac{E_{CD}}{kT}} \tag{II.6.06}$$

where N_C is the so-called effective density of states in the conduction band[2] $2.5 \cdot 10^{19}\ \text{cm}^{-3}(m_{\text{eff}}/m)^{3/2}(T/300°\text{K})^{3/2}$ and E_{CD} is the dissociation energy of a donor.

The concentrations $n_{D^\times}$, n_{D^+}, and $n_\ominus$ are not determined by the law of mass action alone. We must resort to additional equations which depend on the conditions of the experiment. In experiments far enough below the melting point, the total number n_D of donors per unit volume of the crystal is in general independent of temperature:[3]

$$n_{D^\times} + n_{D^+} = n_D \tag{II.6.07}$$

From this total balance (II.6.07) of the donors and from the law of mass action (II.6.04), we obtain

$$n_{D^\times} = n_D \frac{1}{1 + \dfrac{K_D}{n_\ominus}} \tag{II.6.08}$$

$$n_{D^+} = n_D \frac{1}{1 + \dfrac{n_\ominus}{K_D}} \tag{II.6.09}$$

and we see that the charge condition of the donors is dependent on the electron concentration $n_\ominus$ which can be adjusted by experimental conditions that are independently variable. We shall leave this until later and start with the discussion of Eqs. (II.6.08) and (II.6.09). Two cases can be distinguished:

The case of unsaturated donors:

$$n_\ominus \gg K_D \qquad n_{D^\times} \approx n_D \qquad n_{D^+} \approx n_D \cdot \frac{K_D}{n_\ominus} \ll n_D \tag{II.6.10}$$

---

[1] See Eq. (VIII.5.20).

[2] See Eqs. (VIII.4.04) and (VIII.4.17).

[3] At temperatures far below the melting point of the crystal, the atoms and ions change places very seldom. A certain degree of disorder established at higher temperatures is frozen in at the lower temperatures. For further details, see p. 53.

The case of saturated donors:

$$n_\ominus \ll K_D \qquad n_{D^\times} \approx n_D \cdot \frac{n_\ominus}{K_D} \ll n_D \qquad n_{D^+} \approx n_D \qquad (\text{II.6.11})$$

If there are many electrons ($n_\ominus \gg K_D$), many recombination partners $\ominus$ are available to the donors. Therefore most donors are in the associated state, which is the case of unsaturated donors ($n_{D^+} \ll n_D$). If there are only a few electrons ($n_\ominus \ll K_D$), however, most donors do not find recombination partners and are in the dissociated state $D^+ (n_{D^+} \approx n_D)$.

What are then the experimental means by which one can influence the electron concentration $n_\ominus$? If no further impurities or imperfections are present in the semiconductor besides the donors already considered, the electron concentration $n_\ominus$ appears to be fully determined by the neutrality requirement[1] for a crystal. For the number of positive donors D^+ must be equal to the number of negative electrons $\ominus$:

$$n_{D^+} = n_\ominus \qquad (\text{II.6.12})$$

from which follows with (II.6.09)

$$n_{D^+} = n_\ominus = K_D \left[-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{n_D}{K_D}} \right] \qquad (\text{II.6.13})$$

In spite of this it is possible, at least in principle, to establish either the unsaturated or saturated condition by varying the temperature. At low temperatures

$$T \ll \frac{1}{\mathrm{k}} E_{CD} \cdot \frac{1}{\ln \dfrac{N_C}{2n_D}} \qquad (\text{II.6.14})$$

we find $\quad \tfrac{1}{2} N_C \, \mathrm{e}^{-\frac{E_{CD}}{\mathrm{k}T}} = K_D \ll n_D$

and from (II.6.13) one obtains the limiting relation

$$n_{D^+} = n_\ominus \approx K_D \sqrt{\frac{n_D}{K_D}} \gg K_D$$

which corresponds, according to (II.6.10), to the unsaturated impurity case. This leads finally to

$$n_\ominus \approx \sqrt{n_D K_D} \approx \sqrt{n_D N_C}\, \mathrm{e}^{-\frac{\frac{1}{2} E_{CD}}{\mathrm{k}T}} \qquad (\text{II.6.15})$$

The temperature dependence of the electron concentration, and thereby essentially also that of the conductivity, is, in this case, deter-

---

[1] Apart from surface or boundary-layer effects.

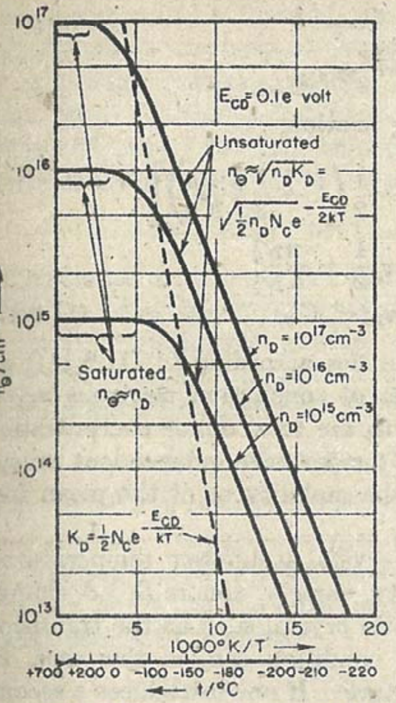


FIG. II.6.1. Concentration $n_\ominus$ of the conduction electrons as a function of the temperature at different donor concentrations n_D.

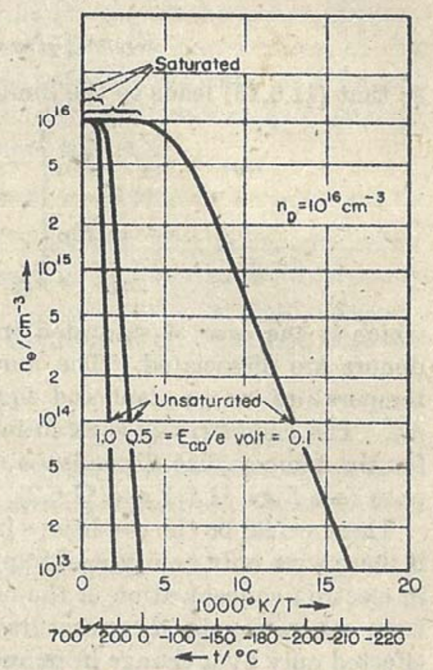


FIG. II.6.2. Concentration $n_\ominus$ of the conduction electrons as a function of the temperature at different donor ionization energies E_{CD}.

mined by *half* the activation energy E_{CD} of the donors (see Figs. II.6.1 and II.6.2). We find, however, that at high temperatures[1]

$$T \gg \frac{1}{\mathrm{k}} E_{CD} \cdot \frac{1}{\ln \dfrac{N_C}{2n_D}} \tag{II.6.16}$$

---

[1] One might easily be misled to the erroneous conclusion that dissociation of the great majority of donors is possible only if the mean energy content kT of a thermal collision is large compared to the dissociation energy E_{CD} of the donors. The condition for donor saturation would then be

$$T \gg \frac{1}{\mathrm{k}} E_{CD}$$

The comparison with (II.6.16) shows, however, that the condition is in reality attenuated by a factor $(1/\ln N_C/2n_D)$. The mistake in this reasoning lies in the fact that the number of the dissociating collisions is not compared with the number of the recombination processes, which would be correct, but rather with the number of *all* collisions, which is quite irrelevant for the final equilibrium condition.

the mass-action constant is

$$K_D = \tfrac{1}{2} N_C \, e^{-\frac{E_{CD}}{kT}} \gg n_D$$

so that (II.6.13) leads to the limiting relation

$$\begin{aligned} n_{D^+} = n_\ominus &= K_D\left[-\frac{1}{2} + \frac{1}{2}\left(1 + \frac{1}{2}4\frac{n_D}{K_D}\right)\right] \\ &= K_D\left[-\frac{1}{2} + \frac{1}{2} + \frac{n_D}{K_D}\right] \\ n_{D^+} = n_\ominus &= n_D (\ll K_D) \end{aligned} \tag{II.6.17}$$

which is the case of saturated impurities according to (II.6.11); all donors are dissociated. The number of conduction electrons $n_\ominus$ is temperature independent and equal to the total donor concentration n_D. The conductivity is essentially temperature independent except for the temperature dependence of the mobility or of the mean free path (see Figs. II.6.1 and II.6.2).

These would be the conditions in a crystal at different temperatures, if there were only one type of impurity, namely, donors D. A change in electron concentration in the neutral crystal such as the transition from saturation to the unsaturated condition can, in this case, be effected only by a change in temperature. If one introduces a second type of impurity into the crystal, however, the dissociation-association conditions of the original impurity type can be influenced by means of the concentration of the second impurity type. This second type of impurity can be represented by donors with a different activation energy or by acceptors which become negatively charged when capturing an electron. A detailed discussion of this problem would consume too much space. We see, a priori, that a second type of impurity introduces two new unknown concentrations into the problem, namely, those of the associated and of the dissociated impurities. The dissociation-association equilibrium of this second type of impurity is also governed by a law of mass action, and the sum of the associated and dissociated impurity concentrations is equal to the total number of the introduced impurities. Thus two new equations are added, and the problem is again fully determined.

The reaction equations

$$2\mathrm{Zn}_{\bigcirc}^{\cdot} + 2\ominus + \mathrm{O}_2^{(\mathrm{gas})} \rightleftarrows 2\mathrm{ZnO} \tag{II.5.01}$$

are also governed by a law of mass action in the case of sufficient "dilution" of the reaction partners. In this case, we can interpret the reaction process from left to right as *association* of two Zn ions in interstitial Zn positions $\mathrm{Zn}_{\bigcirc}^{\cdot}$, two quasi-free electrons $\ominus$, and an oxygen molecule O_2 (gas) of the gas atmosphere. The frequency of such an association is again equal to the product of the concentrations of each single reaction partner. If *two* reaction partners are of the same type, the concentration of this type occurs *twice* as a factor. The coefficients of the reaction equation appear, therefore, as exponents of the concentrations in the law of mass action

$$n^2_{\mathrm{Zn}\,\bigcirc^{\cdot}} \cdot n^2_{\ominus} \cdot n_{O_2}\,(\mathrm{gas}) = K_1\, n^2_{\mathrm{ZnO}}$$

The concentration of the lattice molecules ZnO is incomparably larger than the concentration of all reaction partners, so that it remains practically unchanged during any conceivable reaction process. Therefore, the ZnO concentration is included in the constant, and the law of mass action assumes the form

$$n^2_{\mathrm{Zn}\,\bigcirc^{\cdot}} \cdot n^2_{\ominus} = K_2 \cdot n^{-1}_{O_2}\,(\mathrm{gas}) = K_2 p^{-1}_{O_2}\,(\mathrm{gas}) \qquad \text{(II.6.18)}$$

where instead of the concentration n_{O_2} (gas) of the oxygen molecules in the gas phase the partial oxygen pressure is introduced because it can be obtained directly by experiment.

If the $\mathrm{Zn}\bigcirc^{\cdot}$ are the only imperfection in the ZnO, the quasi-neutrality condition requires

$$n_{\ominus} = n_{\mathrm{Zn}\,\bigcirc^{\cdot}}$$

and from (II.6.18) follows

$$n_{\ominus} = K_2^{1/4} \cdot p_{O_2}^{-1/4}\,(\mathrm{gas})$$

The experiment[1] gives a conductivity proportional to $p_{O_2}^{-\frac{1}{4.3}}{}^{(\mathrm{gas})}$.

Hence, the number of electrons and with it the conductivity decrease with increasing partial oxygen pressure in the adjacent gas phase: ZnO is a "reduction semiconductor."

If the corresponding reasoning is applied to the previously discussed equilibrium between a NiO crystal and an adjacent oxygen gas phase at elevated temperatures, we find[2]

$$n_{\oplus} = K_2^{1/4} \cdot p_{O_2}^{+1/4}\,(\mathrm{gas})$$

This law has also been reasonably well verified by experiment.[3] NiO behaves as an "oxidation semiconductor" whose conductivity increases with increasing partial oxygen pressure.

---

[1] H. H. Baumbach and C. Wagner, *Z. physik. Chem.*, **B22**: 199 (1933).

[2] Concerning the symbol $n_{\oplus}$ in place of p for hole concentration as used elsewhere in this book, see footnote 1, p. 46.

[3] H. H. Baumbach and C. Wagner, *Z. physik. Chem.*, **B24**: 59 (1934).

The two examples show that this difference is causally connected with the electron conduction in ZnO on one hand and with the hole conduction in NiO on the other hand. Thus we obtain the rule:

Electron conductors are reduction semiconductors.
Hole conductors are oxidation semiconductors.

§7. Fundamentals of Atomic Disorder

An unbiased reader who is confronted for the first time with the atomic-disorder phenomena described in this chapter is apt to assume an attitude which can be expressed as follows: "These disorder phenomena cause certainly very interesting effects. However, they are all 'dirt effects.' It would be important to seek a solution to the problem of making crystals with a perfect lattice structure which will then allow the study of ideal crystals."

It may be well to point out that such an escape from the confusion of the disorder phenomena into the simplicity of the perfect lattice is *fundamentally* impossible. A certain degree of atomic disorder is basically unavoidable except at absolute zero, and at finite temperatures the perfect lattice is an improbable and anomalous exception just as a gas in which all molecules have exactly the same velocity.

An exact proof of this fact requires the extensive use of general statistics. We must restrict ourselves to plausibility considerations, which we shall demonstrate with the example of NaCl and AgBr. Both substances shall be of absolute chemical purity. Using this assumption, which is in principle, though not in practice, feasible, we exclude substitutional imperfections (impurities), so that the only possible imperfections are lattice vacancies or interstitials.[1] The following reactions can take place between these imperfections:

$$\mathrm{Na}\ \bigcirc^{\cdot} + \mathrm{Na}\ \square' \rightleftarrows 0 \tag{II.7.01}$$

$$\mathrm{Cl}\ \bigcirc' + \mathrm{Cl}\ \square^{\cdot} \rightleftarrows 0 \tag{II.7.02}$$

$$\mathrm{Na}\ \bigcirc^{\cdot} + \mathrm{Cl}\ \bigcirc' \rightleftarrows \mathrm{NaCl} \tag{II.7.03}$$

Further reactions such as

$$\mathrm{NaCl} + \mathrm{Na}\ \square' + \mathrm{Cl}\ \square^{\cdot} \rightleftarrows 0$$

can be constructed from Eqs. (II.7.01) to (II.7.03) by subtracting the first two equations (II.7.01) and (II.7.02) from the third equation (II.7.03).

---

[1] It is extremely improbable to find in the NaCl lattice a Cl atom or ion in an Na place, or vice versa, so that we need not consider this type of disorder.

The laws of mass action corresponding to Eqs. (II.7.01) to (II.7.03)

$$n_{\mathrm{Na}\,\bigcirc^{\cdot}} \cdot n_{\mathrm{Na}\,\square'} = a \qquad \text{(II.7.04)}$$

$$n_{\mathrm{Cl}\,\bigcirc'} \cdot n_{\mathrm{Cl}\,\square^{\cdot}} = b \qquad \text{(II.7.05)}$$

$$n_{\mathrm{Na}\,\bigcirc^{\cdot}} \cdot n_{\mathrm{Cl}\,\bigcirc'} = c \qquad \text{(II.7.06)}^{1}$$

show that all three mass-action constants a, b, and c must be zero, if all imperfection concentrations are to be zero. This, however, is possible only at the temperature $T = 0$; for the constants in the laws of mass action are generally governed by relations of the following type:[2]

$$a = N^2\, \mathrm{e}^{-\frac{E_a}{kT}}$$

N = total number of $\mathrm{Na^+}$ and $\mathrm{Cl^-}$ lattice positions per unit volume[1] $= 2n_{\mathrm{NaCl}}$.

We summarize: The perfect lattice can exist only at the temperature $T = 0$, where alone all disorder concentrations can vanish. At any temperature $T > 0$, a certain atomic disorder is present at thermal *equilibrium* and the ideal lattice occupancy assumes the character of a special case, which can be realized only in one way and therefore is so extremely rare during the fluctuation processes around the equilibrium condition that it can be ignored for all practical purposes.

The above-mentioned fluctuation processes can take place only if lattice atoms change their position. Such position changes require that lattice atoms overcome energy thresholds using the energy derived from thermal collisions. At low temperatures, the number of thermal collisions of sufficient intensity is very rare. The processes of position change and the fluctuations therefore freeze in at low temperatures. As a result, the degree of disorder corresponding to the equilibrium condition is established only at higher temperatures and a state of excessive disorder is frozen in by cooling to lower temperatures; this is, of course, only a different way of saying that the time required to establish the equilibrium condition corresponding to the lower temperature would be unrealistically long.

We see now that the fundamentally unavoidable degree of disorder associated with the study of thermal equilibrium can in practice be realized only if considerable precautions, such as slow cooling, etc., are taken. While this degree of disorder can be more or less attained through the skill of the experimental scientist, it is fundamentally impossible to reduce it further or even to attain ideal lattice occupancy.

---

[1] The concentration n_{NaCl} of the lattice molecules is again included in the constant c of the law of mass action.

[2] See, for instance, W. Jost, *op. cit.*, particularly p. 61, Eqs. (33) to (35). Concerning the definition of the lattice concentrations, see Jost, pp. 53 and 58.

§8. Problems

1. The calculation in §1 of the ionization energy of a donor does not take into account the fact that the effective mass of an electron is generally different from its free mass. Explain why a smaller effective mass increases the first Bohr orbit, and by what factor. How does this alter the ionization energy? Give numerical values for germanium, assuming[1] $m_{eff} = \frac{1}{4}m_0$.

2. (Geometry of the diamond lattice.) In Fig. II.1.1*a* the geometry of the diamond lattice is shown. How many atoms does every cell contain, if one counts atoms that are shared by several cells only as fractional atoms?

Consider all the atoms shown in Fig. II.1.1*a* as perfect spheres, just touching each other. If the length of the cube shown in the drawing is $2d$, what is the radius of the atomic spheres? Give the numerical value for the germanium atom.

3. (Interstitial positions.) Show that the radius of the largest atom that would fit into an interstitial position in the diamond lattice is equal to the radius of the host atom itself. Which lattice atoms would just touch an interstitial atom of the maximal size if that interstitial atom were put into the center of the cube shown in Fig. II.1.1*a*? Show all the other possible interstitial positions in that cube, including the ones that are shared with adjoining cubes. What is their number and what is the ratio of lattice positions to interstitial positions? What kind of a lattice structure do the interstitial positions themselves have?

4. At what acceptor density does germanium have its highest resistivity, assuming that all acceptors are negatively ionized? At what acceptor density would, under this condition, the conductivity again pass through the intrinsic value? What are the donor or acceptor densities in germanium with a resistivity of 30 ohm-cm? (See Probs. 1 to 3 of Chap. I.)

---

[1] $m_{eff} = \frac{1}{4}m_0$ is an average value given by E. M. Conwell, *Proc. IRE*, **40**: 1330 (1952). For a detailed treatment taking into account the directional dependence of m_{eff} we refer to M. A. Lampert, *Phys. Rev.*, **97**: 352 (1955).

CHAPTER III

The Hole

§1. Introduction

The measurement of the Hall effect is an important source of information in the investigation of the conduction mechanism of a particular solid; for one can derive from it the mobility and the sign (negative or positive) of the charge carriers. The experiment consists of making a current of density **i** flow in the x direction through a block of the material to be investigated while a magnetic field **H** is applied in the z direction. Figure III.1.1 shows that at the beginning of the

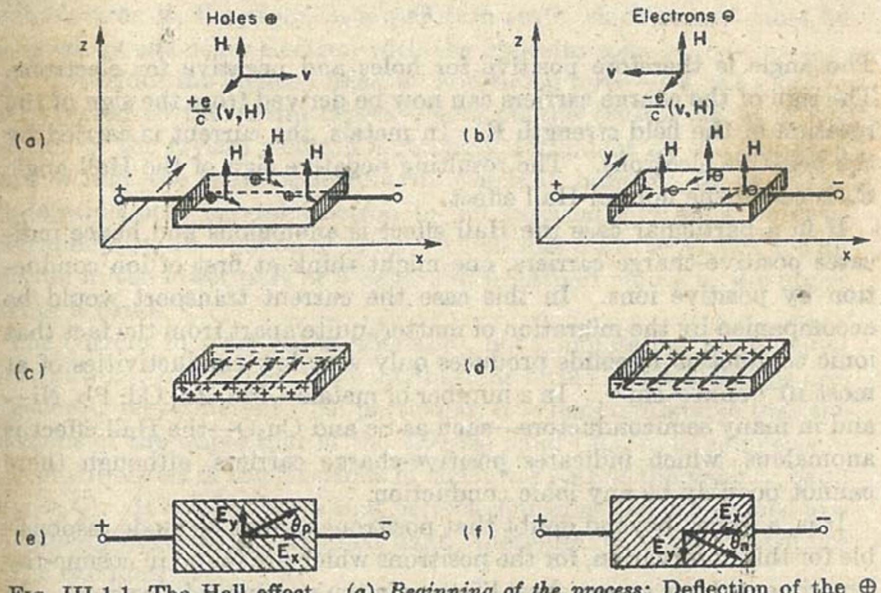


FIG. III.1.1. The Hall effect. (a) *Beginning of the process:* Deflection of the ⊕ in the negative y direction by the Lorentz force. (b) *Beginning of the process:* Deflection of the ⊖ in the negative y direction by the Lorentz force. (c) *Stationary condition:* The charging of the sample sides and the electrical cross field E_y. (d) *Stationary condition:* The charging of the sample sides and the electrical cross field E_y. (e) *Stationary condition:* The rotation of the field strength by the Hall angle. (f) *Stationary condition:* The rotation of the field strength by the Hall angle.

process the Lorentz force $\pm e/c[\mathbf{v}, \mathbf{H}] = \pm e/c[\mathbf{i}/\pm en, \mathbf{H}] = [\mathbf{i}/nc, \mathbf{H}]$ deflects the charge carriers in the negative y direction independently of the positive or negative charge of the particles. The front side of the sample is thus charged positively or negatively with respect to the back side, depending on the sign of the charge carriers, until the transverse electric field E_y just compensates the transverse force $[\pm(e/c)\mathbf{v}, \mathbf{H}]$ of the magnetic field. In the steady-state condition the field strength $\mathbf{E}$ is rotated out of the x axis by an angle which can be calculated with the help of Eq. (I.2.04) for holes

$$\tan \Theta_p = \frac{E_y}{E_x} = \frac{+\frac{1}{c} v \cdot H}{+\frac{1}{\mu_p} \cdot v} = \frac{1}{c} \mu_p \cdot H$$

or for electrons

$$\tan \Theta_n = \frac{E_y}{E_x} = \frac{-\frac{1}{c} v \cdot H}{+\frac{1}{\mu_n} v} = -\frac{1}{c} \mu_n \cdot H$$

The angle is therefore positive for holes and negative for electrons. The sign of the charge carriers can now be derived from the sign of the rotation of the field strength $\mathbf{E}$. In metals, the current is carried by the negative electrons. The resulting negative sign of the Hall angle Θ_n is called the normal Hall effect.

If in a particular case the Hall effect is anomalous and hence indicates positive-charge carriers, one might think at first of ion conduction by positive ions. In this case the current transport would be accompanied by the migration of matter, quite apart from the fact that ionic conduction in solids produces only very low conductivities of at most 10^{-7} ohm^{-1} cm^{-1}. In a number of metals—Mo, Zn, Cd, Pb, Ni—and in many semiconductors—such as Se and Cu_2O—the Hall effect is anomalous, which indicates positive-charge carriers, although there cannot possibly be any ionic conduction.

It is, a priori, beyond doubt that positrons cannot be made responsible for this observation, for the positrons which are found in cosmic-ray experiments have a very short lifetime in the presence of electrons. In the presence of the high electron concentration in a solid, a large positron concentration could be maintained only if positrons were continually created at a high rate. However, the creation of an electron-positron pair requires an energy of at least 10^6 ev; for, according to the Einstein relation, even the rest energy of two particles with the electron mass $m = 9 \cdot 10^{-28}$ g is $2mc^2 = 1.6 \cdot 10^{-6}$ erg $= 1 \cdot 10^6$ ev. On the other hand, the energy available in a solid is only the thermal energy $kT \approx 2.5 \cdot 10^{-2}$ ev (at room temperature), which is not nearly sufficient for the creation of an electron-positron pair.

Hence there is no doubt that the current transport in Se, Cu_2O, etc., is only *apparently* carried by relatively few positively charged electrons and that this can only be the relatively *simple* result of a *complicated* superposition of the contributions of many ordinary negative electrons.

In order to gain insight into this interaction of many negative electrons, we recall a few statements[1] concerning the behavior of a crystal electron in the periodic potential field of the lattice:

1. The energy spectrum of a crystal exhibits the well-known band structure (Figs. I.2.3, I.2.6, I.3.6, and I.3.8).[2] In a semiconducting crystal, the so-called valence band is almost entirely filled and the so-called conduction band is almost entirely empty and contains only the "free" conduction electrons[3] near its lower edge.

2. The Pauli principle must be applied to all electrons of a crystal. Each electron is, therefore, in a quantum state which can at most be occupied by one other electron with the opposite spin.

3. If, besides the lattice forces, an additional force—e.g., an electric field applied to the entire crystal—acts on the crystal electron, the reaction of the crystal electron is strongly dependent on the quantum state within the band spectrum in which it happens to be. The acceleration of a crystal electron by an external additional force is determined by an "effective mass" which has positive values for an electron in the lower part of a band and negative[4] values in the upper part of a band.

The current in a crystal is, therefore, not composed of many equal contributions, but rather, the contributions of the single electrons are, in general, very different. In particular, it is understandable that the electrons of the upper half of the valence band ($m_{eff} < 0$) compensate the contributions of the electrons of the lower half of the valence band ($m_{eff} > 0$). It can be shown that perfect compensation is obtained in a completely filled band. A filled band, therefore, does not contribute to the conductivity.

In a band where a few electrons are missing from full occupancy the compensation is not complete, and such an *almost* filled band does, in

---

[1] See also Chap. I.

[2] See also Chap. VII, §2 and §4.

[3] See also Chap. VIII, §4 and §5.

[4] See also Chap. VII, §6.

fact, furnish a contribution to the current. Heisenberg[1] has demonstrated that this current contribution is the same as if the few empty states of the almost filled band were occupied by electrons with positive charge. The mass of these fictitious particles must be chosen, according to their contribution, equal to the effective mass of an ordinary electron in the corresponding empty crystal state but with opposite sign. As the holes are concentrated at the upper edge of the valence band (see page 22) where the effective electron masses are negative, the effective mass of the holes must be chosen positive, according to the stated rule.

It has been found that the equivalence between the few holes and the incomplete totality of the numerous valence electrons can be demonstrated without calculation. We shall follow this course below. Since the concept of the effective mass and particularly its negative values in the upper half of an energy band are of decisive importance, the behavior of an electron in a crystal under the influence of an external force will be considered in §2 as a starting point. In §3 we shall prove the postulated equivalence between an incomplete band, which is occupied by $N - M$ electrons, and M holes. Then §4 will show clearly that this entire concept of holes and hole conduction is only a symbolic description of a really very complicated situation. We shall show that the validity of this equivalence ceases as soon as interaction between electrons is considered.

§2. The Negative Values of the Effective Mass in the Upper Part of an Energy Band

If an electron moves in a force field, its energy usually changes. If, for instance, the force field **F** and the electron velocity **v** have the same direction, the force does work on the electron and increases its energy. If the force **F** and the electron velocity have opposite directions, the electron is slowed down and its kinetic energy is reduced. An electron in a crystal, in contrast to a free electron, is almost exclusively subject to the forces of the lattice potential, the effect of which is much larger than that of any conceivable externally applied force. Even field strengths of the order of the breakdown field, namely, about 10^5 volts cm^{-1}, do work equal to only about $3 \cdot 10^{-3}$ ev in the displacement of an electron over the distance of one lattice constant, $3 \cdot 10^{-8}$ cm, whereas the lattice potential itself changes by several volts within a lattice constant. This allows us to utilize the force-free

---

[1] W. Heisenberg, *Ann. Physik*, **10:** 388 (1931).