IMPURITIES IN SEMICONDUCTORS: SOLUBILITY, MIGRATION, AND INTERACTIONS

IMPURITIES IN SEMICONDUCTORS: SOLUBILITY, MIGRATION, AND INTERACTIONS

Victor I. Fistul

CRC PRESS

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Fistul, V.I. (Viktor Ilich), 1927–
Impurities in semiconductors : solubility, migration, and interactions / V.I. Fistul ; translated by L.N. Smirnova.
p. cm.
Includes bibliographical references and index.
ISBN 0-415-30831-3 (alk. paper)
1. Semiconductor doping. 2. Semiconductors — Impurity distribution. 3. Semiconductors — Diffusion. 4. Solubility. I. Title.

QC611.6.D6F57 2004
537.6′22 — dc22 2003067460

International Standard Book Number 0-4153-0831-3
Library of Congress Card Number 2003067460
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Preface

Every century bears its own name in History, characterizing its principal achievement. We do not know yet what name our descendants will give to the 20th century. It may be called the nuclear age, owing to the discovery and application of nuclear power. Or, it may be called the age of space flights, because man has overcome the Earth's gravity to go into space and has even visited the Moon. But there is another possible name which might as well be given to the 20th century—the age of electronics! The time that has elapsed since the invention of the radio is a period of major achievements in solid state electronics which has filled up every "pore" of our life. No progress would ever have been possible in nuclear technology, space flights or other technologies without electronics. Electronics, in turn, directly depends on high quality semiconductor materials. Nature has supplied us with very few semiconductor substances, such as germanium, silicon, selenium, and tellurium, whereas other commonly used materials of the III–V and II–VI groups and more complex compounds (III_2–VI_3 and III_2–IV–VI) are man-made.

All semiconductor materials, both natural and synthesized, require much effort for their production in a perfect (defect-free) crystalline state with a negligible background of foreign impurities. Such impurities contaminate the crystal in an uncontrollable or poorly controllable way. On the other hand, all remarkable properties of semiconductors, that paved the way for modern solid state electronics, are due not only to their purification but also to a well-controlled doping dosage.

Today we have at our disposal a nearly complete list of impurities suitable for doping basic semiconductors. This circumstance permits systematization of properties of various semiconductor–impurity systems in one book. However, I did not intend to write a reference book, because fairly complete and good reference books have already been published. A good illustration is the world-known work *Numerical Data and Functional Relationship in Science and Technology. New Series* Ed. K. Hellwey and O. Madelung. Berlin: Springer–Verlag, 1984. V. 17, pp. 652. No doubt it is important to know the properties of a semiconductor doped with an impurity in a definite concentration. For practical applications, it is more important,

however, to know the techniques, procedures, and external effects which can help us to control the distributions of impurity atoms over various crystal positions. This control is the key to a wide controllable application of impurities, whose potentialities are still far from being exhausted.

It was my primary aim to draw the attention of researchers and engineers to impurities which have not yet found a wide application in semiconductor technology: *d*-, *f*-, isovalent, and other types of impurities. I had another aim, too. It seems important to me that the western reader should be introduced to the research done in this area of physics by scientists in the former Soviet Union. I hope this gap will be filled after the publication of this book.

The author's philosophy will inevitably show itself in the material presentation, and the reader may not agree with some of my judgments. Besides, because of the extensive character of the problem, some of its aspects have been left aside, for example, the interaction of impurities with dislocations and stacking faults, or the state and behavior of adsorbed impurities. I hope, however, that this book will appear useful even in its present format. I would be very happy if it could eventually find its place among the books by such outstanding researchers as F.A. Kroger of the Philips Laboratory in Endhoven, R.A. Swalin of the University of Minnesota, A.G. Milnes of the Carnegie–Mellon University, or V.M. Glazov of the Moscow Institute for Electronics Technology.

I would like to express my sincere gratitude to the many people who have stimulated the evolution of my thinking as a scientist and, sometimes, as a human being. Among them are the much lamented professors R.N. Rubinshtein, D.N. Nasledov, and Yu.V. Shmartzev. My gratitude also goes to my numerous colleagues who are presently working actively in physics—professors B.V. Tzarenkov, F.A. Gimelfarb, V.M. Koshkin, D.G. Andrianov, M.G. Milvidsky, N.S. Rytova, S.V. Bulyarsky, P.M. Grinshtein, B.L. Oksengendler, and K.A. Kikoin.

I want to thank the translator of this book L.N. Smirnova, Ph.D., who has brilliantly overcome numerous translation difficulties, and M.A. Smirnova for the camera-ready preparation of the book.

Finally, I am very grateful to the Publisher for their effort in publishing this book.

CONTENTS

Chapter 1

The Semiconductor–Impurity System

1.1 THE SEMICONDUCTOR CRYSTAL AS A THERMODYNAMIC SYSTEM

A semiconductor crystal doped with impurities is usually regarded as a solid solution, in which the semiconductor is the solvent and the ensemble of impurity atoms is the solute.

Sites and interstices in a crystal lattice serve as positions for various structural units—atoms and vacancies. A *chemical potential* can be ascribed only to these units.

A *perfect crystal* consists only of intrinsic (host) atoms and stoichiometric vacancies occupying intrinsic sites in the crystal lattice. Any deviation from crystal perfection is known as a *defect*, and the process that has brought it into life is termed defect formation. In the generally accepted classification [1], impurities and vacancies are referred to as point defects. In this book, the word "vacancy" will be used only for nonstoichiometric vacancies formed after a host atom has left its site.

Normally, point defects are considered to be distributed between two phases—the crystal and its ambient. The former is taken to be an entity, without subdividing it into the variety of positions provided for impurity atoms. This is because a common impurity can usually occupy only one kind of position in a crystal lattice in a wide temperature range, irrespective of other point defects, intrinsic or impurity-type. One exception is amphoteric impurities, which can simultaneously occupy different positions in a lattice

in commensurable concentrations. Such a crystal is then to be regarded as a two-phase system—a site and an interstice, or a cation site and an anion site. Therefore, a thermodynamic analysis should involve, at least, a three-phase system.

It might seem strange, at first glance, to consider the subsystems of sites and interstices as individual phases. But a phase is a homogeneous part of a system having a common boundary with the other parts (phases) and exchanging particles with them. An exchange of particles between sites and interstices gives rise to vacancies in the site subsystem and to intrinsic and impurity interstitials in the interstitial subsystem. Each subsystem may be regarded as an infinite cluster with a possible particle movement inside it [2]. Both subsystems can then be assumed to have a common boundary, which is fractal rather than plane. In this model, the concept of phase is quite applicable to the site and interstitial subsystems of a crystal, as well as to the cation and anion sites in compounds of the $A^{III}B^{V}$ or $A^{II}B^{VI}$ type.

Therefore, a semiconductor crystal is essentially a multiphase thermodynamic system with an actual exchange of atoms, and this should not be ignored in analyzing doping processes. Thermodynamically, a crystal is not a strictly isolated system. Its individual parts interact with each other and with the ambient. To begin with, it is necessary to distinguish between external and internal interactions. External interactions provide the openness of a thermodynamic system and can be classified into the following types [3]:

– *mechanical interactions* occurring under the action of external pressure or force fields creating elastic strain in a crystal;

– *thermal interactions* resulting from energy exchange under the action of a temperature gradient;

– *exchange of atoms* at the boundaries of phases and surfaces, between parts of a crystal and other objects, often described by one word—"sink."

Internal interactions, which change the free energy of a closed thermodynamic system, will be discussed in Chapter 5.

Any interaction results in defect formation. This process, therefore, represents an exchange of structural elements between individual crystal phases and subsystems. These processes disclose the inner structure of the thermodynamic system. The arrival of a host atom at the surface (Schottky disordering) or at an interstice (Frenkel disordering) give rise to the same type of defect—vacancy. However, the crystal free energy changes differently because of different final positions of the atom. For this reason, various quasichemical reactions taking place in a crystal can be considered as processes showing the system "from within." These induce changes in the host lattice and impurities; for example, an atom may change its position in the lattice, moving from a site to an interstice, or producing a stable quasi-

molecule (associate) with another point defect, etc. In any case, one can say that one object has disappeared and another has appeared, because the change of position, association, and similar transformations—all lead to considerable changes in the object's properties.

This kind of system is usually assumed in thermodynamic analyses to be quasi-closed. But even a quasi-closed system is subject to the action of various nonuniform force fields. There are energy and heat flows and numerous chemical reactions occurring in it; besides, associates and precipitates are formed and disintegrate. In other words, a crystal lives its own complicated life due to external and internal interactions. Consequently, one should choose with caution a physical model to describe this or that process and must look critically at the assumptions on which one bases the model.

A crystal can be described in terms of thermodynamic statistics, assuming that the system is in equilibrium if the temperature and pressure have equalized and all kinetic processes have become steady-state. Under these conditions, variations setting the system off balance obey the inequality following from the Klausius inequality [4]:

$$\mathrm{d}U + P\mathrm{d}V - T\mathrm{d}S > 0, \tag{1.1.1}$$

where U is internal crystal energy, P is external pressure, and S is entropy.

There are no variations in internal energy or volume in an isolated system

$$\mathrm{d}V = 0 \quad \text{and} \quad \mathrm{d}U = 0. \tag{1.1.2}$$

It follows from (1.1.1) that

$$\left(\mathrm{d}S\right)_{U,S} < 0, \tag{1.1.3}$$

which is equivalent to the following statement:
In equilibrium, entropy is maximum with respect to all variations, provided that the volume and internal energy remain constant.
The conditions of constant entropy and volume ($\mathrm{d}S = 0$, $\mathrm{d}V = 0$) give

$$\left(\mathrm{d}U\right)_{S,V} > 0, \tag{1.1.4}$$

i.e., in equilibrium, the internal energy of an isolated system takes its maximum value.

Under the conditions of constant pressure and variations setting the system off equilibrium, it is not the internal energy that must remain minimal

but the sum of this energy and potential energy characterizing the relation to external bodies, or enthalpy:

$$H = U + PV. \tag{1.1.5}$$

From (1.1.5), we have

$$\mathrm{d}U = \mathrm{d}H - \mathrm{d}(PV) = \mathrm{d}H - P\mathrm{d}V - V\mathrm{d}P - \mathrm{d}P\mathrm{d}V. \tag{1.1.6}$$

Substituting (1.1.6) into the Klausius inequality (1.1.1), we obtain

$$T\mathrm{d}S - \mathrm{d}H + V\mathrm{d}P + \mathrm{d}P\mathrm{d}V < 0. \tag{1.1.7}$$

A procedure similar to this one yields the equilibrium conditions at constant pressure [3]:

$$(\mathrm{d}S)_{H,P} < 0 \qquad (\mathrm{d}H)_{S,P} > 0. \tag{1.1.8}$$

It is easy to show with (1.1.7) that at constant temperature and pressure, Gibbs free energy must be minimal:

$$G = H - TS \tag{1.1.9}$$

$$(\mathrm{d}G)_{T,P} > 0. \tag{1.1.10}$$

The system free energy rises. Therefore, *Gibbs free energy is a minimum in equilibrium*, and any variation leads to its increase.

This provides us with a method of finding the concentration of point defects—*by minimizing crystal free energy*. The respective procedures will be discussed in Section 1.2.

In addition to the thermodynamic description, a doped crystal can be described in terms of a kinetic model suggesting the study of the system transition from one state to another. Kinetic methods provide more information on the system properties, since one can find kinetic coefficients characterizing the probabilities of processes occurring in the system.

Of importance in both approaches is the possible establishment of thermodynamic equilibrium under the conditions of actual experimental observations of these processes. The thermodynamic method is inapplicable without equilibrium conditions. Reliable information can be obtained from an analysis of kinetic processes only if the initial and final states of the system

are known exactly; but this, again, is possible only in equilibrium. This obvious fact has been emphasized in [3] because the duration of various processes in a semiconductor crystal varies greatly with temperature. This should be taken into account when choosing a model for the description of such processes.

Since the rates of diffusion processes, which determine the moments of time the equilibrium is established, increase exponentially with temperature, equilibrium conditions are mostly applicable at high temperatures, close to those of crystal growth. However, it is always necessary to define the temperature range of equilibrium inside and outside the crystal.

Crystal doping with impurities is often carried out under conditions not differing much from equilibrium conditions, at least, in thin crystal layers adjacent to the ambient. The ambient will be assumed in our theoretical analysis to be a gas or liquid phase, as is usually the case. Both are important because they are widely used in semiconductor doping technologies.

The ambient equilibrium can be maintained if, due to diffusion or via forced mixing, an impurity arrives at the interface, at least, at the same velocity at which it enters the crystal. Forced mixing is used when a semiconductor is doped during its pulling from a melt. The amount of impurity uptake ΔC is small as compared with its total concentration C in the melt at the growing crystal boundary. Therefore, the melt at the boundary can be considered to be in equilibrium conditions. Naturally, this is valid for impurities with the distribution coefficient $K << 1$.

The ambient can also be suggested to be in equilibrium when a crystal is doped in the gas (vapor) phase, because diffusion coefficients of impurity atoms in the intrinsic or any other gas phase are always large, and at the crystal interface we have $\Delta C << C$. Deviations from the ambient equilibrium are especially serious when one uses liquid phase epitaxy.

Denote the growth rate of an epitaxial layer as v and its thickness as L. Then the value of $\Delta C/C$ is defined as [5]

$$\frac{\Delta C}{C} = \left(1 + \frac{vL}{2D}\right)\operatorname{erf}\left[-\frac{1}{2}\left(\frac{vL}{D}\right)^{1/2}\right] + \left(\frac{vL}{\pi D}\right)^{1/2}\exp\left(-\frac{vL}{4D}\right), \qquad (1.1.11)$$

where D is the diffusion coefficient of the impurity in the ambient liquid phase at growth temperature.

Calculations with (1.1.11) show that the necessary rates are $v < 10^{-5}$ cm/s for the typical values of $L = 10^{-3}$–10^{-2} cm and $D = 10^{-6}$–10^{-4} cm^2/s with $\Delta C/C < 0.01$. Since epitaxial growth rates for semiconductor layers are

usually $v = 10^{-5}$–10^{-7} cm/s, liquid phase epitaxy during doping can also be assumed to be in equilibrium.

It is necessary to bear in mind, however, that there are other doping methods used in practice, which are characterized by a complete absence of equilibrium at the interface, such as ion implantation, molecular beam epitaxy, ion–molecular epitaxy, plasma sputtering, diffusion, etc. Here, the impurity concentration is determined by the process parameters and the energy of doping atoms or ions.

The condition for the ambient–crystal equilibrium is characterized by the equality of chemical potentials

$$\mu_l = \mu_s, \tag{1.1.12}$$

where the indices l and s refer to the liquid (ambient) and solid (crystal) phases, respectively.

If we assume that the impurity forms ideal solutions in both phases, the chemical potential for the liquid and solid phases will be described as [6]

$$\mu_l = \mu_{l0} + kT \ln C_l \tag{1.1.13}$$

$$\mu_s = \mu_{s0} + kT \ln C_s + \mu_l. \tag{1.1.14}$$

The quantities C_l and C_s are impurity concentrations in the respective phases:

$$C_{si} = \frac{N_i}{N},$$

where N_i is the number of atoms of the i-th kind, N is the number of sites they can occupy, while μ_{l0} and μ_{s0}, independent of C_l and C_s, are chemical potentials of pure components which consist of one sort of atoms occupying all appropriate sites, or Gibbs energy g_{i0} required for introducing a single defect into a pure component.

Expression (1.1.14) includes chemical potential μ_e of electrons, which reflects the fact that some of the impurity atoms introduced into a crystal may be ionized.

It follows from (1.1.12) and (1.1.14) that in equilibrium

$$C_s = C_l \exp\frac{\mu_{l0} - \mu_{s0}}{kT} \exp\frac{\mu_e}{kT}. \tag{1.1.15}$$

The quantity μ_e can be identified with the Fermi level E_F [7]. It will be counted off from the conduction band bottom E_c, taking $E_c = 0$. Then, for the donor and acceptor states of the impurity, we will have, respectively,

$$\mu_e = E_c - E_F = -E_F,$$

$$\mu_e = E_F - E_v = E_F + E_c - E_v - E_c = E_g + E_F. \qquad (1.1.16)$$

One should keep in mind that $E_F < 0$ and $E_g > 0$. The Fermi level position in a crystal totally determines the concentration of free electrons and holes, respectively:

$$n = N_c \exp(E_F/kT), \qquad p = N_v \exp[-(E_g + E_F)/kT]. \qquad (1.1.17)$$

The opposite is also correct: a change in the free carrier concentration will lead to a change of the Fermi level position. From (1.1.15), this must lead to a change in the impurity solubility. Considering that $C_s = N_n = n = p$ at high temperatures, we will have for donors

$$C_s = C_l^{1/2} N_c^{1/2} \exp[(\mu_{l0} - \mu_{s0})/2kT] \qquad (1.1.18)$$

and for acceptors

$$C_s = C_l^{1/2} N_v^{1/2} \exp[(\mu_{l0} - \mu_{s0})/2kT] \exp(E_g/2kT). \qquad (1.1.19)$$

Therefore, the relation between C_s and C_l is nonlinear because the chemical potential of the crystal electronic subsystem is not constant.

The value of chemical potential can be varied, irrespective of the doping impurity ionization. This can, in particular, be done by a simultaneous doping with two electrically active impurities or by generating intrinsic charged point defects, etc. Hence, one can draw an important conclusion concerning *the possibility of controlling the C_s concentration* by varying factors capable of changing the chemical potential of the free carrier subsystem in a crystal.

On the other hand, it is clear from (1.1.15) that a linear experimental relation is to be observed between C_s and C_l at μ_e = const. This is possible in three situations [5]:

(*a*) $C_s < n_i$, where n_i is an intrinsic concentration of charge carriers, determined by interband electron transitions, rather than by impurity ionization;

(*b*) $C_S < N_{d,a}$ = const, where $N_{d,a}$ is the concentration of other electrically active donors and acceptors which determine the value of E_F, i.e., μ_e;

(*c*) $E_F = \mu_e$ = const at $x = 0$, where x is the crystal–ambient interface *s–l*.

The case (*a*) is typical of a slightly doped or compensated semiconductor.

The case (*b*) is rather trivial and, indeed, occurs quite often, especially when $N_{d,a}$ is determined by shallow hydrogen-like impurities and C_S is a deep impurity.

The case (*c*) is often misinterpreted as relating to an equilibrium between a liquid ambient and a sample surface, rather than to a liquid ambient containing a sample bulk. Nevertheless, it does reflect the equality of chemical potentials in the contacting phases, creating no energy barrier to electrons at the interface. Let us discuss this case in more detail.

As an illustration, consider the contact between a crystal and a liquid phase, assuming it to be purely metallic. There is no doubt that this is valid for liquid phase epitaxy, since epitaxial films are grown from a metal melt containing the necessary non-metal components as dissolved admixtures. For example, $A^{III}B^{V}$ films (GaAs, GaP, GaSb) are usually grown from a Ga melt while InP and InSb are from an In melt.

In the Chokhralski or zone melting techniques, the liquid phase is represented by a semiconductor substance with dopants. A complete metallization of chemical bonding occurs during the melting of most semiconductor materials [8], and so the liquid phase can be treated as being metallic.

Therefore, the *s–l* interface can be considered as the interface between a semiconductor and a metal, or as a Schottky barrier. This approach was first suggested in [5] and later developed by the authors of [9–11].

It is quite clear that if impurity atoms diffuse quickly through the *s*-phase, i.e., if they swiftly pass through the region of space charge W in the Schottky barrier, an equilibrium will be established between the liquid phase and the whole semiconductor bulk. This ordinary bulk equilibrium is established under the condition

$$v = D_S / W, \tag{1.1.20}$$

where v is the growth rate of the solid phase and D_s is the impurity diffusion coefficient in the solid phase at the growth temperature.

Here again, μ_e is independent of C_s, and the function $C_s(C_l^{1/2})$ remains valid. But if the doping impurity diffuses slowly, or the solid phase grows quickly, so that

$$v >> D_s / W, \tag{1.1.21}$$

the impurity appears to be frozen at the interface, and the equilibrium is established only between the liquid phase and the crystal surface. The impurity concentration in a crystal grown under "surface" equilibrium conditions will exceed that in an ordinary bulk equilibrium.

An energy barrier φ_k arises at a metal–semiconductor interface under equilibrium conditions because of equal electron chemical potentials in both phases [12]. So we have at the interface

$$\mu_e = E_F = E_g - e\varphi_k, \tag{1.1.22}$$

and, since E_g and φ_k are independent of the doping impurity concentration C_s, we obtain with (1.1.15)

$$C_s = C_l \exp\left[\frac{\mu_{l0} - \mu_{s0} + E_g}{kT}\right] \exp\left(\frac{e\varphi_k}{kT}\right), \tag{1.1.23}$$

which is a linear relation between C_s and C_l, rather than the square root of (1.1.18) and (1.1.19). This is the criterion for an experimental differentiation between bulk and surface equilibrium conditions.

It is easy to show that this conclusion holds for imperfect impurity solutions in both contacting phases, whereas the temperature dependence $C_s(T)$ proves to be more complicated.

1.2 THERMODYNAMIC DESCRIPTIONS OF IMPURITY SOLUBILITY

There are three approaches to the calculation of point defect concentrations and the parameters that determine them. One is to solve a quantum mechanics problem for the lattice containing a point defect. The problem solution is aimed at finding the energy parameters which characterize the formation of this defect. This method has not yet provided satisfactory results. Only approximate solutions have been suggested based on semi-classical and classical conceptions. However, this approach permits evaluation of some thermodynamic parameters (dissolution enthalpy and entropy of point defects) "from the first principles", so it deserves a special consideration (see Chapter 4).

The other two approaches are based on phenomenological parameters derived from a comparison of theoretical formulas and experimental data. One of them was suggested by Brouwer [13] and described in detail by Kroger [14]. It uses the mass action law and considers defect interactions as quasi-chemical reactions. To understand the principle of this approach, let us represent the interaction between, say, an atom A and an atom B as a reaction producing the product AB:

$$A + B \rightleftarrows AB. \tag{1.2.1}$$

The equilibrium constant K for this reaction is defined from the mass action law as

$$K = [AB] / [A]\,[B], \tag{1.2.2}$$

where the symbol [ ] stands for concentration.

The value of K is determined by enthalpy H of the reaction (1.2.1) under standard conditions:

$$K = K^0 \exp(H^0 / RT), \tag{1.2.3}$$

where K^0 is an entropy factor.

On the other hand, when describing the equilibrium of reaction (1.2.1), we should bear in mind the equality of chemical potentials, following from the additivity principle:

$$\mu_A + \mu_B = \mu_{AB}. \tag{1.2.4}$$

Suppose that μ can be found from (1.1.13); then from (1.2.2), we have

$$\frac{[AB]}{[A][B]} = N_L \exp\left(-\frac{g^0_{AB} - g^0_A - g^0_B}{RT}\right), \tag{1.2.5}$$

which coincides with (1.2.3), if we put

$$g^0_{AB} - g^0_A - g^0_B = H^0. \tag{1.2.6}$$

The other approach to a thermodynamic description of impurity solubility is based on the notion that condition (1.1.12) corresponds to the free energy minimum of a crystal–ambient system. In the description of the Φ function,

which allows this energy minimum to be found, we use the same denotations for Gibbs partial energy as in (1.1.13) and (1.1.4) for a single doping defect (the denotations were first introduced in [15] and employed in our work [3]):

$$\mu_{i0} \equiv g_{i0} = g_{\alpha}^{\beta} - H_{\alpha}^{\beta} - TS_{\alpha T}^{\beta}. \qquad (1.2.7)$$

Here, T is temperature expressed in energy units, the superscript β indicates the defect position in a crystal lattice (site or interstice), and the subscript α stands for the defect type (impurity atom, vacancy, antistructural defect, etc.).

The quantities H_{α}^{β} and $S_{\alpha T}^{\beta}$ in (1.2.7) represent, respectively, the enthalpy of the defect formation and the heat component of entropy, taking into account the change in the crystal phonon spectrum due to the formation of a single defect.

Following the work [15], denote the energy of an electron transfer to the defect (or its escape from it) as $\varepsilon_{\alpha}^{\beta}$. Then the energy due to the defect ionization will be equal to the product of $\varepsilon_{\alpha}^{\beta}$ and the concentration of electron-free acceptors ($N_{\alpha}^{\beta} - n_{\alpha}^{\beta}$), or to the product of $\varepsilon_{\alpha}^{\beta}$ and the number of occupied donors n_{α}^{β}. This is the case for *n*-type crystals, while for *p*-type crystals, the relation will be inverse. The relation inversion in the expressions given below is taken into account by the factor d_{α}^{β} taken to be equal to 1 for donors and to 0 for acceptors.

The Φ function, which allows finding an arbitrary minimum, is [3]:

$$\Phi = G + \sum_{\alpha} \lambda_{\alpha}\varphi_{\alpha} + \sum_{\beta} \lambda^{\beta}\varphi^{\beta} + \lambda_{e}\varphi_{e}, \qquad (1.2.8)$$

where G is Gibbs free energy for a crystal with a defect; λ_{α}, λ^{β} and λ_{e} are Lagrange indeterminate factors minimizing the Φ function.

The expressions for φ are

$$\varphi_{\alpha} = N_{\alpha} - \sum_{\beta} N_{\alpha}^{\beta} = 0, \qquad (1.2.9)$$

$$\varphi^{\beta} = N^{\beta} - \sum_{\alpha} N_{\alpha}^{\beta} = 0, \qquad (1.2.10)$$

$$\varphi_{e} = n - p + \left(\sum_{\alpha,\beta} n_{\alpha}^{\beta} - \sum_{\alpha,\beta} N_{\alpha}^{\beta} d_{\alpha}^{\beta} \right) \left(1 - \delta_{\alpha}^{\beta}\right), \qquad (1.2.11)$$

where δ_α^β is Cronecker's symbol. It was shown in [3] that $\lambda_e = \mu_e$ and $\lambda_\alpha = \mu_\alpha$, whereas λ^β could be excluded from consideration by selecting the energy reference.

Thus, the final expression for the defect concentration is

$$N_\alpha^\beta = N_\beta^\beta A_\alpha^\beta \left(1 + B_\alpha^\beta\right), \tag{1.2.12}$$

where A_α^β and B_α^β represent the reduced functions

$$A_\alpha^\beta = R_\alpha^\beta \exp\left\{\frac{1}{kT}\left[\mu_\alpha - \mu_\beta + g_\beta^\beta - g_\alpha^\beta + \left(\mu_e - \varepsilon_\alpha^\beta\right) d_\alpha^\beta \left(1 - \delta_\alpha^\beta\right)\right]\right\}, \tag{1.2.13}$$

$$B_\alpha^\beta = \frac{r_\alpha^\beta}{R_\alpha^\beta} \exp\left\{\frac{1}{kT}\left[-\mu_e + \varepsilon_\alpha^\beta + \left(1 - d_\alpha^\beta\right)\left(E_g - 2\varepsilon_\alpha^\beta\right)\right]\left(1 - \delta_\alpha^\beta\right)\right\}, \tag{1.2.14}$$

where E_g is the forbidden gap width of a semiconductor and the quantities r_α^β and R_α^β are degeneration factors for the filled and unfilled electronic states of the defect, respectively.

We should add to (1.2.12) an expression for electron concentration

$$n_\alpha^\beta = N_\beta^\beta A_\alpha^\beta B_\alpha^\beta . \tag{1.2.15}$$

The greatest difficulty of this thermodynamic treatment is associated with imperfection of the impurity solution and solutions of other point defects in the solid phase. The liquid (ambient) phase is also imperfect. Even the gas phase often contains multiatomic groups, with the atoms interacting with one another, so that the phase is far from being perfect.

For these cases, the expressions for partial chemical potentials are not as simple as (1.1.14), and the additivity of chemical potentials proves invalid for the crystal as a whole:

$$\mu_s = \sum_i \mu_{is} . \tag{1.2.16}$$

There have been attempts, made in terms of general thermodynamics, to bypass this difficulty by replacing concentrations or mole fractions x by activities a:

$$a = fx, \tag{1.2.17}$$

where f is an activity coefficient which is, in turn, a function of composition. So, instead of (1.1.13), we have for the liquid phase

$$\mu_{\alpha l} = \mu^0_{\alpha l} + RT \ln a_l, \tag{1.2.18}$$

where $\mu^0_{\alpha l}$ is the chemical potential in a pure component melt. In other words, the expression for a partial chemical potential also has a simple form, while the activity coefficient now has a complicated concentration dependence.

Activity is a parameter taking into account the concentration and interaction of any crystal component with other components. Relation (1.2.18) holds true not only for a one-component but also for a multicomponent melt. In particular, if the liquid phase represents an $\alpha\beta$ compound melt, the chemical potential $\mu_{\alpha\beta l}$ will be equal to the sum of the chemical potentials of all constituents:

$$\begin{aligned}\mu_{\alpha\beta l} &= \mu^0_{\alpha l} + RT \ln a_\alpha + \mu^0_{\beta l} + RT \ln a_\beta = \mu^0_{\alpha l} + \mu^0_{\beta l} + RT \ln a_\alpha a_\beta \\ &= \mu^0_{\alpha\beta l} + RT \ln a_{\alpha\beta}.\end{aligned} \tag{1.2.19}$$

The expression

$$\mu_\alpha - \mu_\beta + g^\beta_\beta - g^\beta_\alpha = kT \ln \frac{a_\alpha}{a_\beta} + g^\beta_\beta - g^\beta_\alpha \tag{1.2.20}$$

will be substituted into (1.2.13) to obtain the crystal chemical potential, taking the energy of a perfect defect-free crystal as the reference for all energies:

$$\mu^0_{\alpha s} = \frac{\partial G^0}{\partial N_\alpha} = \frac{\partial \left[\sum_\alpha g^\alpha_\alpha N^\alpha_\alpha - TS \right]}{\partial N_\alpha} = g^\alpha_\alpha = 0. \tag{1.2.21}$$

Instead of (1.2.13), we will have

$$A_\alpha^\beta = R_\alpha^\beta \frac{a_\alpha}{a_\beta} \exp\left\{\frac{1}{kT}\left[-g_\alpha^\beta + \left(\mu_e - \varepsilon_\alpha^\beta\right) d_\alpha^\beta \left(1 - \delta_\alpha^\beta\right)\right]\right\}. \qquad (1.2.22)$$

This expression enters (1.2.12) and (1.2.15), relating the concentrations of any kinds of defect, including soluble impurities, to their activities. The latter represent the ratio of the gas pressure above the crystal melt to that above the pure component at the same temperature [14]:

$$a_\alpha = \frac{P_\alpha}{P_\alpha^0}. \qquad (1.2.23)$$

The activity can also be found from experimental liquidus curves, where the concentrations are defined as

$$\frac{H_f}{R}\left(\frac{1}{T_f} - \frac{1}{T}\right),$$

where H_f and T_f are the melting heat and temperature, and the activity is described as [14]

$$\ln a = \ln fx = \frac{H_f}{R}\left(\frac{1}{T_f} - \frac{1}{T}\right). \qquad (1.2.24)$$

In other words, the activity can be found from an analysis of the phase diagram and thermodynamic quantities—the melting or dissolution heat.

Finally, for some $A^{III}B^{V}$ semiconductors containing impurity *C*, the interaction parameters Ω_{AC}, Ω_{BC}, and Ω_{AB} are known. Then, the activity of an impurity (defect) can be calculated from [17]

$$RT \ln \gamma_C = \Omega_{AC} x_A^2 + \Omega_{BC} x_B^2 + \left(\Omega_{AC} + \Omega_{BC} - \Omega_{AB}\right) x_A x_B. \qquad (1.2.25)$$

Another attempt [14] to allow for the deviation from solution perfection was to introduce what is known as excessive thermodynamic functions ΔG^{exc}, ΔS^{exc}, and ΔH^{exc} equal to the difference between their actual values and those for an ideal solution:

$$\Delta G^{exc} = \Delta G^{mix} - \Delta G_{id}^{mix}, \qquad (1.2.26)$$

$$\Delta S^{\text{exc}} = \Delta S^{\text{mix}} - \Delta S_{\text{id}}^{\text{mix}}, \tag{1.2.27}$$

$$\Delta H^{\text{exc}} = \Delta H^{\text{mix}} - \Delta H_{\text{id}}^{\text{mix}}, \tag{1.2.28}$$

where ΔG^{mix}, ΔS^{mix}, and ΔH^{mix} are the mixing energy, entropy, and enthalpy, respectively.

It is quite obvious that both descriptions of solution deviation from perfection must be equivalent to each other; therefore, the excessive functions and activity coefficients must be interrelated. Both descriptions—the one with activity coefficients and the other with excessive thermodynamic functions—are more or less suitable only at their small values, i.e., at

$$\begin{gathered} f << 1, \qquad \Delta G^{\text{exc}} << \Delta G^{\text{mix}}, \\ \Delta S^{\text{exc}} << \Delta S^{\text{mix}}, \qquad \Delta H^{\text{exc}} << \Delta H^{\text{mix}}. \end{gathered} \tag{1.2.29}$$

Let us compare these phenomenological methods.

The method of quasi-chemical reactions is not flexible enough. If an additional type of defect is introduced into the problem, the number of quasi-chemical reactions is to be increased. One is faced with this situation when treating amphoteric impurity solubility in real crystals. In addition to doping impurity atoms, a semiconductor crystal contains foreign impurities, vacancies, intrinsic interstices, gas admixtures, associate defects of the impurity–vacancy type, etc. Moreover, many of these defects may be in both neutral and ionized states.

Concentrations of all defects are interrelated. Therefore, the concentration of each defect type, including the sought for C_s, can be expressed as a combination of partial K values characterizing individual quasi-chemical reactions. But the main difficulty in using this method is the lack of knowledge about H_i, i.e., the partial enthalpies of quasi-chemical reactions. A general disadvantage of this method is that it neglects practically all types of interaction among defects.

The method of free energy minimization is free from these flaws. On the other hand, one must know the semiconductor–impurity solid solution model, which is unnecessary in the quasi-chemical approach. However, most semiconductor solid solutions are regular, or even perfect. For this reason, the application of the minimization method to calculations of soluble impurity concentrations seems preferable, although involving more cumbersome procedures.

The first general model, based on Gibbs free energy minimization and designed for the calculation of defect formation, was suggested by Vinetzky and Kholodar [18] and developed by Bulyarsky and Oleinikov [15, 16].

Following this method closely, we have derived [3] formulas relating crystal growth conditions to several physical quantities—concentrations of various defects, high temperature concentrations of free carriers, the homogeneity region width, and impurity distribution coefficients. This approach allows one to invoke a greater variety of experimental data, thereby constructing more sophisticated and realistic models. In Section 5.7, we will illustrate the general solution to this problem for a regular semiconductor–impurity solution, taking into account all possible interactions of impurity atoms with one another and with other point defects.

The two methods of thermodynamic description discussed above require a knowledge of the thermodynamic behavior of charge carriers. In the first model, it is described in terms of quasi-chemical ionization reactions of defects (impurities):

$$[\ ]^0 \rightleftarrows [\ ]^+ + n, \qquad [\ ]^0 \rightleftarrows [\ ]^- + p, \tag{1.2.30}$$

where [ ] stands for defect concentration.

In the other model, the carrier thermodynamics is defined by the Fermi level position in the crystal band structure. The electron–hole subsystem of a crystal is related to its atom–defect subsystem. In other words, the concentrations of charge carriers and those of charged point defects are interrelated. Since the crystal as a whole is electrically neutral, this relationship is written as a neutrality condition (equation):

$$n + \sum_k [\ ]_k^- = p + \sum_q [\ ]_q^+ , \tag{1.2.31}$$

where k and q are the number of acceptor and donor defects, respectively.

The concentrations of charged defects are, in turn, related to total concentrations

$$[\ ] = [\ ]^0 + [\ ]^+ \tag{1.2.32}$$

via the quasi-chemical reactions of (1.2.30). Here, a plus superscript refers to a donor defect and a minus one to an acceptor defect.

Thus, the solution of the neutrality equation (1.2.31) is an integral part of a thermodynamic treatment of solubility of any impurity.

1.3 GENERAL CHARACTERISTICS OF IMPURITY CENTERS

There are two general concepts used to describe an impurity center in a crystal—its *state* and *behavior*.

In the generally accepted terminology [19], an impurity state is understood as the electronic structure of a center, its charge states, positions and local symmetry in the host crystal lattice. The word "behavior" covers the energy spectrum, the capture cross sections for charge carriers and photons, the degeneracy factor, and some other parameters of the center, which describe typical features of the system of interest. From here follow quantitative *physicochemical characteristics* of impurity centers. Of primary importance is their solubility, i.e., the equilibrium concentration of impurity atoms in a particular crystal at a definite temperature and pressure. The temperature and pressure dependences of solubility are represented as *P–T–x* diagrams, where *P* is pressure above the solution, *T* is temperature, and *x* is composition expressed as mole fraction or impurity concentration. It is difficult to present a three-dimensional diagram. Moreover, one of the experimental parameters, either *P* or *T*, is maintained constant, reducing the diagram to the functions $x(T)_{P=\text{const}}$ or $x(P)_{T=\text{const}}$. The $x_l(T)$ and $x_s(T)$ functions represent liquidus and solidus curves, i.e., solubilities in liquid and solid semiconductor–impurity solutions.

Naturally, dissolution enthalpy and entropy are also physicochemical characteristics. The former defines the energy required for the dissolution of impurity atoms and the latter characterizes the degree of ordering of a semiconductor–impurity solution.

A parameter important for semiconductor technology is

$$K = \left(\frac{x_S}{x_l} \right)_{T=T_m} \tag{1.3.1}$$

measured at melting temperature. It is known as an equilibrium *distribution coefficient* of an impurity in a particular semiconductor. In real conditions, this ratio is measured at other temperatures, rather than at $T = T_m$. Thus the distribution coefficient K^* differs from K and appears dependent on the growth rate of a doped semiconductor sample, on the mixing rate during the growth from the melt, and on the presence of other point defects because they interact in the liquid and solid phases. The parameter K^* is termed an effective distribution coefficient of an impurity.

Among other physicochemical parameters are the activity a and the activity coefficient f for the solid and liquid phases mentioned above, which describe the degree of deviation from crystal perfection.

Energy characteristics include ionization energies of impurity atoms in a crystal, i.e., the energy levels contributed by impurity atoms to the crystal spectrum. More than one level may be associated with an impurity center, which means that impurity centers may be electrically neutral or singly and multiply charged.

Besides, there is the electron affinity of impurity atoms, E_j. Qualitatively, ionization and electron affinity largely define the donor or acceptor nature of impurities. Calculation of E_i and E_j requires the choice of reference. This is usually the allowed band edge E_c or E_v, but sometimes it is more convenient to take $E_g/2 = 0$ or the Fermi level position $E_F = 0$ as a reference point.

The next important energy characteristic is the cross section of non-equilibrium carrier capture by an impurity center—for electrons σ_n and for donors σ_p. This parameter describes the kinetics of charge carriers, i.e., their recombination by impurities. The quantitative parameters of impurity recombination are the lifetimes τ_n and τ_p of a doped semiconductor.

The energy characteristics of impurity centers manifest themselves in many properties of semiconductors under the action of external factors. For example, light irradiation can provide partial coefficients of light absorption and reflection by "impurity centers" and "free carriers" [21], hydrostatic compression leads to splitting of impurities initially present in a crystal, thereby increasing the content of active doping impurity atoms [22], etc.

Kinetic characteristics are used to describe migration of impurity atoms in the semiconductor bulk and on its surface. A macroscopic parameter is the diffusion coefficient D. Since it is generally defined as [23]

$$D = D_0 \exp(-E_m / kT), \tag{1.3.2}$$

the activation energy for diffusion E_m and the entropy factor D_0 should also be referred to the kinetic characteristics of impurity centers.

It is important to classify impurity centers, but it would be incorrect to base this classification on any one characteristic. There are three criteria generally accepted for this purpose. Impurity centers are classified by the type of crystallochemical position, by the nature of chemical bonding of an impurity atom to host atoms, and by the building-in pattern of impurity levels in the semiconductor energy spectrum. These criteria, denoted I, II, and III in Table 1.1, permit the identification of three types of impurity centers.

Table 1.1 Classification of impurity centers in semiconductors.

I	II	III
1. site (substitution)	1. hydrogen-like	1. shallow
2. interstitial	2. with partly filled electron shells	2. deep
3. antistructural	3. amphoteric	3. resonance
	4. isovalent	4. antiresonance
	5. gas-generating	

The nature of impurities in group I is clear from their names. Those in group II will be described in the next chapter. Here, we will discuss briefly only the impurities of group III.

Shallow impurities are those introducing their energy levels into the semiconductor forbidden gap in the immediate vicinity of the allowed band edges E_c and E_v.

Deep impurities are, on the contrary, those having their energy levels far from E_c and E_v, deep into the forbidden gap. This classification does not draw a sharp energy line between shallow and deep levels. It is generally accepted that the ionization energy E_i of shallow impurities does not exceed $(0.05–0.07)E_g$. Impurities with larger E_i refer to deep impurities. The arbitrary character of this subdivision will be shown in Section 2.1.

The above classification differentiates between resonance and antiresonance levels, depending on their position in the conduction or valence bands, respectively. Hjalmarson and co-workers [24] have shown that they represent intrinsic binding (antibinding) impurity states, as distinguished from impurities, whose binding (antibinding) states produced by hybridization of impurity *s*-electrons with host *p*-electrons are made up of Bloch functions of the crystal atoms. Deep impurities are formed in the latter situation. The formation of impurities listed in Table 1.1 will be discussed in Section 2.2.

REFERENCES

1.1. R.A. Swalin, *Thermodynamics of Solids*. New York: Wiley, 313 p. (1961).

1.2. B.I. Shklovsky, A.L. Efros, *Elektronnye svoistva legirovannykh poluprovodnikov* (Electronic Properties of Doped Semiconductors). Moscow: Nauka, 416 p. (1979) (in Russian).

1.3. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushikh*

defektov v poluprovodnikakh (Thermodynamics and Kinetics of Defect Interactions in Semiconductors). Moscow: Nauka, 352 p. (1997) (in Russian).
1.4. Ter Haar, G. Wergeland, *Elements of Thermodynamics*. Massachusetts–London: Addison–Wesley, 220 p. (1966).
1.5. K.H. Zschfur, A. Vogei, *Proceedings of 3rd Symposium on GaAs, International Conference Series*, No. 9: 100–107 (1970).
1.6. K. Hellwey, O. Madelung, eds., *Numerical Data and Functional Relationship in Science and Technology*, *New Series* **17**. Berlin: Springer–Verlag, 652 p. (1984).
1.7. V.I. Fistul, *Silnoligirovannye poluprovodniki* (Heavily Doped Semiconductors). New York: Plenum Press, 418 p. (1969).
1.8. V.M. Glazov, S.N. Chizhevskaya, N.N. Glagoleva, *Zhidkie poluprovodniki* (Liquid Semiconductors). New York: Plenum Press, 362 p. (1969).
1.9. R.H. Zschaur, *Festkorperproblem*. In: H.J. Quciser, ed., *Braunschweig*, *Wieweg* **15**: 1–20 (1975).
1.10. C.M. Wolfe, G.E. Stilman, *Appl. Phys. Lett.* **27**: 564–574 (1975).
1.11. E. Kuphel, L. Schatachetskii, A. Pocker, *Journ. Appl. Phys*. **17**: 63–72 (1978).
1.12. V.I. Fistul, *Vvedenie v fiziku poluprovodnikov* (Introduction to Semiconductor Physics), 2nd ed., Moscow: Bysshaya shkola, 352 p. (1984) (in Russian).
1.13. G. Brouwer, *Philips Res. Rep*. **9**: 366–377 (1954).
1.14. F.A. Kroger, *The Chemistry of Imperfect Crystals*. Amsterdam: North–Holland Publ. Co., 654 p. (1964).
1.15. S.V. Bulyarsky, V.P. Oleinikov, *Phys. Stat. Sol. B* **141**: 7–10 (1987).
1.16. S.V. Bulyarsky, V.P. Oleinikov, *Phys. Stat. Sol. B* **146**: 439–447 (1988).
1.17. V.M. Andreev, L.M. Dolginov, D.N. Tretyakov, *Zhidkostnaya epitaksia v tekhnologii poluprovodnokovykh priborov* (Liquid-Phase Epitaxy in Semiconductor Device Technology). Moscow: Sov. Radio, 328 p. (1979) (in Russian).
1.18. V.N. Vinetsky, G.A. Kholodar, *Statisticheskie vzaimodeistvia elektronov i defektov v poluprovodnikakh* (Statistical Interactions of Electrons and Defects in Semiconductors). Kiev: Naukova dumka, 187 p. (1969) (in Russian).
1.19. E.M. Omelyanovsky, V.I. Fistul, *Primesi perekhodnykh metallov v poluprovodnikakh* (Transition Metal Impurities in Semiconductors). Moscow: Metallurgia, 192 p. (1983) (in Russian).
1.20. V.M. Glazov, A.M. Pavlova, *Khimicheskaya termodinamika i fazovye ravnovesiya* (Chemical Thermodynamics and Phase Equilibria). Moscow: Metallurgia, 336 p. (1981) (in Russian).
1.21. K. Seeger, *Semiconductor Physics*. Vienna–New York: Springer–Verlag, 615 p. (1973).
1.22. V.I. Fistul, A.R. Turaev, *Phys. Stat. Sol*. A **136**: 337–349 (1993).
1.23. D. Shaw, ed., *Atomic Diffusion in Semiconductors*, London–New York: Plenum Press, 684 p. (1973).
1.24. H.P. Hjalmarson, P. Vogel, *et al*., *Phys. Rev. Lett.* **44**: 810–813 (1980).

Chapter 2

Impurity Behavior in Semiconductors

2.1 HYDROGEN-LIKE IMPURITIES

The hydrogen-like behavior of some impurities follows from a simple mechanism for impurity atoms which form a substitutional solution with a semiconductor crystal. The valences of a substituting atom and a substituted atom differ by ±1. For germanium and silicon, these are group-III and group-V impurity atoms (Figure 2.1). In this case, four outer valent electrons of a host atom are replaced by four valent electrons of an impurity atom. Likewise, hydrogen-like impurities for $A^{III}B^{V}$ compounds are group-II atoms (acceptors), which replace A^{III}-sublattice atoms, and group-VI atoms (donors), which replace B^{V}-sublattice atoms.

It is generally believed that the basic theory of hydrogen-like impurity states (which normally produce shallow levels in semiconductors) was completed about 40 years ago. This was mostly done by Cohn and Lattinger [1, 2] who used the effective mass (EM) method. They showed that the concept of effective mass characterizing the dispersion law of charge carriers in a forbidden gap is also valid for the description of electron states in a nonperiodic force field created by various crystal defects, including impurity centers. On two general assumptions—(1) a slow variation in defect potential perturbing the crystal periodic potential and (2) a lower carrier binding energy (ionization energy) than the forbidden gap width, $E_i << E_g$, i.e., the problem of finding the impurity energy spectrum and the respective wave functions reduces to a one-electron problem described by the equation

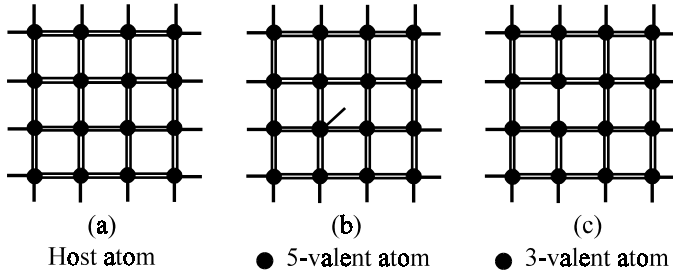

Figure 2.1. Schematic diagram illustrating the formation of donors and acceptors in a doped covalent crystal: (a) – no impurity; (b) – with a donor; (c) – with an acceptor.

$$\left[-\frac{\hbar^2}{2m}\Delta + V(\vec{r})\right]F(\vec{r}) = EF(\vec{r}), \qquad (2.1.1)$$

where $F(\vec{r})$ is what is known as a peak function related to the sought for wave function $\Psi(\vec{r})$ as $\Psi(\vec{r}) = F(\vec{r})U_{n,0}(\vec{r})$. Here, $U_{n,0}(\vec{r})$ is the normalized Bloch function near the conduction band extremum with $k = 0$.

The condition for a slow variation of impurity potential can be written as

$$a_0 = |\nabla V|/|V| << 1, \qquad (2.1.2)$$

where a_0 is the lattice period.

If we represent the potential energy $V(\vec{r})$ of a charge carrier localized in the impurity center field as Coulomb interaction energy, taking into account the crystal static permittivity $V(r) = -e^2/Kr$, equation (2.1.1) will transform to the wave equation for a hydrogen atom. Its solutions are

$$E = -Ry^*/n^2, \qquad (2.1.3)$$

where Ry^* is the ground energy of an impurity in the hydrogen-like approximation (the effective Rydberger is $Ry^* = e^4m/2K^2\hbar^2$, $n = 1, 2, ...$).

The peak function for the ground state is

$$F(\vec{r}) = \frac{1}{(\pi a^3)^{1/2}}\exp\left(-\frac{r}{a}\right), \qquad (2.1.4)$$

where the orbit radius of an impurity electron is $a = h^2K / me^2$, with the value of a much larger than the lattice constant.

These simple relations for the energy and characteristic size of the wave function have been derived on the assumption of an isotropic dispersion $E(k)$ which is valid for many $A^{III}B^{V}$ and $A^{II}B^{VI}$ compounds, with the absolute conduction band maximum lying at the Brillouin zone center at $k = 0$.

The data presented in Table 2.1 were borrowed from [2] to demonstrate that a simple modification of the EM method describes quite satisfactorily experimental results for shallow donors in some crystals. Especially good is the agreement for excited impurity states.

For elemental germanium and silicon crystals having a complex conduction band structure, the ionization energies of the 1*s*-levels[1] of group-V donors calculated from (2.1.3) differ considerably from the experimental values. Nevertheless, the comparison of theoretical and experimental results shows that a simple EM method provides a correct order of magnitude for the energy of the ground *s*-type impurity state and that the ionization energy values for excited *p*-states of hydrogen-like centers are quite close to the experimental values. For direct band semiconductors of the GaAs type, the ionization energies of group-VI and group-IV donors are close to one another (see Table 2.1), showing only a 10% difference from the calculations.

However, the effective mass method failed to account for the dependence of the ground state energy on the doping impurity chemistry. For example, the differences in E_i values for group-V donors in germanium and silicon are as large as 30% and 100%, respectively. This method is also entirely unsuitable for the description of energy spectra of isoelectron and deep impurity centers.

The EM method has been considerably improved over the time that has passed since the publications of Cohn and Lattinger. These improvements have been due to the following findings. The impurity potential in the vicinity of a defect was refined [3–5], and the spatial variation of dielectric screening of the impurity field, $K(r)$, essential in the vicinity of an impurity center, was taken into account [5–8]. Besides, the theory considered the real band structure having both equivalent and additional extrema in the conduction band [9–12].

It should be noted that most authors made their calculations taking account of one factor only, while other important aspects of the problem were left aside. This, certainly, reduced the applicability of theoretical results, indicating, on the other hand, the complexity of the problem.

[1] Similarity in the energy spectra of shallow impurities and hydrogen atoms is reflected, in particular, in the designation of the ground and excited states: 1*s*, 2*s*, 2*p*, etc.

Table 2.1. Ionization energy of shallow impurities in some semiconductors.

Substance	K	m/m_0	E_{1s}, MeV		E_{2p}, MeV
			Theory	Experiment	Theory
GaAs	12.53	0.066	5.67	6.08 (Ge)[1] 5.81 (Si) 5.89 (Se) 6.1 (S) 5.87 (Sn)	1.42[2]
InP	12.60	0.080	6.80	7.28	1.70
CdTe	10.00	0.96	12.96	13.78	3.24
CdSe	9.00	0.110	18.33	–	–

[1] Bracketed are doping impurities for GaAs.

[2] The experimental value for GaAs is $E_{2p} = 1.44$.

A detailed analysis of all refinements introduced in the EM method can be found in [2–13]. So we will present only some results reflecting the progress of this method in the calculation of energy spectra of shallow local states, as was done in our work [14]. This problem can be best illustrated with reference to elemental germanium and silicon, since their shallow donor and acceptor spectra have been studied most thoroughly.

The effective mass of electrons in both semiconductors is anisotropic, the absolute minima are aligned with the [111]-axis for germanium and the [100]-axis for silicon, and the iso-energy surfaces represent ellipsoids of revolution described by the known values of longitudinal and transverse effective masses. The number of equivalent ellipsoids N is prescribed by the cubic crystal symmetry and is equal to 4 and 6 for germanium and silicon, respectively; as a result, the local states appear to be N-fold degenerate. For this case, the wave function of an impurity electron is

$$\Psi(\vec{r}) = \sum_{j=1}^{N} a_j F_j(\vec{r}) U_j(\vec{r}), \qquad (2.1.5)$$

where factors a_j are determined by the impurity state symmetry.

It is essential that the deviation of the impurity potential from the Coulomb potential in the immediate vicinity of a center leads to impurity state splitting, whose nature can be identified from the symmetry considerations alone. For example, the ground 6-fold degenerate donor state in silicon splits into a singlet (A_1), a triplet T_2, and a doublet E state; this splitting is

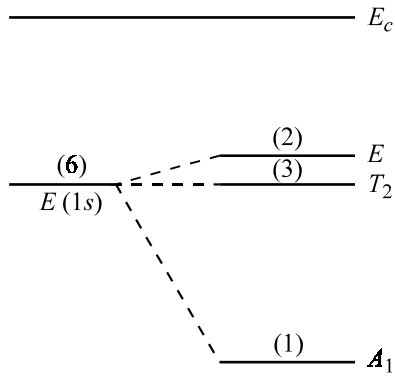

Figure 2.2. Schematic splitting of the ground donor state in silicon. Numbers in brackets indicate the degeneracy multiplicity of energy levels with spin degeneracy not taken into acccount.

illustrated in Figure 2.2. The ground state of a donor in germanium splits into two states—a singlet and a triplet.

Baldereschi [10] has demonstrated the necessity to make allowance for so called intervalley mixing, or intervalley interaction, leading to splitting. He took into account the wave vector dependence of dielectric permittivity, which becomes appreciable at its large values. His calculations and available experimental data for germanium and silicon are given in Table 2.2. The energy spectra of excited 2*s*, 2*p*, 3*s*, etc., donor states in germanium and silicon have been calculated by Faulkner [15], who used the variational method taking account of effective mass anisotropy. These results are presented in Figures 2.3 and 2.4. One can see that the effective mass method satisfactorily describes the spectrum of excited *p*-states of donors, in contrast with the

Table 2.2. Binding energy of splitted levels of the donor ground 1*s*-state in germanium and silicon, meV [7].

Donor	Ge		Si		
	A_1	T_2	A_1	T_2	E
–	9.81*	9.21*	31.27*	20.67*	19.57*
P	12.90	9.90	45.50	33.90	32.60
As	14.17	10.0	57.70	32.60	31.20
Sb	10.32	10.0	42.70	32.90	30.60

* Calculations.

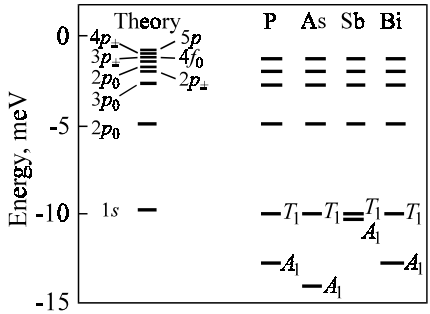

Figure 2.3. The energy spectrum of donors in germanium [15].

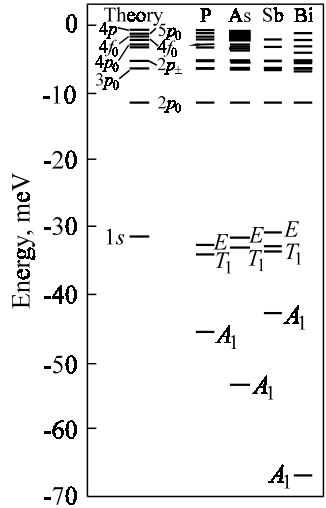

Figure 2.4. The energy spectrum of donors in silicon [15].

ground state. Indeed, the effect of impurity short-range potential is small, and the wave functions of *p*-states vanish in the immediate vicinity of an impurity due to the state symmetry. The agreement between the calculations and experimental values for excited levels proved to be so good that the author [15] was even able to refine the dielectric permittivity values for germanium and silicon at low temperatures.

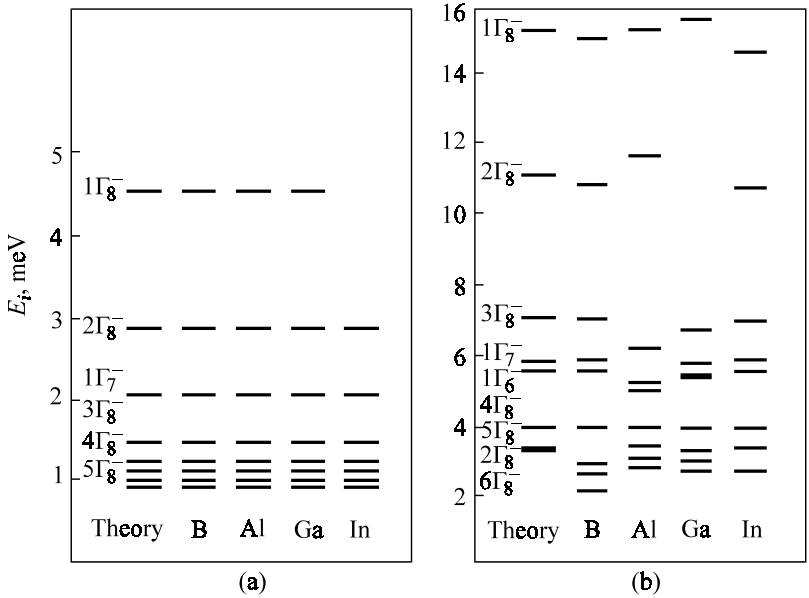

Figure 2.5. The spectrum of excited states of acceptor impurities in germanium (*a*) and silicon (*b*) [2].

The agreement between the experimental and theoretical values for the excited acceptor levels in germanium and silicon was also as good. Without going into details of the theoretical model used, we illustrate in Figure 2.5 the calculated spectra of group-III excited acceptor states, borrowed from [2], together with available experimental data.

As for the ground state energy of impurity centers, the agreement between calculated and experimental values leaves much to be desired. To describe the extent of disagreement, one often uses the value of "chemical shift" $\Delta = E_j - E_1$, where E_j is the binding energy calculated from (2.1.3) with $n = 1$. The chemical shift is primarily due to intervalley interaction and, of course, with the potential (electronic structure) of the impurity center. An important correction to the EM method is that the polarization of the medium in the vicinity of an impurity ion cannot be described by a macroscopic dielectric constant. Indeed, the allowance for spatial dispersion $K(r)$ changes radically the impurity effective potential at the central site.

Experimental energy levels of hydrogen-like impurities in A^{IV} and $A^{III}B^{V}$ semiconductors are summarized in Tables 2.3–2.5; their solubility data will be discussed in Chapter 3.

Table 2.3. Ionization energies E_i (meV) of group-III acceptors and group-V donors in germanium (E_i values are counted from E_v for acceptors and from E_c for donors).

Impurity	Acceptors					Donors			
	B	Al	Ga	In	Tl	P	As	Sb	Bi
E_i (OM)	10.57	10.90	11.07	11.74	13.10	12.76	14.04	10.19	12.68
E_i (EM)	10.4	10.2	10.8	11.2	–	12.0	12.7	9.6	11.5

OM – optical measurement, EM – electrical measurement.

Table 2.4. Ionization energies E_i (meV) of group-III acceptors and group-V donors in silicon. (E_i values are counted from E_v for acceptors and from E_c for donors).

Impurity	Acceptors					Donors			
	B	Al	Ga	In	Tl	P	As	Sb	Bi
E_i (OM)	44.3	68.4	72.3	155.4	–	45.3	53	43	70
E_i (EM)	46	57	65	160	200	44	49	39	69

OM – optical measurement, EM – electrical measurement.

The reader may find it surprising that we have given no summary table for $A^{II}B^{VI}$ semiconductors, but this is due to some specific properties of these crystals. First, the band structure of these semiconductors consists of several valence bands, holes in the lower sub-band having a much lower mobility than in the upper sub-band. So hole transition to the sub-band with a lower mobility increases appreciably the Hall coefficient, e.g., PbTe shows this effect at a temperature above 150 K. Due to this effect, it is practically impossible to identify shallow acceptor levels in *p*-type crystals even if they are introduced deliberately. Second, the high values of static dielectric permittivity and the small effective masses of electrons ($\sim 10^{-2} m_0$) lead to very low values of impurity levels in *n*-type crystals. These levels are so shallow that they become completely ionized even at helium temperatures and merge with the allowed band to form a common band.

Nevertheless, there have been some reports on impurity levels in $A^{II}B^{VI}$

Table 2.5. Ionization energies E_i (meV) of group-II acceptors and group-VI donors in $A^{III}B^{V}$ semiconductors. (E_i values are counted from E_v for acceptors and from E_c for donors).

Semiconductors	Acceptors		Donors		
	Zn	Cd	S	Se	Te
GaP	70.1	102.5	107	102	92.6
GaAs	24...31	21...30	5.84	5.81	3
GaSb	37	–	–	–	–
InP	4	5	–	–	–
InAs	–	11	–	–	–
InSb	7.5	7.5	250	150	50

The GaSb conduction band has three minima: the major Γ-minimum and two minor X- and L-minima at different points of the Brillouin zone; S, Se, and Te impurity atoms form very shallow levels within $E_{g(\Gamma)}$ – 30 meV, the atomic sequence indicating the decreasing depth. As for the minor minima, the atomic sequence is reverse, but the ionization energy lies about 300 times deeper.

semiconductors. If they are found to be shallow, their hydrogen-like nature should be questioned.

2.2 IMPURITIES WITH PARTLY FILLED ELECTRON SHELLS (*d*- AND *f*-IMPURITIES)

It has been shown experimentally [14] that transition metal atoms dissolve at interstices of germanium and silicon crystal lattices and do not interact chemically, in the strict sense of this word, with host atoms. In other words, *d*-electrons do not mix with electrons of germanium and silicon host atoms. In $A^{III}B^{V}$ and $A^{II}B^{VI}$ semiconductors, they, on the contrary, produce substitutional solutions and form chemical bonds. This means that *d*-electrons usually hydridize with *s*- and *p*-electrons of the *A*-sublattice atoms. For this reason, impurity transition metal atoms, which are in an un-ionized state in a semiconductor, are capable, owing to their electrical activity, to give off their *d*-electrons to the conduction band, say, under the action of light:

$$A^0 + \hbar\omega_d \rightarrow A^+ + e_c\,, \qquad (2.2.1)$$

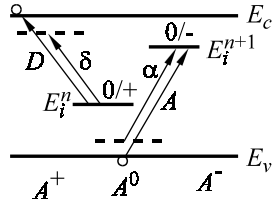

Figure 2.6. Optical transitions in the ionization of donor (D) and acceptor (A) states; transitions with excited donor (δ) and acceptor (α) excitons of amphoteric 3*d*-impurity atoms.

These atoms, however, are also capable of capturing host valent electrons onto the *d*-orbitals to produce holes in the valence band:

$$\mathrm{A}^0 + \hbar\omega_\mathrm{a} \rightarrow \mathrm{A}^- + p_\mathrm{v} \,. \tag{2.2.2}$$

These transitions are indicated by arrows in Figure 2.6. The *d*-shell configuration changes during the transition, which means that the electron population of the shell changes. What one sees in Figure 2.6 are known as Allen diagrams.

The initial configuration of a *d*-atom is determined by its crystallochemical state. This depends on the crystal type and the position of the impurity *d*-atom in the crystal lattice. The rules for the change in the *d*-atom electronic configuration in a semiconductor crystal are described by a model suggested by Ludwig and Woodbury [18], who originally based it on experimental data. Later, the model was substantiated theoretically by Roitzin and Firshtein [19]. So we will use the abbreviation RFLW for it, as was suggested in [14].

When applied to silicon, this model assumes that in the substitution of a crystal lattice site, 4*s*-electrons and some of the missing (up to 4) 3*d*-electrons of a *d*-atom produce bonds with the nearest four silicon atoms. If a *d*-atom occupies an interstice, its 4*s*-electrons are repelled by the electrons of the nearest silicon atoms (ligands) and fill the *d*-shell. Therefore, 4-valent silicon and germanium crystals exhibit the following electronic restructuring of the *d*-atom:

$$\begin{array}{l} \text{(free atom)} \\ 3d^n 4s^m \end{array} \begin{array}{l} \nearrow\ 3d^{n-(4-m)}\ \text{(site)} \\ \searrow\ 3d^{n+m}\ \text{(interstice)}. \end{array} \tag{2.2.3}$$

Many experimental data [14] have supported the applicability of this model to binary $A^{III}B^{V}$ and $A^{II}B^{VI}$ semiconductors as well. The general result of the RFLW model can be represented schematically:

$$3d^n4s^m \text{ (free atom)} \rightarrow \begin{cases} 3d^{n-(V-m)} \text{ (site)} \\ 3d^{n+m} \text{ (interstice)}, \end{cases} \qquad (2.2.4)$$

where V is the valence of the substituted host atom.

The RFLW model is based on the concepts of the well-known crystal field theory [20] assuming the equivalence of spectroscopic units to describe electron terms and of coordination polyhedra (tetrahedra, octahedra, etc.) of the crystal structure. A polyhedron is considered as consisting of negatively charged ligand ions located at its vertices and a positively charged *d*-metal ion located at its center.

The main feature of the crystal field theory is the neglect of the electronic structure of ligands, which means that ligands are identified with electrical point charges. Their function is reduced to inducing an electric field, termed a crystal field.

The crystal field symmetry is determined by the symmetry of the polyhedron which makes up the crystal structure. In A^{IV}, $A^{III}B^{V}$, and $A^{II}B^{VI}$ semiconductors, one should consider only two types of symmetry—tetrahedral symmetry (T_d) and octahedral symmetry (O_h) to account for the second coordination sphere. The transformation mechanism for electron terms of the central *d*-ion in a ligand crystal field has been discussed in the book [14].

Table 2.6. Types of interaction described by crystal field theory parameters.

Parameters of crystal field theory	Types of interaction
$\Delta = 10D_q$	Interaction of a *d*-ion with crystal field H_{cr}; it is the measure of field strength and determines the splitting of *d*-ion terms in the crystal.
A	Interelectron interaction in a *d*-ion; A = const for all d^n-configurations.
B, C	Interelectron interaction in a *d*-ion H_{ee}; it determines the separation of terms due to Coulomb repulsion of electrons.
λ	Spin-orbital interaction H_{LS}; it is the interaction measure of orbital and spin moments of a *d*-ion in the crystal, leading to additional splitting of its levels.

Table 2.7. Valent states and splitting of *d*-ion terms in a tetrahedral crystal field.

Free *d*-ions	*d*-ions in crystal field								
Ni $3d^8 4s^2$					Ni^{4+}	Ni^{3+}	Ni^{2+}		
Co $3d^7 4s^2$					Co^{3+}	Co^{2+}	Co^{+}		
Fe $3d^6 4s^2$							Fe^{2+}	Fe^{+}	Fe^{0}
Mn $3d^6 4s^2$	Mn^{7+}	Mn^{6+}	Mn^{5+}	Mn^{4+}	Mn^{3+}	Mn^{2+}	Mn^{+}	Mn^{0}	Mn^{-}
Cr $3d^5 4s^{+}$	Cr^{6+}	Cr^{5+}	Cr^{4+}	Cr^{3+}	Cr^{2+}	Cr^{+}	Cr^{0}		
V $3d^3 4s^2$	V^{5+}	V^{4+}	V^{3+}	V^{2+}	V^{+}	V^{0}			
Ti $3d^2 4s^2$	Ti^{4+}	Ti^{3+}	Ti^{2+}	Ti^{+}	Ti^{0}				
Sc $3d^1 4s^2$	Sc^{3+}	Sc^{2+}	Sc^{+}	Sc^{0}					
Electronic configuration	$3d^0$	$3d^1$	$3d^2$	$3d^3$	$3d^4$	$3d^5$	$3d^6$	$3d^7$	$3d^8$
Number of d^n-electrons	0	1	2	3	4	5	6(4)	7(3)	8(2)
Spin $S = n/2$	0	1/2	1	3/2	2	5/2	2	3/2	1
Term ^{2s+1}L	^1S	^2D	^3F	^4F	^5D	^6S	^5D	^4F	^3F

The parameters in the crystal field theory are the Pack parameters: *A*, *B*, *C* and parameters D_q and λ. The types of interaction they describe are listed in Table 2.6, and the term splitting in a tetrahedral field following from group theory is shown schematically in Table 2.7.

It is easy to see that the schemes for the *d*-configuration in a tetrahedral field are identical to those for d^{10-n}-configurations in an octahedral field. It follows from the crystal field theory [20] that the splitting of terms of the d^1, d^4, d^6, and d^9 configurations is described only by parameter Δ. For ions, the d^2, d^3, d^7, and d^9 splittings are described by parameters D_q, *B*, and *C*. For this case, the expressions for interlevel energy are presented in [14].

A complete semi-quantitative picture of term splitting for ions with different d^n-configurations in a tetrahedral field (in an octahedral field for d^{10-n}-configurations) is represented by what is known as Tanabe–Sugano diagrams which can be found in [20] (Figure 2.7). These diagrams give a general picture of the behavior of all *d*-ion levels in a crystal field: the arrangement of the levels and the relative differences between them. Tanabe–Sugano diagrams are usually taken as initial data in experimental data processing.

It is necessary to point out the limits of the diagram applicability, which are largely associated with two circumstances: (1) the diagram is built only for one value of *B* and *C*; (2) the values of *B* and *C* refer only to free *d*-ions. However, the values of *B* and *C* for *d*-ions in crystals may differ considerab-

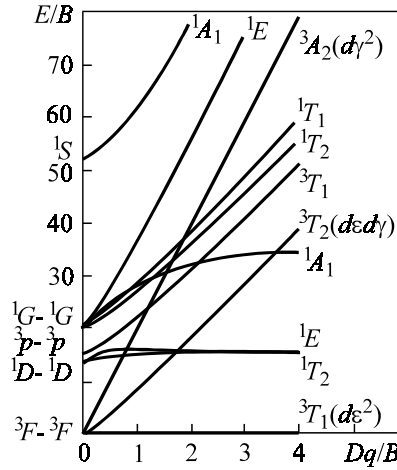

Figure 2.7. Energy levels for the d^n-configuration of $3d^8$ ions in a tetrahedral field.

ly. Moreover, the diagrams ignore spin-orbital and other fine interactions. These limitations indicate their qualitative, rather than quantitative, significance.

There are also more general limitations of the crystal field theory following from its initial postulates.

The major limitation is the concept of ligands as structureless point charges which do not exchange electrons with the central *d*-ion. This model corresponds to the limit case of 100% ionic bonding. For this reason, researchers were skeptical even about a qualitative application of the crystal field theory to describe *d*-ions in chemical compounds or in the A^{IV} or $A^{III}B^{V}$ types of crystals having a large proportion of covalent bonds.

Extensive experimental material, however, has shown that the crystal field theory can describe fairly well the qualitative character of level splitting in semiconductors, their sequence and relative energy intervals between them [14]. But in spite of its successful application to the description of term splitting for transition metal impurities, the theory gives no answer to two main questions: how the system of splitted levels is related to the allowed band edges E_c and E_v and what type (acceptor or donor) of the ground state these levels have.

Hjalmarson and co-workers [21] have demonstrated an important feature of *d*-impurities, namely, the fact that the deep levels introduced by them in the forbidden gap are intrinsic levels. They produce binding states built up primarily by wave functions of uncompleted *d*-shells and have, therefore, the *d*-symmetry, rather than the *s*- or *p*-symmetry. The electron wave function

then appears to consist of the Bloch component Ψ_{iB} and the *d*-function component:

$$\Psi_i = \Psi_{iB} + \Psi_{id}. \tag{2.2.5}$$

Thus, this wave function consists of the central core Ψ_{id}, retaining all characteristics of an atomic wave function but with an angular component transformed in accordance with the crystal point symmetry, and the Bloch trace Ψ_{iB} of the superposition of the *p*- and *d*-components of the Bloch wave functions of the valence and conduction bands. It was shown in [22] that

$$\Psi_{iB} = \hbar^2 / (2m_a E_a), \tag{2.2.6}$$

where m is a mean effective mass of carriers in the band a and E_a is the deep level energy counted from the band edge.

The more shallow the level, the larger the contribution of the trace to the wave function. The impurity potential of a *d*-atom turns out to have a resonance character [23] and is described as

$$U_d \sim A / (E - E_\Gamma), \tag{2.2.7}$$

where E_Γ is the intrinsic energy level of a *d*-electron renormalized by the crystalline medium, E is the energy of a scattered electron, and A is a function defined by the wave functions of *d*-electrons in (2.2.5).

If common potential scattering is negligible, as compared with resonance scattering, the deep level energy $E_{i\Gamma}$ will be

$$E_{i\Gamma} = E_{\Gamma 0} + M(E_{i\Gamma}), \tag{2.2.8}$$

where $E_{\Gamma 0}$ is the intrinsic level energy of a *d*-atom in the valence band and $M(E_{i\Gamma})$ is a function describing the effect of the covalent medium, i.e., the level renormalization.

It is this renormalization that "pushes" the intrinsic *d*-level out of the valence band to the forbidden gap. What is important is the fact that the *d*-*p*-hybridization occurs with the *E*- and T_2-states differently, being more significant with the latter. This can be interpreted in terms of the chemical bonding theory: ligands form stronger σ-bonds with the T_2-states and weaker π-bonds with the *E*-states. In [23, 24], this fact was interpreted from purely symmetrical considerations: the symmetry of cubic crystals allows the mixing of

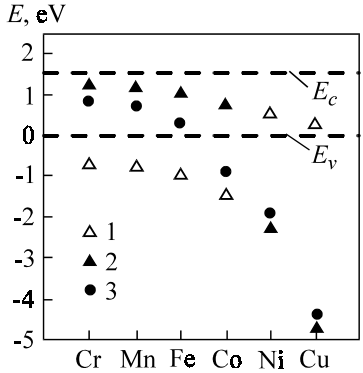

Figure 2.8. Energy levels of 3*d*-impurities in gallium arsenide (cluster calculations) [27]: 1 – t_2 (DBH); 2 – t_2 (CFR); 3 – *e*.

$T_2(p)$- and $T_2(d)$-states of different parity. This leads to a dominant contribution of the *p*-states of the valence band to the renormalization of the $T_2(d)$-states, while the *E*-states hybridize only with states of the higher conduction bands of the same kind. Such symmetry differences affect differently the formation of the trace Ψ_{iB} and the renormalization of the T_2-level $M(E_{i\Gamma})$. States in the conduction band make practically no contribution to the renormalization (2.2.8). This approach develops further the well-known ligand field theory [26] which takes into account the overlapping of the wave functions of the central (impurity) ion and neighboring ions of a semiconductor host.

This modification of the theory has been successfully applied to *d*-impurities and yielded the major parameter of term splitting Δ [25]:

$$\Delta = E_{iE} - E_{iT_2} = \Delta_0 + M\left(E_{iE}\right) - M\left(E_{iT_2}\right), \qquad (2.2.9)$$

where Δ is the splitting value in the crystal field theory and the other terms are the results of covalent renormalizations.

A more rigorous theory [24] takes into account potential and resonance scattering, so that the existence of two types of T_2-symmetry levels becomes possible: resonance states due to the *d*-levels of impurity electrons and "hybrids with dangling bonds" [26] formed primarily in potential scattering by an impurity *d*-shell. The former are usually termed crystal field resonance (CFR) levels and the latter dangling bond hybrid (DBH) levels.

The ligand field theory stimulated the development of cluster methods for

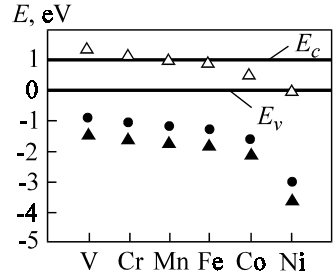

Figure 2.9. Energy levels of 3*d*-impurities in silicon [28].

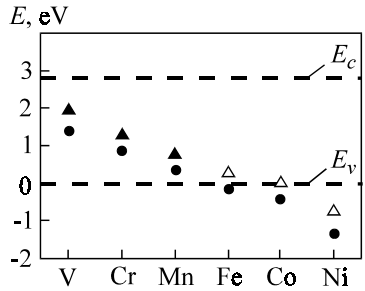

Figure 2.10. Energy levels of 3*d*-impurities in ZnSe [32].

the calculation of impurity center levels. Figures 2.8 and 2.9 illustrate such calculations for GaAs and Si and Figure 2.10 for ZnSe, showing the three types of levels T_2^{CFR}, T_2^{DBH}, and *E*. One can clearly see the difference in the behavior of *d*-impurities in these semiconductors. In $A^{III}B^{V}$ and $A^{II}B^{VI}$ compounds, the intrinsic *d*-level of impurities, except for nickel and copper, is in the forbidden gap. In silicon, as well as in germanium [29], the T_2-level appears to be in the valence band, while the *E*-level, split off the valence band and having the *p*- rather than *d*-symmetry, lies in the forbidden gap.

A limitation of this theory is its one-electron character. In reality, however, interelectron interactions play a great role. If we ignore their contribution, we will not be able to find the *d*-level position in the energy spectrum of a semiconductor. It is shown in [28, 29] that the energy of a multi-electron ion d^n containing rT_2-orbitals and $(n–r)E$-orbitals is described as

$$E_{\mathrm{i}}\left(d^{n}\right) = E_{\mathrm{i}\Gamma} + E_{MC}\left(B, C\right), \qquad (2.2.10)$$

where $E_{i\Gamma}$ defined in (2.2.8) has the form

$$E_{i\Gamma} = rE_{iT_2} + (n-r)E_{iE}, \tag{2.2.11}$$

where E_{MC} are corrections for multi-electron interactions.

It is demonstrated in [30, 31] that these corrections represent linear combinations of Racah's parameters *A*, *B*, and *C* (see Table 2.6). Then, if the initial (un-ionized) state of an impurity atom is d^n, and it is changed by ionization to d^{n-1} (donor transition),

$$d^n + \hbar\omega \to d^{n-1} + e_c, \tag{2.2.12}$$

the only way to define deep levels is to use the expression

$$E_{iD} = E_c - E_{i\Gamma}\left(T_2^r E^{n-r}\right) + E_{i\Gamma}\left(T_2^{r'} E^{n-1-r'}\right) + \Delta E_v + \Delta E_{MC}(B, C, B', C'), \tag{2.2.13}$$

where r' is the number of T_2-orbitals in the d^{n-1}-configuration; ΔE_v is the change in the valence band states due to the change of the scattering type when the valence transforms from d^n to d^{n-1}; B′ and C′ are the Racah parameters for the d^{n-1}-configuration.

For the acceptor transition, i.e., for electron capture, or hole emission to the valence band

$$d^{n-1} + \hbar\omega \to d^n + p_v, \tag{2.2.14}$$

we have

$$E_{iA} = E_{i\Gamma}\left(T_2' E^{n-r}\right) - E_{i\Gamma}\left(T_2^{r'} E^{n-1-r'}\right) - E_v + \Delta E_v + \Delta E_{MC}(B, C, B', C'). \tag{2.2.15}$$

The first transition will be designated as 0/+ and the second as 0/–, where zero stands for the initial un-ionized state and plus and minus in the deno-minator mean the *d*-ion charge resulting from the ionization. It is these tran-sitions which represent deep *d*-levels in Allen's diagrams (see Figure 2.6).

The E_{MC} corrections prove to be irregular functions of the *d*-shell popula-tion, i.e., of the *Z*-serial number of the *d*-atom. On the other hand, $E_{i\Gamma}$ repre-

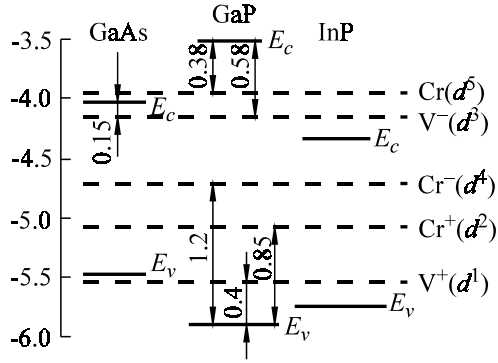

Figure 2.11. Energy levels in some $A^{III}B^{V}$ semiconductors, counted from the vacuum energy level.

sents a smooth function of Z. On the whole, therefore, (2.2.10) has an irregular character, and the multi-electron theory gives a piecewise linear function $E_i(d^n)$ of n (or of Z), well supported by experiments (Figure 2.11). The fundamental question as to whether these levels are related to the semiconductor band structure was answered by Ledebo and Ridley, who suggested introducing a common reference—the vacuum level, for both the lowest d-impurity levels and the semiconductor band structure. The position of the valence band edge relative to vacuum is to be found from the semiconductor photoionization energy (work function), while the d-level position with respect to vacuum is easy to find if the level position relative to this edge is known from experiments or calculations.

With this procedure, this idea has been verified for all impurities, from vanadium to copper, in basic $A^{III}B^{V}$ compounds [33] (Figure 2.11) and in $A^{II}B^{VI}$ semiconductors [34] (Figure 2.12).

The calculations of positions for CFR and DBH energy levels of 3-d impurities at silicon sites and interstices [35, 36] are illustrated in Figures 2.13 and 2.14, respectively. Although these results were obtained within one-particle theory, their qualitative agreement with experiments is fairly good. They show that it is only Zn, Cu, and Ni that produce energy levels in the forbidden gap, whereas the other d-levels correspond to interstices.

The general characteristics of d-impurity levels in basic semiconductors are summarized in Table 2.8 and the values of their energy levels E_i in Table 2.9. Impurities with partly filled 4-d (Ru, Rh, Pd) and 5-d shells (Os, Ir, Re) have not been studied in as much detail as 3-d impurities. Moreover, they were studied only in silicon [37, 38]. These impurities exhibit

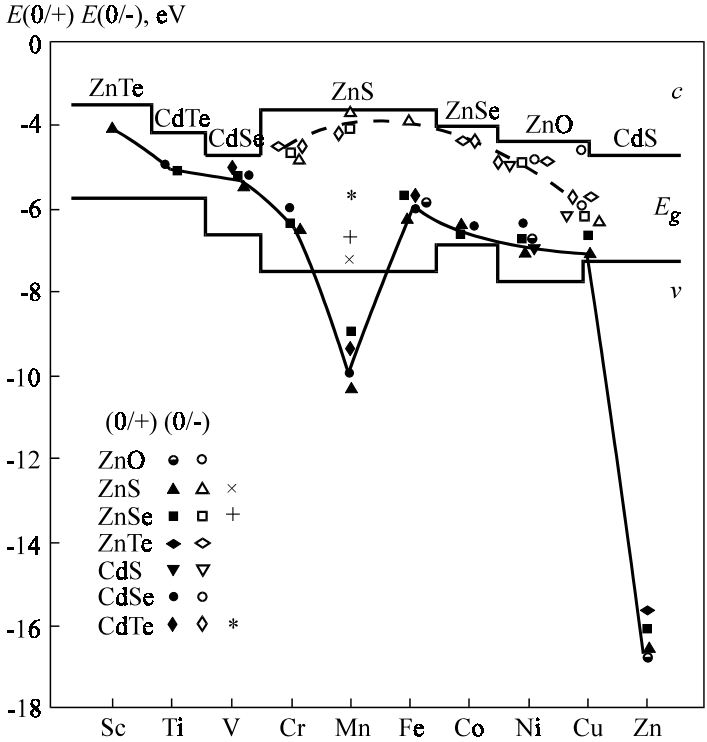

Figure 2.12. Energy levels of *d*-impurities in $A^{II}B^{IV}$ compounds, counted from the vacuum level [34].

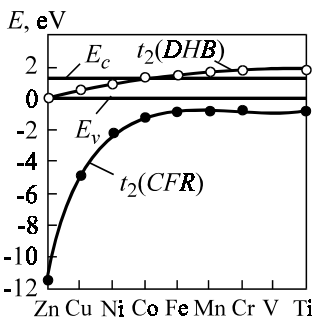

Figure 2.13. Energy levels of transition metal atoms at silicon sites, calculated in the one-particle approximation [35].

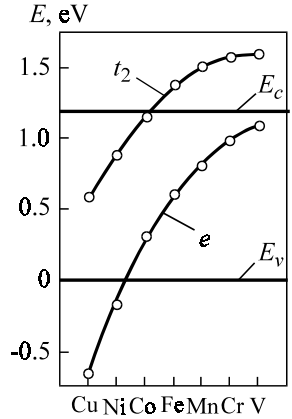

Figure 2.14. Energy levels of interstitial transition metal atoms in silicon, calculated in the one-particle approximation [36].

Table 2.8. Types of 3-*d* impurity levels in the forbidden gap of basic semiconductors.

Semiconductors	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu
GaAs			*a*	*A*					
GaP			*A*						
InP			*d*						
ZnO						*d*		*d*	*a*
ZnS	*d*		*d*	*A*	*Ax*	*A*	*d*	*A*	*A*
ZnSe		*d*	*d*	*A*	*Ax*	*d*	*d*	*A*	*Ax*
ZnTe				*a*			*a*	*a*	*a*
CdS								*A*	*a*
CdSe		*d*	*d*	*d*	*d*	*d*		*A*	*a*
CdTe			*d*	*a*	*Ax*	*d*	*a*	*a*	*a*
Si							*A*	*x*	*x*

d – CFR-type donor (0/+), *x* – DBH-type donor (0/+), *a* – CFR-type acceptor (0/–), *A* – amphoteric impurity.

amphoteric properties to be discussed in the next section. Here we will restrict our description to their states in the silicon crystal lattice.

It was established in the studies of diffusion characteristics, decomposition of silicon–impurity solid solutions [37], and electron spin and ENDOR

Table 2.9. Ionization energies of 3*d*-impurity atoms in basic semiconductors.

Impurities	E_i, eV				
	Si		GaP	GaAs	InP
	$E_c - E_i$	$E_v + E_i$	$E_v + E_i$	$E_v + E_i$	$E_v + E_i$
Sc	0.27*d*, 0.35*d*, 0.5*d*	0.35*a*, 0.45*d*		0.57, 0.85	
Ti	0.26*d*	0.1*a*, 0.29*d*		1.05	
V	0.3*d*, 0.4*d*, 0.5*d*	0.45*d*	0.85, 1.15	1.38, 0.73	0.95
Cr	0.23*d*, 0.4*d*	0.31*d*, 0.39*d*		1.85	1.64
Mn	0.3*d*; 0.5*d*		0.4, 1.93*	0.113	0.22
Fe	0.13*d*, 0.5*d*	0.4*d*	0.86, 1.27*	0.49, 0.86	0.8, 1.15
Co	0.22*a*, 0.3*a*, 0.53*a*	0.35*a*, 0.4*d*, 0.5*a*	0.41, 0.97*, 1.3*, 1.92*	0.14, 0.64, 1.0, 1.53	0.24, 0.71, 1.03
Ni	0.4*a*	0.2*a*	0.51, 1.74*	0.2	0.35
Cu	0.49*a*	0.4*a*	0.66, 0.82, 0.17	0.14, 0.46, 0.24	1.18, 1.06, 1.02
Zn	0.55*a*	0.4*d*	0.64	0.14, 0.29	0.31

resonances [38] that all atoms of these groups of elements can occupy only interstitial positions in crystals.

The authors of [37, 38] believe that isolated 4*d*- and 5*d*-atoms cannot exist at an interstice, because they form weak bonds with the lattice. Such states may be interstitials on bonds, split bonds or produce complexes with the neighboring vacancies. The most probable crystallochemical model of Pt and Pd impurity localization in the silicon lattice [38] is shown in Figure 2.15. In this model, an impurity atom is bonded by two silicon atoms in a unit cell. A hole is shared by a *d*-atom and a bond between the other two silicon atoms. So, the *d*-atom behaves as a singly charged acceptor, Pt^-.

The specific features of these impurities are a nonspherical electronic configuration and a strong localization. These two chemical properties are characteristic of coordination compounds with asymmetric distortions known as the Yan–Teller effect. In the work mentioned above, the analysis of this effect was reduced to symmetry identification of the basic term of local im-

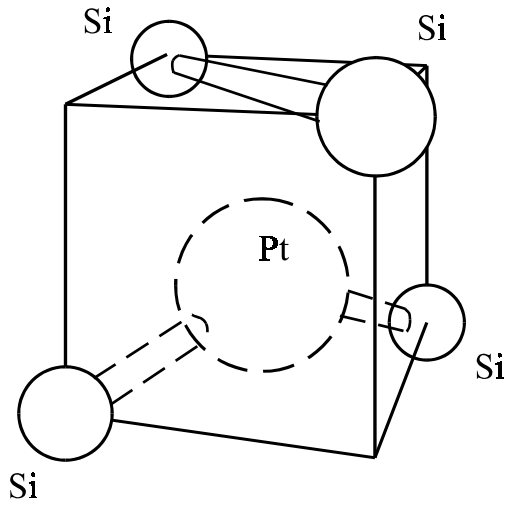

Figure 2.15. A hypothetical distribution of Pt (Pd or Ni) atoms in the Si (and Ge) crystal lattices [38].

purity centers having different signs. The group theoretical analysis included the following algorithm: a classification of one-electron states of molecular orbitals by irreducible point group representations of the local center; symmetry identification of the upper filled molecular orbital; mapping of multielectronic states from the spin-orbitals of the upper filled molecular orbital, and their classification by total spin and the point group symmetry representations of the local center. The result was the conclusion about the existence of the Yan–Teller effect in a unit cell consisting of an impurity atom and four nearest silicon atoms. This effect was found to vary with the charge state of the impurity atom (Table 2.10).

The Yan–Teller type of displacement has been observed in many semiconductor–deep impurity systems. In addition to Si<Os>, Si<Pt>, and Si<Ir>, these are InAs<Mn> [40], InSb<O> [41], GaAs<Cu> [42]. It appears that displacements of atoms, like the Yan–Teller effect or crystal lattice relaxation (Figure 2.16) during ionization and de-ionization are a common phenomenon inherent, to some extent, in all deep level centers. This is due to a strong electron localization at the impurity center, as compared with a shallow impurity. A strongly localized electron interacts only with the nearest host atoms.

It is interesting that an electron of a 3*d*-impurity in gallium arsenide is localized on the second coordination sphere [14], which is clear from the

Table 2.10. Yan–Teller distortions of the Pt-group impurity atoms in a tetrahedral field [39].

Charge state	Total number of electrons	Electronic configuration	Basic term	Distortion
		Rutenium, osmium		
+	11	E^3	^2E	Tetragonal
0	12	E^4	$^2A_1$	No distortion
–	13	$E^4(T_2^*)^2$	$^2T_2$	Trigonal
		Rodium, iridium		
+	12	E^4	$^1A_1$	No distortion
0	13	$E^4(T_2^*)^1$	$^2T_2$	Trigonal
–	14	$E^4(T_2^*)^2$	$^3T_2$	Trigonal
		Platinum, palladium		
+	13	$E^4(T_2^*)^1$	$^1T_2$	Trigonal
0	14	$E^4(T_2^*)^2$	$^3T_1$	Trigonal
–	15	$E^4(T_2^*)^2$	$^4A_2$	No distortion

EPR line broadening $\Delta H/\beta^B < \Delta H/\beta^A$, where β^B and β^A are the magnetic moments of atomic nuclei in the first and second coordination spheres.

Neighboring atoms are not static entities but they oscillate with different frequencies. These oscillations, or phonons, also interact with a deep center. Therefore, an impurity center must be treated in combination with a certain number of host atoms rather than as a single center, or as a multi-atomic quasi-molecule.

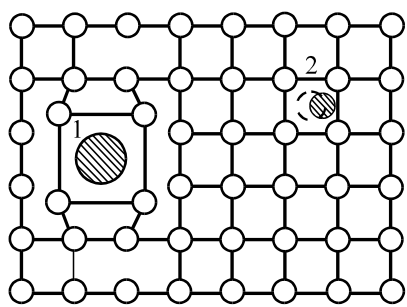

Figure 2.16. Displacements of impurity centers at crystal lattice interstices: 1 – symmetrical distortions; 2 – asymmetrical distortions.

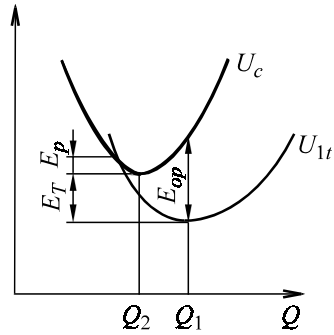

Figure 2.17. The configuration coordinate diagram of a deep center with a strong electron–phonon interaction.

The model center involved in a strong electron–phonon interaction can be conveniently represented as a configuration coordinate plot (Figure 2.17). The ordinate in the plot is the total energy of a quasi-molecule U and the abscissa is a combined coordinate which describes the average difference between the deep center and its neighboring host atoms.

For simplicity, the energy of oscillating atoms can be taken to be that of a harmonic oscillator:

$$U = E_c(Q) + \frac{1}{2} M\omega^2 Q^2 , \qquad (2.2.16)$$

where M is the mass of oscillating atoms and ω is the oscillation frequency of the oscillator.

Then, the curves in Figure 2.17 reflect the parabolic character of the second term in (2.2.16). The upper parabola U_c describes a "molecule" with an ionized impurity center, or an impurity electron in the conduction band. The first term is constant, since atomic oscillations do not affect the energy of free electrons in the conduction band. If the impurity atom is not ionized, the first term in (2.2.16) will also depend on Q. A system tending to have minimum energy must take a new equilibrium position Q_1 different from Q_0; the respective parabola U_{1t} will be different, as is shown in Figure 2.17. One can also see that the optical transition (ionization by light) requires energy E_{op} by virtue of the Franck–Condon principle stating that an optical transition within a system (molecule) occurs without changing the coordinate. The values of thermal ionization energy E_T appear to be lower (Figure 2.17). Note that the curvature of both parabolas must also be

different, since the bonding of an atom to its nearest neighbors is different for an ionized and un-ionized center.

It is clear from the configuration coordinate plot that the necessity to obey the Franck–Condon principle entails a restructuring of the crystallographic configuration in the vicinity of an impurity center when it is ionized or excited, as is indicated in Figure 2.16 by symmetric (lattice relaxation) and asymmetric (Yan–Teller effect) distortions of the unit cell.

The displacement $q = Q_1 - Q_0$ in Figure 2.17 is described as [43]

$$q = a\sqrt{\hbar / M\Omega}\,, \tag{2.2.17}$$

where a is a constant of electron–phonon bonding and Ω is the frequency of the phonon involved in the lattice restructuring.

It is important that Ω does not necessarily characterize the crystal. Rather, it is a local oscillation of an atom together with its nearest neighbors affected by the electron localization.

One can see from (2.2.17) that in the case of hydrogen-like centers, for which electron localization extends to several dozens of lattice periods, a "quasi-molecule" involves hundreds of neighboring atoms, i.e., M is very large and $q \cong 0$. For deep impurity centers, M is small and $q >> 0$.

The value of q is found experimentally by comparing the experimental curve for the optical photoionization cross section $\sigma(h\nu)$ of a deep center and the theoretical expression allowing for the electron–phonon interaction [44]:

$$\sigma(h\nu) = \frac{\sigma_0}{y}\frac{1}{(\pi\Theta)^{1/2}}\int \exp\left[-\frac{(x-y)^2}{\Theta}\frac{(x-1)^{1/2}}{x^2}\right]dx\,, \tag{2.2.18}$$

where

$$x = E/E_{\text{op}}\,, \qquad y = h\nu/E_{\text{op}}\,,$$
$$\Theta = \left(\frac{a\hbar\Omega}{E_{\text{op}}}\right)^2 \text{cth}\,\frac{\hbar\Omega}{2kT}\,. \tag{2.2.19}$$

In addition, we have the relation

$$d_{\text{FC}} = E_{\text{op}} - E_{\text{T}}\,, \qquad d_{\text{FC}} = \frac{1}{2}a^2\hbar\Omega\,. \tag{2.2.20}$$

The value of Θ is found from the comparison of the $\sigma(h\nu)$ curve and (2.2.18); then, a and $\hbar\Omega$ are found from (2.2.19) and (2.2.20), while q is calculated with (2.2.17).

Let us consider two consequences of the crystal lattice distortions in the vicinity of a deep impurity atom.

The appearance of a recombination barrier. It is clear from Figure 2.17 that an electron will overcome the recombination barrier E_p when it goes back from the conduction band to a deep center during recombination. The consequence of this transition is a larger lifetime τ of the electron:

$$\tau = \tau_0 \exp\left(E_p / kT\right), \tag{2.2.21}$$

where τ_0 is the electron lifetime in a crystal containing no deep impurity.

Estimations made with E_p = 0.06 eV give the value $\tau/\tau_0 = 10^4$ at 77 K, which means that the existence of a recombination barrier leads to a tremendous (10^4-fold) gain in lifetime and, hence, to a great increase in the sensitivity of photocells made from such crystals.

It is worth noting that a similar increase in the electron lifetime arises in double doping of a semiconductor by a deep N_T and a shallow N impurity simultaneously [45]. For this, it is necessary that the following relation be valid

$$N_T > N > \left(N_d - N_a\right), \tag{2.2.22}$$

where N_d and N_a are shallow background impurities.

But now, the impurity background becomes practically unimportant, and the uncompensated deep impurity fraction at a concentration ($N_T - N$) will remain essentially un-ionized due to the high ionization energy. The compensation degree of the semiconductor will be close to unity. This is used to obtain semi-insulating samples with a very high electrical resistivity.

If both shallow and deep impurities have a very high solubility in the semiconductor, the crystal will be heavily doped. Heavy doping and strong compensation produce large-scale fluctuations of impurity potential and, hence, corrugated energy bands [46], as shown in Figure 2.18. One can see a recombination barrier arising here, denoted as φ_{rec}. In this case, the electron lifetime will also be defined by expression (2.2.21); but physically, τ_0 means the lifetime in the absence of strong compensation. This effect was observed in GaAs<Cr,Sn> crystals [45]; the details of the theory can be found in [47].

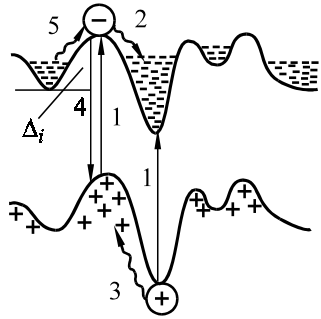

Figure 2.18. Generation (1, 2, 3) and recombination (4, 5) transitions in a heavily doped and strongly compensated (disordered) semiconductor.

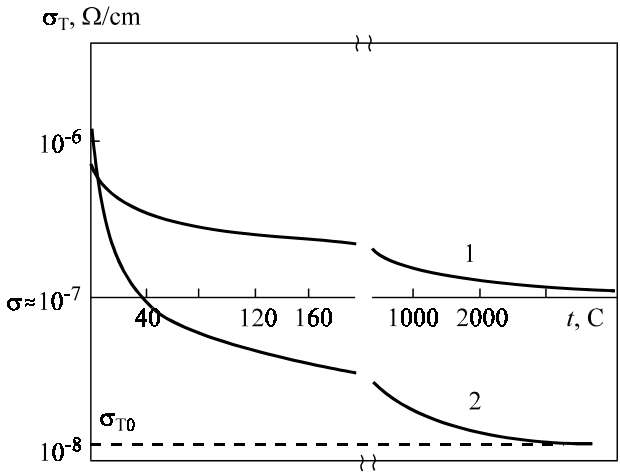

Figure 2.19. Photoconductivity kinetics on illumination in recombination in heavily doped and strongly compensated GaAs<O>: 1 – $h\nu$ = 0.52 eV, 2 – $h\nu$ = 1.1 eV.

Long-term photocurrent relaxation. This effect (Figure 2.19) is associated with the relaxation barrier just described and the longer lifetime of electrons. For this reason, one should choose with caution the type of deep impurity to be used for the doping of crystals for IR- photocells, because it determines the recombination barrier value in a particular semiconductor and the operation temperature of the device. A lower temperature will increase the photocell sensitivity but it will also increase its response time.

To conclude this section, we will discuss basic information on the behavior of *f*-impurities in semiconductors [48, 92]. These impurities are second-phase inclusions in silicon and are produced during crystallization due to a high affinity of rare-earth atoms to oxygen, carbon, and other chemical admixtures to be described in Section 2.5. This is the reason why *f*-atoms are used for gettering the melt to produce crystals with a low impurity background. But the doping of a semiconductor by *f*-impurities becomes more difficult for the same reason. These difficulties do not permit a reliable identification of the energy levels registered experimentally in silicon and other semiconductors containing *f*-impurities.

Samarium impurity. The effect of samarium on silicon is manifested after thermal treatment of Si<Sm> at 1000–1100°C as the appearance of the energy levels $E_c - 0.28$ eV and $E_v + 0.45$ eV at a concentration of 1.5×10^{14} cm^{-3}. Levels $E_v + 0.1$ eV and $E + 0.3$ eV were also identified, but the latter level was found to be unstable, becoming deeper with time.

Gadolinium impurity. It has been found that silicon doping with gadolinium produces the acceptor levels +0.045, +0.34, +0.1 eV counted from the valence band. Their concentration varies from 10^{13} to 5×10^{14} cm^{-3}. After thermal treatment of Si<Gd> in the temperature range 900–1100°C, these levels are transformed to produce other levels unstable in time.

Golmium impurity. The behavior of this impurity in silicon, as far as level instability is concerned, is similar to that of gadolinium. The most stable levels are of the acceptor type found to be $E_v + 0.066$ and $E_v + 0.35$ eV.

Difficulties associated with silicon doping with rare-earth elements from a melt stimulated the application of ion implantation for silicon doping with *f*-impurities. This method provided energy levels for neodime (Nd) and terbium (Tb), equal to $E_c - 0.33$ and $E_c - 0.29$ eV, respectively. But these levels cannot be attributed with certainty to single impurity atoms.

The complicated behavior of *f*-impurities in silicon is mostly associated with the lack of our knowledge about their states in a semiconductor lattice, the solubility thermodynamics, and the migration mechanism. As a result, it is still difficult to find applications for *f*-impurities. However, there is a report [49] of the application of Si<Er> for the production of efficient silicon light diodes operating at the light wavelength of 1.54 μm at 300 K. This is an encouraging result showing that research into the behavior of *f*-impurities in various semiconductors may become more active.

Among III-V semiconductors doped with *f*-impurities, GaAs and InP doped with Er and Yb have been studied most intensively. These studies are focused on recombination processes involving these impurities, especially on Auger recombination [48].

2.3 AMPHOTERIC IMPURITIES

2.3.1 General concepts

The concept of amphoteric impurity, as applied to semiconductors, was first introduced by Dunlap in 1955, when analyzing electrical properties of germanium doped with gold. It was found that gold could behave as a donor or as an acceptor in the same semiconductor crystal. This behavior fully agrees with the definition of amphoteric behavior as the ability to produce positive and negative ions.

Note that a sequence of two processes—ionization and de-ionization—is always possible in semiconductors. In the former, an electron is detached and in the latter it is re-attached. But de-ionization neutralizes an impurity center, which comes to equilibrium without charge reversal. It is clear then that amphoteric impurities are impurities with charge Z_0 prior to ionization (most often, $Z_0 = 0$), which can acquire, depending on the ionization conditions, charges $Z_1 = Z_0 + |\Delta Z|$ and/or $Z_2 = Z_0 - |\Delta Z|$, where $|\Delta Z|$ is the absolute charge change in the ionization (most often, $|\Delta Z| = 1$).

According to up-to-date chemical concepts, any element may become amphoteric, since atoms can give off their electrons but they can also attach electrons because of a certain affinity to the electron. Indeed, as far back as 1881, Helmholtz suggested that an atom might possess different charges in different compounds. A typical example is hydrogen charged differently in LiH and HCl, whose decomposition reactions in electrolysis are $LiH \rightarrow Li^+ + H^-$ and $HCl \rightarrow Cl^- + H^+$.

There are many examples of this kind in chemistry. But if we consider the amphoteric nature of an impurity in a semiconductor crystal in terms of its behavior in the same crystal, it will appear that such impurities are not numerous but their number is large enough to regard them into a special class of impurities. Available data permit classification of all amphoteric impurities by the crystallochemical principle, i.e., by their arrangement in the crystal lattice (Table 2.11).

Impurity atoms that can behave as donors and as acceptors in one of the positions—a site or interstice—will be referred to the first two types and termed amphoteric site centers and amphoteric interstitial centers—A_s and A_i, respectively. Impurity atoms located at sites are acceptor-type and those at interstices are donor-type; they are the third type of site/interstitial or dissociative amphoteric centers denoted as A_{si}.

Table 2.11. Amphoteric impurity centers in semiconductors (*D* is for a donor and *A* is for an acceptor).

Type of amphoteric center	Position in crystal	Center symbol	Type of center
Site	Site	A_s	*D* and *A*
Interstitial	Interstice	A_i	*D* and *A*
Dissociative (site/interstice)	Site, interstice	A_{si}	*D*
Cation–anion	Cation site, anion site		*A*
Associative	Site or interstice in interaction with other point defects	A_{ca} A_a	*D* *A* *D* and *A*

It is necessary to stress here that, in principle, there may be such dissociative amphoteric centers which manifest a donor behavior at sites and an acceptor behavior at interstices. Experimentally, this was observed only for an interstitial impurity in UO_2 [50].

The fourth type of impurity centers manifest amphoteric properties in the substitution of different sublattices in semiconductor compounds. These will be termed cation–anion centers and denoted as A_{ca}.

Finally, amphoteric impurities, as any other impurities, may produce complex associative defects in interactions with one another or with other point defects. Moreover, there may be associates of point defects in a crystal, which show an amphoteric behavior, whereas the individual components have no amphoteric properties. Complex amphoteric centers produced by interactions of various point defects will be called associative amphoteric centers and denoted as A_a.

The first three types of centers manifest their amphoteric properties in elemental semiconductors, such as Ge and Si, and all five types are found in $A^{III}B^{V}$ and $A^{II}B^{VI}$ semiconductor compounds.

The first two types show their amphoteric nature in one and the same position in a crystal. Here, an impurity center has a chance to give off an electron from one of its electron shells and contribute it to the electron ensemble of a host atom, or, on the contrary, an electron can be captured by this shell. Obviously, both possibilities can be realized by an atom with the initial charge state Z_0 only if its operating electron shell is uncompleted. These may be atoms of *d*- and *f*-elements only, and this idea is supported by experimental data. In many situations, transition metal impurities do not show amphoteric properties. Doping *f*-atoms are often electrically inactive.

The questions as to whether *d*- and *f*-atoms will show an electrical activity in this or that semiconductor and which of them will turn out to be amphoteric require a special analysis. The answers cannot be derived from the trivial consideration of donor and acceptor properties of impurity atoms, based on the similarity of geometry and electrochemical characteristics of the impurity atom and the position it occupies in the semiconductor. According to these concepts, the position and charge of an impurity atom are determined by the similarity in the radii and negative charge values of the impurity atom and the host atom. None of these criteria are applicable to transition metal atoms, because these two parameters are variable in them [14].

Dissociative amphoteric impurities often occur in elemental semiconductors. This has been pointed out in many publications on impurity levels in A^{IV}, $A^{III}B^{V}$, and $A^{II}B^{VI}$ semiconductors. It is clear from general principles that equal probability for an impurity atom to occupy a site or an interstice requires approximately equal energies for its incorporation into both positions. But the calculation of these energies encounters difficulties associated with the different interactions of an impurity atom with the neighboring host atoms. At a site, an impurity atom produces chemical bonds with the nearest host atoms via a hybridization of electrons from both types of atoms. At an interstice, electrons of the host atoms only repel those of an impurity atom, producing no chemical bonds between the host and impurity atoms.

In these cases, the concepts of radii and negative electrical charges are inapplicable to most impurity atoms as constant atomic characteristics. For this reason, the calculation of energy necessary for an impurity atom to occupy a site or an interstice and, hence, the prediction of dissociative amphoteric impurity centers may be possible only in terms of a rigorous theory of impurity solubility at sites and interstices. The current state of this problem and the available approaches to its solution, including the author's conception, will be discussed in Chapter 4.

Cation–anion amphoteric centers can occur only in semiconductor compounds.

An impurity atom can be localized in different sublattices if it has an intermediate valence relative to the valences of the other constituents. This feature is especially characteristic of group-IV atoms in $A^{III}B^{V}$ semiconductors. In this case, an A^{IV} atom exhibits donor properties when substituting a cation A^{III} site, while an anion B^{V} site shows acceptor properties. The well-known valence rule $\Delta V = \pm 1$ becomes valid, as for hydrogen-like substitutional impurities.

2.3.2 Carrier thermodynamics in semiconductors with amphoteric impurities

Two problems arise in the treatment of amphoteric impurity states in semiconductors. In one problem, the unknown quantity is solubility [A] which varies, other conditions being equal, with the charge state distribution of the amphoteric impurity (0, "+", "–") in a crystal. In the other problem, the total concentration of the amphoteric impurity [A] is taken to be known while the unknown quantity is its charge state distribution.

Therefore, both problems require the knowledge of the statistical charge state distribution of an amphoteric impurity. This issue has been treated by several workers with reference to various particular cases [51] and was generally considered by Shockley and Hast [52].

Figure 2.20 shows model energy levels corresponding to an amphoteric impurity, and Table 2.12 gives its charge states Z, the donor and acceptor behavior as a function of concentration N and ionization energy E, as well as the number of excessive electrons r in an impurity atom, as compared with the neutral state with $r = 0$. The indices m and n stand for the last impurity atom states localized in the forbidden gap. This means that if there are $(m + 1)$ or $(n + 1)$ states, their ionization energies will lie in the allowed spectrum (Figure 2.20) and become unobservable, at least, in conventional Hall measurements.

An important feature of the model represented in Figure 2.20 is that all levels belong to the same impurity center. This means that the sequence of amphoteric impurity levels cannot be arbitrary. Indeed, a center must first give off one electron, spending for this energy E_{d1}, and only after that can it

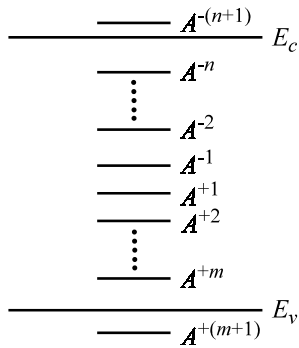

Figure 2.20. Energy levels of amphoteric impurities in various charge states.

Table 2.12. Symbols for amphoteric impurity atoms in various charge states (see Figure 2.20).

Impurity atom states	Ionization energy of level, ${}^2E_i$	Concentration, N	Number of excess electrons, r	Impurity ion charge, Z
A^{n+}	$-E_{dm}$	N_{dm}	$-m$	$+m$
⋮	⋮	⋮	⋮	⋮
A^{2+}	$-E_{d2}$	N_{d2}	–2	+2
A^{1+}	$-E_{d1}$	N_{d1}	–1	+1
A^0	–	${N_A}^0$	0	0
A^{1-}	$+E_{a1}$	N_{a1}	+1	–1
A^{2-}	$+E_{a2}$	N_{a2}	+2	–2
⋮	⋮	⋮	⋮	⋮
A^{n-1}	$+E_{an}$	N_{an}	$+n$	$-n$

Plus means that the energy is counted from the valence band top E_v and minus from the conduction band bottom E_c.

give off another electron, spending energy $E_{d2} > E_{d1}$. Similarly, the transition of one electron to the level A^{-1} requires energy E_{a1}, and this electron will repel the next one, resulting in $E_{a2} > E_{a1}$. Besides, the electron attachment requires a higher energy than the electron detachment because of Coulomb repulsion, i.e., $E_a > E_d$. As a result, the level localization pattern for an amphoteric center and, generally, for any multivalent center, must be such as is shown in Figure 2.20. This general rule is, of course, valid for site and interstitial amphoteric impurities, which exhibit amphoteric properties when the impurity atom occupies one crystallochemical position. For other amphoteric impurities, occupying different positions, the condition $E_a > E_d$ may not be satisfied with the relations

$$E_{dm} > \ldots E_{d2} > E_{d1} \qquad E_{an} > \ldots E_{a2} > E_{a1}. \tag{2.3.1}$$

The interrelation of several energy levels of the same impurity atom is an obstacle to finding their electron populations with the simple Fermi function. This requires the use of an expression derived from a more general Gibbs distribution, which relates the population on the next r level to that on the previous $(r–1)$ level, as

$$\frac{N_r}{N_{r-1}} = \frac{g_{r-1}}{g_r} \exp\left(\frac{\mu_e - E_i}{kT}\right), \tag{2.3.2}$$

where μ_e is electron chemical potential (Fermi level) and g is the g-factor representing the spin degeneration degree of the levels. It follows from simple considerations that g may be equal to 1 or 2. If an impurity center captures an electron, the latter can occupy one of the two possible spin states (spin up or spin down); then $g_r = 2$ and $g_{r-1} = 1$. If, on the contrary, an electron is detached from an impurity center, then $g_r = 1$ and $g_{r-1} = 2$. Such simple situations are seldom feasible, so one has to find g_r and g_{r-1} from experiments. Below, we will preserve the subscripts for the g-factors. With (2.3.2) and the symbols in Table 2.12, we will represent the electron population for the first acceptor level A^{-1}:

$$N_{a1} = \frac{g_0}{g_{a1}} N_0 \exp\left(\frac{\mu - E_{a1}}{kT}\right). \tag{2.3.3}$$

The population of the second acceptor level A^{-2} will be

$$N_{a2} = \frac{g_{a1}}{g_{a2}} N_{a1} \exp\left(\frac{\mu - E_{a2}}{kT}\right). \tag{2.3.4}$$

Or, substituting (2.3.3) into (2.3.4), we will have

$$N_{a2} = \frac{g_0}{g_{a2}} N_0 \exp\left(\frac{\mu - E_{a1}}{kT}\right) \exp\left(\frac{\mu - E_{a2}}{kT}\right). \tag{2.3.5}$$

Similarly, we find

$$N_{an} = \frac{g_0}{g_{an}} N_0 \prod_{n=1}^{n} \exp\left(\frac{\mu - E_{an}}{kT}\right). \tag{2.3.6}$$

For the donor states, the preceding level will be the ($r = -1$) level and the next one will be the ($r = 0$) level. Therefore, with (2.3.2) and the symbols of Table 2.12, we will find

$$N_0 = \frac{g_{d1}}{g_0} N_{d1} \exp\left(\frac{\mu - E_{d1}}{kT}\right), \tag{2.3.7}$$

from which we have

$$N_{d1} = \frac{g_0}{g_{d1}} N_0 \exp\left(\frac{E_{d1} - \mu}{kT}\right). \qquad (2.3.8)$$

Using similar procedures, we can find

$$N_{d2} = \frac{g_0}{g_{d2}} N_0 \exp\left(\frac{E_{d1} - \mu}{kT}\right)\exp\left(\frac{E_{d2} - \mu}{kT}\right) \qquad (2.3.9)$$

and

$$N_{dm} = \frac{g_0}{g_{dm}} N_0 \prod_{m-1}^{m} \exp\left(\frac{\mu - E_{dm}}{kT}\right). \qquad (2.3.10)$$

These are general expressions for the electron populations on the donor and acceptor levels of an amphoteric impurity center. Therefore, the sums $N_{a1} + N_{a2} + ... + N_{an}$ and $N_{d1} + N_{d2} + ... + N_{dm}$ enter the neutrality equation (1.2.31). But since a crystal may contain other impurity atoms or intrinsic point defects, the more general sums $\sum_k \left[\ \right]_k^-$ and $\sum_q \left[\ \right]_q^+$ appear in expression (1.2.31).

By substituting these sums into (1.2.31) taking account of (2.3.6) and (2.3.10) and using n and p expressed through μ:

$$n = N_c e^{\mu/kT} \quad \text{and} \quad p = N_v e^{\left(E_g - \mu\right)/kT}, \qquad (2.3.11)$$

we will get the neutrality equation with one unknown quantity μ (the g-factors are taken to be known). By solving this neutrality equation, we find μ; hence, the solubility of an amphoteric impurity can be found from (2.3.6) and (2.3.10). Or, conversely, if we know the total concentration N_a of an amphoteric impurity, we can find its charge state distribution from the values of μ. The general solution to equation (1.2.31) can be obtained only by numerical computations. For some particular cases, when the amphoteric impurity levels are much separated and the Fermi level lies between two neighboring levels, equation (1.2.31) is simplified, and the problem becomes similar to that for independent impurity centers [51]. Then, analytical solutions to the neutrality equation are also possible. Using these procedures, we will now analyze the states and behavior of some amphoteric impurities.

The statistics of charged states discussed above yield the basic relation for the charged state concentrations in (2.3.2) and general expressions for the concentrations N^0, N^-, and N^+ in (2.3.7), (2.3.3), and (2.3.8), respectively. In

other words, we have obtained the site and interstitial solubilities of a charged impurity as a function of the Fermi level position. For this, of course, the energy levels and degeneracy factors must be known.

2.3.3 Amphoteric impurity distribution in elemental semiconductors

Of interest is the relation between the measured quantities n and A. The former is found from the Hall coefficient and the latter by a direct method, say, by atom labeling, radioactive analysis, mass-spectrometry, etc.

This relation can be derived from the neutrality equation

$$n + V_s^- + A_s^- = p + A_I^+ . \tag{2.3.12}$$

Consider first a particular case when charged vacancies are absent:

$$A_I^+ - A_s^- = n - p . \tag{2.3.13}$$

Since all amphoteric centers are ionized at high temperatures, we have

$$A_I^+ + A_s^- = A . \tag{2.3.14}$$

By dividing (2.3.14) by (2.3.13), we get

$$A = (n - p)\frac{1 + K_A (n/n_i)^2}{1 - K_A (n/n_i)^2} . \tag{2.3.15}$$

In this expression, the values of n and p refer to high temperature, but usually they are measured at room temperature denoted here as T_0. Then, the carrier concentrations will be denoted as n_0 and p_0.

Let us assume that the distribution of an amphoteric impurity becomes "frozen" at high temperature T close to the doping temperature. The validity of this statement increases with decreasing diffusion coefficient of an impurity in a solid crystal at this temperature. With this assumption, the neutrality equation (2.3.13) remains unchanged at room temperature T_0, i.e.,

$$n_0 - p_0 = n - p . \tag{2.3.16}$$

For further analysis, we must choose the type of conductivity of the semiconductor containing an amphoteric impurity. For the *n*-type conductivity, the conditions are

$$n_0 >> p_0, \qquad n > p, \qquad K_A = A_s^- \big/ A_I^+ < 1, \tag{2.3.17}$$

and for the *p*-type conductivity, these are

$$p_0 >> n_0, \qquad p > n, \qquad K_A = A_s^- \big/ A_I^+ > 1. \tag{2.3.18}$$

The treatment is simplified and generalized if, instead of $K_A = A_s^-/A_I^+$, we take for *p*-type semiconductors the inverse quantity A_I^+/ A_s^-, denoting it K_A, as before. Then, we have $K_A < 1$, irrespective of the type of conductivity. Which of the true values of K_A must be taken into account in a particular case will be clear from the context.

By solving (2.3.16) separately for *n* or *p* with the conditions (2.3.17) and (2.3.18) and the equality $np = n_i^2 = p_i^2$, we get

$$n = \frac{1}{2}\left[n_0 + \sqrt{n_0^2 + 4n_i^2} \right] (n\text{-type}), \tag{2.3.19}$$

$$p = \frac{1}{2}\left[p_0 + \sqrt{p_0^2 + 4n_i^2} \right] \quad (p\text{-type}). \tag{2.3.20}$$

The substitution of these expressions into (2.3.15) gives the final expressions for *n*-type crystals

$$A = n_0 \frac{1 + \dfrac{K_A}{4}\left[\dfrac{n_0}{n_i} + \sqrt{4 + \left(\dfrac{n_0}{n_i} \right)^2} \right]^2}{1 - \dfrac{K_A}{4}\left[\dfrac{n_0}{n_i} + \sqrt{4 + \left(\dfrac{n_0}{n_i} \right)^2} \right]^2}, \tag{2.3.21}$$

and *p*-type crystals

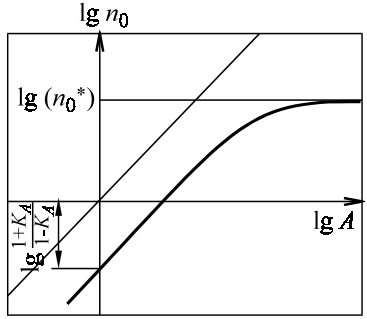

Figure 2.21. Electron concentration versus total amphoteric impurity concentration at room temperature.

$$A = p_0 \frac{1+\frac{K_A}{4}\left[\frac{p_0}{n_i}+\sqrt{4+\left(\frac{p_0}{n_i}\right)^2}\right]^2}{1-\frac{K_A}{4}\left[\frac{p_0}{n_i}+\sqrt{4+\left(\frac{p_0}{n_i}\right)^2}\right]^2}. \qquad (2.3.22)$$

Note that n_0 and p_0 change at room temperature and n_i at doping temperature.

The sought for dependence $n_0(A)$ in the form of (2.3.21) was obtained by the authors of [53]. The dependence presented in Figure 2.21 will be discussed qualitatively, as in [53]. The key factor here is the ratio n_0/n_i, or p_0/n_i for p-type crystals.

For low carrier concentrations

$$n_0/n_i << 1 \quad (\text{or } p_0/n_i << 1), \qquad (2.3.23)$$

the relationship between A and n_0 (or p_0) is linear:

$$A = n_0\left(1+K_A\right)/\left(1-K_A\right), \qquad (2.3.24)$$

$$A = p_0\left(1+K_A\right)/\left(1-K_A\right). \qquad (2.3.25)$$

In these expressions, K_A for n- and p-type conductivities represent inverse ratios of the site and interstitial components of an amphoteric impurity.

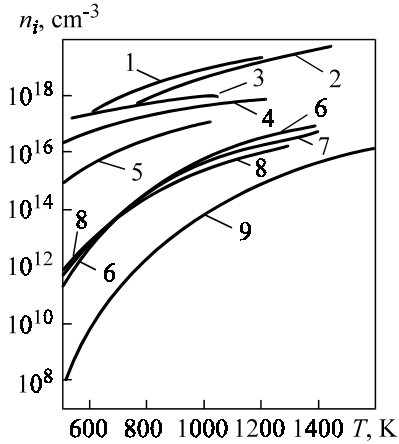

Figure 2.22. Regions of constant chemical potential of electrons (below the $n_i(T)$ curves) in doped semiconductors: 1 – Ge; 2 – Si; 3 – InSb; 4 – InAs; 5 – GaSb; 6 – AlSb; 7 – GaAs; 8 – InP; 9 – GaP.

In order to assess the feasibility of the condition in (2.3.23), Figure 2.22 gives the temperature dependence $n_i(T)$ for basic semiconductors. At $K_A << 1$, the ratios (2.3.24) and (2.3.25) are close to the equality $A = n_0$ ($A = p_0$). Or, from Figure 2.21 we find that, at low concentrations, an amphoteric impurity occupies primarily one of the possible crystallochemical positions. At $K_A \leq 1$, the curve is shifted away from $A = n_0$ (see Figure 2.21), and the empirical value of K_A in (2.3.24) or (2.3.25) can be found from the shift along the lg n_0 (or lg p_0) axis.

In the range of high concentrations n_0 (or p_0), the $n_0(A)$ and $p_0(A)$ curves go beyond the saturation region. The concentrations n_0^* or p_0^* in this region are found when the denominator in (2.3.21) and (2.3.22) vanishes:

$$n_0^* \left(\text{or } p_0^*\right) = n_i \left(\frac{1}{\sqrt{K_A}} - \sqrt{K_A} \right). \qquad (2.3.26)$$

The conditions $K_A << 1$ and $K_A \leq 1$ may become practically feasible at high concentrations n_0 (or p_0). The first condition is valid for an extensive concentration range, for which $(n_0/n_i) >> 1$. Then, (2.3.21) is simplified as

$$A = n_0 \frac{1 + K_A (n_0/n_i)^2}{1 - K_A (n_0/n_i)^2} . \qquad (2.3.27)$$

A similar expression with p_0 and K_A will describe a *p*-crystal. The second condition (at $K_A < 1$) contains no $n_0/n_i >> 1$ regions in *n*- or *p*-type crystals, and the experimental $A(n_0)$ and $A(p_0)$ curves are to be treated using general formulas (2.3.22) and (2.3.23).

Return now to the neutrality equation (2.3.12) which ignored the presence of vacancies. To take them into account, one should merely suggest that vacancies are univalent acceptors and that their concentration at room temperature is equal to the equilibrium concentration at a high doping temperature (*T*). The latter does not occur in actual reality, but this suggestion allows us to elucidate qualitatively the effect of vacancies. A more rigorous theory will be considered below. The neutrality equation for a *n*-type crystal at room temperature will then appear as

$$n_0 + V^- = A_I^+ - A_s^- , \qquad (2.3.28)$$

while the equation for the $A(n_0)$ curve will be

$$A = \left(n_0 + V^-\right) \frac{1 + K_A (n/n_i)^2}{1 - K_A (n/n_i)^2} . \qquad (2.3.29)$$

Expression (2.3.29) will be valid for a *p*-type crystal, if n_0 is replaced by p_0, n by p, and K_A is understood as the A_I^+/A_s^- ratio. Compensation by vacancies will not change the curve shape in the range of high concentrations *A* or the value of n_0 on the shoulder (2.3.26). However, for the range of low concentrations $n_0/n_i << 1$, we will have, by analogy with (2.3.24), the expression

$$A = \frac{\left(n_0 + V^-\right)\left(1 + K_A\right)}{1 - K_A} , \qquad (2.3.30)$$

in which

$$V^- = V^0 \frac{N_v}{n_i} \exp\left(-\frac{\Delta E_a}{T}\right) \qquad (2.3.31)$$

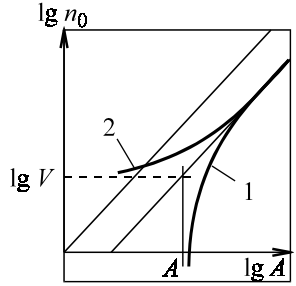

Figure 2.23. The $n_0(T)$ dependence in the low concentration region, n_0, in the presence of foreign acceptor (1) and donor (2) compensation.

is independent of n_0 .

It is clear from (2.3.30) that there must be a deviation from the linear function $A(n_0)$ in the region of comparable concentrations n_0 and V^-. This is illustrated by curve 1 in Figure 2.23. The asymptote, to which the curve tends in the range of small n_0 values, permits finding an "arbitrary" concentration A^* defined as

$$A^* = V^- \frac{1+K_A}{1-K_A}, \tag{2.3.32}$$

while the length n_0^* along the ordinate shows the concentration V^-.

It should be noted for the sake of generality that if vacancies have a donor character, i.e., V^+ are positively charged, the $A(n_0)$ function will look like curve 2 in Figure 2.2. In this case, the ordinate will show the concentration V^+, and A^* will look exactly as in (2.3.32) with V^+ instead of V^-. The opposite is true of *p*-type crystals.

The behavior of an amphoteric impurity in a real crystal is also complicated by the fact that, in addition to vacancies, there are background or deliberately introduced donors and acceptors in it: N_d and N_a. So we generalized the above simple theory on the following assumptions [54]. The concentrations N_d and N_a were considered to be constant throughout the process of crystal cooling from the doping temperature. Vacancies were assumed to be so mobile that their concentration followed the temperature change to become "frozen" at a certain temperature T^*, which was lower than the doping temperature T but higher than room temperature T_0, at which n_0 (or p_0) and A were measured.

The value of T^* varies with the cooling rate and approaches T as the rate increases. Besides, we suggested in [54] a complete ionization of vacancies and all impurity centers present in the crystal.

With the above assumptions, the neutrality equations corresponding to the temperatures T, T^*, and T_0 are

$$n + V^- - n_i^2 / n = A_I - A_s + N_d - N_a , \tag{2.3.33}$$

$$n^* + V^{*-} - n_i^{*2} / n^* = A_I - A_s + N_d - N_a , \tag{2.3.34}$$

$$n_0 + V^{*-} = A_I - A_s + N_d - N_a . \tag{2.3.35}$$

In these equations, we have used the condition $pn = n_i^2$ but ignored the concentration p_0 in (2.3.35) because we are considering a n-type crystal. The concentrations of charged vacancies in the neutrality equations can be written as [55]

$$V^- = V_i(P_{sm}, T) n(T) / n_i(T) , \tag{2.3.36}$$

where $V_i(P_{sm}T)$ is the equilibrium concentration of vacancies in an intrinsic semiconductor. It should also be taken into account that the intrinsic concentration n_i decreases quickly with decreasing temperature (see Figure 2.22); therefore, we can take $n_0 >> 2n_i^*$ at T^* and, hence, put

$$n^* = n_0 . \tag{2.3.37}$$

With these assumptions, the relation between n and n_0 has been found to be

$$n = \frac{1}{2}\left[n_0 \frac{1+x^*}{1+x} + \sqrt{n_0^2 \left(\frac{1+x^*}{1+x} \right)^2 + \frac{4n_i^2}{1+x}} \right], \tag{2.3.38}$$

with the denotations

$$x = V_i / n_i \quad \text{and} \quad x^* = V_i^* / n_i^* \tag{2.3.39}$$

and the sought for function $n_0(A)$ as

$$A = \left[N_a - N_d + n_0\left(1 + x^*\right)\right]\frac{1 + K_A\left(n/n_i\right)^2}{1 - K_A\left(n/n_i\right)^2}, \qquad (2.3.40)$$

where n is defined by equality (2.3.38).

It is easy to see that with small and large values of n_0, the $A(n_0)$ functions will transform, respectively, to

$$A = \left[n_0\left(1 + x^*\right) + N_a - N_d\right]\frac{1 + x + K_A}{1 + x - K_A} \qquad (2.3.41)$$

and

$$\left(n_0\right)_{sh} = \frac{n_i}{1 + x^*}\left(\frac{1 + x}{\sqrt{K_A}} - \sqrt{K_A}\right). \qquad (2.3.42)$$

The complete $n_0(A)$ function is represented by curve 1 in Figure 2.24. The curve shift relative to the bisectrix $n_0 = A$ is

$$b = \log\left(1 + x^*\right)\left(1 + x + K_A\right)/\left(1 + x - K_A\right). \qquad (2.3.43)$$

The asymptotic value of A^* will be

$$A^* = \left(N_a - N_d\right)\left(1 + x + K_A\right)/\left(1 + x - K_A\right). \qquad (2.3.44)$$

The ordinate length in this case is expressed as

$$n_0^* = \left(N_d - N_a\right)/\left(1 + x^*\right). \qquad (2.3.45)$$

If the vacancies are positively charged (V^+), the $n_0(A)$ function will look as curve 2 in Figure 2.24.

The $p_0(A)$ function for a p-type crystal will be represented by the same curves as in Figure 2.24, but one should bear in mind that n_0 should be replaced by p_0, N_a by N_d, and, conversely, N_d by N_a. The quantity K_A should be understood as A_I^+/A_s^-.

The theory considered demonstrates a qualitative similarity of the results (Figures 2.23 and 2.24), irrespective of the presence of this or that type of defect in the crystal. The type of defect, however, essentially determines the

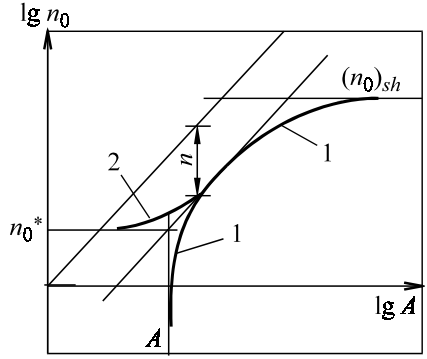

Figure 2.24. The $n_0(A)$ dependence in a rigorous theory allowing for the compensation by foreign point defects.

shape of the $n_0(A)$ curve, or the quantitative parameters A^*, b, and $(n_0)_{sh}$. Obviously, for theoretical and experimental $n_0(A)$ curves to be comparable, the equilibrium conditions at T^* must be met. So measurements must be made in annealed samples at preset P_{sm} and T followed by hardening; then, $T^* = T$. The presence of numerous point defects which do not correspond to this pressure and temperature can make the interpretation of experimental data impossible.

If a direct experimental measurement of the total concentration A of an amphoteric impurity turns out to be too complicated, one can find the total concentration of all ionized centers N_i, determined from charge carrier mobility measurements. The function to be found from the theory [55] is

$$N_i = n_0\left[x^* + \left(1+x^*\right)\frac{1+x+K_A}{1+x-K_A}\right] + \frac{2\left[N_A\left(1+x\right) - N_d K_A\right]}{1+x-K_A}. \qquad (2.3.46)$$

This function is illustrated by curves 1 and 2 in Figure 2.25 for n-type crystals containing V^- and V^+, respectively. The concentration values on the shoulders of curves 1 and 2 are defined by the same expression (2.3.42) as for the curves in Figure 2.24. The shift of the b curves (Figure 2.25) relative to the bisectrix, $n_0 = N_i$, is

$$b = \log\left[x^* + \left(1+x^*\right)\frac{1+x+K_A}{1+x-K_A}\right]. \qquad (2.3.47)$$

The asymptotic value of N_i^* in this case is described as

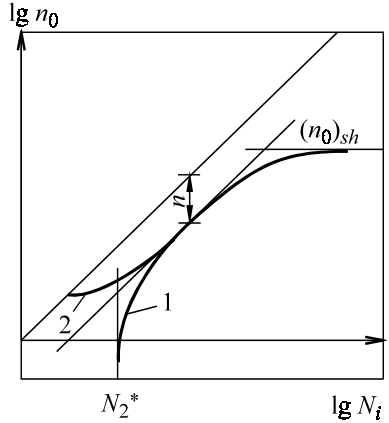

Figure 2.25. The concentration dependence of charge carriers on ionized centers in semiconductors with an amphoteric impurity and compensating centers.

$$N_{\mathrm{i}}^{*} = 2\frac{N_{\mathrm{a}}(1+x) - N_{\mathrm{d}}K_{\mathrm{A}}}{1+x-K_{\mathrm{A}}}, \qquad (2.3.48)$$

and the length on the ordinate has no longer a clear physical meaning. This is easy to see if one makes N_{i} in (2.3.46) tend to zero.

For a *p*-type crystal, expressions (2.3.46) through (2.3.48) are valid, with appropriate substitutions similar to those just described for $n_0(A)$.

2.3.4 Amphoteric impurity distribution in semiconductor compounds

The thermodynamic theory discussed in the previous sections can be extended to semiconductor compounds. Consider first a binary semiconductor of the A_1B_1 type, such as the commonly used $A^{III}B^{V}$ and $A^{II}B^{VI}$ crystals. For simplicity, we will ignore for the time being the possible interstitial dissolution of an amphoteric impurity in both sublattices. In other words, the impurity atoms will be assumed to occupy only sites A_{A_1} and A_{B_1}. It is also necessary to choose the compound constituent possessing a greater volatility. For the semiconductors mentioned, such a constituent is B_1 (As, P, Sb, etc.).

Therefore, the expressions for an A_1B_1 crystal are those of the *A* distribution between the sublattices, which describe the functions $n_0(A)$ and $n_0(N_i)$

found in Section 2.3.3, but with a different physical sense of the constants K_A, which can now be presented as:

Type of crystal	A_1	A_1	A_1B_1	A_1B_1
Type of conductivity	n	p	n	p
K_A	A_S^- / A_I^+	A_I^+ / A_S^-	$A_{B_1}^- / A_{A_1}^+$	$A_{A_1}^+ / A_{B_1}^-$

In addition to binary semiconductors, modern semiconductor electronics widely use multicomponent solid solutions on their base.

Let us discuss the distribution of an amphoteric impurity between the sublattices, with reference to $A^{III}B^{V}$ solid solutions [56]. In contrast to the other sections of this book, the component number will be denoted here by a superscript and its content, expressed as a mole fraction, by a subscript. Thus, we will have a pure (undoped) solid solution with the anion substitution

$$A^1B^1_{1-x-y-\ldots-z}B^2_xB^3_y\ldots B^n_z\,,$$

containing n components of group-V elements, and an undoped solid solution with the cation substitution

$$A^n_z\ldots A^3_yA^2_xA^1_{1-x-y-\ldots-z}B^1\,,$$

containing n components of group-III elements (superscripts III and V are omitted for simplicity).

The further treatment requires the following assumption to be made:

$$x+y+\ldots+z<1\,. \qquad (2.3.49)$$

In fact, this means that the sum of all additional components is a small value in the A^1 or B^1 sublattices. In other words, we assume that only single vacancies of components A^1 and B^1 with concentrations V_{A^1} and V_{B^1}, respectively, may be intrinsic point defects in multicomponent solutions. Component B^1 is considered, as before, to be volatile.

These assumptions allow us to write simple equations, as in Section 2.3.2, for processes reflecting point defect equilibrium inside a solid solution

$$V^0_{A^1}+V^0_{B^1}\rightleftarrows 0 \qquad (2.3.50)$$

and, with volatile B_1 vapor, outside it

$$mV^0_{B^1} + B^1_m \rightleftarrows mB^1_{B^1}, \tag{2.3.51}$$

where m is the number of atoms in a B^1 molecule in the gas phase.

Equations (2.3.50) and (2.3.51) should be supplemented by the condition of an equal number of sites in both sublattices. For an anion solution, this condition is

$$A^1_{A^1} + V_{A^1} \rightleftarrows B^1_{B^1} + B^2_{B^1} + B^3_{B^1} + \ldots + B^n_{B^1} + V_{B^1}, \tag{2.3.52}$$

where $V_{A^1} = V^0_{A^1} + V^-_{A^1}$ and $V_{B^1} = V^0_{B^1} + V^-_{B^1}$, i.e., we consider the vacancies in both sublattices to be of the acceptor type. Note that the knowledge of the exact charge sign of intrinsic defects is of no importance to our presentation, since we are discussing a general approach; the details are to be established from experimental data.

Equations (2.3.5) and (2.3.51) satisfy the following relations between the chemical potentials:

$$\mu\left(V^0_{A^1}\right) + \mu\left(V^0_{B^1}\right) = 0, \tag{2.3.53}$$

$$m\mu\left(V^0_{B^1}\right) + \mu\left(B^1_m\right) = m\mu\left(B^1_{B^1}\right). \tag{2.3.54}$$

The low concentrations of point defects in common sublattices suggest their solution in a crystal to be a perfect dilute solution. Then, we can write

$$\mu\left(V^0_{A^1}\right) = g\left(V^0_{A^1}\right) + T\ln\frac{V^0_{A^1}}{A^1_{A^1} + V_{A^1}}, \tag{2.3.55}$$

$$\mu\left(V^0_{B^1}\right) = g\left(V^0_{B^1}\right) + T\ln\frac{V^0_{B^1}}{B^1_{B^1} + B^2_{B^1} + \ldots B^n_{B^1} + V_{B^1}}. \tag{2.3.56}$$

The gas phase, too, is a perfect gas, and its chemical potential is

$$\mu\left(B^1_m\right) = \mu^0(T) + T\ln P_{(B^1)m}. \tag{2.3.57}$$

The substitution of expressions (2.3.55) through (2.3.57) into (2.3.53) and (2.3.54) yields

$$\frac{V^0_{A^1}V^0_{B^1}}{\left(A^1_{A^1}+V_{A^1}\right)\left(B^1_{B^1}+V_{B^1}+B^2_{B^1}+\ldots B^n_{B^1}\right)}=\exp\left[-\frac{g\left(V^0_{A^1}\right)+g\left(V^0_{B^1}\right)}{T}\right]\equiv K_1'(T), \tag{2.3.58}$$

$$\frac{V^0_{B^1}P^{1/m}_{B^1}}{B^1_{B^1}+V_{B^1}+B^2_{B^1}+\ldots+B^n_{B^1}}=\exp\left[-\frac{g\left(V^0_{B^1}\right)+\mu^0(T)/m-\mu\left(B^1_{B^1}\right)}{T}\right]\equiv K_2'(T). \tag{2.3.59}$$

In the first approximation, the energy of defect formation can be considered to be dependent only on temperature:

$$g\equiv g(T). \tag{2.3.60}$$

The composition of a solid solution defines chemical potential $\mu\left(B^1_{B^1}\right)$:

$$\mu\left(B^1_{B^1}\right)=\mu^0\left(B^1_{B^1}\right)+T\ln\left[1-\frac{x+y+\ldots+z}{2}\right]. \tag{2.3.61}$$

Therefore, K_1' will be only a function of temperature, while K_2' will also vary with the solid solution composition. The constant K_2' can be represented as

$$K_2'=K_2''\left[1-\frac{x+y+\ldots+z}{2}\right], \tag{2.3.62}$$

where K_2'' coincides with the function for a binary A^1B^1 compound. Taking this into account and considering the number of vacant sites in each sublattice to be much smaller than that of occupied sites, it is easy to get

$$V^0_{B^1}=\left(1-\frac{x+y+\ldots+z}{2}\right)_2(T)P^{-1/m}_{\left(B^1\right)_m}, \tag{2.3.63}$$

$$V_{A^1}^0 = \frac{1}{1-\dfrac{x+y+\ldots+z}{2}} \frac{K_1(T)}{K_2(T)} P_{(B^1)_m}^{1/m} . \qquad (2.3.64)$$

Here, $K_1 = N_L K_1'$ and $K_2 = K_2' N_L$, where N_L is the concentration of sites in each sublattice, slightly varying with the solution composition.

The theoretical treatment of point defect formation in a solid solution with the cation substitution is simplified, because one can suggest that a change in the A^1-sublattice composition will not lead to the dependence of $(B_{B^1}^1)$ on x, y, ..., z. Indeed, in this approximation, the defect solutions in both sublattices are taken to be perfect, which means that the defects do not interact with one another. Then, V_{A^1} and V_{B^1} will be described by the same expressions as a binary A^1B^1 compound. As a result, we have obtained the following expressions for solid solutions with the anion substitution [54]:

$$\frac{A^-}{A^+} = \left(1-\frac{x+y+\ldots+z}{2}\right)^2 K_3' \frac{K_2^2}{K_1} P_{(B^1)_m}^{-2/m} \frac{N_v}{N_c} \frac{n^2}{n_i^2} \exp\left[\frac{E_{A^+} - E_{A^-}}{T}\right] \qquad (2.3.65)$$

and with the cation substitution:

$$\frac{A^-}{A^+} = K_3' \frac{K_2'}{K_1} P_{(B^1)_m}^{-2/m} \frac{N_v}{N_c} \frac{n^2}{n_i^2} \exp\left[\frac{E_{A^+} - E_{A^-}}{T}\right], \qquad (2.3.66)$$

where

$$K_3' = \exp\left[-\frac{g^0(A^-) - g^0(A^+) + g(V_{A^-}^0) - g(V_{B^1}^0)}{T}\right] \qquad (2.3.67)$$

$g(A^-)$ and $g(A^+)$ are standard chemical potentials relating to acceptor and donor amphoteric atoms; E_{A^+} and E_{A^-} are the energy differences between the donor and acceptor levels, on the one hand, and the respective allowed band edges, on the other; N_v and N_c are electron state densities in the valence and conduction bands.

It is clear from (2.3.65) and (2.3.66) that the distribution of an amphoteric impurity in a multicomponent solid solution of anion substitution is determined by both the volatile component gas pressure above the solid solution and the crystal composition. In the case of cation substitution, it is determined only by gas pressure. The relation between n_0 and A has the same

form as in (2.3.21), but the constant K_A will differ from that in the expressions for binary $A^{III}B^{V}$ or $A^{II}B^{VI}$ semiconductors.

For the anion substitution, this constant will be

$$K_i = \left(1 - \frac{x + y + \ldots + z}{2}\right)^2 K_3' \frac{K_2^2}{K_1} P_{(B^1)_m}^{-2/m} \frac{N_v}{N_c} \exp\left(\frac{E_{A^+} - E_{A^-}}{T}\right), \quad (2.3.68)$$

and for the cation substitution it is

$$K_i = K_3' \frac{K_2'}{K_1} P_{(B^1)_m}^{-2/m} \frac{N_v}{N_c} \exp\left(\frac{E_{A^+} - E_{A^-}}{T}\right). \quad (2.3.69)$$

One could make the final judgment about the distribution pattern of an amphoteric impurity in semiconductor compounds if one had reliable data on the dependence of volatile component vapor on solid solution composition. We have to state with regret that there are no such data for $A^{III}B^{V}$ or $A^{II}B^{VI}$ compounds at present.

Similarly, the limit concentrations of additional components, for which condition (2.3.49) remains valid, can be found only experimentally.

However, we can make an *a priori* statement that small admixtures of A^n or B^n can change but slightly the free vacancy concentrations and, hence, are unable to redistribute appreciably the amphoteric impurity positions. Indeed, the only reason for such a redistribution in a perfect solution would be displacement entropy. While the atomic fractions of additional components remain small, the changes in vacancy concentrations will also be small.

This is what follows from classical thermodynamics. But the study of isovalent doping has shown that there are significant deviations from the classical concepts at low contents of isovalent impurities ($x \leq 0.01$). This problem will be discussed in Section 2.4.

2.3.5 Data on amphoteric impurity states and behavior

Here, we will consider only the most informative data on amphoteric impurities of various kinds. More detailed information on various amphoteric impurities can be found in the books [54, 57].

Amphoteric d-impurities. One can see from Figures 2.11 and 2.12 which of the *d*-impurities are expected to show amphoteric properties. These are impurities, whose levels, 0/+ and 0/–, appear to be in the forbidden gap. For

Table 2.13. Different denotations for Cr impurity states in GaAs.

Configuration at a Ga site	Chemical symbols		Physical denotations		Amphoteric symbols
	Oxidation degree	Ion charge	Ion charge	Donor acceptor	
d^3	III	Cr^{3+}	Cr^0	–	A^0
d^2	IV	Cr^{4+}	Cr^+	d^+	A^+
d^4	II	Cr^{2+}	Cr^-	a^-	A^-
d^5	I	Cr^+	Cr^{2-}	$a^=$	$A^=$

both levels to "be accommodated" in it, the forbidden gap must be wide enough and the Coulomb repulsion of electrons in a pseudo-ion, representing a *p-d* hybrid, must be weak. This, in turn, requires a maximum mixing of *p*-electrons of the valence band and *d*-electrons of the impurity.

It is seen from Table 2.8 that *d*-impurities exhibit a slight amphoteric property in $A^{II}B^{VI}$ compounds, while in $A^{III}B^{V}$ semiconductors this property is exhibited only by chromium in GaAs and by vanadium in GaP (Figure 2.11), showing type-I amphoteric behavior. In elemental silicon, the type-I behavior is exhibited only by cobalt.

Consider the properties of these systems.

Properties of the GaAs<Cr> system. Numerous experimental studies, summarized in [14, 58, 59], indicate that chromium impurity atoms occupy gallium sites in GaAs. In the RFLW scheme (2.2.4), the chromium ground state Cr^0 has a d^3 electronic configuration, and the ionized states are, respectively, d^2 (donor) and d^4 (acceptor). One finds in the literature different denotations for charge states of chromium ions and other impurities. To avoid confusion, we are giving, as an illustration, all generally accepted denotations for Cr in GaAs (Table 2.13). The chemical symbols are used in inorganic chemistry, as well as in optical and EPR spectroscopy. The physical denotations are commonly used for the description of energy levels, transfer processes, and recombination in semiconductors.

A detailed experimental evidence for the presence of all configurations in GaAs<Cr> (Table 2.13) can be found in [14, 57].

Two models of the $Cr^{3+}(d^3)$ impurity center have been suggested on the basis of numerous experimental data, mostly on EPR spectra observed under illumination at low temperature [60] and uniaxial crystal compression [61]. In one model, the center is located on the cube axes directed away from the tetrahedron center toward the nearest neighbors at Yan–Teller sites. In the other model [59], the $Cr^{3+}(d^3)$ center forms an associate with another center [59], with the bonding along the [100]-axis.

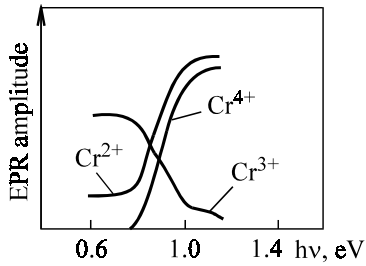

Figure 2.26. EPR intensities from various Cr impurity states in GaAs.

Irrespective of the state model, a Cr^{3+} center with the d^3-configuration does exist in GaAs crystals, showing amphoteric properties. An A^- state with the d^4-configuration, negatively charged relative to the semiconductor, was also detected by the EPR technique [62, 63]. The authors of this work found the d^4-state of chromium impurity in the crystal field with the T_d-symmetry. This result was supported independently by phonon ballistic absorption data [64] and absorption measurements of monochromatic ultrasonic waves [65].

The donor $Cr^{4+}(d^2)$ state was detected by the EPR technique in *p*-GaAs<Cr> samples, whose Fermi level was located in the vicinity of the valence band top [66, 67]. The d^2-state in the T_d crystal field must split into a set of levels, but the isotropic EPR line is so wide even at 4.2 K that the superfine structure of the EPR signal is unresolvable. As a consequence, the splitting parameters of these center levels have not yet been determined. Moreover, the d^2-configuration itself was found from the EPR data by comparing Cr concentrations in *p*-GaAS and the calculated EPR line intensity (area).

One can hardly doubt the existence of $Cr^{4+}(d^2)$ ions, since the three configurations d^3, d^4, and d^2 are observable in the EPR spectra of samples illuminated by light quantum energy equal to or larger than the forbidden gap width. More sensitive to EPR light in high resistance (semi-insulating) samples is GaAs<Cr> with the Fermi level lying close to the forbidden band center. The changes in the EPR signal intensity are shown schematically in Figure 2.26 [58]. One can see that a signal from the donor d^2-state appears at $\hbar\omega \geq 0.8$ eV with a simultaneous increase of the signal from the acceptor d^4-state and a decrease of that from un-ionized chromium with the d^3-configuration.

These findings suggest that the photoionization process occurs as a sequence of reactions (Figure 2.27):

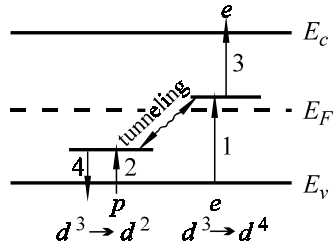

Figure 2.27. Schematic electron transitions in the energy level system of amphoteric Cr states in GaAs.

$$d^3 + e_v \rightarrow d^4 + p_v \quad \text{at} \quad h\nu_1, \tag{2.3.70}$$

$$d^3 + p_v \rightarrow d^2 \quad \text{at} \quad h\nu_2. \tag{2.3.71}$$

A light quantum first ionizes the $Cr^{3+}(d^3)$ atom, which captures an electron from the valence band under illumination, and transforms to $Cr^{2+}(d^4)$, i.e., reaction (2.3.70). Then, the free hole produced in the valence band is captured by another un-ionized $Cr^{3+}(d^3)$ center, transforming it to the donor $Cr^{4+}(d^2)$ state, i.e., reaction (2.3.71). This transition occurs at $h\nu_2 \cong 0.45$ eV and the first acceptor capture takes place at $h\nu_1 \cong 0.8$ eV. At zero illumination, the captured carriers recombine, and the EPR spectrum shows the return to the initial concentration of d^3-centers.

The d^4 and d^2 concentrations decrease exponentially with the same time constant [58], in agreement with the photogeneration model of amphoteric d^4- and d^2-Cr states. However, the recombination mechanism is not quite clear. Since after reactions (2.3.70) and (2.3.71) the valence band contains no vacant sites, one can suggest a direct interimpurity recombination via tunneling, as was proposed by the authors of [68]. This implies the minimum separation between the two d^3-centers, which may happen only in a correlated, rather than random, distribution of impurity atoms throughout a crystal. This question remains to be answered, as well as the question of a random or controlled doping impurity distribution in semiconductors.

The diagram of energy levels proposed in [54] for amphoteric transitions of the $Cr^{3+}(d^3)$ center generally accounts for the electrical, optical, and photoluminescence properties of the GaAs<Cr> system.

The energy levels of all amphoteric chromium centers in gallium arsenide can be conveniently represented as a configuration diagram (Figure 2.28); the respective energy values are given in Table 2.14.

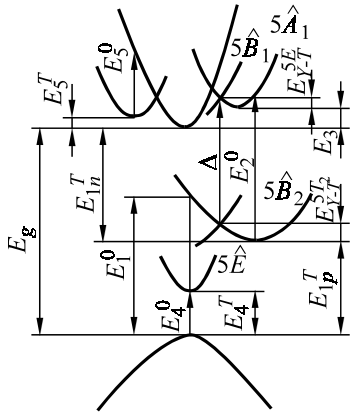

Figure 2.28. A configuration diagram of energy levels for various amphoteric Cr states in GaAs (for numerical values, see Table 2.14).

Table 2.14 also includes the doubly charged acceptor state $Cr^{2-}(d^5)$. To study this state, the Fermi level must be shifted much higher than the middle of the forbidden gap. This is done using double doping with chromium N_{Cr} and shallow donor N_d, simultaneously. If impurity chromium remains singly charged in its acceptor state A^-, we will have the concentration equality:

$$n = N_d - N_{Cr}, \qquad (2.3.72)$$

Table 2.14. Summary of Cr energy levels in GaAs.

States	Energy characteristics (Figure 2.28)	Energy values, eV
$Cr^-(d^4)$	E_n^{opt}	0.8
	E_{1n}	0.67–0.68
	E_{1p}^T	0.76
	Δ	?
	E_2^{opt}	0.9
	$E_{Y-T}^{5T_2}$	0.062
	E_3	?
	E_{Y-T}^{5E}	0.007
$Cr^+(d^2)$	E_4^{opt}	0.45
	E_4^T	?
$Cr^{2-}(d^5)$	E_5^{opt}	0.5
	E_5^T	0.055

where n is free electron concentration defined by the Hall coefficient.

If impurity chromium is in the A^--state, we have the equality

$$n = N_d - 2N_{Cr} . \quad (2.3.73)$$

The largest errors in these experiments vary with the accuracy of independent measurements of concentrations N_d and N_{Cr}. The authors of [69] used chromium and tin and determined the concentrations with radioactive labels. In [70], the value $N_d = N_{Ga}^{Si} - N_{As}^{Si}$ was found from the local modes of optical absorption by silicon atoms at gallium and arsenic sites; the value of N_{Cr} was measured by mass-spectrometry. The results of these experiments do not coincide, which seems to be due to the complexity of impurity concentration measurement. In any case, the data of [70] better agree with relation (2.3.73).

Of interest are experiments on the study of properties of the GaAs<Cr> system under hydrostatic pressure [71, 72]. At first, the samples with $N_{Cr} \geq n$ behaved as described above, i.e., they exhibited light absorption. At pressure ~1 GPa, the 0.9 eV absorption peak associated with intra-center transitions of the d^4-center disappeared, increasing the sample conductivity. These data were interpreted as being due to the d^5-level position above the conduction band bottom. With increasing pressure, the band bottom goes higher, exposing the d^5-level, and the absorbed light is nearly totally spent for the transition of the second electron to $Cr^{2+}(d^4)$, or for the formation of $Cr^+(d^5)$. The d^5-level was found in these experiments to lie at $E^T = 0.055$ eV above the conduction band bottom at atmospheric pressure and 77 K. Ionization light energy of the level $d^4 \rightarrow d^5$ at ~1 GPa is $E^0 \cong 0.5$ eV, which is shown in Figure 2.28 and Table 2.14.

Properties of the $A^{III}B^V$<V> system. The amphoteric vanadium levels in $A^{III}B^V$ semiconductors were obtained in a general form from numerous experimental data (Figure 2.29) by the authors of [73]. They used double doping with vapors of V and Se, V and Si, V and Zn to obtain GaAs<V> samples with varying Fermi level positions—practically from E_c to E_v. The level E_c – 0.15 eV, depleted by electron escape to the conduction band on heating, was found only in samples with the Fermi level in the range $E_c > \mu > 0.15$ eV.

The 0.15 eV level was identified as the A^- acceptor state of impurity vanadium, produced in the reaction

$$A^0\left(d^2\right) + e_v^- \rightarrow A^-\left(d^3\right) + p_v^+ . \quad (2.3.74)$$

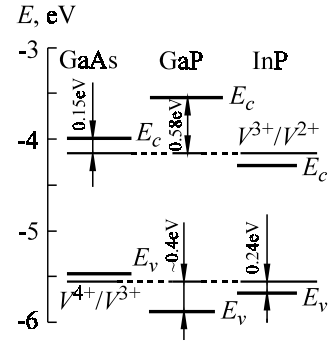

Figure 2.29. Energy levels of amphoteric vanadium in basic $A^{III}B^{V}$ compounds [73].

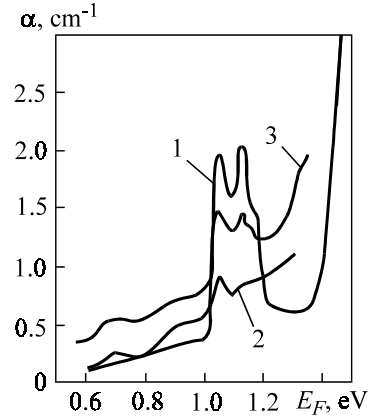

Figure 2.30. Optical absorption spectra of GaAs<V> samples at T = 5 K [146]: 1 – p-crystal; 2 – high resistance n-crystal at $N_V < N_d$; 3 – n-crystal at $N_V > N_d$ (N_d is background donor concentration).

If a vanadium atom occupies an A^{III} site, the d^3-state is in a tetrahedral crystal field and splits into the ground $^4T_1$ and two excited $^4T_2$ and $^4A_2$ states. The level E_c – 0.15 eV seems to belong to the lower of these $^4T_1$ states.

That this level belongs to the $V^-(d^3)$ center is supported by optical absorption data. Figure 2.30 shows spectra from three groups of samples differing in the Fermi level position [73].

The first group includes GaAs<V> samples with the Fermi level located in the lower half of the forbidden gap. These samples either had a high resistance with the Fermi level close to the forbidden band center owing to the

natural impurity background or close to the valence band top in simultaneous doping with V and Zn. The absorption spectra of both sample subgroups are identical (curve 1 in Figure 2.30). The right-hand side of the curve indicates a fast absorption beginning at $h\nu$ = 1.35 eV and corresponding to E_g – 0.15 eV. So, this level was identified from reaction (2.3.74). The rest of the spectrum must then reflect the light absorption by intra-center electron transitions in un-ionized $A^0(d^2)$ states of vanadium atoms. The position of these transitions, as interpreted in [73–75], is shown in Figure 2.30. Here, the $^3A_2 \rightarrow {}^3T_2$ transition was identified from the photoluminescence spectrum [73] which has a phonon-free line at 0.74 eV, reflecting the transition of an electron from the lowest of the three excited states to the ground state. The identification of the spectrum represented by curve 1 in Figure 2.30 was supported by direct EPR studies of the electronic configuration [76–78].

The second group includes *n*-type samples with the Fermi level lying above E_c – 0.15 eV at $N_V < N_d$, where N_d is the concentration of doping shallow donors. The spectrum is represented by curve 2 in Figure 2.30 and corresponds to the transitions inside a $A^-(d^3)$ center, shown in Figure 2.31. The changes in the spectrum intensities and the 100 K peak in the DLTS spectrum of Figure 2.32 correlate well for samples with different concentrations of vanadium centers. This is why the transition $V^0(d^2) \rightarrow V^-(d^3)$ was attributed to the 100 K peak in the DLTS spectrum.

Mixed light absorption by the d^2- and d^3-states of vanadium centers (curve 3 in Figure 2.30) occurs in samples of the third group with the N_V/N_d ratio corresponding to the Fermi level E_c – 0.15 eV. The spectral curve in this case is nothing else but the superposition of spectra 1 and 2.

The appearance of the $A^+(d^1)$ donor state of vanadium impurity could be expected to be due to the same mechanism as for the GaAs<Cr> system.

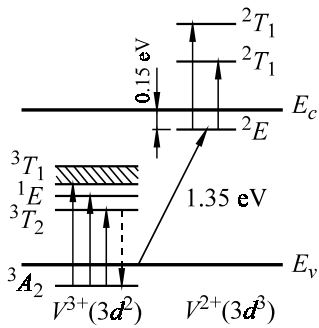

Figure 2.31. The main vanadium energy levels in GaAs [73].

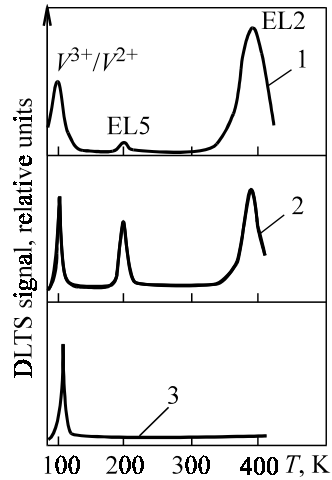

Figure 2.32. DLTS spectra from GaAs<V> samples grown by the following methods [146]: 1 – Chokhralski method; 2 – Bridgeman (horizontal) method; 3 – electroliquid epitaxy.

Since reaction (2.3.74) produces a free hole, it can be captured, similarly to (2.3.71), by another $V^0(d^2)$ center to produce a d^1-state. However, such a level could not be identified by the authors of [73] during the Fermi level transition across almost the whole GaAs band. It may be concluded that the level $V^3(d^2) \rightarrow V^{4+}(d^1)$ is localized either in the forbidden gap near the valence band top or even inside it. The authors of [73] consider the second interpretation more probable, following the general scheme of d^n-level prediction shown in Figure 2.11. We think, however, that the other interpretation should not be entirely discarded, because a shallow level located near E_v is hard to detect experimentally. In any case, for this level to be detected, the Fermi level must be located between it and the valence band top. This probably cannot be done by heavy doping with an ordinary acceptor impurity because of the formation of an impurity band (the tail of the density of states) merging with E_v. Of course, the vanadium donor level may be expected to be in resonance with the impurity band state and to manifest itself in electroluminescence spectra of diodes made, say, from GaAs<V,Zn>.

To draw the final conclusion concerning the energy position of the vanadium donor state, it would be necessary to perform a series of studies of samples subjected to hydrostatic compression in order to broaden E_g and to shift the $V^+(d^1)$ level away from the edge E_v. The general scheme of amphoteric vanadium levels in GaAs, as was formulated in [73], is given in Figure 2.31.

The behavior of a vanadium impurity atom in GaP is much less clear. If vanadium atoms form, like other *T*-atoms, a substitutional solution in GaP and occupy Ga sites, their electronic structure must be $V^{3+}(3d^2)$. The ground state A^0 of this electronic configuration has a zero orbital momentum with the effective spin $S = 1$; therefore, the EPR signal of these centers should be observable. But no EPR spectrum could be detected in *n*-type crystals with a shallow donor background at about 10^{17} cm^{-3} [79]. All vanadium impurity was found to be ionized as V^{2+} with the $3d^3$-configuration. Donor electrons filled the lowermost level, producing a non-zero orbital momentum, so no EPR spectrum could be observed. But a feature was likely to appear in the optical absorption spectrum due to the term splitting of V^{2+} ions by the T_d-crystal field. Such a feature was, indeed, detected in the optical absorption spectrum of GaP<V> [80] and interpreted as being due to electron transitions inside the *d*-shell of a V^{2+} ion, which are similar to those considered above for the GaAs<V> system.

The conductivity type of GaP<V> was changed to the hole conductivity by double doping with vanadium and manganese [79]. The samples exhibited EPR spectra for manganese and vanadium ions in the $3d^2(V^{3+})$ state. The latter spectrum consisted of two fine lines. Since the nuclear spin of the V^{51} isotope is equal to 7/2 (99.76% occurrence), the EPR spectrum was expected to contain eight superfine lines. But the experiment did not reveal these lines [79], which was interpreted as being due to their broadening associated with the superfine interaction between the *d*-electrons of the V^{3+} ion and its nucleus.

Recent data have shown that the energy levels of vanadium ions in GaP exhibit only acceptor properties and lie at 0.58 eV below the conduction band [81]. They were found in *n*-type samples and reflected the V^{3+}/V^{4+} transition.

Generally, *p*-type GaP<V> samples are to have the V^{3+}/V^{4+} state, but there have been no reports of such observations. The general diagram of vanadium levels relative to the vacuum levels in $A^{III}B^{V}$ allows prediction of the V^{3+}/V^{4+} donor level in GaP at $E_v + 0.4$ eV, which is clear from Figure 2.29 based on experimental findings [73]. The vanadium donor levels identified in this work for InP were found to be $E_v + 0.24$ eV, against $E_v + 0.21$ eV in other reports.

Properties of the $A^{II}B^{VI}$<Ni> system. Nickel impurities in $A^{II}B^{VI}$ crystals are located in the A^{II} sublattice, in which the $Ni^{2+}(d^8)$ state is un-ionized relative to the semiconductor. The transitions between the ground T_1 state and three excited states, observed in optical absorption and photolumi-

nescence spectra, reliably identify the $Ni^{2+}(d^8)$ state in various $A^{II}B^{V}$ crystals[1].

In addition to intra-center transitions, $A^{II}B^{VI}$ crystals exhibit wide absorption bands. The long wavelength edge of these bands in ZnS<Ni> and ZnSe<Ni> were unambiguously identified as being due to the transition of the un-ionized $Ni^{2+}(d^8)$ state to the acceptor $Ni^{+}(d^9)$ state, as in the reaction

$$Ni^{2+}\left(d^8\right) + e^- \rightarrow Ni^{+}\left(d^9\right) + p_v . \qquad (2.3.75)$$

This process has been confirmed for ZnSe<Ni> by several researchers who used different experimental techniques: photocapacitance measurements [83], photo-EPR studies [84], and others. Similarly, this has also been established for the ZnS<Ni> system [84, 85].

The energy levels of donor ($A^0 \rightarrow A^+$) and acceptor ($A^0 \rightarrow A^-$) transitions of nickel in $A^{II}B^{VI}$ crystals are given in Table 2.15. One can see that four crystals of this group (ZnSe, ZnS, CdSe, and CdS) show amphoteric properties of the *d*-state of impurity nickel.

Properties of the Si<Co> system. A combined investigation of the Hall coefficient and deep level positions by the DLTS technique has recently revealed [86] the donor Co^0/Co^+ level $E_v + 0.21 \pm 0.02$ eV and the acceptor Co^0/Co^- level $E_c - 0.41 \pm 0.02$ eV, although Table 2.9 gives different values for cobalt levels in silicon.

Experiments with diffusion saturation of silicon samples with cobalt show that the level concentrations of amphoteric cobalt in *n*- and *p*-samples are close; therefore, they belong to the same cobalt state. The authors of [54] give preference to interstitial cobalt in the Co^0 state.

Table 2.15. Energy levels (eV) of donor ($A^0 \rightarrow A^+$) and acceptor ($A^0 \rightarrow A^-$) transitions in $A^{II}B^{VI}$<Ni> crystals*.

Energy level	ZnSe	ZnS	ZnO	ZnTe	CdSe	CdS	CdTe
Donor level $E_c - E_i$	2.74	3.5	–	–	1.51	2.27	–
Acceptor level $E_v + E_i$	1.9	2.6	–	1.0	1.8	2.33	0.92

* The position of E_i in the forbidden gap is indicated as transitions *D*, *A*, δ, and *a* in Figure 2.6.

[1] Intra-center transitions to $Ni^{2+}(d^8)$ were detected only in CdTe.

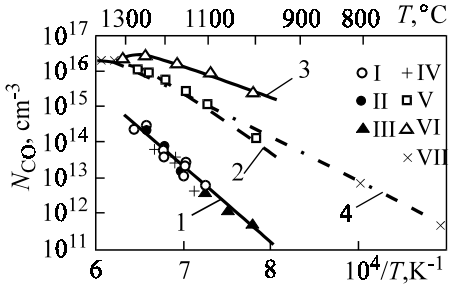

Figure 2.33. Cobalt solubilities in silicon: 1 – concentration of electrically active Co; 2, 3, 4 – total Co concentrations. (From the following reports: I – [87] *n*-type; II – [87] *p*-type; III – [88]; IV – [89]; V – [90]; VI – [94]; VII – [91]).

This is consistent with the fact that this state captures an electron to form Co^- in *n*-silicon and gives off an electron to form Co^+ in *p*-silicon. However, the solubility curve of electrically active cobalt (curve 1 in Figure 2.33) based on the data of [87–89] is two orders of magnitude lower than the total solubility curve for impurity cobalt (curve 2 in Figure 2.33). The latter curve was detected using radioactive isotope ^{59}Co [90–93]. So the amphoteric cobalt component should be ascribed to the site state Co_s with the d^5-configuration in an un-ionized state. When an electron is captured in *n*-type silicon, the amphoteric center Co_s changes to the d^6-state; when a hole is captured in *p*-type silicon, it changes to the d^4-state. It is the middle filling of the *d*-shell of an un-ionized $Co_s(d^5)$ center which determines its "unstable" state possessing an equal probability to accept or give off an electron, which is in full agreement with the above conception of transition metal behavior in silicon.

Thus, cobalt atoms exhibit amphoteric impurities of the site-type in cilicon, although their solubility is low.

The attempts to find manifestations of amphoteric properties of interstitial cobalt have been unsuccessful. The reason may be that Co_i is not a purely interstitial state but, rather, an associate of interstitial cobalt with other point defects. This is indirectly supported by the fact that the forbidden gap is "populated" by numerous levels (Table 2.9) attributed by many authors to impurity pairs containing cobalt atoms.

Poorly explored amphoteric d-impurities in silicon.

Titanium in silicon is shown by the calculations of solubility enthalpies [54] to occupy both ($Ti^0[3d^3 4s]$) sites and ($Ti^+[3d^3]$) interstices in a crystal lattice. The number of titanium atoms in one position must be close to that in the other position, which is due to the small difference in the solubility en

Table 2.16. Electronic configurations and energy levels of poorly explored interstitial amphoteric impurities in Si.

Impurity	Charge and electronic states	Amphoteric states	Energy levels, eV	References
Ti	Ti^{-}/Ti^{0}; $3d^5/3d^4$	A^-	$E_c - 0.08$	[96–98]
	Ti^{0}/Ti^{+}; $3d^4/3d^3$	D^+	$E_c - 0.28$	
	Ti^{+}/Ti^{2+}; $3d^3/3d^2$	D^{2+}	$E_v + 0.25$	
V	V^{-}/V^{0}; $3d^6/3d^5$	A^-	$E_c - 0.16$	[97, 99, 100]
	V^{0}/V^{+}; $3d^5/3d^4$	D^+	$E_c - 0.45$	
	V^{+}/V^{2+}; $3d^4/3d^3$	D	$E_v + 0.30$	
Mn	Mn^{-}/Mn^{0}; d^8/d^7	A^-	$E_c - 0.11$	[97, 100]
	Mn^{0}/Mn^{+}; d^7/d^6	D^+	$E_c - 0.42$	
	Mn^{+}/Mn^{2+}; d^6/d^5	D^{2+}	$E_v + 0.25$	

thalpies. Experimental data, however, show that only Ti interstitial states (Table 2.16) are amphoteric.

The presence of the $Ti^{+}[3d^3]$ state has been detected by the EPR and DENR techniques [93], as well as by diffusion studies [95].

Table 2.16 presents data on vanadium and manganese which, according to some workers, also show amphoteric properties in silicon. Both impurities are interstitial amphoteric centers.

Among other impurities in silicon, we will mention scandium and hydrargium.

Scandium is the least explored *d*-impurity in silicon. The authors of [101] studied the electrical properties of the Si<Sc> system and arrived at the conclusion that it had amphoteric characteristics. Scandium atoms behaved like donors in *p*-Si and like acceptors in *n*-Si. The former is typical of transition metal atoms at interstices and the latter at silicon sites. However, the lack of data on its solubility and diffusion, on the charge states and electronic configurations makes it difficult to draw final conclusions concerning the amphoteric nature of scandium in host silicon.

Hydrargium forms in silicon two acceptor levels $E_c - 0.31$ and $E_c - 0.36$ eV and two donor levels separated from the valence band top by 0.25 and 0.33 eV [102]. The author of [103] believes that since Hg occupies in the periodic table an intermediate position between Tl and Au, which are substitutional impurities in silicon, Hg should also occupy lattice sites in silicon and behave as an amphoteric impurity. This suggestion is to be tested experimentally.

2.3.6 Amphoteric excitons bound by *d*-impurities

The amphoteric properties of d^n-impurities were shown above to be described by reactions (2.2.1) and (2.2.2). A generated carrier may have a dual fate in the allowed band. One possibility is to be captured by another d^n-center. This was described with reference to impurity Cr in an un-ionized $A^0(d^3)$-state in GaAs when a free hole generated by the reaction $A^0 \rightarrow A^-(d^4)$ according to (2.3.70) is captured by another d^3-center which is to transform to $A^+(d^2)$ in accordance with (2.3.71).

The alternative is that an ionized *d*-center retains a free carrier, so that instead of (2.2.1) and (2.2.2) we will have

$$A^0 + \hbar\omega_\delta \rightarrow \left[\left(A^0 + p\right)e\right] = A^+ e\,, \tag{2.3.76}$$

$$A^0 + \hbar\omega_\alpha \rightarrow \left[\left(A^0 + e\right)p\right] = A^- p\,. \tag{2.3.77}$$

The activation energies of these reactions are

$$\hbar\omega_\delta = \hbar\omega_D - \varepsilon_\mathrm{p}\,, \tag{2.3.78}$$

$$\hbar\omega_\alpha = \hbar\omega_A - \varepsilon_\mathrm{e}\,, \tag{2.3.79}$$

differing from regular ionization energies of charged centers by the energy of the carrier binding to such a center. Electron transitions producing excitons α and δ are shown in Figure 2.6 by dashed arrows.

The products of reactions (2.3.76) and (2.3.77) are excitons bound by a neutral (relative to the semiconductor) *d*-center. Here, one can really see the well-known analogy with an exciton captured by an isoelectron trap, say, by a nitrogen impurity in GaP. In both cases, the first carrier is retained by a short-range potential and the other, having an opposite sign, by a long-range Coulomb potential. Excitons bound by *d*-impurities will be termed, like in [104], donor and acceptor excitons, in accordance with (2.3.76) and (2.3.77), i.e., in accordance with the sign of the captured hydrogen-like carrier. In this terminology, an isoelectron donor captures a donor exciton (2.3.76), while an isoelectron acceptor captures an acceptor exciton (2.3.77).

It is clear from these reactions that a *d*-impurity in the A^0-state can capture both donor and acceptor excitons. This is a manifestation of the essential difference between a *d*-impurity and a simple isoelectron (isovalent) impurity of the nitrogen type in GaP. Therefore, impurities may show

amphoteric properties not only in the formation of charged centers but also in the capture of excitons.

Theoretically, the amphoteric capture of an exciton by a *d*-impurity was predicted in [105], and donor and acceptor excitons were detected experimentally in ZnSe<Ni> by the authors of [106, 107]. A quantum mechanics theory of impurity excitons bound by *d*-centers in semiconductors is discussed in [104]. Such excitons are shown to be excitations intermediate in their properties between the Frenkel and Mott excitons, but both can be excited by an electromagnetic field. The two excitation systems are interrelated due to the exchange and spin-orbital interaction between an electron and a hole. A simple formula was derived in [107] for the probability ratio of donor (W_δ) and acceptor (W_α) excitons. With the account of (2.2.4), it can be written for a site impurity as

$$\frac{W_\delta}{W_\alpha} = \left(\frac{m_e^*}{m_{pp}^*}\right)^3 \frac{E_c - C_i\left(d^{n-(V-m)+1}\right)}{E_i\left(d^{n-(V-m)-1}\right) - E_v}, \qquad (2.3.80)$$

where m_e^* and m_{pp}^* are the effective masses of electrons and heavy holes, respectively; E_i are the *d*-level energies of the impurity center, counted off from the band edges E_c and E_v and representing the level $A^0 \to A^-$ in the numerator of formula (2.3.80) and the level $A^0 \to A^+$ in its denominator.

Therefore, the formation of excitons "follows" the appearance of the donor or acceptor level of a *d*-impurity in the crystal forbidden gap. This means that amphoteric excitons may be formed only if amphoteric *d*-levels already exist in the forbidden gap, which can be easily established from a comparison of data in Tables 2.17 and 2.15.

It has been shown that amphoteric excitons in $A^{III}B^{V}$ crystals occur very rarely [104] but are normally found in GaAs<Cr> and GaP<Cr> crystals and in many $A^{II}B^{VI}$ semiconductors (Table 2.17).

Unlike excitons bound by a simple isovalent impurity (nitrogen in

Table 2.17. Basic lines of donor and acceptor excitons in $A^{II}B^{VI}$ crystals*.

Energy level	ZnSe	ZnS	ZnO	ZnTe	CdSe	CdS	CdTe
Donor exciton	2.64	3.39	–	–	–	2.19	–
Acceptor exciton	1.82	2.44	1.6	–	1.7	–	0.966

* The positions of $\hbar\omega$ in the forbidden gap are indicated as the transitions *D*, *A*, δ, and α in Figure 2.6.

gallium phosphide), *d*-excitons can be identified only from the intracenter transitions. For this reason, the capture of excitons, α and δ in Figure 2.6, must produce satellite lines near the impurity absorption band edge. This fine structure was first observed by Kazansky and Ryskin [82] and ten years later by other researchers [107]. The fine structure consists of the basic line and its phonon reproductions.

An experimental registration of narrow weak lines against the background of a strong photoionization absorption spectrum is a very difficult task. So, a step forward was the application of electroabsorption, which is light absorption when an external electric field is applied to the sample. The first destroys the hydrogen-like bond in an exciton, reducing its lifetime τ. Variation in the lifetime changes the absorption linewidth. Besides, because of the Stark effect, the electric field shifts the absorption line toward lower energies. However, it practically does not affect the intra-center transitions, since *d*-electrons are rather strongly localized, i.e., they are strongly bound to their nucleus, or, at least, they are bound in a small space limited by the nearest ligands. The reader can find in [104] the theory and treatment of optical electroabsorption line broadening data, the information on donor and acceptor excitons, as well as on bound excitons of different types in various $A^{II}B^{VI}$ crystals with *d*-impurities.

2.3.7 Dissociative amphoteric impurities

Impurity copper in germanium. The charge state of impurity copper can be predicted from the general model of *d*-impurity behavior in semiconductors (Section 2.2). The electronic structure of a copper atom in a free state is $3d^{10}4s^1$.

When dissolved at a site or an interstice, a copper atom acquires one of the electronic configurations presented in Table 2.18. It is clear from this table that a donor state localized at a site is to have the d^6-configuration, which is very unlikely because the removal of seven out of ten electrons of the *d*-shell would require much energy, whereas the filling of the *d*-shell up to ten electrons, i.e., to equilibrium, can be done easily. Therefore, copper must be a triply charged acceptor at germanium sites.

In the RFLW model, an interstitial electron is pushed out of the *s*-shell to the *d*-shell. But the *d*-shell in copper is filled up, and so its un-ionized state retains the $3d^{10}4s^1$ configuration which can be changed by giving off rather than by acquiring electrons. Therefore, interstitial copper can have only Cu^0 and Cu^+ states which have been registered experimentally. For example, Cu_s

Table 2.18. Electronic configurations and charge states of Cu impurity centers in Ge.

Charge state	Electronic configuration	
	Site solubility	Interstitial solubility
Cu^0	$3d^7$	$3d^{10}4s^1$
Cu^-	$3d^8$	–
$Cu^=$	$3d^9$	–
$Cu^{\equiv}$	$3d^{10}$	–
Cu^+	$3d^6$	$3d^{10}$

forms three acceptor levels $E_{a1} = E_v + 0.04$ eV, $E_{a2} = E_v + 0.33$ eV, and $E_{a3} = E_c - 0.26$ eV [108], attributed to Cu^-, $Cu^=$, and $Cu^{\equiv}$, respectively. The presence of interstitial copper in the germanium crystal lattice has been confirmed by experiments on its diffusion and on decomposition kinetics of the Ge<Cu> solution.

The amphoteric behavior of copper reveals itself clearly in the decomposition mechanism in germanium samples containing, in addition to copper, other shallow doping impurities. This was reliably established in experiments with the Ge<Cu,Sb> system [109, 110]. Copper precipitation in the presence of donors was attributed (with the allowance for the charge state of reacting defects) to the following quasichemical reactions:

$$Cu_s^- \underset{K_2}{\overset{K_1}{\rightleftarrows}} Cu_i^+ + V^{2-} \qquad (2.3.81)$$

$$Cu_s^- \underset{K_4}{\overset{K_3}{\rightleftarrows}} Cu_i + V^- + e^- \qquad (2.3.82)$$

$$D^+ + nV^{2-} \underset{K_6}{\overset{K_5}{\rightleftarrows}} Z^{-(2n-1)} \qquad (2.3.83)$$

$$mCu_i^+ + Z^{-(2n-1)} \underset{K_8}{\overset{K_7}{\rightleftarrows}} Z' \qquad (2.3.84)$$

$$\mathrm{Cu_i^+} \xrightarrow[K_{10}]{K_9} \text{sink} \qquad (2.3.85)$$

$$V \rightarrow \text{sink}. \qquad (2.3.86)$$

Therefore, the study of decomposition of the complex solid solution Ge–Cu–Sb indicates the possible existence of copper atoms in two crystallochemical positions: at sites and at interstices. In the former, a copper atom is a multiply charged acceptor and in the latter it has a positive charge and is, therefore, a donor. But its ionization energy level appears to be located in the allowed conduction band spectrum, so it is always in the inactive Cu_i state.

Impurity copper in silicon. The above considerations concerning the electrical behavior of copper in germanium will also be valid for silicon. This follows from the general RFLW model for these semiconductors. Like in germanium, copper in silicon is a triply charged acceptor at a site. The generally accepted values for the ionization energy are E_v + 0.24 eV, E_v + 0.37 eV, and E_v + 0.52 eV, in accordance with the acquired number of negative charges. The difference is that the first ionized state Cu_s^- in silicon is located deep in the forbidden gap and does not contribute significantly to the electrical properties of silicon samples.

Impurity gold in silicon. In the late 1950s, Collins and co-workers carried out an investigation of the electrical properties of Si<Au> [111] and identified two Au levels in the forbidden gap: a donor level $E_D = E_v + 0.35$ eV and an acceptor level $E_A = E_c - 0.54$ eV. The electron filling of these levels in silicon samples doped and undoped with shallow impurities is shown schematically in Figure 2.34.

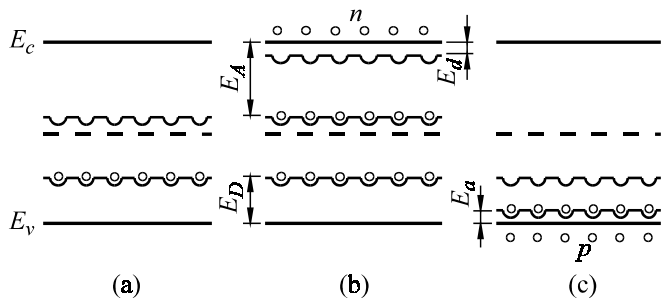

Figure 2.34. Schematic filling of gold levels in silicon: (a) – without foreign impurities, (b) – with a foreign donor impurity; (c) – with a foreign acceptor impurity (N_d, $N_a >> N_{Au}$).

At low temperatures, the Au levels are separated by $\Delta E >> kT$ and can be regarded as being independent, which considerably simplifies the solution of the neutrality equation, allowing one to identify the ionization energy of the levels with the slope of the temperature dependence of Hall concentration. The details of this treatment can be found in [54]. Here, we will mention only the basic (recombination) properties of Si<Au>. These can be easily understood in terms of the filling diagram of energy levels in Figure 2.34.

At $N_d >> N_{Au}$, the crystal has the *n*-type conductivity, and the Fermi level lies high above both Au levels which appear to be completely filled by electrons. The donor level is un-ionized and the acceptor level is negatively charged Au^-. It has a characteristically large value of hole capture cross section σ_{pA} because of the Coulomb attraction of holes from the valence band. The capture of a hole will immediately lead to the attraction of an excess electron from the conduction band. Therefore, the Au acceptor level E_A must act as an effective recombination center.

In a *p*-type crystal, i.e., at $N_a >> N_{Au}$, the acceptor level will, on the contrary, be un-ionized, while the donor level will be totally ionized, effectively capturing electrons from the conduction band with a large capture cross section σ_{eD}.

The experimental lifetimes of minority carriers in Si<Au> were indeed found to be small. The respective values of capture cross sections are presented in Table 2.19.

Although the data of different authors differ by a factor of 5 or 7, they demonstrate a considerable excess of the capture cross sections for centers of opposite signs, as was expected from the amphoteric behavior of gold impurity in silicon. Direct evidence for gold as an effective recombination center was obtained in the work [112] by measuring the relaxation time of a diode made from Si<Au>.

Table 2.19. Cross sections for hole capture by the acceptor level (σ_{pA}) in *p*-Si<Au> and electron capture (σ_{eD}) in *n*-Si<Au>.

Measurement temperature, *K*	σ, cm^2 in *n*-Si		σ, cm^2 in *p*-Si		References
	Charge state of recombination center				
	Au^0	Au^-	Au^0	Au^+	
300	5×10^{-16}	1×10^{-15}	10^{-16}	3.5×10^{-15}	[113]
	1.7×10^{-16}	1.1×10^{-14}	2.4×10^{-15}	6.3×10^{-15}	[114]
77	3×10^{-15}	1×10^{-13}	3×10^{-15}	6×10^{-14}	[113]
	5×10^{-16}	2.3×10^{-13}	?	1×10^{-13}	[113]

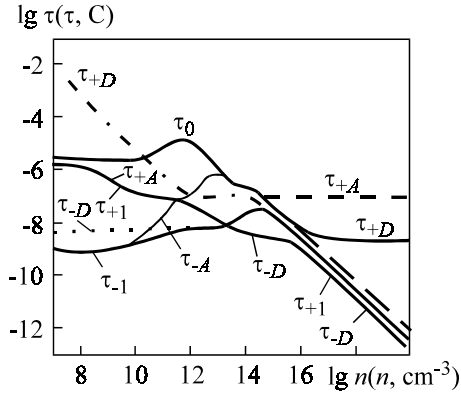

Figure 2.35. Recombination times τ_{+1}, τ_{-1}, and τ_0 in combined recombination via amphoteric Au levels in silicon. The times of independent recombination (τ_{-A}, τ_{+A}, τ_{-D}, τ_{+D}) via Au donor and acceptor levels are given for comparison.

These facts indicate that the recombination at both levels may be considered to occur independently in the first approximation. Nevertheless, we would like to mention the work [115] which shows with a more general model that the recombination in Si<Au> must be described by three lifetimes: τ_0, τ_1, and τ_2. Their dependence on the carrier concentration (holes) calculated in [115] is shown in Figure 2.35. One can see *p* ranges, in which τ_{+1} and τ_{-1} coincide with τ_- or τ_+ in separate recombination via gold amphoteric levels.

Silver in silicon behaves as other group-I impurities. Silver atoms produce donor (E_v + 0.32 eV) and acceptor (E_c – 0.29 eV) levels in the silicon forbidden gap [95, 249]. These values can be regarded as "established" ones, but many workers have found other energy values: –0.22, –0.36, and –0.59 eV for acceptor states and +0.26, +0.33, and +0.40 eV for donor states. It still remains to be found which of these levels are really associated with silver atoms and to which states they correspond.

That silver atoms belong to dissociative amphoteric impurities follows from the dissociative nature of their diffusion [116].

Amphoteric 5d-impurities in silicon were identified by Yunusov and coworkers [117]. Using these data, Table 2.20 summarizes the basic properties of impurities considered to be amphoteric. The diffusion parameters D_0 and ΔE_m refer to the temperature range from 1000 to 1250°C and describe the total concentration of the doping impurity. The dissolution enthalpy ΔH_s also refers to the total concentration, while the decay activation energy E_a characterizes the decay of each individual level.

Table 2.20. Characteristics of dissociative amphoteric 5*d*-impurities in silicon.

Impurity	D_0, cm^2/s	ΔE_m, eV	ΔH_s, eV	E_a, eV	Donor level, eV	Acceptor level, eV
Ir	4×10^{-2}	1.3±0.1	2.2±0.1	0.89 0.61 1.18	$E_c - 0.3$ (*i*)	$E_c - 0.18$ (*k*) $E_c - 0.5$ (*s*)
Rh	1.6×10^{-2}	1.2±0.2	1.6±0.2	–	$E_c - 0.3$ (*i*)	$E_c - 0.55$ (*s*)
Ru	–	–	3.35±0.02 2.3±0.02	– –	$E_c - 0.45$ (*i*)	$E_c - 0.22$ (*s*)
Os	–	–	–	–	$E_v + 0.18$(*k*)	 $E_c - 0.18$ (*k*) $E_c - 0.54$ (*s*)

s – site, *i* – interstice, *k* – unidentified associate.

A specific feature of these centers is the energy line broadening which looks more like a band, so that the values in Table 2.20 are a kind of "center of mass" of the bands. The broadening is 0.02–0.03 eV, and it is unclear whether it is due to the choice of centers slightly differing in the ionization energy or to the close vicinity of the 5*d*-atom excited states.

2.3.8 Cation–anion amphoteric impurities in semiconductor compounds

It follows from Section 2.3.1 that cation–anion amphoteric impurities in $A^{III}B^{V}$ compounds are to be atoms of group-IV elements in the periodic table. Indeed, this suggestion has been supported by numerous experimental investigations (Figures 2.36 and 2.37). Moreover, the theory of amphoteric impurities described in Section 2.3.4 can predict the impurity distribution between the crystal sublattices, which varies with the impurity concentration, temperature, and pressure of the volatile component. This result has generally been confirmed experimentally. But in some cases, the actual behavior of group-IV impurities appears more complicated and depends on the presence of other defects and their interaction with the amphoteric impurity.

The behavior of gallium arsenide doped with silicon is understood much better than that of other systems. This seems to account for the wide application of GaAS<Si>. Epitaxial films made from this material possess both *n*- and *p*-type conductivity, and this can be achieved in the same technological process. This approach is used for the fabrication of *p*-*n*-structures, for example for light-emitting devices.

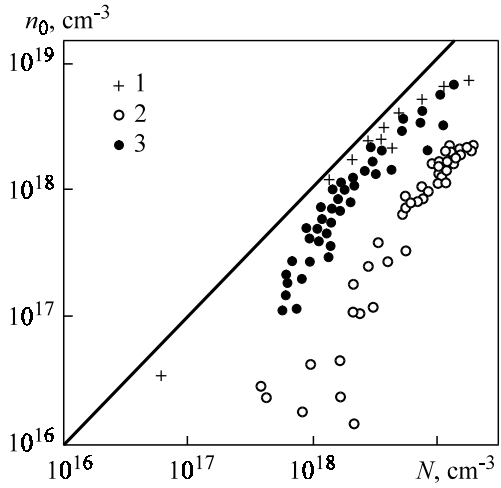

Figure 2.36. Electron concentration versus group-IV amphoteric impurity content in GaAs: 1 – Si [118]; 2 – Ge [258]; 3 – Sn [259].

Germanium in gallium arsenide shows a greater amphoteric activity than silicon. When grown from stoichiometric melts, GaAs<Ge> crystals possess the *n*-type conductivity, but the degree of self-compensation N_D/N_A in them is lower than in GaAs<Si>. However, the compensation of GaAs samples doped with germanium is more sensitive to variation in arsenic vapor pressure—there is an additional compensation due to the formation of associates with germanium atoms [122].

When grown by liquid-phase epitaxy, GaAs<Ge> films have *p*-conductivity with a small compensation degree. This means that the amphoteric behavior of germanium shifts toward the acceptor side with decreasing temperature, i.e., germanium atoms tend to dissolve in the As-sublattice rather than in the Ga-sublattice. The inversion temperature (*n*-to-*p* transition temperature) becomes higher than the usual epitaxial growth temperature and lies within the range of the time–temperature regime normal for GaAs<Si>.

Doping with tin always produces *n*-type gallium arsenide crystals. This means that the amphoteric behavior of this impurity is strongly shifted toward the Ga-sublattice.

Information on the behavior of impurities in other $A^{III}B^{V}$ compounds is more scarce. This concerns physicochemical, electrophysical, and optical properties of these compounds doped with group-IV elements.

The data on the energy levels of group-IV impurities in $A^{III}B^{V}$ com-

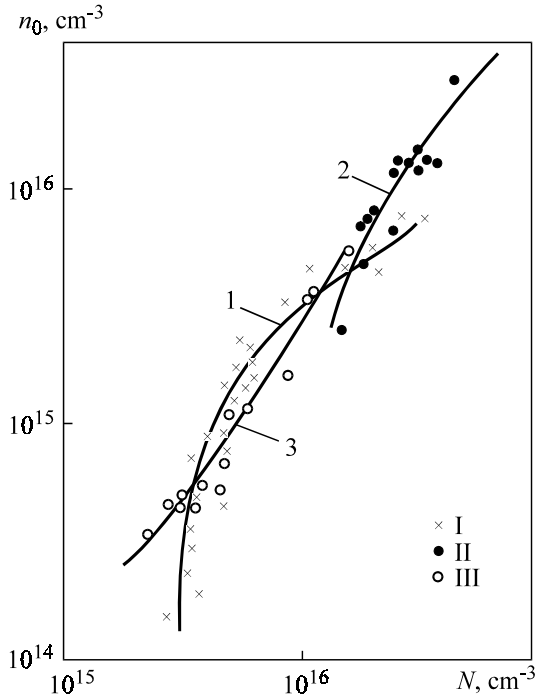

Figure 2.37. The $n_0(N)$ dependence in undoped GaAs layers grown by gas-phase epitaxy [121]. Substrate orientation: I – (100); II – (111)*B*; III – (311)*A*. Solid lines – calculations from formula (2.3.29) with the values:

I – $K_A = 0.38$; $[V^-] = 7\times10^{14}$ cm^{-3};
II – $K_A = 0.23$; $[V^-] = 3.7\times10^{15}$ cm^{-3};
III – $K_A = 0.74$; $[D^+] = 2\times10^{16}$ cm^{-3}.

pounds are given in Table 2.21. One can notice certain regularities here. For example, group-IV impurities produce shallow levels only in direct band $A^{III}B^{V}$ semiconductors; they produce deep levels in indirect band gallium phosphide.

Shallow levels can be considered as hydrogen-like only with many reservations. Their hydrogen-like behavior is supported by practically identical ionization values of their donor states in gallium arsenide and by close values in indium phosphide, as is seen from Table 2.21. Their difference in gallium phosphide, however, makes one doubt their hydrogen-like nature. Table 2.21 contains a fairly large number of blank spots in the energy level measurements.

Table 2.21. Energy levels (eV) of group-IV impurities in basic $A^{III}B^{V}$ compounds [123, 124].

Impurity	GaAs		GaP		InP	
	Donor	Acceptor	Donor	Acceptor	Donor	Acceptor
C		0.019–0.046	–	0.041 0.046 0.048	0.020	–
Si	≤0.006	0.026–0.030	0.080–0.082	0.200–0.202	0.005–0.007	–
Ge	≤0.006	0.035–0.038	0.195–0.201	0.253–0.257	≤0.010	–
Sn	0.006	0.17–0.20	0.060	–	≤0.010	–
Pb	–	0.12	–	–	–	–

Let us consider briefly the behavior of cation–anion impurities in particular $A^{III}B^{V}$ compounds.

Amphoteric impurities in GaAs. It is first interesting to see how much the experimental $n(N_A)$ curve for GaAs agrees with the theory developed in Section 2.3.1. The theory gives a general expression (2.3.21) for the description of the $n(N)$ function. However, when films were grown by gas-phase epitaxy in the temperature range 970–1080 *K* [121], the intrinsic carrier concentration $n_i \cong 10^{17}$ cm^{-3} was much higher than the maximum electron concentration in the films, $n_0 \cong (2–3)\times10^{16}$ cm^{-3}. Since $n = n_i + n_0$, then $n \cong n_i$, in which case the condition of (2.3.21) changes to a more simple expression (2.3.29) used for the data comparison [260] illustrated in Figure 2.37. Note that the quantity N_I plotted on the abscissa represents the sum of A from (2.3.29) and V from (2.3.31). Qualitatively, the curves in Figure 2.37 look very much like theoretical curve 1 in Figure 2.23. The authors obtained a quantitative agreement at the values of K_i and $[V^-]$ indicated in the caption to Figure 2.37.

It follows from the amphoteric impurity theory that K_i decreases with a change in arsenic vapor pressure P_{As} as $P_{As}^{-1/2}$. This theoretical prediction was confirmed experimentally in [121].

However, curve 3 in Figure 2.37 shows a deviation from linearity in the high concentration region not toward the shoulder, as it is required by the theory, but toward higher electron concentrations. The amphoteric theory

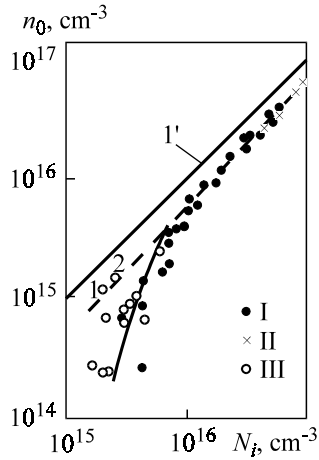

Figure 2.38. Concentration dependence of ionized centers in GaAs epitaxial layers with various contents of silicon and oxygen impurity atoms [126]: 1' – calculation without compensation; I – undoped layers; II – Si-doped layers; III – layers with fixed oxygen content (numbers at the points, cm^{-3}): 1 – 1.2×10^{16}; 2 – 1.5×10^{16}; 3 – 5.8×10^{17}; 4 – 10^{18}.

requires, in this case, the assumption concerning generation of additional donor centers D^+. On this assumption, curve 3 can be fitted to the theory, taking $K_A = 0.74$ and $[D^+] = 2\times10^{16}$ cm^{-3}. Of course, the origin of additional D^+ donors remains unclear.

The study of epitaxial layers deliberately doped with silicon has shown [125, 126] that the simple version of amphoteric impurity theory (Chapter 1) is valid only in the range of low impurity concentrations $N_A < 10^{17}$ cm^{-3}. At high concentrations, an essential contribution is made by other background impurities. The main background impurity is oxygen, whose effect is clearly demonstrated in Figure 2.38. At $N_i > 8\times10^{15}$ cm^{-3}, there is a linear dependence $n_0(N_i)$ demonstrating a stable compensation degree $K = N_a/N_d = 0.25$. This value agrees well with the above value of 0.23 obtained by gas-phase epitaxy for the same substrate orientation.

The experimental points obtained for silicon-doped and undoped films [126] fit well in the same curve, which allows considering $K = K_i = 0.25$ as a self-compensation parameter of amphoteric silicon. In the low concentration range N_i (< 8×10^{15} cm^{-3}), the compensation degree strongly depends on oxygen content, beginning with the value ≥ 10^{18} cm^{-3}. Without going into a debate concerning the nature of compensating centers, we can conclude from

Figure 2.38 that the amphoteric theory also agrees with the experiment for silicon-doped GaAs samples grown by liquid-phase epitaxy.

More direct evidence for the cation–anion amphoteric behavior of silicon in gallium arsenide was obtained by Spitzer and co-workers [127] by measuring local modes in optical absorption spectra. The far-infrared region was found to contain absorption bands associated with the donor state Si_{Ga} ($\nu = 384$ cm^{-1}) and the acceptor state Si_{As} ($\nu = 399$ cm^{-1}); the band intensities were found to vary with the compensation degree.

Germanium has a well pronounced amphoteric behavior in GaAs among group-IV impurities. This is manifested as a higher degree of self-compensation (curve 2 in Figure 2.36), which essentially affects the electrical properties of GaAs<Ge> because of the lower electron mobility than the ion mobility in GaAs<Si>.

Evidence for the amphoteric nature of germanium in gallium arsenide was obtained from thermal treatment of GaAs<Ge> samples [128]: the *p*-type conductivity changed to the *n*-type conductivity when the samples were heated in arsenic vapor under elevated pressure. The number of arsenic vacancies decreased and the Ge_{Ga} concentration increased.

Tin, in contrast to germanium, has less pronounced amphoteric properties. It retains its *n*-conductivity up to a very high concentration in GaAs<Sn> samples (curve 3 in Figure 2.36). But some workers have also identified an acceptor level located deep in the forbidden gap (Table 2.21). All doubt concerning the amphoteric nature of tin in gallium arsenide and other $A^{III}B^{V}$ was removed after γ-resonance (the Mossbauer effect) of ^{119}Sn nuclei was observed during their implantation into $A^{III}B^{V}$ crystals [128]. This work also showed that the donor state represented a single substitutional tin atom in the *A*-sublattice, while the acceptor state was a defect with a neutral tin atom associated with the nearest point defect at a *B*-sublattice site.

Among other properties of GaAs doped with group-IV elements, we would like to mention some specific features in photoluminescence spectra, in particular, the presence of long wavelength bands. Such bands are known to exist in undoped gallium arsenide samples. Most workers have concluded that they are due to recombination centers involving intrinsic point defects. Doping of gallium arsenide with group-IV impurities gives rise to long wavelength bands: at 1.0 eV ($n < 10^{18}$ cm^{-3}) and 1.15–1.3 eV in GaAs<Si> and GaAs<Sn>; at $h\nu_m = 1.4$ eV and $h\nu_m = 1.17$ eV in GaAs<Ge>. It is known from the general recombination radiation theory that band intensities depend on the presence of nonradiative centers. An investigation of this problem[2] revealed an important difference between the doping with ampho-

[2] M.N. Kamalov, "The formation of electrically and recombinationally active centers in doped *n*-GaAs crystals". Abstr. dissert. Tashkent, Izd. FTI AN Uz. SSR, 1981 [in Russian].

teric group-IV impurities and that with non-amphoteric impurities, for example, with group-VI elements. In the latter case, the concentration of nonradiative centers, N_s, in the range $n = 1\times10^{17}$–2×10^{18} cm^{-3} did not depend on the charge carrier concentration and began to show this dependence only in the heavy doping range. With silicon or tin, the concentration of nonradiative centers varies over the impurity concentration range.

This is an essentially fundamental result. It supports the idea [129] of neutral defect generation by group-IV impurities, which was suggested from the study of electrical activity of these impurities in gallium arsenide. The idea is that electrical properties, in particular, the $n_0(N)$ curves theoretically predicted as resulting from the impurity dissolution in both $A^{III}B^{V}$ sublattices, have a different interpretation. They are accounted for by the transition of, say, silicon atoms to the neutral state. This transition depends on the concentration of volatile impurity (e.g., silicon) and on the crystal growth conditions which prescribe the ensemble of intrinsic point defects. These defects determine the transition of a group-IV impurity to the neutral state because they are either involved in neutral associates or accelerate the loss of electrical activity by the group-IV impurity.

The main advantage of this model is the self-consistency of the problem: the concentration of an electrically active group-IV impurity depends on the defect composition of a crystal, while the defect composition itself varies with the impurity concentration.

This syneresis deserves special attention. It partly follows from the studies of GaAs<IV> compounds, but if it is supported by data on other $A^{III}B^{V}$ compounds, the amphoteric impurity theory will require further sophistication.

Amphoteric impurities in gallium phosphide. The $n_0(N)$ curve for amphoteric impurities in GaP has the characteristic shape shown in Figure 2.39. There is no statistically reliable distribution of group-IV impurities in the GaP sublattices, but the measurements of electrical resistivity [130, 131] and photoluminescence [132] indicate the building-in of, say, germanium atoms into both GaP sublattices. The activation energies of Ge donor and acceptor levels are found to be 0.36 and 0.45 eV, respectively. These values seem to be overestimated, since they were obtained without consideration of the electron–phonon interaction of deep levels inherent in germanium.

Amphoteric impurities in indium phosphide. The electrical activity of group-IV impurities is discussed in detail in [134] with reference to samples grown by the Chokhralsky method. The total concentration of impurity atoms was found from the atom absorption analysis with an error ±10%. The greatest difficulty was caused by silicon doping.

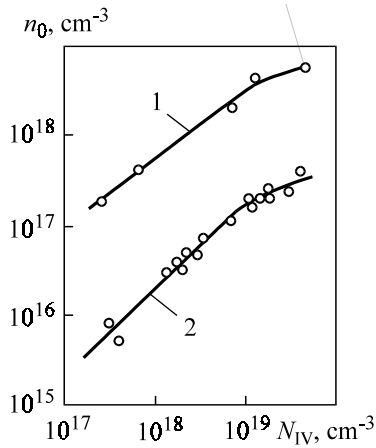

Figure 2.39. The $n_0(N)$ dependence in GaP<IV> crystals: 1 – GaP<Si> [133]; 2 – GaP<Ge> [131].

The authors of this work could grow single crystals only with a silicon content in the melt of less than 0.15% (mass.) (1.5×10^{20} cm^{-3}).

The electron concentration in InP<Si> crystals was less than 5×10^{16}. No ways of obtaining samples with higher silicon concentrations were found. This radical difference between InP and other $A^{III}B^{V}$ crystals doped with silicon remains could not be explained, so the authors of [134] focused mostly on the behavior of germanium and tin in InP crystals.

The same samples were used to measure the Hall effect and total impurity concentration after the removal of contacts. The electron concen-

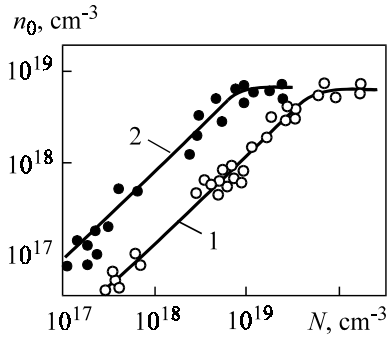

Figure 2.40. The $n_0(N)$ dependence in InP<IV> crystals [134].

tration variation with impurity content in the solid phase is presented in Figure 2.40. One can see that germanium is more amphoteric than tin. Calculations show that about 90% of tin atoms are localized in the In-sublattice and only 53–55% in the Ge-sublattice. The analysis of the binding energy of impurity and host atoms in InP and of their tetrahedral radii indicates [135] that the behavior of germanium and tin cannot be accounted for by a chemical interaction. One must also consider the deformation energy which causes the remarkable amphoteric behavior of germanium, since its tetrahedral radius lies approximately halfway between the tetrahedral radii of indium and phosphorus.

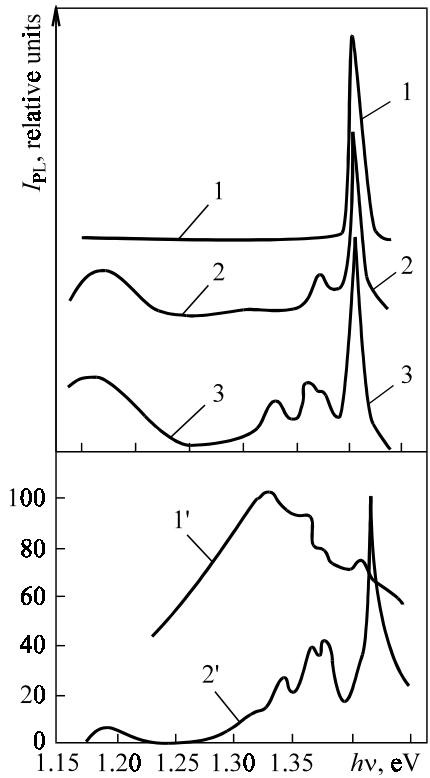

Figure 2.41. Photoluminescence spectra of undoped InP samples grown at 77 K by the following methods [135]: 1 – from the gas phase ($n = 3\times10^{14}$ cm^{-3}); 2 – from a solution melt; 3 – from a stoichiometric melt ($n = 3\times10^{16}$ cm^{-3}); 4, 5 – photoluminescence spectra of InP<Si> samples cut out of various parts of an ingot ($n = N_d - N_a$): 1' – 1.7×10^{13} cm^{-3}; 2' – 5.4×10^{16} cm^{-3}.

Doping with group-IV impurities (Figure 2.41) changes considerably the photoluminescence spectra of doped samples, as compared with undoped crystals. When a strong compensation is present (curve 1), the 1.33 eV line has the same intensity as the 1.41 eV line. The origin of the former, like the origin of the other lines except for 1.41 eV, is unclear and may reflect the ability of silicon to interact with other impurities. This idea is supported by a considerable change in photoluminescence intensity with the doping technology used, as was pointed out in [135].

The photoluminescence spectra of indium phosphide doped with Ge and Sn are shown in Figure 2.41. Clearly, these impurities contribute similarly to the optical properties as well. Their spectral lines at ~1.37–1.36 eV seem to belong to a group-IV impurity in one of the amphoteric states, most likely, in the acceptor state. In other words, these are radiation lines arising during the transition of an electron from a shallow donor level (or from the conduction band) to the acceptor state of amphoteric germanium or tin.

The longer wavelength bands at 1.16–1.15 eV are likely to belong to associates of the A_{IV}–V_p-type of donor–acceptor pairs, but this is to be proved experimentally.

Amphoteric impurities in gallium antimonide. The compound GaSb has a high concentration of defects and is normally of the *p*-type with the concentration of "natural" acceptors of 10^{18} cm^{-3}.

Of interest is GaSb doping with group-IV elements, whose amphoteric behavior has a strong influence on the ensemble of intrinsic defects in both the cation and anion sublattice. In combination with a high compensation degree under heavy doping conditions, this promises new luminescence properties of these materials. Photoluminescence studies of slightly doped GaSb<Si>, GaSb<Ge>, and GaSb<Sn> did not show a significant difference from the doping ($p \sim 10^{17}$ cm^{-3}) with conventional shallow impurities of group-II and group-VI elements.

The authors of [136] studied photoluminescence of heavily doped [$p = (2–5)\times10^{19}$ cm^{-3}] and strongly compensated (K_i = 0.6–0.8) epitaxial GaSb<Sn> *p*-layers. Epitaxial layers grown from the melt and single crystals show a discrepancy between the carrier concentration (p) and the number of tin atoms (N_{Sn}), indicating the amphoteric behavior of the impurity. The compensation degree K_i was found to be 0.6–0.8. It was calculated from the layer composition data obtained analytically with the neutrality equation on the assumption of a complete ionization of impurity atoms at 300 K.

In addition to the edge line ($h\nu_m$= 0.799 eV), the representative photoluminescence spectra had a line corresponding to the first charge state of a natural acceptor at 0.775 eV and bands at $h\nu_m$ = 0.736 eV (*B*), 0.695 eV (*C*), and 0.650 eV (*D*). Of these bands, only *B* and *C* could be reliably attributed

to amphoteric tin, because the investigators [137] used the effect of amphoteric impurity redistribution between the sublattices when the doping with an additional impurity shifted the Fermi level and reduced the number of substituting positions in one of the sublattices. The acceptor impurity used for this purpose was Cd. It was found from the photoluminescence spectra and the compensation degree that the tin concentration in the Ga-sublattice (band *C*) increased but its concentration in the Sb-sublattice (band *B*) decreased. Unfortunately, there are no other reliable reports of the cation–anion amphoteric behavior of other group-IV impurities in GaSb.

Cation–anion amphoteric impurities in CdSb. The behavior of tin is understood better than that of other group-IV impurities. Electrical properties of CdSb samples with known impurity contents were studied experimentally in order to identify the states of tin atoms [138]. It was found that only about 16% of the total amount of impurity atoms in the depletion region were ionized. It was natural to suggest from the diffusion and solubility data that most tin atoms were localized at interstices and that these atoms were electrically neutral, while tin atoms substituting Sb had acceptor properties.

An increase of cadmium vapor pressure decelerated tin diffusion, which was interpreted by the authors of [138] as being due to a higher concentration of vacancies in the Sb-sublattice and to the "pumping" of interstitial tin to the positions V_{Sb}. Obviously, tin atoms must be doubly charged acceptors in these positions and doubly charged donors in the *V* positions. However, there is no experimental evidence for the existence of tin atoms in the Cd-sublattice, and this circumstance does not permit the conclusion to be made that this and other group-IV impurities in CdSb possess cation–anion amphoteric properties.

Cation–anion amphoteric impurities in SiC. According to the data of [139, 140], SiC crystals doped with nitrogen exhibit properties that should be treated in terms of the amphoteric nature of this impurity.

Every silicon or carbon atom in the silicon structure is tetrahedrally surrounded by four atoms of the other sublattice. The SiC structure has a polytype character. The first coordination spheres have a completely identical arrangement of atoms in all SiC polytypes, but in higher spheres, beginning with the second one, there are differences in the arrangement of neighboring atoms. For example, 4*H*–SiC has two unequivalent positions of atoms in the host lattice. The lattices of 6*H*–SiC and 15*R*–SiC have, respectively, three and five unequivalent positions, and so on.

Substitutional impurity atoms will also occupy unequivalent positions in the SiC lattice. EPR data indicate that impurity nitrogen mainly substitutes sites in the Si-sublattice, but there is also a probability, though not very high,

that nitrogen atoms can occupy the C-sublattice sites. It follows from simple considerations that a nitrogen atom capable of substituting atoms in both sublattices must be a donor. Indeed, donor levels *N* have been found to be 101, 158, and 163 meV, counted from the conduction band bottom.

It is quite fair to mention an alternative model, in which a donor center, producing one level at ~100 MeV and another at 150 MeV, is ascribed to background impurity oxygen. Note that the choice of interpretation of experimental levels determines the validity of amphoteric nitrogen content data, since it is the unequivalence of substituted positions which underlies amphoteric properties [139, 140]. The properties are due to the different depths of the energy levels. Of course, the impurity does not change its sign, as is the case with group-IV impurities in both $A^{III}B^{V}$ sublattices. But since the level depth is $E_i = f(Z)$, where Z is the impurity effective charge, then the difference in the ionization energies of the levels is an evidence for different values of Z for impurity nitrogen in different unequivalent positions in the C-sublattice of SiC.

Impurity nitrogen appears to show a similar amphoteric behavior in the Si-sublattice. The values of the three donor levels counted off from the conduction band were found to be 159, 247, and 255 meV [139, 140].

It should be emphasized that further investigations are necessary because the "attribution" of observable lines to foreign impurities or to amphoteric behavior implies different applications of SiC<N> crystals. The former interpretation will require the development of additional deep purification techniques and the latter will require different ways of charge exchange of the levels.

2.4 ISOVALENT IMPURITIES

2.4.1 General concepts

The investigation of isovalent impurity states in semiconductors was initiated in the 1960s [141]. Since that time, interest in the behavior of this class of impurities has been slowly but steadily increasing. The earlier studies were reviewed in the publications [142–144], which describe the basic experimental data and theoretical concepts concerning isovalent impurity effects on electron energy spectra.

To begin with, isoelectron states can be produced in several ways. One way is by heavy doping of semiconductor compounds, for examples, $A^{III}B^{V}$

crystals with ordinary shallow donors and acceptors. Acceptors, say Zn atoms, substitute the A^{III} element and donors, say Te atoms, substitute the B^{V} element. If the number of both types of atoms is large, there is a high probability for them to be located at neighboring lattice sites to form a donor–acceptor pair. Then, the total valence of the pair will be equal to the sum of the valences of the substituted A^{III} and B^{V} atoms.

Another way is to produce a neutral vacancy in a crystal lattice. A single vacancy can be regarded as a kind of "quasi-atom" possessing the same absolute valence as the substituted host atom. Of course, the vacancy's coordination number will be different from that of a normal lattice atom. For this reason, vacancies represent a special type of isoelectron state. It should be mentioned that the role of vacancies in semiconductors has not been considered at this angle.

Finally, there is a more obvious way of creating an isoelectron situation— by doping a crystal with isovalent impurities. Then, the impurity atoms belong to the same group in the periodic table as the host atoms they substitute. This isovalent doping (substitution) is the simplest way of creating isoelectron states, because this is an elemental doping which does not change the coordination number of the substituted crystallographic position.

Today, isovalent impurities are classified into two groups [142, 145]. One group consists of impurities introducing local states into the forbidden gap: either single energy levels or excitonic levels bound by an isoelectron impurity center. The other group includes isovalent impurities, whose energy states are involved in the formation of allowed energy bands. In this case, we deal with what is known as continuous solid solutions with a band spectrum monotonically changing with the concentration of isovalent impurities.

The latter are capable, at the beginning, of producing quasi-local levels in allowed bands [146]. These are so called resonance levels possessing individual properties, like any local levels, but also belonging to a continuous band spectrum, or, more exactly, being in resonance with it.

The ionization of an isovalent impurity can be described by the same reactions as that of any other impurity center [145]:

$$X_0 \xrightarrow{E_I} X^+ + e^- \text{ (donor)}, \qquad X^0 \xrightarrow{E_I} -X^- + e^+ \text{ (acceptor)}, \qquad (2.4.1)$$

where X^0, X^+, and X^- are neutral and charged states of a center; e^- and e^+ are a free electron and a free hole; E_I is the center ionization energy necessary for the reaction (2.4.1) to occur.

The binding energies of charge carriers, E_{bd}, for a donor and an acceptor are described by the neutralization reactions

$$X^+ \xrightarrow{E_{bd}} X^0 + e^+ \text{ (donor)}, \qquad X^- \xrightarrow{E_{bd}} X^0 + e^- \text{ (acceptor)}. \qquad (2.4.2)$$

In these reactions, a charged impurity becomes neutral. The combination of the ionization reaction and the neutralization reaction yields the ionization reaction of the host semiconductor: $0 \xrightarrow{E_g} e^+ + e^-$, where E_g is the forbidden gap width. Hence, we have $E_g = E_I + E_{bd}$; isovalent impurities have low values of E_{bd}. For example, when a phosphorus atom is substituted by a Bi atom in GaP, holes are bound by the energy E_{bd} = 0.038 eV [147]. Since E_g = 2.338 eV in GaP, a Bi_P atom in GaP is a very deep donor with the ionization energy E_I = 2.3 eV. The impurity N_P in GaP binds an electron using the energy E_{bd} = 0.008 eV [148]. Therefore, this impurity is a very deep acceptor with the ionization energy E_I = 2.33 eV. These illustrations show that the problem of isovalent impurities is part of the more general problem of deep impurity centers in semiconductors. Therefore, most of the unsolved problems of deep levels are equally relevant to isovalent impurities.

The isovalent impurity levels are so deep that they can be regarded, in many situations, as coinciding with the allowed band edges E_c and E_v. Thus, isovalent impurities do not practically contribute additional energy levels to the forbidden gap of the host semiconductor, and this is their specific feature.

Historically, there were two distinct periods in the study of isovalent impurities. During the initial period (1963–1977), much research effort was focused on the manifestation of such impurities in nonequilibrium effects observed in a limited number of semiconductor materials. More recently, the focus was on equilibrium characteristics of semiconductors doped with isovalent impurities.

As early as 1963, Gross and co-workers [141] detected a series of narrow lines in photoluminescence spectra of GaP<N> crystals (Figure 2.42). The spectral lines were shown to be associated with the presence of nitrogen in GaP crystals, and its evolution with increasing nitrogen concentration was followed.

The basic idea underlying the interpretation of the GaP<N> spectra was the formation of an exciton bound by an isovalent impurity. This process was represented as consisting of two stages [148]. A neutral isovalent atom first captures an electron (or a hole) in the short-range impurity field and becomes charged. There is already a long-range impurity potential in this state as screened Coulomb potential which pulls out a hole (or an electron).

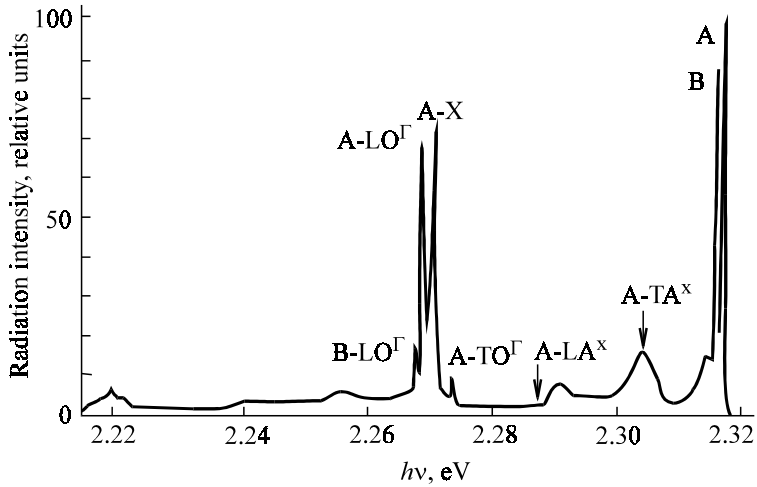

Figure 2.42. Photoluminescence spectra of GaP samples slightly doped with nitrogen at a concentration less than 5×10^{17} cm^{-3}, T = 4.2 K [141].

As a result, the isovalent impurity again becomes neutral, but this new state is an excited one. The electron–hole complex around an isovalent impurity was called an exciton bound by an isoelectron trap [149]. The energy released in exciton annihilation produces narrow *A* and *B* lines in luminescence spectra (Figure 2.42). The other spectral lines represent phonon reproductions. In the first approximation [150], the annihilation energy is the sum of binding energies of both carriers. The binding energy of a carrier in a Coulomb field is often known; so, the luminescence measurements can be used to find the binding energy of the first carrier in the short-range impurity potential field. This is the way the binding energies of most isovalent impurities were determined [142, 143]. For example, the energy of exciton binding by N_P in GaP is 21 meV [151] and by Bi_P in GaP about 107 meV [152].

A similar type of luminescence (recombination radiation) is observed in other semiconductors with isovalent impurities (Table 2.22).

Isovalent impurity pairs capable of binding excitons can arise as the impurity content in the crystal increases. An illustration is GaP heavily doped with nitrogen. An increased nitrogen concentration leads to a shorter N–N distance, and the respective energy lines shift toward the long wavelength region (Figure 2.43).

Table 2.22. Experimental data on the isovalent impurity effect on electron energy spectra in semiconductors.

Type of spectral change	Semiconductor–impurity system	References
Continuous change of $E_g(x)$	$GaAs_{1-x}Sb_x$	[153–155]
	$Ga_{1-x}In_xAs$	[153–155]
	$Ga_xAl_{1-x}As$	[153]
Singularities in $E_g(x)$	$Ge_{1-x}Si_x$	[156]
	CdS<Te>	[153]
Levels inside E_g	ZnTe<O>	[148]
	Ge<Sn>	[157]
	GaP<Bi>; GaP<N>	[148]
	InP<Bi>	[157]
	GaP<Zn,O>	[158]
	GaP<Cd,O>	[159]
	GaP<Li,O>; Si<C>	[160]
Resonance levels	$GaAs_{1-x}Sb_x$	[161]
Bound excitons	CdS<Te>	[162]
	ZnTe<O>; GaP<N>	[148]
	$GaP_{1-x}As_x$<N>	[163]
	AgBr<J>; AgCl<J>	[143]

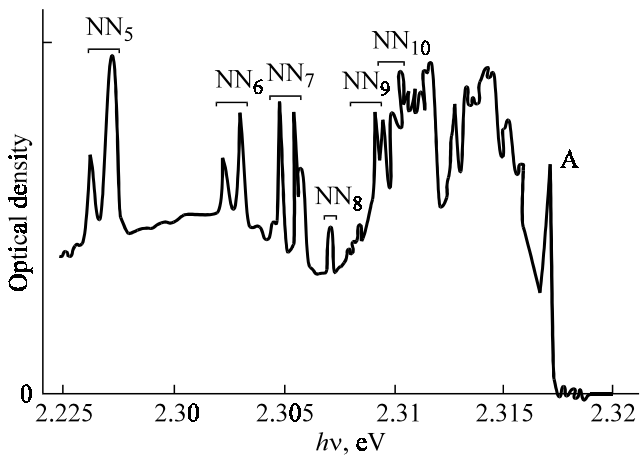

Figure 2.43. Photoluminescence spectra of GaP heavily doped with nitrogen at a concentration 5×10^{18} cm^{-3} at $T = 4.2$ K [144].

Luminescence studies have shown that isovalent impurities can bind not only single excitons but also excitonic molecules [164] consisting of several electrons and holes, but this phenomenon remains poorly understood.

The above two-stage model of exciton binding by isovalent impurities was successful in interpreting the luminescence and optical absorption spectra in various semiconductor–isovalent impurity systems. Nevertheless, many aspects of the impurity involvement in the formation of nonequilibrium properties are still unclear. For example, it is not clear whether an isovalent impurity state can exist with one carrier. In any case, there is no direct evidence for the localization of only one type of carrier at an impurity, even in the well studied GaP<N> system; such transitions, however, seem to occur in GaP<Bi> crystals [146].

Further, it is not clear whether an exciton is formed in a two-stage or one-stage process, i.e., whether the capture of both carriers occurs simultaneously or with a time delay at different potentials. Ryvkin and co-workers [165] suggested a one-stage process of exciton formation in their study of thermal attenuation of the *A* band (Figure 2.42). But later, these authors reconsidered this idea in favor of a two-stage model. The reason for the long duration of afterglow following a temperature increase in GaP<N> crystals is also obscure [167]. Thermal decay of an exciton in these crystals is known to begin at 12 K, but excitonic radiative transitions dominate in the spectrum at room temperature, too.

This phenomenon stimulated the designing of commercial green luminescence emitters on GaP<N> crystals, although many aspects of isovalent impurity behavior still remain unclear.

Recombination radiation studies have revealed bound isovalent impurity states in the forbidden gap in some semiconductors. Of interest is the appearance of an isoelectron situation after double doping in GaP<Zn,O> and GaP<Cd,O>. The bound states here are produced by impurity atoms at neighboring lattice sites. A more complicated isoelectron situation arises in the formation of Li_2O in GaP<Li,O> crystals. It was shown in [168] that an O_P atom and one Li_{Ga} atom substitute neighboring sites, while the other Li atom is in an interstitial position. A complex isovalent impurity with an interstitial C atom was found in silicon [169]. Among special states, which an isovalent impurity can produce, are resonance and antiresonance levels in the allowed electronic spectrum. Experimental observation of such states promises their manifestations in optical phenomena. Today, however, there is only one report [170] of an additional feature at $\hbar\omega = 2.1\,\text{eV}$ in the photoconductivity spectrum of $GaAs_{1-x}Sb_x$ at $x \leq 0.02$, which is interpreted by the authors as a resonance level of isovalent Sb at 0.6 eV in the valence band. This value coincides with the theoretical value obtained by the same authors. Still, the

identification of this spectral feature as a resonance level in GaAs<Sb> requires additional evidence.

Considering the effect of isovalent impurities on electron energy spectra, one should not ignore possible singularities in the band structure. The study of optical absorption spectra in $Ge_{1-x}Si_x$ ($x = 9\times10^{-4}$–1.2×10^{-3}) revealed a discontinuity in the $E_g(x)$ function. Detailed investigations of $E_g(x)$ at very small x in $GaAs_{1-x}Sb_x$ and in $Ga_{1-x}In_xAs$ [154, 155] detected no singularities but, on the contrary, supported the validity of equations of the type $E_g = E_{g0} - ax + bx^2$ [153, 171] describing a monotonic variation of $E_g(x)$ over the whole range of x values. Investigators of equilibrium effects in semiconductors should bear in mind that isovalent impurities produce very deep bound states, if at all. For this reason, their identification from the temperature dependence of equilibrium density of charge carriers, for example, from Hall coefficient measurements, does not seem to be possible. This is an important distinction of isovalent impurities from all other types of impurities. In spite of the fact that isovalent impurities can scatter free charge carriers and, thereby, be detected from the mobility changes, from differential thermal e.m.f. and other transfer phenomena, a lower mobility was actually observed only in GaP<N> crystals [142]. This effect could not be detected reliably in other semiconductor–isovalent impurity systems mostly because of the difficulty of separating it from side effects.

In our view, the isovalent impurity effect on scattering cannot be expected to be strong, since isovalent atoms have a characteristic short-range potential, so they must scatter carriers like neutral rather than ionized impurities. On the other hand, the concentration of isovalent impurity atoms in a crystal is one or two orders of magnitude higher than that of the usual shallow, ionized impurities. So, it is only quantitative evaluations that can help us to answer the question of whether isovalent atoms can scatter charge carriers. The scattering cross section can be approximately taken to be equal to R^2, where R is the extent of the electron wave function. Figure 2.44 shows that the wave function maxima of an isovalent atom and an ordinary impurity are shifted. The R_{IVI}^2/R^2 ratio is less than 1/30. This means that the concentration of isovalent atoms can be increased by a factor of 30, as compared with that of an ordinary impurity, to make both types of atoms scatter carriers identically. When recalculated as the isovalent impurity content, these values show that an additional scattering by isovalent impurities can be expected to manifest itself at $N_{IVI} > 0.1$ at. %.

Isovalent impurity atoms do not usually produce energy levels in the forbidden gap of a doped semiconductor but they are involved in chemical and quasichemical reactions in the liquid phase, from which a crystal is grown,

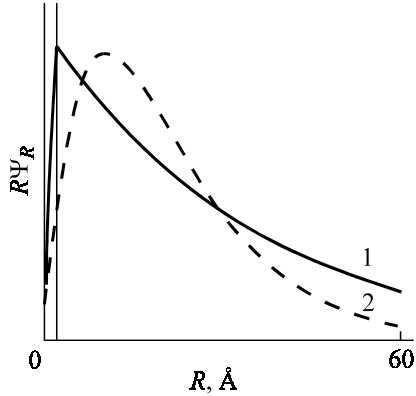

Figure 2.44. The radial wave function of an electron with m_e = 0.35, localized in a potential well of 2 E width and 10 meV depth (1) and of an electron bound by a charged donor with the binding energy 100 meV (2): 1 – short-range potential; 2 – Coulomb potential; vertical line – the potential well boundary [149].

and in the solid phase during the crystal cooling or thermal treatment. Thereby, isovalent impurities strongly affect the solubility and distribution coefficients of other impurities, the homogeneity region boundaries of a semiconductor compound, and the formation of defect associates. As a result, isovalent impurities change the arrangement of background impurities and defects in a crystal.

Below, we discuss mechanisms and give illustrations of this "indirect" influence of isovalent impurities on the properties of semiconductors.

2.4.2 Empirical models of isovalent impurities

Because of the absence of a rigorous theory, various empirical models have been suggested to describe the electron spectrum of deep impurities, in particular of isovalent impurities. For example, carrier binding by such an impurity was considered to be due to the difference in the electronegative charge $\Delta\chi$ of the substituted and isovalent atoms. At $\chi_{IVI} > \chi_a$, an impurity atom must bind an electron to become a deep acceptor. But at $\chi_{IVI} < \chi_a$, a hole will be bound, and the isovalent atom must become a donor.

Table 2.23 summarizes electronegative charges of atoms measured by three workers. The differences in these data are of little importance to this

Table 2.23. Electronegative charges of atoms, as measured by different authors [172].

Atom	Pauling	Phillips	John	Atom	Pauling	Phillips	John
Al	1.5	1.18	1.48	Ge	1.8	1.35	1.59
Ga	1.6	1.13	1.46	Sn	1.8	1.15	1.40
In	1.7	0.99	1.32	Pb	1.8	1.09	1.41
Te	1.8	0.94	1.37	N	3.0	3.00	2.99
C	2.5	2.50	2.50	P	2.1	1.64	1.87
Si	1.8	1.41	1.64	As	2.0	1.57	1.71
B	2.0	2.00	2.00	Sb	1.9	1.31	1.52
				Bi	1.9	1.24	1.46

review, although John's data obtained from nonlocal atom pseudopotentials are more accurate.

According to the data of Table 2.23, $\chi_N > \chi_P$, so N_P in GaP must be a deep acceptor. On the contrary, $\chi_{Bi} < \chi_P$ and $\chi_{Sb} < \chi_P$, so Bi_P and Sb_P in GaP must be deep donors. These predictions agree with experiments. However, As_P in GaP and P_{As} in GaAs are expected to be a donor and an acceptor, respectively, but experimental studies do not reveal any bound states in these two cases at all. Similarly, the behavior of N, Bi, and Sb in GaAs is expected to be the same as in GaP. However, neither nitrogen [144], nor Bi [173] nor Sb [174] show bound states in GaAs.

The donor states of Bi in InSb reported by some researchers [175, 176], were later re-identified as manifestations of interstitial bismuth [176]; therefore, they have nothing to do with the states to be predicted from electronegative charges.

A limitation of the electronegative charge approach is that it completely ignores the genetic affinity of the bound states of an isovalent impurity to the conduction band for an acceptor and to the valence band for a donor. This drawback can be removed by comparing the atomic pseudopotential depths [177] for *s*- and *p*-states, instead of electronegative charges. The data analysis [178] shows that if the impurity atom is heavier (or lighter) than the lattice atom, the isovalent impurity can bind a hole (or an electron). It appears, in fact, that the use of atomic pseudopotentials contributes nothing new, as compared with the electronegative charge rules. The demerit of both approaches is the complete neglect of electron polarization and local lattice distortions in the defect vicinity [179]. The character and value of local lattice distortions are, in turn, determined by the nature of forces binding an electron to an impurity atom with a short-range potential.

If we use the Koster–Slater approximation and approximate the impurity potential by a rectangular potential well, as was done in [149], then the real space wave function ψ of a bound electron outside the well can be written as

$$\Psi_{\mathrm{IVI}} \sim \frac{1}{R}\exp\left(-\gamma_1 R\right), \tag{2.4.3}$$

where R is the vector radius of an electron;

$$\gamma_1 = \frac{1}{\hbar}\left(2m_e U_{\mathrm{IVI}}\right)^{1/2}, \tag{2.4.4}$$

where m_e is the effective mass of an electron and U_{IVI} is the potential well depth.

If, however, we deal with an ordinary donor retaining an electron due to Coulomb potential, the wave function Ψ_d is

$$\Psi_d \sim \exp\left(-\gamma_2 R\right), \tag{2.4.5}$$

where γ_2 has the same form as in (2.4.4), except that U_{IVI} is replaced by the energy of electron binding to the donor nucleus.

Figure 2.44, borrowed from [149], shows the radial wave function for an electron localized in a potential well of 2 E wide and 10 meV deep. Although the binding energy for an electron bound by an ordinary donor is an order of magnitude higher than that for an electron bound by an isovalent impurity, the localization of the impurity is, on the contrary, higher than that of the ordinary donor. Due to the strong localization, the excess electron (negative) charge is concentrated in the central unit cell, whereas the neighboring atomic skeletons are charged positively. As a result, the isovalent atom and the nearest host atoms of the same sublattice have to come closer to one another. The influence of the other sublattice atoms remains unclear.

The key role of atomic shifts in the vicinity of an isovalent impurity atom was also emphasized in [180]. Phillips, in contrast, believes [181] that local lattice distortions reduce the difference $\Delta U = U_{\mathrm{IVI}} - U_a$ until the potential well disappears completely. But then, isovalent impurities will bind neither electrons nor holes. The fact that some isovalent impurities do form bound states makes one suggest the presence of some repulsive forces leading to $\Delta U \neq 0$. One can see, therefore, that the isovalent impurity problem involves the problem of band structure and that of crystal lattice dynamics.

There are other empirical approaches to the explanation of energy level spectra of isovalent impurities.

Allen [182, 183] tried to find ΔU from the difference in the elastic properties of an isovalent impurity and a substituting atom. Eventually, this approach reduces to the consideration of the difference in the sizes of both atoms, and so it cannot be regarded as satisfactory either.

Zakharov and Shcherbak [180] tried to find ΔU as the difference in the forbidden band widths at the band edges $E_g(x)$, where x is a mole fraction of the isovalent impurity. This value was normalized to a unified unit cell volume of the initial crystal ($x = 0$) as

$$\Delta U = \Delta E_g - \frac{\partial E_g(x=1)}{\partial \ln V}\frac{\Delta V}{V}, \tag{2.4.6}$$

where $\Delta E_g = \Delta E_g|_{x=0} - E_g|_{x=1}$ and $\Delta V/V$ is a relative change in the crystal volume at the transition from $x = 1$ to $x = 0$.

The condition for an energy level to appear in the forbidden gap is

$$\Delta U \frac{2}{E_v}, \tag{2.4.7}$$

where E_v is the valence band width. Seven cases (GaP : Sb; ZnTe : Cd; GaP : In; AlP : In; PbSe : Te; CdTe : S; CdSe : Te) out of 45, calculated in [180], should be expected to exhibit local levels inside the forbidden gap. Unfortunately, these systems have not yet been studied experimentally.

The mechanism of production of isovalent bound states is clear qualitatively. The difference in the electronegative charges or ion pseudopotentials of the two atoms, their screening by valent electrons and local lattice distortions give rise to a short-range potential allowing an isovalent atom to capture a charge carrier. This potential must exceed a certain threshold value determined by the kinetic energy of a free carrier in the band, in which the bound state originated. This is the conduction band for deep acceptors and the valence band for donors.

The threshold energies E for deep acceptors or deep donors depend only on the semiconductor band structure and can be calculated by the empirical pseudopotential method [184]. These energies for isovalent impurities which bind holes (deep donors) are shown in Figure 2.45 as a function of ionic charge of the chemical bonding [185, 186] in binary semiconductors. One can see that the conditions for the hole binding by isovalent impurities are

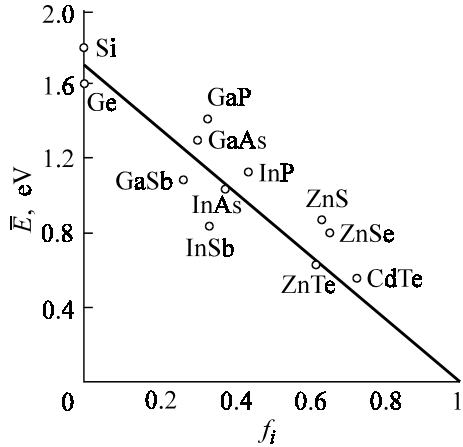

Figure 2.45. The threshold energy E for isoelectron donors as a function of ionic binding f_i for some semiconductors; the straight line – $E = 1.7(1 - f)$ eV.

facilitated in ionic crystals having a narrower valence band. With a 10% error, the data in Figure 2.45 fit well the dependence $E = 1.7(1 - f_i)$ eV [178].

Since isovalent impurity potential $U(r)$ in diamond or zincblend types of crystal has the point group symmetry T_d, the impurity potential matrix proves to be diagonal in the base of three Vannier functions in the conduction band, i.e., $\overline{U}_{\alpha\beta} = \overline{U}\,\delta_{\alpha\beta}$. For this reason, the condition for bound states of deep donors does not differ from that of deep acceptors, except for $\overline{U} < 0$ and $\overline{E} < 0$ for deep acceptors and $\overline{U} > 0$ and $\overline{E} > 0$ for deep donors.

Table 2.24. Matrix elements U (eV) for various cases [189].

Doped semiconductor	U_{LS}	$\overline{U}$ (X=0) unscreened	$\overline{U}$ (X=0) screened	$\overline{U}$ (X=1) screened	X	Bound states theory	Bound states exp.
Si<Ge>	0.04	0.32	0.08	0.54	0.41	No	No
Si<Sn>	0.14	1.96	0.45	2.34	0.36	No	–
Si<Pb>	0.43	2.45	0.72	3.63	0.35	No	–
GaP<As>	0.06	1.24	0.30	0.76	0.43	No	No
GaP<Sb>	0.20	4.49	1.07	2.80	0.37	Yes	No
GaP<Bi>	0.72	5.60	1.71	4.12	0.35	Yes	Yes
ZnS<Se>	0.11	4.81	1.20	1.01	0.51	No	No
ZnS<Te>	0.28	9.46	2.41	2.98	0.48	Yes	Yes

To calculate $\overline{U}$ for deep donors, one should take into account

(1) the difference in ionic pseudopotentials of the two atoms [178];
(2) the screening of this difference by the crystal valent electrons [187];
(3) local distortions of the lattice [182];
(4) the effect of spin-orbital interaction [184].

The latter circumstance can be easily taken into account as the difference between model relativistic pseudopotentials of the two atoms [188], because the respective contribution to $\overline{U}$ is independent of the local lattice distortions. The contributions to $\overline{U}_{LS}$ for some deep donors are presented in Table 2.24. For the binding by isovalent impurities, only symmetrical local distortions of the lattice are important [178].

Generally, the distance d between an impurity atom and the nearest lattice atom differs from the unperturbed bonding length d_i, defined by the sum of covalent radii and from the distance d_k between neighboring atoms in a perfect crystal. An elastic strain field is induced near the impurity atom.

The contribution of the elastic strain field to $\overline{U}$ can be accounted for by assuming that the shifts of remote atoms are described by elastic continuum theory [190]. Then, the elastic strain field is fully defined by prescribing the distance between the impurity atom and its nearest neighbor or by prescribing an equivalent parameter $X = (d - d_k)/(d_i - d_k)$, which can vary from $X = 0$ in the absence of local distortions to $X = 1$ for maximum lattice distortions near the impurity atom.

The value of $\overline{U}$ at $X = 0$ is estimated as the locally screened difference between the ionic pseudopotentials of the impurity and host atoms [182]. One can see from Table 2.24 that the screening of the impurity ionic potential considerably decreases $\overline{U}$, so that the conditions for the binding by an isovalent impurity are not satisfied, except for Bi_P in GaP and Se_S and Te_S in ZnS ($\overline{E}$ values are given in Figure 2.45). At $X = 0$, i.e., in the absence of local lattice distortions, Se_S in ZnS must bind a hole, but this contradicts experimental data.

The value of $\overline{U}$ at $X = 1$ can be roughly estimated from the crystal pseudopotential difference, with the pseudopotential pertaining to the central unit cell [178]. One can make use of the empirical pseudopotentials from [184] to obtain the data of Table 2.24. The behavior of $\overline{U}$ in the intermediate region is a linear extrapolation of the data for $X = 0$ and $X = 1$, as is done in Figure 2.46 with the primes marking the threshold energy for hole binding.

It is clear that local lattice distortions stimulate the capture of holes by isovalent impurity atoms. The same is true of the capture of electrons. The calculation of the equilibrium value of X for each particular case requires

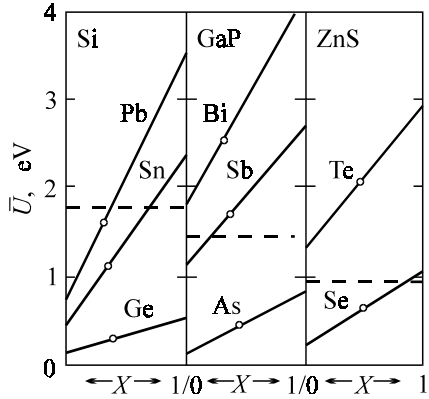

Figure 2.46. The dependence of a host element $\overline{U}$ for isoelectronic donors on the X parameter of local lattice distortions: points – X values in equilibrium conditions; dashed lines – threshold energies E.

minimization of the total energy of a crystal containing an isovalent impurity. The equilibrium X values were estimated in [178, 182] by minimizing the elastic energy in terms of the known force constants for the Martin adiabatic potential [191]. These values are marked by dots in Figure 2.46 and presented in Table 2.24. We see that the agreement between experiment and theory for the hole binding by isovalent impurities is quite satisfactory. The conclusion about a higher binding energy between a quasiparticle and an isovalent impurity due to local lattice distortions is valid only for neutral isovalent impurities in quasi-equilibrium conditions, when the impurity empty level for an electron is much higher. Under nonequilibrium conditions, when an impurity captures a quasiparticle, the binding energy may decrease because of the local distortions, as described by Phillips [182]. The mechanism suggested by Phillips is based on the assumption that the lifetime of a quasiparticle bound by an isovalent impurity is short, as compared with the lattice vibration period. Then, the hole binding by an impurity may be affected by asymmetrical local distortions (Yan–Teller effect). The conditions for quasiparticle binding, of course, imply its long lifetime.

The problem of localized states in crystals containing isovalent impurities was first formulated generally by Lifshitz [192]. The Hamiltonian $H^0 = T + V^0$ of a perfect crystal with the kinetic energy operator T and potential energy V^0 has a quasicontinuous spectrum of elementary excitations, but the intrinsic Hamiltonian states $H = T + V$ in a similar crystal cannot be numbered with a wave vector. An impurity atom introduces a localized

perturbation $U = V - V^0$, which may give rise to discrete energy levels in the forbidden gap (bound states) or to an essential rearrangement of the levels in the quasicontinuous spectrum (virtual states).

The most detailed description can be achieved with the operator $G(E) = (E - H)^{-1}$ [191–193]. However, it is more convenient to introduce another operator, $Q(E) = 1 - G^0U$, with the Green function for a perfect crystal $G^0(E)$ and impurity potential U. In the presence of perturbation, the spectrum is identified by solving the equation $\mathrm{Re}\Delta(E) = \mathrm{Re}\ \det Q(E) = 0$. Lifshitz [192] was the first to point out that the impurity potential matrix $U(\boldsymbol{r})$ can often be approximated by a finite order matrix, so that the order of the determinant $\Delta(E)$ appears finite for isovalent impurities.

The choice of the base for $Q(E)$ representation is of no principal importance. One can use unlocalized base functions (Bloch functions [194] or plane waves [195]). The order of ΔE in this case will be defined by the intrinsic state extent. More convenient are localized base functions (Vannier functions [196–199], atomic orbitals [200], and generalized Vannier functions [201]). Here, the order of ΔE is determined by the impurity potential extent. For bound impurity states, the wave function covers a larger number of unit cells than the impurity potential, so the use of localized base functions is preferable [192].

The wave function of the virtual state $|\nu>$ with energy E_ν satisfies the ordinary Lippman–Swinger equation

$$|\nu> = |k> + G^0(E_\nu)U|\nu>,$$

where the Bloch function $|k>$ corresponds to the state of a perfect crystal with energy $E^0(k) = E_\nu$, and $G^0(E)$ is taken to be $\lim G^0(E + i\delta)$ at $\delta \to 0$. The density of electron states in a doped crystal

$$\rho(E) = -\frac{1}{\pi}\mathrm{Im}\,Sp\,G(E) \tag{2.4.8}$$

differs from the density of states $\rho^0(E)$ in a perfect crystal by the value

$$\Delta\rho(E) = \rho(E) - \rho^0(E) \cong \frac{1}{\pi}\mathrm{Im}\frac{\mathrm{d}}{\mathrm{d}E}\ln\Delta(E) \tag{2.4.9}$$

and satisfies the following condition [179]:

$$\int \mathrm{d}E\Delta\rho(E) = 0\,. \tag{2.4.10}$$

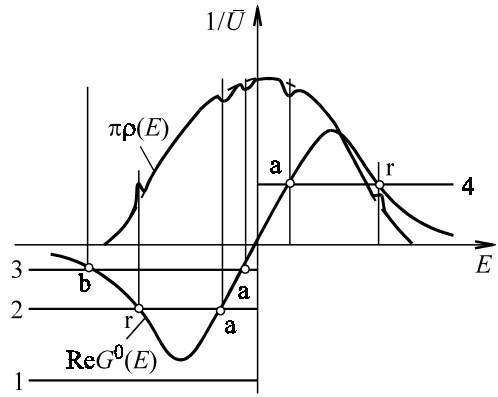

Figure 2.47. Bound (b), resonance (r), and antiresonance (a) levels in a semiconductor doped with an isovalent impurity.

The wave function of a bound state satisfies another equation:

$$|\nu>=G^0(E_\nu)U|\nu>, \tag{2.4.11}$$

where $G^0(E_\nu)$ is the real function of energy E_ν. For bound states, $\mathrm{Re}\Delta(E_\nu) = \mathrm{Im}\Delta(E_\nu) = 0$, and from (2.4.9) we have $\Delta\rho(E) = \delta(E - E_\nu)$. If there are only N_δ bound states, expression (2.4.10) can be re-written as $dE\Delta\rho(E) = N_\delta$, where the integral is taken only with respect to the quasicontinuous spectral energies. Clearly, the appearance of bound states is always accompanied by a rearrangement of quasicontinuous spectral levels.

The quantity $\Delta\rho(E)$ often proves to be large only in certain spectral regions known as resonances and antiresonances. For resonances, $dE\Delta\rho(E)$ is equal to a positive integer and for antiresonances it is a negative integer. In the absence of bound states, the number of resonances is the same as that of antiresonances. The number of bound states is smaller than that of antiresonances exactly by the number of resonances. Antiresonance levels are often referred to as hole resonances.

Figure 2.47 illustrates a typical situation taking place in a semiconductor with an isolated band, density of states $\rho^0(E)$, and the Gilgert transformant

$$\mathrm{Re}\,G^0(E) = \int \frac{dE'\rho^0(E')}{E - E'}. \tag{2.4.12}$$

If there is a deep donor with average potential $\overline{U} < 0$, the straight line $1/\overline{U}$ will not intercept $\mathrm{Re}G^0$ at any point at small $|\overline{U}|$ (line 1). In this case, there will be neither bound nor virtual levels. At higher $|\overline{U}|$, a resonance state (p) and an antiresonance state (*a*) arise (line 2), and a peak and a valley appear in the density of states line $\rho(E)$. By increasing $|\overline{U}|$ further (line 3), one can obtain a bound state (c) and an antiresonance state (*a*). It is seen that there is a critical value of $|\overline{U}|$ separating the regions of bound and resonance states.

The foregoing is equally applicable to deep acceptors with $\overline{U} > 0$. Line 4 in Figure 2.47 describes the case when an acceptor produces one resonance level (p) and one antiresonance level (*a*) in the isolated band. Again, there is a critical value of $|\overline{U}|$ separating the regions of bound and resonance states. In principle, we can imagine a semiconductor with $\mathrm{Re}G^0(E) > 0$ in the energy region of interest [198, 199]. Of course, localized states can arise in this semiconductor only in the presence of deep acceptors.

This illustration is not an exact reproduction of a real situation; it is known as the Koster–Slater one-band/one-site model [196]. This model results from the use of Vannier functions as the basis, ignoring all nondiagonal matrix elements of impurity potential, i.e.,

$$< a_n^0(r - R_i)\left|U(r)\right|a_m^0(r - R_{i'}) >= U\delta_{n0}\delta_{m0}\delta_{i0}\delta_{i'0}\,, \qquad (2.4.13)$$

where 0 indicates the isolated band number and the central impurity cell number, simultaneously. It is within the framework of the Koster–Slater one-band/one-site model that the applicability of the quasiclassical approximation to isovalent impurities can be substantiated and the conditions for the existence of bound states formulated.

The subdivision of localized isovalent impurity states into bound, resonance, and antiresonance states is absolutely rigorous. For resonances and antiresonances, $\mathrm{Im}\Delta(E_v) \neq 0$, and the variation in the density of electron states is approximately described as

$$\Delta\rho(E) = \frac{1}{\pi}\frac{\Gamma}{(E - E_v)^2 + \Gamma^2}\,, \qquad (2.4.14)$$

where Γ, characterizing the resonance width ($\Gamma > 0$) or the antiresonance width ($\Gamma < 0$), is defined as

$$\Gamma = \frac{\mathrm{Im}\,\Delta(E)}{\mathrm{Re}\,\Delta'(E)}\,. \qquad (2.4.15)$$

Here, the prime indicates a derivative with respect to energy.

In a simple case, Γ may be independent of U and have the form

$$\Gamma = -\pi\rho^0(E_v)\Big/\frac{\mathrm{d}\,\mathrm{Re}\,G^0(E)}{\mathrm{d}(E)}. \qquad (2.4.16)$$

The $\hbar/2|\Gamma|$ ratio is a lifetime measure of the virtual state. [198]. Large lifetimes are characteristic of sharp resonances which normally lie at the wings $\rho^0(E)$, where Re $G^0(E)$ varies rapidly with energy.

The Koster–Slater model is unable to describe adequately all aspects of the isovalent impurity spectrum. For example, Faulkner [150] failed to account for the energies of excitons bound by NN_i pairs in terms of this approximation using potential U fitted to the energy of exciton binding by an isolated nitrogen atom. Nevertheless, Faulkner's one-band multisite model [150], accurate within the limits of very long-range potentials, provides a nearly correct series for impurity potential as a Gaussian well. Still, this series does not approach the A line of an exciton bound by an isolated impurity as fast as the observable excitonic lines bound by atomic pairs (Figure 2.43). Therefore, a very short-range potential of an isovalent impurity in a semiconductor does not permit an adequate description of all aspects of this problem. Clearly, effective potential must allow for the crystal lattice relaxation and electron polarization. These effects are of a longer-range type than the impurity ionic potential. They will decrease the binding energy sensitivity to average potential energy and make the interband matrix elements weaker, which may result in splitting off the excited states from the allowed band edges. However, there must be a strong short-range impurity potential, too, because it provides a strong optical absorption in GaP<N> crystals.

Jaros and co-workers [194, 202] calculated the local self-consistent impurity potential N_P in GaP in terms of the multiband model and found that the electron polarization led to an oscillating potential with short-range attraction components and long-range repulsion components. In this situation, the binding energy of an electron bound by N_P in GaP is close to zero due to the compensation of short- and long-range potentials. This compensation was neglected by Faulkner [150] who ignored the valence band effect and the potential repulsion components.

But Brand and Jaros [202] failed to find a satisfactory explanation for the excitonic series NN_i in GaP, although the predicted electron binding energies are close in their order of magnitude to experimental values. The discrepancy is due to the neglect of local lattice distortions and electron–hole correlation. Cohen and Sturge [203] subtracted from the observable exciton binding

energy the hole binding energy E_h found by extrapolation over excited excitonic states. For N_P in GaP, they obtained $E_h \sim 40$ meV even for remote NN_i pairs. This fact confirms the applicability of the phenomenological bound exciton model suggested by Hopfield and co-workers [148] but questions Allen's model [204].

Some workers [205] have suggested that the potential of a nitrogen atomic pair in GaP contains two components: a short-range electron attraction owing to the difference in the ionic pseudopotentials and electron polarization, and a long-range component due to local lattice distortions. In contrast to Phillips [181] who considered the long-range part as repelling an electron, they assumed the attraction. Thus the model impurity potential was approximated by

$$U(r) = -U_0 \quad \text{for } r \le a_0 ,$$
$$U(r) = -U_0 \left(\frac{r}{a_0} \right)^3 \quad \text{for } r > a_0 . \tag{2.4.17}$$

The width a_0 and the depth U_0 were to be found from experimental data on the energies of electron binding by NN_i in GaP.

The electron states bound by NN_i at distance R_i are described, in a quasiclassical approximation, by the Hamiltonian [204, 205]

$$H = -\frac{m}{\overline{m}} \nabla^2 + U(r) + U(r + R_i) , \tag{2.4.18}$$

where the effective mass m does not coincide with that of a carrier in the vicinity of the conduction band maximum. The energies of bound states were determined by the variational analysis with the test function

$$\Phi_i(r) = \left[2(1 + \Delta_i) \right]^{-1} \left[U_i(r) + U_i(r + R_i) \right], \tag{2.4.19}$$

where

$$U_i(r) = \left(a_i^3 / \pi \right)^{1/2} \exp(-a_i r) , \quad \Delta_i = \left(1 + \alpha_i R_i + \alpha_i^2 R_i^2 / 3 \right) \exp(-\alpha_i R_i)$$

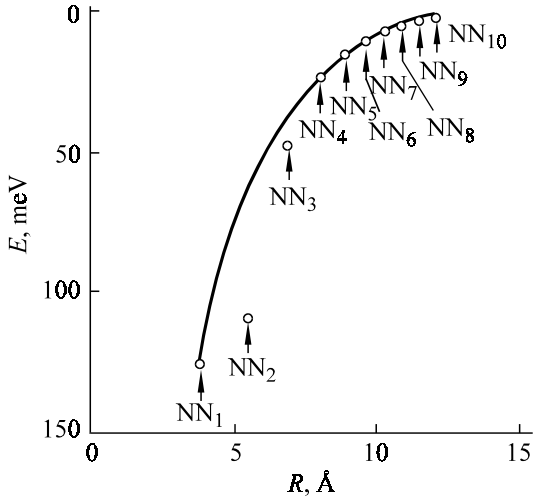

Figure 2.48. The binding energy E of an electron bound by a pair of nitrogen atoms NN_i in GaAs versus interatomic distance R.

and α_i is a variational parameter. One can see from the data in Figure 2.48 that an agreement with experiment was achieved at $(ma_0^2U_0)^{-1} = 0.88$, so that with $a_0 = 1.185$ E, we will have $U_0 = 3.08$ at $\overline{m} = m$.

This shows that the calculation of electron binding by NN_i in GaP must take into account all components of impurity potential, including the ionic pseudopotential difference, electron polarization, and local lattice distortions. Moreover, one should bear in mind that local distortions increase the electron binding energy, rather than decrease it. It follows from the above quasiclassical calculation that isolated nitrogen atoms, namely NN_∞ pairs, cannot bind electrons at all, but they, of course, are able to bind excitons. Generally, this calculation does not contradict the suggestions made by Baldareschi and Hopfield [178, 182] but rejects the applicability of Phillips' model considerations [181] concerning the role of local distortions.

The problem of binding quasiparticles by atomic pairs of the Zn–O, Cd–O, and Mg–O types in GaP crystals was discussed in [206], ignoring local lattice distortions. The symmetry group for these defects is the point group C_{3V} containing nondegenerate σ-like A_1 states and doubly degenerate π-like E states. The latter arise from p-like T_2 states of an isolated impurity and have energies close to those of T_2 states because of their π-like nature. But A_1 states are of the σ-type, so that their energy can vary considerably with the kind of metal atom in the atomic pair.

Coupling of a metal atom with oxygen can considerably change the energy of electron binding by O_P in GaP, depending on the electron energy difference in isolated atoms. The bound states appears to be of the resonance type for the larger energy difference.

Although it is clear from the foregoing that the theory of isovalent impurities in semiconductors is not yet completed, it is as clear how it can be developed further. Exciton binding by isovalent impurities seems to be interpreted satisfactorily in terms of Hopfield's model [144], in which an isoelectronic acceptor first captures an electron by short-range forces of the central cell, and the induced Coulomb field then captures a hole. The exciton binding energy is given by the sum of binding energies of the electron and the hole [150]. Since the hole binding can be described by the effective mass method, the main problem that remains is to describe the electron binding by short-range forces of the central cell. The difficulty is to calculate the hole binding energy for an isolated donor.

The short-range forces of the central cell are determined by three factors: ionic impurity potential, electron polarization, and local lattice distortions. The Koster–Slater one-band/one-site approximation can take into account only the first factor. Both the electron polarization and local distortions extend the range of impurity potential, so the Koster–Slater approximation proves unsuitable for the calculation of binding energies of electrons and holes. On the other hand, this approximation with empirical impurity potential satisfactorily describes optical properties of isovalent impurities.

It is quite obvious that the binding of an electron or a hole by an isovalent impurity atom requires a self-consistent theory of the type suggested in [207, 208] for a neutral vacancy. Such a theory can yield a self-consistent impurity potential for any fixed configuration of atoms and vary the energy of the impurity ground state in a crystal to determine the equilibrium configuration. It may turn out that the stage of the impurity potential self-consistency will appear unnecessary, because Lannoo [209] obtained similar results by means of a simple screening function.

The principal difficulties are associated with the equilibrium configuration calculation. In principle, evaluation of the band structure contribution to the ground state energy does not present a problem, but the contribution of ion–ion repulsion and the change in the exchange–correlation corrections are not easy to consider. Fortunately, the contribution and corrections determine the short-range adiabatic potential, so one can use Kiting's phenomenological adiabatic potential, whose parameters are to be selected from experimental data. This has not been done yet, but the approach seems quite promising for isovalent impurities in semiconductors.

Energy variation of the center ground state with its charge state defines its ionization energy. For an isovalent impurity, ionization energies are high, about the forbidden gap width, while bound states are fairly extensive. In these conditions, one should not expect an essential contribution of exchange and correlation to the ionization energy (in contrast to multicharge centers). For this reason, we have to recognize the important role of local lattice distortions in the formation of bound impurity states. While only symmetrical local distortions seem to be essential for isoelectronic acceptors, asymmetric local distortions should not be neglected when dealing with isoelectronic donors. The possible Yan–Teller effect was emphasized by Morgan [210].

Electron polarization and local lattice distortions leading to long-range components of impurity potential must be treated in terms of a multiband isovalent impurity theory. It is quite likely that such a theory could be developed in the spirit of Kleiman's one-site model [211], which considers long-range potential by renormalizing the perfect crystal spectrum. Although this model contains some flaws, it is capable of explaining the presence of excited states of isovalent impurities quite clearly.

Little attention has so far been given to virtual impurity states which are to occur with a higher probability than bound states. Experimental observation of such states is possible only in sharp resonances lying near the forbidden gap of a semiconductor. If this ever happens, one should expect to find specific features in the photoconductivity and optical absorption spectra. The effect will certainly depend on the details of the resonance state, the symmetry and localization degree of the wave function, as well as on the interaction with the lattice and the presence of other impurities. Experimental study of virtual states of isovalent impurities may contribute much to our understanding of the nature of isovalent substitution.

2.4.3 Physicochemical behavior of the host–IVI system

Elemental isovalent impurities belong to the same groups in the periodic table as the elements comprising the host semiconductor. For this reason, the physicochemical behavior of such impurities have been studied with reference only to binary (for Ge and Si) and quasibinary (for $A^{III}B^{V}$, $A^{II}B^{VI}$, $A^{IV}B^{VI}$, etc.) systems characterized mostly by infinite substitutional solid solutions of the cation and anion types. The fundamental characteristics of such solid solutions—the crystal lattice period and forbidden gap width—were assumed to vary monotonically with the composition. This refers to solutions with macrocontents (from units to dozens at. %) of isovalent atoms. Thermodynamic studies of solid solutions with a low content (less

than 1–2 at. %) of one of the components and the physicochemical analysis of semiconductor–isovalent impurity systems have demonstrated some principal differences between a diluted and a concentrated solution. These differences show that it is impossible to extend the physicochemical concepts of donor and acceptor impurities to the behavior of isovalent impurities at their "microconcentrations".

Briefly, the specific aspects of the physicochemical behavior of isovalent impurities can be stated as follows.

(1) Experiments using the electromotive force of a galvanic cell and differential calorimetry have revealed sign reversal deviations from the Raoult law [212], negative for diluted, and positive for concentrated InSb-Bi solutions.

(2) Measurements of vapor pressure above Bi and InP–Bi liquid solutions have shown positive deviations from the Raoult law and a tendency for liquid stratification [213].

(3) Measurements of electrical conductivity, viscosity, and density of InSb-GaSb and InSb-GaAs liquid solutions have shown that the liquid phase contains well-ordered regions corresponding to "triple chemical compounds" of the Ga_2InSb_3 and In_2GaSb_3 types [214].

(4) High precision measurements of the crystal lattice period and X-ray diffuse scattering intensities have demonstrated that isovalent impurities could, under certain conditions, occupy sites and interstices simultaneously in crystal–isovalent impurity solid solutions. Among these are GaAs<Sb>, GaAs<In>, InSb<Bi>, and some others [176].

The most important feature of the physicochemical behavior of isovalent impurities is their dual role in semiconductors. On the one hand, they are ideal solvents (Ga, In, Bi, InBi, In_2Bi, Sn, etc.) possessing a low melting temperature and a low vapor pressure. On the other hand, they are suitable as doping impurities in liquid phase epitaxy. The latter decreases the crystallization temperature and allows the growth of semiconductor films with a low defect concentration.

2.4.4 Possible mechanisms of the isovalent impurity effect

Isovalent impurities can affect a semiconductor in the liquid and solid phases.

Through the liquid phase, an impurity admixture increases the number of constituents, so that the crystal grows from $(n + 1)$ components, as compared with n components in the absence of impurity. The quantity n also includes various background impurities. A typical example is epitaxial GaAs crystal

growth. The basic components are GaAs and Ga with Si as the main background. The degree of Si incorporation into the growing crystal is described by the distribution coefficient $K_{Si} = (C_s/C_l)_{Si}$. When In or Sb is added to the liquid phase, the distribution coefficient of silicon impurity incorporated into the crystal changes because of the interaction between In or Sb with silicon. Of course, the distribution coefficients of the isovalent impurity itself will also vary with its content in the liquid phase and with temperature.

In other words, isovalent impurity doping changes the heterogeneous equilibrium of the melt-layer system. Therefore, it can be stated that the interaction between background and doping impurities will be most effective only if the impurity has the highest thermodynamic interaction parameter [215]. This qualitative conclusion will be supported by quantitative evaluations in Section 2.4.5.

The effect of isovalent impurities on the point defect ensemble in the solid phase is associated with their interaction with vacancies. Since isovalent impurities are uncharged, their interaction is likely to be purely elastic. This issue was treated theoretically in detail in [216]. A point defect is regarded by the authors of this work as a source of internal stresses. The principal theoretical result of their consideration is the analysis of interaction energy $E(\tilde{R})$ of defects in an isotropic crystal:

$$E\left(\tilde{R}\right) \sim \tilde{R}^{-6}, \tag{2.4.20}$$

where $\tilde{R}$ is the distance between two defects.

It is clear from (2.4.20) that one must first take into account the interaction of isovalent impurities with other defects, for example, with vacancies located in the first coordination sphere. The probability for defects to encounter one another at neighboring sites will be described as [217]

$$W_{V-IVI} = 12aN_V N_{IVI} D_V, \tag{2.4.21}$$

where a is the crystal lattice period, D_V is the coefficient of the most mobile defect—vacancy, and N_V and N_{IVI} are the concentrations of vacancies and isovalent impurities, respectively. For a common electrically charged impurity, for instance silicon, the probability of the encounter with a vacancy is

$$W_{V-Si} = 12lN_V N_{Si} D_V, \tag{2.4.22}$$

where l is the Bohr orbital radius.

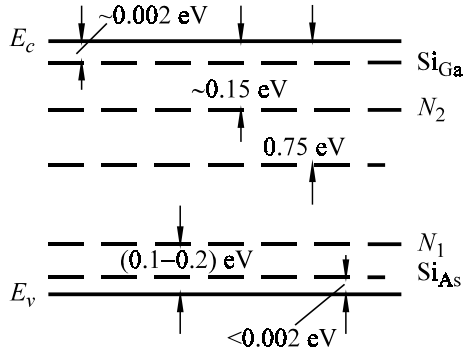

Figure 2.49. Energy diagram of basic defects in undoped GaAs.

For the estimation, we will take GaAs with l = 83.3 E and a = 5.5 E. Taking $N_{Si} \cong 10^{17}$ cm^{-3} and $N_V \cong 5\times10^{19}$ cm^{-3}, we will get

$$W_{V\text{-}IVI}/W_{V\text{-}Si} >> 1 \qquad (2.4.23)$$

for both gallium and antimony vacancies. Therefore, the probability of defect encounter appears to be higher when vacancies interact with isovalent impurities rather with silicon.

The intensity of complexation involving isovalent and other impurities available in the crystal varies with the nature and concentration of the latter. The concentration ratio of two complexes Z_1 and Z_2 produced by impurities C_1 and C_2 can be written as

$$Z_1/Z_2 = \left(C_1/C_2\right)\exp\left(Q_1 - Q_2\right)/kT\,, \qquad (2.4.24)$$

where Q_1 and Q_2 are the binding energies in the complexes. The dominant process will be the formation of complexes with impurity atoms having a higher binding energy or a higher concentration. The above illustration of *V*-isovalent impurity and *V*-Si complexes is just characteristic of the case when the concentration of the isovalent impurity is higher than that of silicon.

Another illustration is GaAs with In or Sb as an isovalent impurity at a concentration of 10^{18}–10^{20} cm^{-3}. Undoped GaAs is characterized by the presence of associated defects N_1 and N_2 (Figure 2.49), which include silicon atoms, a Ga vacancy in the N_2 associate, and an As vacancy in the N_1 associate [218, 219]. The doping of GaAs with Sb decreases the concentration of N_2, while the doping with In produces the opposite effect [220, 221]. Similar data were obtained for the N_1 center [220, 221]. These results indicate a de-

crease in the concentration of Ga vacancies for Sb doping and its increase for In doping. This conclusion was supported by a uniform distribution of the amphoteric germanium between the GaAs sublattices in the presence of isovalent impurities [219].

In addition to the binding of vacancies, isovalent impurities are capable of generating vacancies. This process is associated with the concept of an isovalent impurity as a source of elastic strain in the crystal lattice. Due to this strain, the defect formation energy at the impurity site differs by a dozen angstroms from that in an undoped crystal. As a result, the average concentration of defects in a crystal changes. It has been shown theoretically [224] that the total concentration of vacancies V in the case of isovalent impurity doping is expressed as

$$V = V_0\left(1 + N\Omega_0 \exp Q'\right), \tag{2.4.25}$$

where V_0 is the concentration of vacancies in an undoped crystal, N is the concentration of impurities, Ω is the volume of a defect potential well, created by impurity atoms in an elastic field [225], and Q' is the binding energy of an impurity–vacancy complex. This result was supported experimentally, for example, in Si<Sn> [226] and in GaAs<B> [227].

2.4.5 Isovalent doping effects

Transformation of the homogeneity region. The effect of the crystallization medium composition on the homogeneity region of the compound being crystallized, i.e., on the equilibrium of intrinsic point defects, can be conveniently analyzed with reference to $A^{III}B^{V}$ compounds produced from melts containing isovalent impurities. A quantitative evaluation of the intrinsic defect equilibrium can be made in terms of equilibrium of the quasi-chemical reactions which describe the production of such defects.

Assuming that the main types of intrinsic point defects are vacancies of $A^{III}(V_A)$ and $B^{V}(V_B)$ components, as well as interstitial $B^{V}(B_i)$ atoms, and that isovalent impurities produce substitutional solid solutions, we can write these reactions as follows:

$$A_I \overset{K_1}{\rightleftarrows} A_A + V_B \tag{2.4.26}$$

$$B_l \overset{K_2}{\underset{\leftarrow}{\rightarrow}} B_B + B_i + V_A \qquad (2.4.27)$$

$$A_l^* \overset{K_3}{\underset{\leftarrow}{\rightarrow}} A_A^* + V_B \qquad (2.4.28)$$

$$B_l^* \overset{K_3^*}{\underset{\leftarrow}{\rightleftarrows}} B_B^* + V_A, \qquad (2.4.29)$$

where A^* and B^* are impurities isovalent to the elements A and B, respectively; K are equilibrium constants of direct quasi-chemical reactions.

A possible interaction of intrinsic point defects in the solid phase is described as

$$B_i + V_B \overset{K_4}{\underset{\leftarrow}{\rightarrow}} B_B. \qquad (2.4.30)$$

The concentrations of point defects in atomic fractions can be expressed using equations (2.4.26) through (2.4.30) as

$$[V_A] = 8K_1K_2K_4a_A^l a_B^l \qquad (2.4.31)$$

$$[V_B] = 2K_1a_A^l \qquad (2.4.32)$$

$$[B_i] = 0.25/K_1K_4a_A^l \qquad (2.4.33)$$

$$[A_A^*] = K_3a_{A^*}^l/2K_1a_A^l \qquad (2.4.34)$$

$$[B_B^*] = K_3^*a_{B^*}^l/8K_1K_2K_4a_A^l a_B^l\ . \qquad (2.4.35)$$

These five equations show that concentrations of all defects are in an unstable equilibrium varying with a_i^l and K. The former are calculated from the known ratios for regular solutions, and the equilibrium constants of the reactions are found with the algorithm suggested in [229].

The knowledge of the values of a_i^l and $K_1 - K_4$ is necessary for the calculation of intrinsic defect concentrations in AB<A^*> or AB<B^*> as a function of temperature and isovalent impurity concentration. Such calculations have

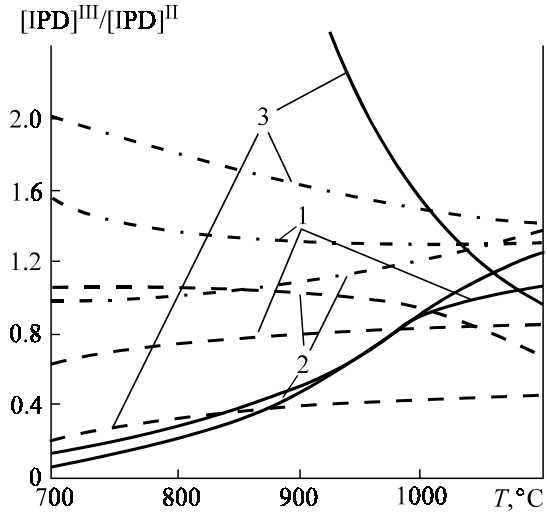

Figure 2.50. Calculated temperature dependences of IPD relative concentration variation in GaAs during crystallization from a 3-component (Ga-As-isovalent impurity) liquid phase: solid lines – [V_{Ga}]; dashed lines – [V_{As}]; broken lines – [As_i]; isovalent impurities: 1 – In, 2 – Sb, 3 – Bi; calculation for In and Sb at C_{IVI}^{sol} = 4.4×10^{20} cm^{-3}, for Bi at C_{IVI}^{sol} corresponding to solubility limit; index II is for the two-component Ga-GaAs system.

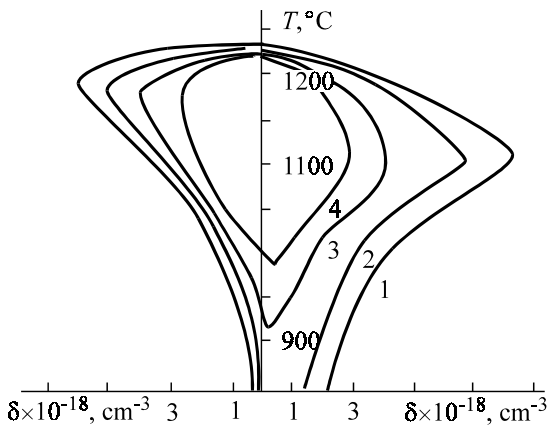

Figure 2.51. Polythermal cross sections of homogeneity regions in GaAs<Bi> for C_{Bi}^{sol}, cm^{-3}: 1 – 0; 2 – 1×10^{18}; 3 – 3×10^{18}; 4 – 5×10^{18}.

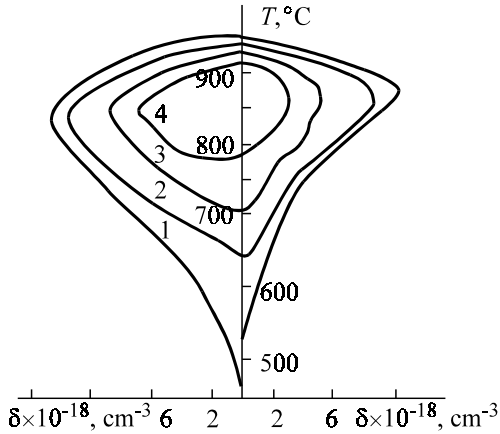

Figure 2.52. Polythermal cross sections of homogeneity regions in InAs<Bi> for C_{Bi}^{sol}, cm^{-3}: 1 – 0; 2 – 2×10^{18}; 3 – 8×10^{18}; 4 – 1.4×10^{19}.

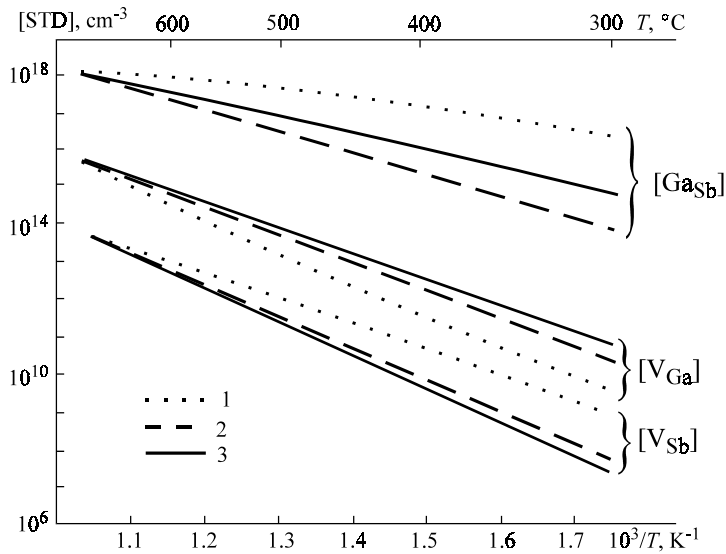

Figure 2.53. Temperature dependences of IPD concentrations in GaSb obtained from solutions: 1 – Ga; 2 – Sn; for crystallization from Bi-solutions, the dependence is close to line 2; 3 – calculations neglecting the activity of components in the liquid phase.

been made for GaAs<IVI> [230], InAs<IVI> [231], and GaSb [232]. For illustration, we present Figures 2.50 through 2.52. It should be noted that an essential role in GaSb is played by antistructural defects. In this case, the author of [232] referred the Ga_{Sb} type of defects to the intrinsic defects listed above. The calculations are illustrated in Figure 2.53. All the results show that the effect of isovalent impurities on the concentration of intrinsic point defects increases with decreasing temperature and that Bi has the greatest perturbative effect on defect equilibrium.

Amphoteric impurity redistribution between the $A^{III}B^{V}$ sublattices. It was shown in Section 2.3 that cation–anion amphoteric impurities have a definite thermodynamic distribution between the crystal sublattices under equilibrium conditions. Since doping with isovalent impurities leads to a transformation of the homogeneity region, one should also expect a sublattice redistribution of an amphoteric impurity in the presence of an isovalent impurity. Indeed, such a redistribution was established experimentally in [219] and is illustrated in Figure 2.54.

"Purification" of a semiconductor by isovalent doping. The effect of isovalent impurities on the distribution coefficients of background impurities in the liquid phase can decrease their concentration in a growing crystal, i.e., it

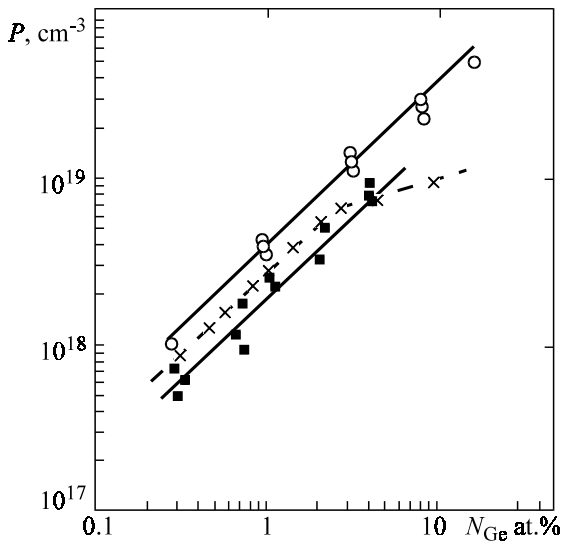

Figure 2.54. Hole concentration in GaAs:IVI:Ge epitaxial layers versus Ge content in the solution melt: x – GaAs:Ge; o – $GaAs_{1-x}Sb_x$:Ge, $x = 0.006$; ■ – $Ga_{1-x}In_xAs$: Ge, $x = 0.006$.

can produce a kind of purifying action to remove the background. It was mentioned in Section 2.4.4 that the greatest effect should be expected from isovalent impurities with the maximum thermodynamic parameter in the interaction with a background impurity. Let us consider this problem quantitatively now.

The interaction parameter in the liquid phase, ω^{l}, can be evaluated for a system with an infinite solubility in the liquid phase and a practically zero solubility in the solid phase, such as the isovalent impurity-Si systems, using the following relation [215]:

$$\omega^{l} = \frac{\Delta H_{Si}^{f}\left(1 - T/T_{Si}^{f}\right) + RT \ln x_{Si}^{l}}{1 - x_{Si}^{l}}, \qquad (2.4.36)$$

where T_{Si}^{f} and ΔH_{Si}^{f} are the melting temperature and enthalpy of a background impurity and x is the atomic fraction of silicon in the liquid phase.

The estimations of ω^{l} for the interaction of silicon with an isovalent impurity in GaAs are given in Figure 2.55. It is seen that the Si–Bi system has the maximum value of ω^{l}, so the experimental checkup of isovalent purification was undertaken in [173] during the GaAs epitaxial growth from a bismuth melt. The experimental data show a considerable decrease in electron concentration and an increase in electron mobility, while the impurity back-

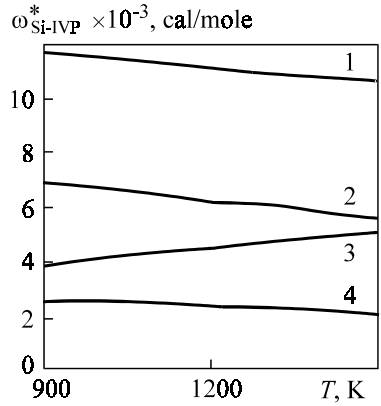

Figure 2.55. The temperature dependence of the interaction parameter in the liquid phase for Si–IVI systems: 1 – Si<Bi>, 2 – Si<In>, 3 – Si<Sb>, 4 – Si<Ga>.

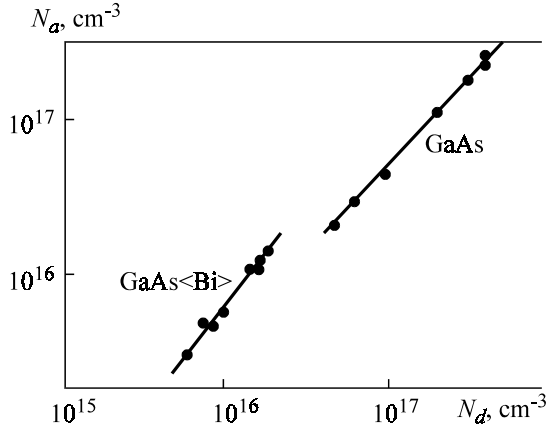

Figure 2.56. The relationship between donor and acceptor concentrations in GaAs and GaAs<Bi> samples.

ground (Figure 2.56) was reduced to 8×10^{15}–3×10^{16} cm^{-3}, as compared with 8×10^{16}–6×10^{17} cm^{-3} in control samples containing no Bi.

Another striking example of the purification effect of isovalent doping was reported in [174] from a photoluminescence study of GaAs epitaxial layers doped with In and Sb. The results presented in Figure 2.57 show that the In doping considerably reduced the radiation band at 1.4 eV, and Figure 2.49 demonstrates that this reduction corresponds to the radiation transition of electrons from the N_1 center levels to the conduction band. In other words, indium impurity has lowered appreciably the concentration of these background centers.

The effect on the compensation degree of a semiconductor. A direct consequence of isovalent purification is a changed compensation of shallow impurities in the semiconductor. It is clear from the curve slopes in Figure 2.56 that the compensation degree of GaAs<Bi>, i.e., the component $k = N_a/N_d = 0.8$, appears to be higher than $k = 0.5$ for GaAs samples containing no bismuth. With the assumption of the compensation being due to amphoteric background silicon, this result also indicates a redistribution of the amphoteric impurity between the Ga and As sublattices in the presence of bismuth. We showed in Section 2.3 (see expressions (2.3.65) and (2.3.66)) that $k \sim P_{As}^{-2}$; therefore, the experimentally observable increase in the compensation degree unambiguously indicates a decrease in P_{As} above the crystal grown from a bismuth melt, as compared with the conventional growth from a gallium melt. This was established by direct measurements in [233].

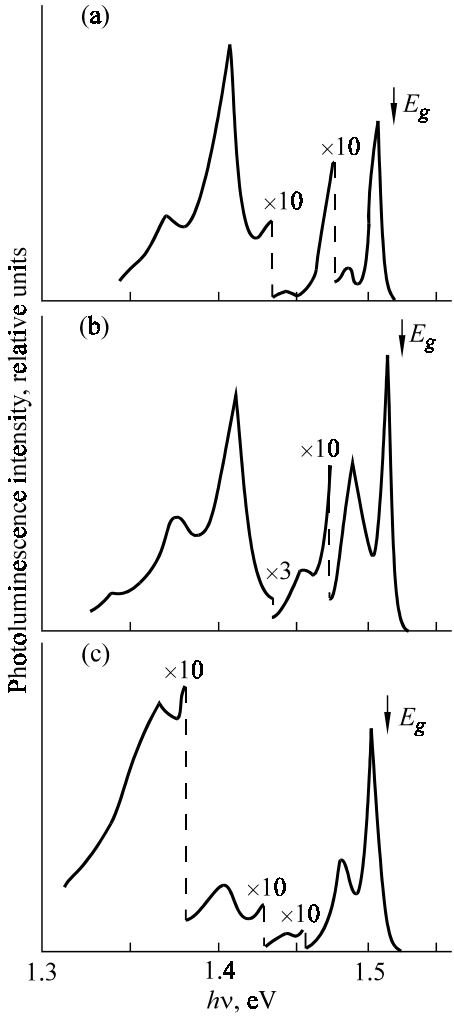

Figure 2.57. Photoluminescence spectra at T = 2K: (a) – $GaAs_{1-x}Sb_x$ (x = 0.0034); (b) – GaAs; (c) – $Ga_{1-x}In_xAs$ (x = 0.007).

Variation in the compensation degree due to isovalent doping was analyzed from photoluminescence spectra [234] and from the electrical behavior of crystals doped with isovalent impurities [221].

The effect on the dislocation structure. The effect of impurities on the macroscopic elasticity and dislocation mobility was considered, before the

application of isovalent doping, as a manifestation of impurity electrical activity. But there was an alternative view, according to which elastic interaction of impurity atoms with dislocations dominated at $T \geq 0.7T_f$. The use of isovalent impurities, which are uncharged in a crystal, unambiguously confirmed the latter point of view in studies of single crystals doped with In or Sb [236].

The effect on solid solution decomposition. The elastic deformation of a crystal lattice produced by doping has also an effect on defect formation due to decomposition of oversaturated solid solutions. If an impurity "expands" a crystal lattice, its decomposition is slowed down to liberate components with a larger specific volume than that of the host crystal. An illustration of this effect is the slower decomposition of an oversaturated oxygen solid solution in silicon single crystals doped with isovalent Ge and Sn [237].

2.5 VOLATILE IMPURITIES

2.5.1 Hydrogen

All available data indicate that hydrogen atoms occupy interstitial positions in the crystal lattice of diamond-like semiconductors. Debatable is only the type of interstice. For silicon and germanium, the preferable position seems to be the "antibinding" position along the <111> axis at a distance of ~0.16 nm from the nearest regular atom, with a slight dispersion (~0.01 nm) in the transverse direction. This conclusion follows from the channeling effect in silicon samples with implanted deuterium [238]. The study of infrared absorption spectra in the same samples [239] led to the finding of a considerable binding energy of about 2.0 eV between deuterium atoms (hence, of hydrogen) and silicon atoms. This is sufficient evidence for their strong chemical bonding.

A similar bonding occurs between hydrogen and other impurity atoms in a crystal. An example is the interaction between hydrogen and copper impurities in germanium [240]. This effect can be identified from the change in the energy spectrum of the interacting components, similar to that observed in copper–lithium complexation in germanium, when some copper levels in the forbidden gap disappear and others are shifted.

Hydrogen in germanium and silicon interacts actively with intrinsic point defects. Some models of this interaction are discussed in [241]. Any point defects can produce complex defects, whose energy spectrum is similar to that of hydrogen-like impurities but differs in the impurity center symmetry

from common hydrogen-like substitutional impurities. In contrast to single hydrogen atoms, these complex centers produce numerous shallow donor and acceptor levels at $E_c - 12.3$ and $E_v + 11.3$ eV.

It has been established reliably by the MSR technique that germanium and silicon contain atomic muonium [242, 243]. Since muonium and hydrogen are actually isotopes, everything established for muonium will be valid for hydrogen. The most important result is the "swelling" of a muonium atom in germanium and silicon crystals, as compared with the muonium size in vacuum (5.32 nm). For example, the muonium sizes in germanium and silicon are 6.4 and 7.2 nm, respectively [244]. However, the observable increase in the size cannot be described in terms of the effective mass theory.

Kittel and co-workers [245] have explained theoretically the experimentally detected increase of the Bohr orbital radius of muonium in germanium and silicon crystals. Their conclusion is that muonium, and, therefore, hydrogen, are to be deep donors with the ionization energies 1.52÷1.58 and 1.58–1.70 eV, respectively. These values agree satisfactorily with the above value of 2.0 eV. Such high ionization energies account for the fact that atomic hydrogen does not show electrical properties in germanium and silicon.

Among more complex semiconductors, only SiC was found to have a similar increase in muonium size [246], which was 14% of its vacuum value. Besides, SiC crystals were subjected to hydrogen ion implantation followed by a photoluminescence spectral analysis of one of the polytypes, 6*H*–SiC [247]. The author suggested a model of a center consisting of a hydrogen atom at a vacant silicon site. In this position, a hydrogen atom is surrounded by four carbon atoms and is shifted toward one of them. No data are available on the state and behavior of hydrogen impurity in other semiconductor compounds.

When discussing hydrogen behavior in semiconductors, one should not ignore its effect on crystal surface properties. Although this behavior is still poorly understood, we do know that atomic hydrogen forms fast recombination centers on a germanium surface [248]. Some effort has been made to study hydrogen saturation of amorphous semiconductors, mostly silicon [249], in which hydrogen atoms saturate dangling chemical bonds of the host atoms, thereby stabilizing the semiconductor properties. This problem, however, does not involve the behavior of hydrogen in a bulk crystal. For this reason, we do not discuss it in this book.

2.5.2 Oxygen

Oxygen is abundant in semiconductors, because it possesses a high diffusion coefficient in melts and solid crystals. It finds its way to a growing crystal

from the quartz chamber walls, the substrates used and the ambient atmosphere. For these reasons, oxygen content in a semiconductor depends to a great extent on the crystal growth technique used. This is especially typical for silicon. The concentration of oxygen atoms in crystals grown from a melt may be as high as 10^{18} cm^{-3} and even more, but in samples produced by zone melting it is less than 10^{16} cm^{-3}. It is quite difficult to measure the oxygen content in silicon. There are various ways of doing this, but the most reliable method is radioactivational analysis aimed at determining the isotope ^{18}O after irradiation by accelerated ^{3}He ions [250]. The radioactive isotope ^{18}O has a short half-decay time, which greatly decreases the feasibility of the method. On the other hand, this method removes from analysis the adsorbed surface oxygen, yielding absolute values. An alternative is an optical technique using the characteristic absorption band in the infrared spectral region at $\lambda = 9.1$ μm (the frequency 1106 cm^{-1}) and the proportionality of absorption coefficient K_m at the band peak and the oxygen concentration:

$$N_O = AK_m .$$

At the same time, spectral studies show complex oxygen behavior in silicon. Oxygen is involved in different types of bonding, which readily transform to one another. Table 2.25 shows eight bands due to ogygen in silicon.

The three bands at 1106, 1205, and 515 cm^{-1} are due to the nonlinear configuration of Si–O–Si [251, 252]. The first band was identified as an antisymmetric external oscillation ν_1 (Figure 2.58) and the other two as symmetrical valent oscillation ν_2 and deformation oscillation ν_3. The other absorption bands due to oxygen presented in Table 2.25 are associated with

Table 2.25. Infrared absorption bands in silicon with impurity oxygen.

Type of center	Si–O_i–Si*	*V*–O_i *A*-center	C_i–O_i *X*-center	Si_i–O_i–C	O_i	SiO	Si–O_i–Si**	Si–O_i–Si***
Absorption band at ν, cm^{-1}	515	830	865	890	935	1000	1106	1205

* symmetrical bending oscillations, ** antisymmetrical stretching oscillations, *** symmetrical stretching oscillations.

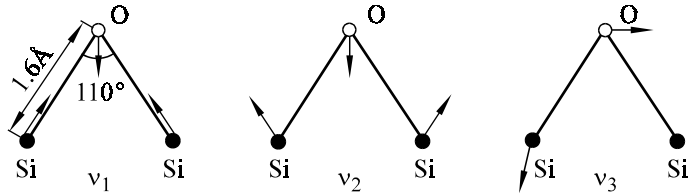

Figure 2.58. Nonlinear configurations of the Si_2O molecule, corresponding to absorption at K, cm^{-1}: $\nu_1 = 1106$; $\nu_2 = 1205$; $\nu_3 = 515$.

thermal treatment of silicon at 1000°C for several days. This considerably reduced the main band intensity at 1106 cm^{-1}.

The main feature of oxygen behavior in silicon is the formation of what is known as thermal donors produced during thermal treatment. Today, the general view, based on the data of [253] presented in Figure 2.59, is that there are two kinds of thermal donors. One, represented by the low temperature peak in Figure 2.59, is attributed to complexes consisting of four oxygen atoms [253] and the other, corresponding to the high temperature peak in the same figure, is due to oxygen precipitates.

A model was suggested in [251] to account for the formation of groups of the O_n type. If the formation of, say, "pairs" and "triplets" is in equilibrium, the following reactions occur:

$$\begin{aligned} O_2 &= \alpha_1[O_2] \\ O_3 &= \alpha_2[O_2][O]. \end{aligned} \qquad (2.5.1)$$

The formation kinetics of "quartets" is described as [251]

$$\frac{d}{dt}[O_4] = K[O]\left\{\alpha_1\alpha_2[O]^3 - [O_4]\left(1+\frac{1}{\alpha_3[O]}\right)\right\}. \qquad (2.5.2)$$

The solution to this kinetic equation is

$$[O_4](t) = [O_4]^{max}\left[1-\exp(-t/\tau)\right] \qquad (2.5.3)$$

with

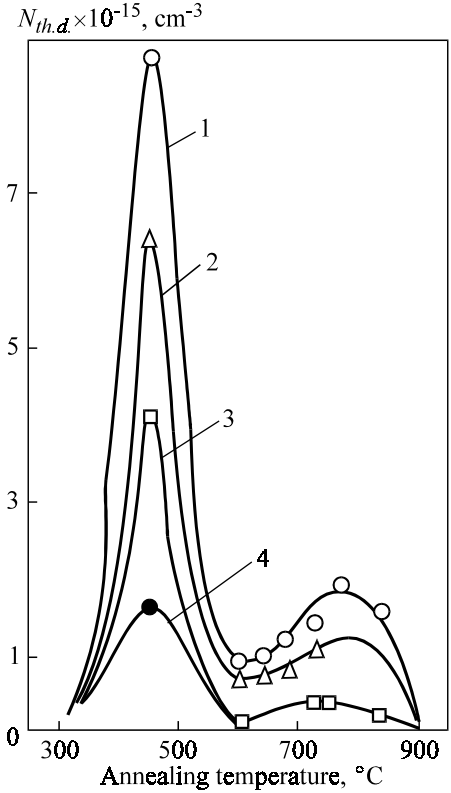

Figure 2.59. Relationship between thermal donor concentration $N_{th.d.}$ and annealing temperature for Si samples with various oxygen concentrations N_O, cm^{-3}: $1 - 1\times10^{18}$; $2 - 1.85\times10^{18}$; $3 - 1.6\times10^{18}$; $4 - 1.2\times10^{18}$.

$$[O_4]^{max} = \frac{\alpha_1\alpha_2[O]^2}{1+\dfrac{1}{\alpha_3[O]}},$$

$$\frac{1}{\tau} = K[O]+\frac{K}{\alpha_3} = K[O]\left(1+\frac{1}{\alpha_3[O]}\right). \qquad (2.5.4)$$

It is easy to see that the formation kinetics of thermal donors are described satisfactorily by equation (2.5.2) at

$$\alpha_3[O] >> 1. \qquad (2.5.5)$$

Hence, the reaction producing quartets from triplets is essentially non-equilibrium. Quartets are more likely to form from quintets.

Higher temperature donors are formed only following a low temperature treatment [246]. This work describes a model, in which neutral complexes containing more than four oxygen atoms are produced at 300–500°C, in addition to electrically charged O_4. During the treatment at 600–800°C, the decomposition of a supersaturated oxygen solution occurs actively and O_m complexes dissociate into simpler ones. The electrical properties of oxygen tetrahedra begin to manifest themselves again. This model was supported by the temperature dependences of concentrations of electrons which go from the first and second thermal donor levels to the conduction band [246]. It was found that both have the same energy level at $E_c - 0.02$ eV.

Oxygen in silicon also interacts with vacancies to produce the well-known *A*-center [117]. The *A*-center has an acceptor level in the silicon forbidden gap at $E_c - 0.16$ eV. This complex is stable and is annealed at ~600 *K*. Besides, an oxygen atom forms two unstable associates, $(O–V)_{<111>}$ and $(O–V)_{<100>}$, which are annealed at 100 and 120 K, respectively [254], with the second associate transforming to a stable *A*-center.

When an oxygen atom captures a vacancy, it is displaced and becomes localized nearly at the vacant site (Figure 2.60). When an *A*-center captures an electron, it becomes negatively charged and paramagnetic, which permits its study using the spin and double resonance (ENDOR) method. This problem has been studied extensively, so we will only mention a recent review [255].

Oxygen atoms in silicon are “polymerization” centers for vacancies. The following types of centers have been reported [156, 257]: $(V_2–O)$, $(V_2–O_2)$,

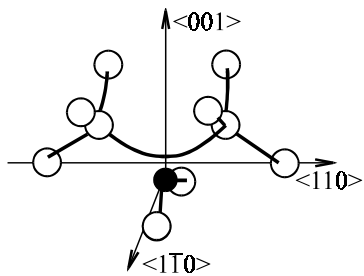

Figure 2.60. The model *A*-center (vacancy + oxygen atom) in silicon: the black circle – an oxygen atom with closed bonds of neighbors around a vacant site.

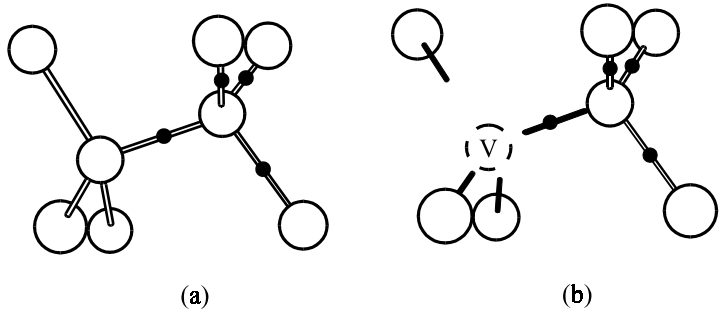

Figure 2.61. The model of an electrically inactive structural group (a) and its electrically active oxygen complex (b) in germanium: black circle – impurity oxygen atoms, without allowance for the lattice relaxation.

(V_3–O), (V_3–O_2), and (V_3–O_3). The stability of such complexes increases with their "size." The ability of oxygen atoms to form associates with vacancies, intrinsic and doping impurity atoms stems from the interstitial position of oxygen atoms on the Si–Si bonds in the silicon lattice.

Oxygen behavior in germanium is similar to that in silicon. For example, the vibrational modes shown in Figure 2.58 are also valid for germanium [258, 259], with the only difference that the ν_3 mode corresponds to the 11.7 μm band in the infrared absorption spectrum, instead of the 9.1 μm band.

The interstitial oxygen model for germanium is shown in Figure 2.61 [260]. One can see that tetrahedral GeO_2 is made up of GeO_4 groups. Such defect associations are electrically inactive, because valent electrons of oxygen atoms are bound by intrinsic crystal atoms. If GeO_4 centers capture vacancies (Figure 2.61b), they become electrically charged. Further attachment of the second and third vacancy shifts the donor level of the complex into the forbidden gap.

The existence of *A*-centers was also suggested for germanium, but this question still remains debatable (see, for example, [117]). The formation of thermal donors in germanium doped with oxygen has been reported [261], but there are no detailed studies of their formation and transformation. There has been much interest in oxygen doping of III-V semiconductors, but it has been investigated only in GaAs and GaP.

The principal difficulty in the investigation of these crystals doped with oxygen is that it is impossible to remove the oxygen background during oxygen doping. The main sources of oxygen contamination are surface oxide films on the initial $A^{III}B^{V}$ components, the oxygen dissolved in the A^{III} bulk, quartz, etc. It was found that the use of hydrogen stimulates the reaction

$$SiO_2|_s + H_2|_\Gamma \rightleftarrows SiO|_\Gamma + H_2O|_\Gamma \qquad (2.5.6)$$

producing oxygen contamination of the growing crystal via SiO.

Vacuum also produces oxygen contamination of $A^{III}B^{V}$ compounds because of the interaction between A^{III} and quartz [263]:

$$3SiO_2|_s + 2Ga_l \rightleftarrows Ga_2O_3|_\Gamma + 3SiO|_\Gamma. \qquad (2.5.7)$$

The introduction of oxygen or Ga_2O_3 into the reaction region can decrease the contamination via SiO and increase the oxygen content in material. It was shown for GaP [264] that the introduction of Ga_2O_3 into the melt decreased Si content, increasing oxygen concentration.

A high oxygen background is an obstacle to obtaining reliable data on oxygen solubility in $A^{III}B^{V}$ compounds at various temperatures. For GaP<O>, for example, no solidus curve has been found, but there are only data indicating that the total oxygen content in GaP may be as high as 2×10^{19} cm^{-3}. For GaAs<O>, the solidus curve was found in [265]. The solubility has a retrograde character with a maximum of 3×10^{19} cm^{-3} at 1100°C. Note that the doping was carried out in two ways: by adding Ga_2O_3 or As_2O_3 into the melt. The light absorption spectrum for such crystals has a peak in the doping through As_2O_3 [266]. The authors of this work even demonstrated that this peak shifted on dilution of As_2O_3 with the isotope ^{18}O. They interpreted their data as follows. The doping through Ga_2O_3 produces approximately 10^{17} cm^{-3} oxygen atoms at an As site, i.e., this is substitutional oxygen. The doping through As_2O_3 produces about 10^{17} cm^{-3} interstitial oxygen, which is a neutral impurity. The total oxygen content was found to be 10^{19} cm^{-3}.

In spite of the lack of data on oxygen contents in samples, there are many reports on the electrical, optical, and photoelectrical properties of GaP<O> and GaAs<O>.

For example, there was an appreciable EPR signal detected in GaP<O> [267, 268], which permitted a partial determination of oxygen abundance at phosphorus sites. The Hall measurements made in these samples revealed a level $E_c - 0.89 \pm 0.02$ eV (Figure 2.62). Indirect evidence for this level being an O_p state is the coincidence of concentrations found from Hall measurements and EPR studied. Direct evidence could be obtained from photostimulated EPR, by illuminating electrons by light at $h\nu = 0.9$ eV into the conduction band, thereby changing the number of charged 0.9 eV centers and following this change from the EPR line intensity variation. Such an approach, however, has not yet been used. The application of GaP for the fabri-

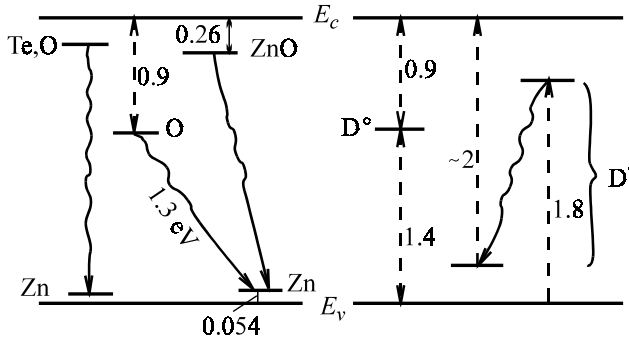

Figure 2.62. Impurity levels and radiative recombination transitions (arrows) in impurity gas.

cation of efficient emitters has stimulated the investigation of GaP<O> luminescence. Luminescence spectra of such samples clearly show the tendency for oxygen to produce impurity pairs and complexes with other impurity atoms.

The formation of impurity pairs and complexes with oxygen will be discussed in detail in Chapter 6. Here, we will only mention the principal difference between them. The partners in a pair are merely bound by Coulomb forces, while a complex involves a chemical bonding changing it to a kind of quasimolecule. In the pair interaction, the energy levels of the partners and their shift toward the allowed band edges are preserved. When a complex is formed, new levels arise in the semiconductor forbidden gap, which have nothing to do with the partner levels (E_c – 0.26 eV).

Pair luminescence was observed in GaP<O,Zn> [269], which had a lower intensity than other group-VI impurities. The maximum intensity of the pair luminescence (1.3 eV), the zinc level position (0.064 eV), and the forbidden gap value in GaP (2.12 eV) allowed identification of the level of the second partner in a pair (0.9 eV) shown in Figure 2.62, which coincides with the O_p level.

Figure 2.62 shows another recombination radiation transition with a maximum at ~1.8 eV, which was interpreted in [270] as being due to pair luminescence between a ZnO-type complex and a single Zn level. Similar radiation was observed in GaP<O,Cd>.

When discussing the behavior of oxygen in GaP, one should not ignore the formation of what is known as D^--centers. A D^--center is one of the three possible O_p states: D^+ is an ordinary ionized state with the transition of an electron to the conduction band; D^0 is a neutral state, or a deep center having

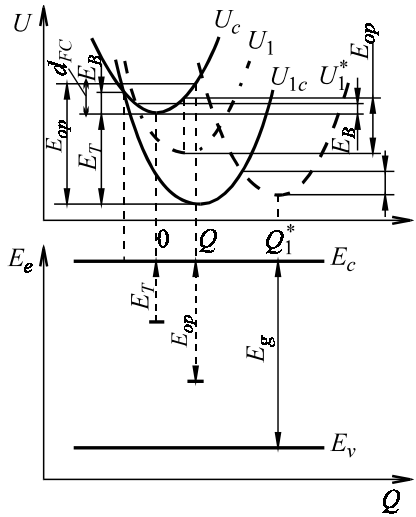

Figure 2.63. The configuration coordinate diagram of an oxygen-containing E_1-center in GaAs; Q – configuration coordinate.

an electron with the optical ionization energy of 0.9 eV; D^- is a state with a second, attached electron. The optical ionization energy of D^- was found experimentally [271, 272] to be 2 eV below the conduction band (Figure 2.62). Interestingly, a D^--state appears only under the action of 0.4 optical pumping. For this reason, some workers believe that illumination by this light makes an electron jump onto the level E_c – 0.4 eV, after which the optical ionization energy E_c becomes equal to 2.0 eV due to a strong electron–phonon interaction during the lattice relaxation. An alternative model is a mere doubling of the binding energy of each of the two 0.9 eV electrons.

Both models raise many questions. The formation of D^--states of oxygen-containing centers was also observed in GaAs<O> [273]. It has been established in many investigations that oxygen atoms in GaAs produce two levels: $E_1 = E_c – (\sim 0.69)$ eV and $E_2 = E_c – 0.18$ eV. Both values correspond to thermal ionization. The study of photoelectric properties and the use of optically induced light absorption modulation (for details, see Section 5.4 of this book and [22]) have yielded electronic configurations for both centers, E_1 and E_2, shown in Figures 2.63 and 2.64. In Figure 2.63, curve U_{1t} represents the energy E_1 of an oxygen-containing center in a singly charged state D^0, curve U_1^* is the metastable excited state of this center, and curve U_1^- is the state of

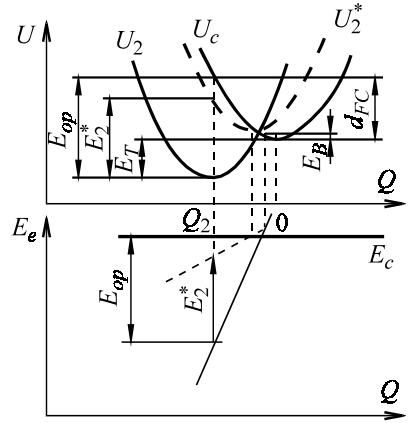

Figure 2.64. The configuration coordinate diagram of an oxygen-containing E_2-center in GaAs; Q – configuration coordinate.

a D^--center. The quantitative characteristics for the diagrams shown in Figure 2.63 are given in Table 2.26.

One can see from Table 2.26 a remarkable Frank–Condon shift for the E_1-center, which is likely to be due to the Yan–Teller effect, when an oxygen atom is displaced from a B^V site, leaving behind part of the vacancy. This is the way an associative defect is produced, which can be treated as a kind of quasimolecule. It seems quite likely that the second electron can be captured only by such a quasimolecule. In any case, this model accounts for the existence of D^--states when the sample is illuminated by light, i.e., under non-equilibrium conditions, both in GaAs and GaP.

Table 2.26 also presents configuration diagram parameters for another oxygen-containing center in GaAs. Both diagrams in Figures 2.63 and 2.64 show a similarity of the two oxygen-containing centers in GaAs. It seems surprising that the D^--state should exist only for the E_1-center. This is probably due to the fact that E_1 belongs to an O_i–V_{As} molecule and E_2 to an O–Si molecule, as is stated in the publications.

Table 2.26. Some configuration parameters of E_1- and E_2-centers in GaAs [22].

Parameter	E_T	E_{op}	d_{FC}	E_B	E_{op}^*	E_B^*	E_T^-	E_{op}^-	E_B^-
E_1-center, eV	0.69	0.83	0.14	0.08	–	0.2	0.31	0.55	–
E_2-center, eV	0.18	0.43	0.25	0.06	0.36	–	–	–	–

Available data on the oxygen behavior in other $A^{III}B^{V}$ compounds are too ambiguous to be discussed in this book.

2.5.3 Carbon

Most information on carbon states in silicon was obtained by EPR techniques [274, 275]. A positively charged interstitial state C_i^+ producing a dumbbell pair with a silicon atom along the <100> direction was identified. When the samples were heated to 65°C, the spectrum of the pair disappeared for 30 min to be replaced by a more stable C_s–C_i pair spectrum along the <111> axis. Such pairs were stable up to 520 K.

The three types of pairs were electrically active. The transitions between them correspond to the following levels: $C_i^0 \rightarrow C_i^+$: $E_v + 0.28$ eV; $C_i^- \rightarrow C_i^0$: $E_c - 0.12$ eV, and C_s–C_i pairs have $E_v + 0.36$ eV levels. The light absorption spectrum of silicon doped with carbon contains several bands, whose energy positions and identification are given in Table 2.27.

Among other behavioral features of carbon in silicon is a slower formation of thermal donors due to the presence of carbon. The mechanism of this process is unknown, but the production of CO_n-type associates was established with certainty. This process leads to a lower concentration of oxygen atoms constituting thermal donors.

The carbon state in $A^{III}B^{V}$ compounds may be considered to be practically unknown. There is only a report [276] of the activation energy of carbon in GaAs equal to $E_v + 0.019$ eV and in GaP equal to $E_v + 0.041$ eV and $E_v + 0.048$ eV [277, 278].

Since carbon belongs to group-IV elements, one may expect this impurity to show cation–anion amphoteric properties in $A^{III}B^{V}$ semiconductors. However, there have been no direct experimental observations of this kind, probably because of the "small" size of the carbon atom, resulting in its instability at $A^{III}B^{V}$ lattice sites.

Table 2.27. Infrared light absorption bands in silicon doped with carbon.

Type of center	C–Si	^{12}C	SiC*	C_i–O_i**	Si_i–O_i–C	C_i	C–O
Absorption band ν, cm^{-1}	604	607	820	865	890	921	1111

* precipitates observed in Si<C> crystals, ** *X*-center.

2.5.4 Nitrogen

It is very difficult to introduce nitrogen into silicon by the conventional techniques of pulling from the melt or zone melting. This is associated with the low distribution coefficient in solid silicon and because nitrogen is more liable to form silicon nitride than to be incorporated by a crystal lattice as an impurity. For this reason, nitrogen is introduced by ion implantation followed by sample annealing. It is concluded [279] that a nitrogen atom occupies an interstitial position in silicon at the level $E_c - 0.14$ eV. The light absorption spectrum of Si<N> shows a peak at 10.6 μm. Nitrogen in $A^{III}B^{V}$ semiconductors was discussed in Section 2.4 with reference to GaP, for which it is an isovalent impurity.

As an impurity, nitrogen has proved to be most important in silicon carbide, to which it may be incorporated, substituting sites in both SiC sublattices. Table 2.28 presents the energy levels of impurity nitrogen atoms in SiC crystal positions.

Table 2.28. Energy levels E_i (eV) of impurity nitrogen atoms in various SiC crystal positions.

C-sublattice	Si-sublattice	Interstice
$E_c - 101$	$E_c - 159$	$E_c - 94$ $N^0 \rightleftarrows N^+$
$E_c - 158$	$E_c - 247$	$E_c - 192$ $N^+ \rightleftarrows N^{2+}$
$E_c - 163$	$E_c - 255$	$E_c - 307$ $N^{2+} \rightleftarrows N^{3+}$
		$E_c - 501$ $N^{3+} \rightleftarrows N^{4+}$
		$E_c - 633$ $N^{4+} \rightleftarrows N^{5+}$

The energy values given in Table 2.28 cannot be taken to be absolutely reliable. There are alternative data on the energy levels of nitrogen in SiC. The ambiguity is due to the difficulty of identifying optical transitions because of the presence of oxygen and the difficulties in the identification of the polytype and preparation of single crystal samples. As an impurity, nitrogen plays an important role in semiconductor diamond.

REFERENCES

2.1. W. Kohn, *Solid State Physics* **5**: 257 (1957).
2.2. S. Pantelides, *Rev. Mod. Phys.* **50**, No. 4: 797–858 (1978).

2.3. S. Balasubraminau, *J. Phys. C*. **1**: 1774 (1968).
2.4. M. Jaros, *J. Phys. C*. **4**: 1162 (1971).
2.5. P.K. Katana, Sh.D. Turen, N.V. Dernovich *et al*., In: *Fizika primesnykh tzentrov* (The Physics of Impurity Centers). Tallinn: Izd. AN Eston. SSR, 81 (1970) (in Russian).
2.6. A. Morita, M. Azuma, H. Nara, *J. Phys. Soc. Jap*. **17**: 1570 (1962).
2.7. H. Nara, *J. Phys. Soc. Jap*. **20**: 778 (1965).
2.8. T.I. Kucher, K.B. Tolpygo, *FTP* **1**: 77 (1967) (in Russian).
2.9. T.H. Ning, C.T. Sah, *Phys. Rev*. B **84**: 3468–3482 (1971).
2.10. A. Baldereschi, *Phys. Rev*. *B* **1**: 4673 (1970).
2.11. H. Kaplan, *J. Phys. Chem. Sol*. **21**: 1593 (1963).
2.12. G.P. Bir, *FTT* **13**: 460 (1971) (in Russian).
2.13. F. Bassani, G. Iadonisi, B. Preziozi, *Rev. Progr. Phys*. **37**: 1099–1210 (1974).
2.14. E.M. Omelyanovsky, V.I. Fistul, *Primesi perekhodnykh metallov v poluprovodnikakh* (Transition Metal Impurities in Semiconductors). 192 p. (1984) (in Russian).
2.15. R.A. Faulkner, *Phys. Rev*. **184**: 713 (1969).
2.16. S.T. Pantelides, C.T. Sah, *Phys. Rev*. *B* **10**: 621–637 (1974).
2.17. M. Altarelli, W. Hsu, R. Satatini, *J. Phys*. *C*. **10**: 605–609 (1977).
2.18. G.W. Ludwig, H.H. Woodbury, *Phys. Rev. Lett*. **5**: 98–103 (1960).
2.19. V.S. Vavilov, A.E. Kiv, O.P. Niyasova, *Mekhanizmy obrazovaniya i migratzii defektov v poluprovodnikakh* (Mechanisms of Defect Formation and Migration in Semiconductors). Moscow: Nauka, 386 p. (1981) (in Russian).
2.20. S. Sugano, Y. Tanabe, Y. Kamimura, *Multiplets of Transition Metal Ions in Crystals*, New York–London: Academic Press, 350 p. (1970).
2.21. H.P. Hjalmarson, P. Vogel *et al*., *Phys. Rev. Lett*. **44** : 810–813 (1980).
2.22. V.I. Fistul, ed., *Fizika i materialovedenie poluprovodnikov s glubokimi urovnyami*. (The Physics and Materials of Deep Level Semiconductors). Moscow: Metallurgia, 231 p. (1987) (in Russian).
2.23. K.A. Kikoin, V.N. Fleourov, *J. Phys. C*. **10**: 4295–4308 (1977).
2.24. A. Zunger, U. Lindefelt, *Phys Rev*. **27**: 1191–1227 (1983).
2.25. K.A. Kikoin, V.N. Fleourov, *J. Phys. C*. **17**: 2357–2373 (1984).
2.26. I.B. Bersuker, *Elektronnoe stroenie i svoistva koordinatzionnykh soedinenii* (The Electronic Structure and Properties of Coordination Compounds). Leningrad: Khimiya, 350 p. (1976) (in Russian).
2.27. L.A. Hemstreet, *Phys Rev*. *B* **22**: 4590–4599 (1980).
2.28. G. De Leo, G.D. Watkins, J. Fowler, *Phys. Rev. B* **23**: 1851–1856 (1981).
2.29. B.V. Shulichenko, *Izv. AN Est. SSR, Ser*. *Fiz*. **33**:: 243–248 (1984) (in Russian).
2.30. H. Katayama-Yoshida, K. Shindo, *Sol. St. Commun*. **44**: 999–1002 (1982).
2.31. Yu. Kagan, K.A. Kikoin, *Pis'ma v JETF* **31**: 361–367 (1980) (in Russian).
2.32. L.A. Lebedo, B.K. Ridley, *J. Phys. C*. **15**: P. L961–L964 (1982).
2.33. G. Bremond, G. Guillot *et al*., *J. Appl. Phys*. **59**: 2038–2043 (1986).
2.34. V.I. Sokolov, *FTT* **29**: 1848–1852 (1987) (in Russian).
2.35. P. Vogl, J. Baranowski, *Phys. Rev. B* **29**: 841–852 (1985).
2.36. A. Zunger, *Phys. Rev*. *B* **28**: 3628–3634 (1983).
2.37. M.S. Yunusov, *Fizicheskie yavleniya v kremnii legirovannom elementami platinovoi gruppy* (Physical Phenomena in Silicon Doped with Platinum-Group Elements). Tashkent: Izd. FAN, 112 p. (1983) (in Russian).

2.38. A.B. Van Oosten, *The Electronic Structure of Deep Impurity Centers in Silicon* (Thesis). Amsterdam: Amsterdam Univ. Press, 150 p. (1989).
2.39. M.S. Yusunov, *Povedenie 4-d i 5-d primisei v kremnii* (The Behavior of 4-*d* and 5-*d* Impurities in Silicon). In: *Fizika i materialovedenie poluprovodnikov s glubokimi urovnyami* (The Physics and Materials of Deep Level Semiconductors). Moscow: Metallurgia, 231 p. (1987) (in Russian).
2.40. D.G. Andrianov, A.S. Savelyev, S.M. Yakubenya *et al*., *FTP* **16**, No. 8: 1365–1370 (1982) (in Russian).
2.41. S. Porovski, *Phys. St. Sol*., *B* **73**: K131–K133.
2.42. T.K. Ashirov, E.N. Boborykin, A.A. Gutkin, In: *Phizika soedinenii $A^{III}B^{V}$* (The Physics of $A^{III}B^{V}$ Compounds). Leningrad: Izd. LPI: 83–85 (1979) (in Russian).
2.43. T.H. Keil, *Phys. Rev*. **40**, No. 2A: 601–612 (1965).
2.44. A.A. Kopylov, A.N. Pikhtin, *FTP* **10**: 15–21 (1976) (in Russian).
2.45. M.A. Messerer, E.M. Omelyanovsly, A.N. Pantyukhov *et al*., *FTP* **8**: 2279–2284 (1974) (in Russian).
2.46. A.L. Efros, B.I. Shkovsky, *Elektronnye svoistva legirovannykh poluprovodnikov* (Electronic Properties of Doped Semiconductors). Moscow: Nauka, 416 p. (1979) (in Russian).
2.47. M.A. Messerer, E.M. Omelyanovsky, L.Ya. Pervova *et al*., *FTP* **10**: 851–859 (1976) (in Russian).
2.48. V.A. Kasatkin, V.P. Savelyev, *Sov. Phys. Semicond*. **18**: 1022 (1984).
2.49. *Appl. Phys. Lett*. **69**: 2077–2079 (1966).
2.50. F.A. Kroger, *The Chemistry of Imperfect Crystals*. Amsterdam: North–Holland Publ. Co., 654 p. (1964).
2.51. J.S. Blakemore, *Semiconductor Statistics*. London: Pergamon, 302 p. (1962).
2.52. W. Shockley, J.T. Hast, *Phys. Rev*. **107**: 293–296 (1957).
2.53. N.S. Rytova, V.I. Fistul, *FTP* **4**: 1109–1116 (1970) (in Russian).
2.54. V.I. Fistul, *Amfoternye primesi v poluprovodnikakh* (Amphoteric Impurities in Semiconductors). Moscow: Metallurgia, 240 p. (1992).
2.55. R.L. Longini, R.F. Greeni, *Phys. Rev*. **102**: 992–999 (1956).
2.56. V.I. Fistul, *FTP*, **17**: 1107–1110 (1983) (in Russian).
2.57. K.A. Kikoin, *Elektronnye svoistva ipimesei perekhodnykh metallov v poluprovodnikakh* (Electronic Properties of Transition Metal Impurities in Semiconductors). Moscow: Energoatomizdat, 304 p. (1991) (in Russian).
2.58. U. Kaufman, J. Schneider, In: *Advances in Electronics and Electron Physics* ed. C. Marton, New York–London: Academic Press **58**: 118–141 (1982).
2.59. V.F. Masterov, *FTP* **18**, No. 1: 3–23 (1984) (in Russian).
2.60. J.W. Allen, In: *Deep Centers in Semiconductors*, ed. S. T. Pantelides. New York: Gordon & Breach Sci. Publ., 755 p. (1986).
2.61. J.J. Krebs, G. H. Strauss, *Phys. Rev*., *B* **15**: 17–22 (1977).
2.62. G.H. Strauss, J.J. Krebs, *Phys. Rev*., *B* **22**: 2050–2056 (1980).
2.63. J.J. Krebs, G.H. Strauss, *Phys. Rev*., *B* **16**: 971–976 (1977).
2.64. J.J. Krebs, G.H. Strauss, *Phys. Rev*., *B* **20**: 795–801 (1979).
2.65. V. Narayanamurti, M.A. Chin *et al*., *Appl. Phys. Lett*. **33**: 481–486 (1978).
2.66. U. Kaufman, J. Schneider, *Appl. Phys. Lett*. **36**: 747–753 (1980).
2.67. G.H. Strauss, J.J. Krebs, S.H. Lee *et al*., *Phys. Rev*., *B* **22**: 3141–3146 (1980).
2.68. A.M. White, J.J. Krebs, G.H. Strauss, *J. Appl. Phys*. **51**: 419–423 (1980).

2.69. G.K. Ippolitova, E.M. Omelyanovsky, L.Ya. Pervova, *FTP* **9**: 1308–1312 (1975) (in Russian).
2.70. D. Bois, P. Pinard, *Phys. Rev., B* **9**: 4171–4176 (1974).
2.71. A.M. Hennel, W. Szuszkiewicz, G. Martinez *et al*., Phys. Rev. B 9: 4228 (1974).
2.72. A.M. Hennel, G. Martinez, *Phys. Rev., B* **25**: 1039–1044 (1982).
2.73. A.M. Hennel, C.D. Brandt, K.Y. Ko *et al., J. Appl. Phys.* **62**: 163–170 (1987).
2.74. M. Galdas, S.K. Figueirodo, A. Fazzio, *Phys. Rev., B* **33**: 7102–7107 (1986).
2.75. H. Katayama-Yoshida, A. Zunger, *Phys. Rev., B* **33**: 2961–2966 (1986).
2.76. B. Clerjaud, C. Naud, B. Deveaud *et al., J. Appl. Phys.* **58**: 4207–4213 (1985).
2.77. U. Kaufman, J. Schneider, *Festkerperprobleme* **20**: 87–98 (1980).
2.78. A.V. Vasilyev, G.K. Ippolitova, E.M. Emelyanovsky *et al., FTP* **10**: 571–575 (1976) (in Russian).
2.79. B.E. Samorukov, V.K. Sobolevsky *et al., FTP* **12**: 522–533 (1978) (in Russian).
2.80. S.A. Abagyan, G.A. Ivanov, Yu.N. Keznetzov *et al., FTP* **8**: 1691–1694 (1974) (in Russian).
2.81. W. Ulrici, L. Eaves, K. Fridland *et al., Proceedings of 14th International Conference on Defects in Semiconductors*. Paris (1986).
2.82. S.A. Kazansky, A.I. Ryskin, *Optika i spektroskopiya* (Optics and Spectroscopy), 31: 618–624 (1971) (in Russian).
2.83. S.A. Kazansky, A.I. Ryskin, *FTT* **13**: 3633–3636 (1971) (in Russian).
2.84. G. Roussos, G. Nagel, H.J. Schultz, *Z. Phys. B* **53**: 95–107 (1983).
2.85. S.A. Kazansky, A.I. Ryskin, *FTT* **13**: 3775–3781 (1971) (in Russian).
2.86. H. Kitagawa, H. Nakashima, K. Hashimoto, *Met. Fac. of Eng. Kyushu Univ*. **46**: 129–130 (1986).
2.87. H. Kitagawa, H. Nakashima, K. Hashimoto, *Jap. J. Appl. Phys*. **24**: 373–374 (1985).
2.88. H. Nakashima, H. Tomokage, H. Kitagawa *et al., Jap. J. Appl. Phys*. **23**: 776–778 (1984).
2.89. H. Tomokage, H. Kitagawa, K. Hashimoto, *Met. Fac. of Eng. Kyushu Univ*. **41**: 59–61 (1981).
2.90. H. Kitagawa, K. Hashimoto, *Jap. Appl. Phys*. **16**: 857–860 (1977).
2.91. N. Wiehl, U. Herpers, E. Weber, *J. Radioanal. Chem*. **72**: 69–78 (1982).
2.92. S. Zainabidinov, *Fizicheskie osnovy obrazovaniya glubokihk urovnei v kremnii* (The Physical Basis for the Formation of Deep Levels in Silicon). Tashkent: Izd. FAN Uz.SSR, 160 p. (1984) (in Russian).
2.93. D.A. Van Veser, *A Magnetic Resonance Study of 3-d Transition Metals and Thermal Donors in Silicon* (Thesis). Amsterdam, 128 p. (1986).
2.94. M.K. Bakhadyrkhanov, B.I. Boltaks, G.S. Kulikov, *FTT* **12**: 181–189 (1970) (in Russian).
2.95. V.P. Boldyrev, I.I. Pokrovsky, S.G. Romanovskaya *et al., FTP* **11**: 1199–1201 (1977) (in Russian).
2.96. J.W. Chen, A.G. Milnes, A. Rohatgi, *Sol. St. Electron*. **22**: 801–808 (1979).
2.97. K. Graff, H. Pieper, *Semiconductor Silicon*, ed. H.R. Huff. N.Y. Electrochem. Soc., 331–346 (1981).
2.98. A.M. Salama, L.J. Cheng, *J. Electrochem. Soc*. **127**: 1164–1168 (1980).
2.99. E. Ohta, M. Sakata, *Sol. St. Electron*. **23**: 759–764 (1980).
2.100. H. Lemke, *Phys. St. Sol., A* **64**: 549–553 (1981).

2.101. Sh.I. Askarov, G.K. Azimov, *Effekty vliyania vneshnikh vozdeistvii v poluprovodnikakh i poluprovodnikovykh priborakh* (External Effects on Semiconductor Materials and Devices). Tashkent: Izd. Tashk. GU: 76–84 (1985) (in Russian).
2.102. Yu.A. Zibutz, L.G. Paritzky, S.M. Ryvkin *et al*., *FTT* **8**: 2549–2557 (1966) (in Russian).
2.103. A.G. Milnes, *Deep Impurities in Semiconductors*. New York–London: Wiley, 562 p. (1973).
2.104. K.A. Kikoin, V.I. Sokolov, *Soviet Sci. Rev., Sec. A* **12**, part 3. London: Harwood Acad. Publ., 149 p. (1990).
2.105. V.N. Fleourov, K.A. Kikoin, *Sol. St. Commun*. **42**: 353–357 (1982) (in Russian).
2.106. V.I. Sokolov, A.E. Nikiforov, V.V. Chernyaev *et al*., *Pis'ma v JETF* **33**: 189–192 (1981) (in Russian).
2.107. K.A. Kikoin, V.I. Sokolov, V.N. Fleourov *et al*., *JETF* **83**: 2335–2347 (1982).
2.108. K. Hellwey, O. Madelung, eds., *Numerical Data and Functional Relationship in Science and Technology. New Series*. Berlin: Springer–Verlag, **17**, 652 p. (1984).
2.109. V.I. Fistul, V.N. Tzygankov, A.G. Yakovenko, *FTP* **6**: 1396–1398 (1972) (in Russian).
2.110. V.I. Fistul, V.N. Tzygankov, A.G. Yakovenko, *FTP* **7**: 2037–2041 (1973) (in Russian).
2.111. C.B. Collins, R.O. Carlson, C.J. Gallagher, *Phys. Rev*. **105**: 1168–1173 (1957).
2.112. A.E. Baranovski, J.H. Forster, *Bell Syst. Techn. J*. **39**: 87–94 (1960).
2.113. G. Bemsky, *Phys. Rev*. **111**: 1515–1522 (1958).
2.114. J.M. Fairfield, B.V. Gokhale, *Sol. St. Electron*. **8**: 685–689 (1965).
2.115. C.T. Sah, *Proceedings IEEE* **55**: 672–684 (1967).
2.116. B.I. Boltaks, Sue-Shi-In, *FTT* **2**: 2677–2684 (1960) (in Russian).
2.117. V.V. Emtzov, T.V. Mashovetz, *Primesi i tochechnye defekty v poluprovodnikakh* (Impurities and Point Defects in Semiconductors). Moscow: Radio i svyaz, 248 p. (1981) (in Russian).
2.118. J.M. Whelan, J.D. Struthers, J.A. Ditzenberger, *Proceedings Intern. Conference on Semiconductor Physics.* Prague: Chekh. Acad. Sci., 943 p. (1961).
2.119. F.E. Rosztoczy, K.B. Wolfstirn, *J. Appl. Phys*. **42**: 426–429 (1971).
2.120. I.A. Bobrovnikova, M.D. Vilisova, L.P. Porokhovnichenko *et al*., *Izvestiya vuzov. Fizika* **33**: 37–41 (1990) (in Russian).
2.121. H. Ikoma, M. Nakagawa, *Jap. J. Appl. Phys*. **11**: 338–342 (1972).
2.122. V.V. Karataev, M.G. Milvidsky, L.M. Morgulis, *Legirovannye poluprovodniki* (Doped Semiconductors). Moscow: Nauka: 110–114 (1975) (in Russian).
2.123. L.F. Zakharenkov, B.E. Samorukov, *Tekhnologiya polucheniya i elektricheskie svoistva soedinenii $A^{III}B^{V}$* ($A^{III}B^{V}$ Technology and Properties). Leningrad: LPI, 15–21 (1981) (in Russian).
2.124. A.A. Kopylov, A.N. Pikhtin, *FTP* **11**: 867–872 (1977) (in Russian).
2.125. M.G. Milvidsky, E.V. Solovyeva, *FTP* **13**: 553–557 (1979) (in Russian).
2.126. E.V. Solovyeva, L.D. Sabanova, M.G. Milvidsky, *FTP* **12**: 394–395 (1978) (in Russian).
2.127. W.P. Allred, G. Cumming, J. Kung *et al*, *Proceedings of International Symposium on GaAs*. New York–London: Reading Inst. of Phys.: 66–72 (1968).

2.128. J.O. McCaldin, R. Harada, *J. Appl. Phys*. **31**: 2065–2068 (1960).
2.129. S.P. Grishina, M.N. Kamalov, L.I. Kolesnik, *FTP* **15**: 1449–1454 (1981) (in Russian).
2.130. M.Yu. Putilovskaya, B.E. Samorukov, S.V. Slobodchikov, *FTP* **15**: 353–356 (1981) (in Russian).
2.131. A.M. Zykov, B.E. Samorukov, *Izvestiya AN SSSR, ser. neorg mat*. **11**: 819–823 (1975) (in Russian).
2.132. A.E. Yunovich, In: *Izluchatelnaya rekombinatziya v poluprovodnikakh* (Radiative Recombination in Semiconductors). Moscow: Nauka, 224–304 (1972) (in Russian).
2.133. L.F. Zakharenkov, A.M. Zykov, B.E. Samorukov, *Izvestiya AN SSSR, ser. neorg. mat*. **21**: 10–13 (1985) (in Russian).
2.134. J.B. Mullin, A. Royle, B.W. Straughan *et al*., *J. Cryst. Growth* **13**/**14**: 640–646 (1972).
2.135. R.W. Cunnigham, E.E. Harrap, W.M. Billis, *Report of Zuter. Conference on Phys. Semicond*. Roma, 722 p. (1982).
2.136. V.S. Zemskov, A.I. Lebedev, A.I. Strelnikova *et al*., In: *Legirovanie poluprovodnikov* (Semiconductor Doping). Moscow: Nauka, 77–81 (1982) (in Russian).
2.137. A.E. Bochkarev, L.I. Kolesnik, A.M. Loshinsky *et al*., *FTP* **16**: 1886–1888 (1982) (in Russian).
2.138. A.N. Arshavsky, V.V. Arbenina, S.I. Skakovsky *et al*., *FTP* **20**:101–104 (1986) (in Russian).
2.139. O.V. Vakulenko, O.A. Guseva, *Ukrain. Phys. J*., **23**: 1360–1364 (1978) (in Russian).
2.140. O.V. Vakulenko, O.A. Guseva, *FTP* **15**: 1528–1530 (1981) (in Russian).
2.141. E.F. Gross, D.S. Nedzevetzky, *DAN SSSR* **152**, No. 2: 309–312 (1963) (in Russian).
2.142. P.J. Dean, *J. Luminescence* **1**, No. 1–2: 398–415 (1970).
2.143. W. Czaja, *Festkorperproblem* **5**, No. 11: 65–85 (1971).
2.144. D.Z. Garbuzov, *Izluchatelnaya rekombinatziya cherez izoelektronnye lovushki*. (Radiative Recombination via Isoelectronic Traps). In: *Mater. V Zimnei Shkoly po Fizike poluprovodnikov* (Matrials of 5th Winter School on Semiconductor Physics). Leningrad, 219–256 (1973) (in Russian).
2.145. J.J. Van Vechten, C.D. Thurmond, *Phys. Rev., B* **14**, No. 8: 3539–3550 (1976).
2.146. V.K. Bazhenov, I.K. Doicho, A.G. Petukhov, *FTP* **14**, iss.1: 7–12 (1980) (in Russian).
2.147. P.J. Dean, J.D. Cuthbert, R.T. Lynch, *Phys. Rev*. **179**, No. 3: 754–763 (1969).
2.148. J.J. Hopfield, D.J. Thomas, R.T. Lynch, *Phys. Rev. Lett*. **17**, No. 6: 312–315 (1966).
2.149. A.A. Bergh, R. Dean, *Light-Emitting Diodes*. Oxford: Clarendon Press, 689 p. (1976).
2.150. R.A. Faulkner, *Phys. Rev*. **175**, No. 3: 991–1009 (1968).
2.151. D.G. Thomas, J.J. Hopfield, *Phys. Rev*. **150**, No. 2: 680–689 (1968).
2.152. F.A. Trumbore, M. Gershenson, D.G. Thomas, *Appl. Phys. Lett*. **9**, No. 1: 4–6 (1966).
2.153. A.N. Pikhtin, *FTP* **11**, iss. 3: 425–455 (1977) (in Russian).

2.154. Yu.F. Biryulin, N.V. Ganina, V.V. Cheldyshev, *FTP* **15**, iss. 9: 1849–1852 (1981) (in Russian).
2.155. Yu.F. Biryulin, N.V. Ganina, M.G.Milvidsky *et al*., *FTP* **17**, iss. 1: 108–114 (1983) (in Russian).
2.156. E.A. Balagurova, A.L. Petrov, E.N. Khabarov, *FTP* **15**, iss. 5: 985–987 (1981) (in Russian).
2.157. J.M. Feldman, K.M. Hergenrother, *J. Appl. Phys.* **42**, No. 13: 5563–5566 (1971).
2.158. T.N. Morgan, B. Welber, R.N. Barghara, *Phys. Rev.* **166**, No. 3: 751–753 (1968).
2.159. C.H. Henry, P.J. Dean, J.D. Cuthbert, *Phys. Rev.* **166**, No.3: 754–756 (1968).
2.160. P.J. Dean, *Phys. Rev.* **4**, No. 4: 2596–2612 (1971).
2.161. E.V. Solovyeva, A.G. Petukhov, V.K. Bazhenov, *FTP* **15**, iss. 4: 768–771 (1981) (in Russian).
2.162. G.B. Stringfellow, T.H. Hall, G.P. Burmeister, *J. Appl. Phys.* **46**, No. 7: 3006–3011 (1975).
2.163. G.G. Kleiman, M.F. Decker, *Phys. Rev., B, Separatas*, S-009 (1979).
2.164. J.L. Merz, R.A. Faulkner, P.J. Dean, *Phys. Rev.* **188**, No. 3: 1228–1239 (1969).
2.165. B.M. Ashkinadze, S.M. Ryvkin, N.D. Yaroshetsky, *FTP* **3**, iss. 4: 535–540 (1069) (in Russian).
2.166. V.V. Evstropov, B.V. Tzarenkov, *FTP*, **4**, iss. 5: 923–932 (1970) (in Russian).
2.167. R.Z. Bachrach, O.G. Lorimor, *J. Appl. Phys.* **43**, No. 2: 500–507 (1972).
2.168. P.J. Dean, *Phys. Rev., B* **4**, No. 4: 2596–2612 (1971).
2.169 J. Weber, W. Schmidt, R. Sauer, *Phys. Rev. B* **21**, No. 6: 2401–2414 (1980).
2.170. V.K. Bachenov, J. Bauman, A.G. Petukhov, *Phys. St. Sol., B* **100**, No.1: K105–K108 (1980).
2.171. R.E. Nahory, M.A. Pollack, J.C. De Winter *et al*., *J. Appl. Phys.* **48**, No. 4: 1607–1614 (1977).
2.172. J.R. Chelikovsky, J.C. Phillips, *Phys. Rev., B* **17**, No. 6: 2453–2477 (1978).
2.173. N.V. Ganina, V.B. Ufimtsev, V.I. Fistul, *JTF*, **8**, iss. 10: 620–623 (1982).
2.174. Yu.F. Biryulin, N.V. Ganina, M.G. Milvidsky *et al*., *FTP* **17**, iss.1: 108–114 (1983) (in Russian).
2.175. Yu.A. Bratashevsky, V.M. Vasilkov, N.A. Doroshenko, *FTP* **11**, iss. 11: 2214–2216 (1977) (in Russian).
2.176. R.Kh. Akchurin, V.G. Zinovyev, V.B. Ufimtzev *et al*., *FTP* **16**, iss. 2: 202–206 (1982) (in Russian).
2.177. A.O. Animalu, V. Heine, *Phil. Mag.* **12**, No. 120: 1249–1270 (1965).
2.178. A. Baldareschi, *Luminescence* **7**, No. 1: 79–91 (1973).
2.179. S.T. Pantelides, *Rev. Mod. Phys.* **50**, No. 4: 707–722 (1978).
2.180. A.Yu. Zakharov, Ya.Ya. Shcherbak, *FTP* **10**, iss. 4: 775–777 (1976) (in Russian).
2.181. J.C. Phillips, *Phys. Rev., B* **1**, No. 4: 1545–1548 (1970).
2.182. A. Baldereschi, J.J. Hopfeld, *Phys. Rev. Lett.* **28**, No. 3: 171–174 (1972).
2.183. J.W. Allen, *J. Phys. C.* **4**: 1936–1944 (1971).
2.184. V. Heine, M.L. Cohen, D. Weaire, *The Pseudopotential Concept*, New York–London: Academic Press, 557 p. (1970).
2.185. J.C. Phillips, *Rev. Mod. Phys.* **42**, No. 3: 317–356 (1970).

2.186. V.K. Bazhenov, V.I. Soloshenko, V.V. Timofeenko, *Neorg. mater*. **12**, No. 6: 981–985 (1976) (in Russian).
2.187. S.T. Pantelides, *Festkorperproblem* **15**, No. 1: 149–161 (1975).
2.188. A.O. Animalu, *Phil. Mag*. **13**, No. 121: 53–69 (1966).
2.189. V.K. Bazhenov, V.I. Fistul, *FTP* **18**, iss. 8: 1347–1362 (in Russian).
2.190. G. Leibfried, N. Brener, *Point Defects in Metals*. Berlin–New York: Springer–Verlag, 429 p. (1978).
2.191. R.M. Martin, *Phys. Rev., B* **1**, No. 10: 4005–4011 (1970).
2.192. E.M. Lifshitz, *JETF* **17**, iss: 1017–1028 (1947) (in Russian).
2.193. F. Bassani, G. Iadonisi, B. Preziosi, *Phys. Rev*. **186**, No. 3: 735–746 (1969).
2.194. M. Jaros, S. Brand, *Phys. C*. **12**, No. 3: 525–539 (1978).
2.195. S.G. Louie, M.S. Dchluter, J.R. Chelikovsky *et al*., *Phys. Rev. B* **13**, No. 4: 1654–1663 (1976).
2.196. F.G. Koster, J.C. Slater, *Phys. Rev*. **96**, No. 5: 1208–1223 (1954).
2.197. J. Gallaway, A.J. Huges, *Phys. Rev, B* **156**, No. 3: 860– 876 (1967).
2.198. A. Mauger, J. Friedel, *Phys. Rev., B* **12**, No. 6: 2543–2546 (1975).
2.199. G.I. Kharus, N.G. Shelushinina, *FTP* **12**, iss. 3: 431–438 (1978) (in Russian).
2.200. M. Lannoo, P. Lenglart, *J. Phys. Chem. Sol*. **30**, No. 10: 2409–2418 (1969).
2.201. J.G. Gay, J.R. Smith, *Phys. Rev., B* **9**, No. 10: 4151–4164 (1974).
2.202. S. Brand, M. Jaros, *J. Phys. C*. **12**, No. 12: 2789–2795 (1979).
2.203. E. Cohen, M.D. Sturge, *Phys. Rev., B* **15**, No. 2:1039–1051 (1977).
2.204. J.W. Allen, *J. Phys. C*. **1**, No. 4: 1136–1148 (1968).
2.205. F. Thuselt, K. Kreher, H.J. Wunsche, *Sol. St. Commun*. **36**, No. 6: 563–566 (1980).
2.206. O.F. Sankey, H.P. Hjalmarson, J.D. Dow *et al*., *Phys. Rev. Lett*. **45**, No. 20: 1656–1659 (1980).
2.207. G.A. Baraff, M. Schluter, *Phys. Rev., B*, **19**, No. 9: 4965–4972 (1979).
2.208. J. Bernholc, N.O. Lipari, S.T. Pantelides, *Phys. Rev., B*, **21**, No. 8: 3545–3562 (1981).
2.209. M. Lannoo, G. Alan, *Sol. St. Commun*. **33**, No.3: 293–297 (1980).
2.210. T.N. Morgan, *Phys. Rev. Lett*. **24**, No. 16: 887–890 (1970).
2.211. G.G. Kleiman, *Phys. Rev., B* **19**, No. 6: 3198–3214 (1979).
2.212. A.V. Ermakov, V.V. Egorkin, V.B. Ufimtzev, *JFKh* **27**, No. 6: 1544–1546 (1982) (in Russian).
2.213. S.B. Evgenyev, *JNKh* **32**, No. 11: 2781–2784 (1987) (in Russian).
2.214. V.M. Glazov, A.S. Timoshin, V.B. Ufimtzev, *DAN SSSR* **218**, No. 5: 1097–1099 (1974) (in Russian).
2.215. R.A. Swalin, *Thermodynamics of Solids*. Massachusetts–London: Addison–Wesley Publ. Co., 220 p. (1966).
2.216. A.M. Kosevich, *Osnovy mekhaniki kristallicheskoi reshetki* (The Fundamentals of Crystal Lattice Dynamics). Moscow: Nauka, 280 p. (1972) (in Russian).
2.217. V.I. Fistul, *Raspad peresyshchennykh poluprovodnikovykh rastvorov* (Decomposition of Oversaturated Semiconductor Solutions). Moscow: Metallurgia, 240 p. (1977) (in Russian).
2.218. N.S. Rytova, E.V. Solovyeva, M.G. Milvidsky, *FTP* **16**: 1491–1494 (1982) (in Russian).

2.219. E.V. Solovyeva, N.S. Rytova, M.G. Milvidsky *et al*., *FTP* **15**: 2141–2146 (1981) (in Russian).
2.220. Yu.F. Biryulin, N.V. Ganina, V.V. Chaldyshev, *FTP* **15**: 1348–1351 (1981) (in Russian).
2.221. E.V. Solovyeva, M.G. Milvidsky, N.V. Ganina, *FTP* **16**: 1810–1815 (1982) (in Russian).
2.222. G.G. Kleiman, M.F. Decker, *Phys. Rev*., *B* **17**: 924–928 (1978).
2.223. P. Cappelletti, G.F. Cerofolini, G.U. Pignatel, *J. Appl. Phys*. **54**: 853–856 (1983).
2.224. M.G. Milvidsky, N.S. Rytova, E.V. Solovyeva, *Tzvet. met*., No. 8: 23–26 (1991) (in Russian).
2.225. N.S. Rytova, E.V. Solovyeva, *FTP* **20**: 1380–1387 (1986) (in Russian).
2.226. E.V. Solovyeva, B.M. Leiferov, A.G. Lototsky, *FTP*, **18**: 1573–1576 (1984) (in Russian).
2.227. L.B. Ta, H.M. Hobgood *et al*., *Appl. Phys. Lett*. **41**: 1091–1093 (1982).
2.228. R.Kh. Akchurin, T. Lill, *JFKh* **64**: 1906–1919 (1990) (in Russian).
2.229. R.Kh. Akchurin, *JFKh* **62**: 1764–1770 (1988) (in Russian).
2.230. R.Kh. Akchurin, *Kristallografia* **34**: 520–523 (1989) (in Russian).
2.231. R.Kh. Akchurin, I.O. Donskaya, S.I. Dulin *et al*., *Kristallografia* **33**: 464–470 (1988) (in Russian).
2.232. R.Kh. Akchurin, *Kristallografia* **38**: 198–204 (1993) (in Russian).
2.233. R.Kh. Akchurin, Le Din Kao, D.N. Nishanov *et al*., *Izvestia AN SSSR, ser. Neorg. mat.* **22**: 9–12 (1986) (in Russian).
2.234. Yu.F. Biryulin, V.V. Chaldyshev *et al*., *FTP* **19**: 1104–1107 (1985) (in Russian).
2.235. M.G. Milvidsky, V.B. Osvensky, S.S. Shifrin, *J. Cryst. Growth* **52**, No. 3: 396–403 (1981).
2.236. Yu.N. Bolsheva, Yu.A. Grigiryev, M.G. Milvidsky, *Kristallografia* **27**: 722–728 (1982) (in Russian).
2.237. Yu.M. Boritsky, N.I. Gorbacheva, P.M. Grinshtein, *FTP* **18**: 1303–1312 (1982) (in Russian).
2.238. S.T. Picraux, F.L. Vook, *Phys. Rev., B* **18**, No. 5: 2066–2077 (1978).
2.239. S.T. Picraux, F.L. Vook, H.J. Stein, *Defects and Radiation Effects in Semiconductors 1978, Conference Series No. 46*, Bristol–London: IOPP, 31–44 (1979).
2.240. E.E. Haller, *Defects and Radiation Effects in Semiconductors, Conference Series 1978, No. 46*, Bristol–London: IOPP, 205–211 (1979).
2.241. R.L. Kleinhenz, Y.H. Lee *et al*., *Defects and Radiation Effects in Semiconductors 1978, Conference Series No. 46*, Bristol–London: IOPP, 200–204 (1979).
2.242. D.G. Andrianov, G.G. Myasishcheva, Yu.V. Obukhov *et al*., *JETF* **56**: 1195 (1969) (in Russian).
2.243. K.M. Crove *et al*., *Bull. Amer. Phys. Soc*. **17**: 594 (1972).
2.244. D.G. Andrianov, L.A. Goncharov *et al*., *FTP* **10**: 1167–1172 (1976) (in Russian).
2.245. J. Shy-Yih Wang, Ch. Kittel, *Phys. Rev., B* **7**: 713 (1973).
2.246. D.G. Andrianov, G.G. Myasishchev *et al*., *FTP* **12**: 202–204 (1978) (in Russian).

2.247. W.J. Choyke, *Radiation Effects in Semiconductors 1976, Conference Series No. 31*. Bristol–London: IOPP, 58–69 (1977).
2.248. O.V. Romanov, P.P. Konorov, T.A. Kotova, *FTP* **3**: 124 (1969) (in Russian).
2.249. M.N. Brodsky, ed., *Amorphous Semiconductors*. Berlin: Springer–Verlag (1979).
2.250. Y. Yatsurugi, N. Akiyama *et al*., *J. Electrochem. Soc*. **120**: 975 (1973).
2.251. W. Kaiser, P.H. Keck, C.F. Lange, *Phys. Rev*. **101**: 1264–1268 (1956).
2.252. H.J. Hrostoski, W. Kaiser, *Phys. Rev*. **107**: 966–971 (1957).
2.253. V. Cazcarra, P. Zunino, *J. Appl. Phys*. **51**: 42 06 (1980).
2.254. G.D. Watkins, *Lattice Defects in Semiconductors 1974, Conference Series No. 23*. London–Bristol: IOPP, 1–22 (1975).
2.255. H.H. Bekman, *Oxygen Aggregation in Semiconductors Study by Magnetic Resonance*, Thesis. Amsterdam: Amsterdam Univ. (1991).
2.256. K.L. Brower, *Radiative Effects*. **8**: 213–219 (1971).
2.257. G.I. Aleksandrova, L.A. Goncharov *et al*., *Zav. lab*. **42**, No. 9: 1079–1081 (1976) (in Russian).
2.258. D.R. Bosomworth, W. Hayes *et al*., *Proc. Royal Soc*. **317**, No. 1528: 133–152 (1970).
2.259. J.W. Corbett, R.S. McDonald, G.D. Watkins, *J. Phys. Chem. Soc*. **25**, No. 8: 873–879 (1964).
2.260. V.V. Emtsev, L.A. Goncharov, T.N. Dostkhodjaev, *FTP* **12**: 139–144 (1978) (in Russian).
2.261. C.S. Fuller, F.H. Doleiden, *J. Phys. Chem. Soc*. **19**, No. 1–2: 251–260 (1961).
2.262. C.J. Frosch, C.D. Thurmond, H.G. White, *Trans. Met. Soc*. **239**: 365 (1967).
2.263. C.N. Cochran, L.M. Foster, *J. Electrochem. Soc*. **109**: 142 (1962).
2.264. A.M. Zykov, B.E. Samorukov, *Izvestiya vuzov, ser. Fizika* **11**: 157 (1970) (in Russian).
2.265. L.A. Borisova, Z.L. Akkerman, A.N. Dorokhov, *Izvestiya AN SSSR, ser. Neorg. mat*. **13**: 908 (1977) (in Russian).
2.266. Z.L. Akkerman, L.A. Borisov, A.F. Kravchenko, *FTP* **10**: 997 (1976) (in Russian).
2.267. S. Toytomi, K. Morigaki, *J. Phys. Soc. Japan* **29**: 800 (1970).
2.268. N.P. Ilyin, V.F. Masterov *et al., FTP* **10**: 1581 (1976) (in Russian).
2.269. P.J. Dean, C.H. Henry, C. Frosch, *Phys. Rev*. **168**: 812 (1968).
2.270. M. Gerschenson, F. Trumbore, R. Mikulyak, *J. Appl. Phys*. **37**: 483 (1966).
2.271. H.G. Grimmeiss *et al*., *Proc. 12th Intern. Conf. on Semiconductor Physics*. Stuttgart, 386 p. (1974).
2.272. T.N. Morgan, *Phys. Rev. Lett*. **40**, No. 3: 190 (1978).
2.273. R.A. Malinauskas, L.Ya. Pervova, V.I. Fistul, *FTP* **16**: 466–469 (1982) (in Russian).
2.274. K.L. Brower, *Phys. Rev., B* **9**, No. 6: 2607–2617 (1974); *Erratum. Phys. Rev., B* **17**, No. 10: 4130 (1978).
2.275. G.D. Watkins, K.L. Brower, *Phys. Rev. Lett*. **36**, No. 22:1329–1332 (1976).
2.276. S.M. Sze, J.C. Irvin, *Sol. St. Electron*. **11**: 599 (1968).
2.277. D.P. Bortfield, B.J. Curtis, H. Meier, *J. Appl. Phys*. **43**: 1293 (1972).
2.278. P.J. Dean, E.A. Schonherr, R.B. Zetterstrom, *J. Appl. Phys*. **41**: 3475 (1970).
2.279. W.J. Kleinfelder, *Stanford Univ. Electronics Lab*., Report K701-1, SU-SEL-67-015.

Chapter 3

Impurity Solubility in Semiconductors (a Macroscopic Approach)

3.1 RETROGRADE SOLUBILITY OF IMPURITIES

The solubility of most impurities in semiconductors has a retrograde character. This means that maximum solubility occurs above eutectic melting temperature and decreases with decreasing temperature. This phenomenon is also known as negative solubility. In most metallic systems, the maximum solubility of one metal in another metal occurs at eutectic temperature. The solubility in such systems had been studied long before this was done in semiconductors. It is probably for this reason that maximum solubility at eutectic temperature was considered as a common rule and retrograde behavior as an exception to this rule.

A possible appearance of a negative solubility region was pointed out by Van Laar in 1908, but his work did not attract the researchers' attention at that time. It was only forty years later, after the solubility curves were studied in detail, that this problem received due attention. Mejering was one of the first to derive general expressions for this solidus region [1]. The principal results of the calculations are given below.

Impurity solubility is calculated in macroscopic thermodynamics from the equilibrium condition of the liquid and solid phases. With the activity coefficients, it is written as follows:

$$\mu_{\alpha l}^{0} + kT \log \gamma_{\alpha l} x_{\alpha l} = \mu_{\alpha s}^{0} + kT \log \gamma_{\alpha s} x_{\alpha s}. \qquad (3.1.1)$$

It should be noted that chemical potentials must, in reality, refer to an idealized state, rather than to a pure substance, when this component is present under conditions similar to those in the solution [2]. In the solid phase, for example, the ideal state should have a crystal lattice corresponding to a tetrahedral semiconductor. To avoid these complications, one normally uses chemical potential values for a pure component α having its own structure, while all deviations are taken into account by introducing activity coefficients.

The coefficient of impurity distribution between the solid and liquid phases is defined by the impurity concentration ratio in both phases:

$$\log K_{\alpha}^{\beta} = \log \frac{x_{\alpha s}^{\beta}}{x_{\alpha l}} = \frac{\mu_{\alpha l}^{0} - \mu_{\alpha s}^{0}}{kT} + \log \frac{\gamma_{\alpha s}}{\gamma_{\alpha l}}, \qquad (3.1.2)$$

where $x_{\alpha s}{}^{\beta}(x_{\alpha l})$ is the fraction of α substance in the solid (liquid) phase, γ is an activity coefficient, and superscript β indicates the dissolved impurity position in a crystal.

The first term in (3.1.2) can be expressed with sufficient accuracy through the melting enthalpy $\Delta H_{\alpha}{}^{f}$ and entropy $\Delta S_{\alpha}{}^{f}$ of the dissolved component, taken at the melting temperature of the pure component. Their temperature dependence, which is the heat capacity difference in both phases, is neglected. As a result, one gets the equation first derived by Thurmond and Struthers [3] for the treatment of solidus lines for various impurities in germanium and silicon:

$$\log K_{\alpha}^{\beta} = \frac{\Delta H_{\alpha}^{f} - \Delta H_{\alpha}^{\beta}}{kT} - \frac{\Delta S_{\alpha}^{f} - \Delta S_{\alpha}^{\beta}}{R} + \log \gamma_{\alpha l}. \qquad (3.1.3)$$

This formula is convenient for a comparison with experimental data. This is a straight line equation in the coordinates $\log K = f\,(10^3/T)$, and the line slope is used to find the impurity dissolution enthalpy.

The dissolution enthalpy of a substance is involved in (3.1.3) as a parameter, and the temperature dependence of the distribution coefficient is purely exponential. It is easy to see, therefore, that negative solubility arises when the dissolution enthalpy is higher than the melting enthalpy of a pure substance. This result will become clearer if equation (3.1.3) is combined with that for the liquidus line of a perfect solution [2]:

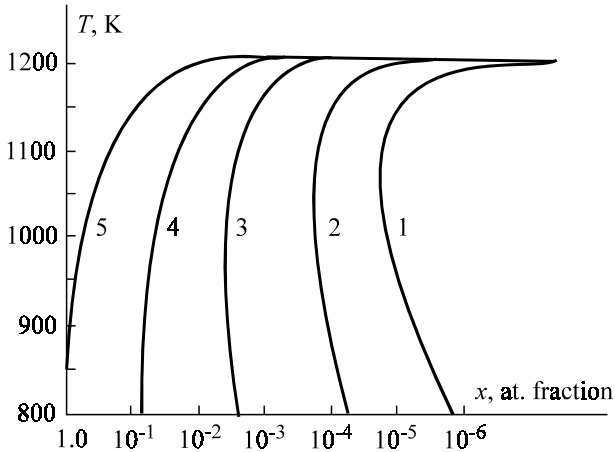

Figure 3.1. Solidus lines for impurity solubility in material with melting heat ΔH_m = 33.5 kJ×mole and melting entropy ΔS_m = 12.6 J/(g×atom×degrees K), calculated for various dissolution enthalpy values: 1 – 92; 2 – 69; 3 – 46; 4 – 23; 5 – 15 kJ/mole.

$$\log(1-x_{\alpha l}) = \frac{\Delta H_\beta^f}{k}\left(\frac{1}{T_\beta^f}-\frac{1}{T}\right), \tag{3.1.4}$$

where ΔH_β^f and T_β^f refer to solvent melting. From (3.1.3) and (3.1.4), we have [4]

$$N_\alpha^\beta = N^\beta\left(1-\exp\frac{T\Delta S_\beta^f-\Delta H_\beta^f}{kT}\right)\exp\frac{\Delta H_\alpha^f-\Delta H_\alpha^\beta-T\Delta S_\alpha^f}{k}. \tag{3.1.5}$$

The results of calculations obtained with this formula are given in Figure 3.1 [4]. They support the above reasoning concerning the appearance of retrograde solubility.

Hall [5] used the above equations in a linear approximation of the temperature dependence of enthalpy

$$\Delta H = \Delta H_0 - \alpha T. \tag{3.1.6}$$

They drew the conclusion that there was a common point in the distribution coefficient curves when these were extrapolated to a high temperature

region. This permitted a correction to be made of the available experimental findings.

However, the linear dependence in (3.1.6) and the existence of interception points of the curves was subjected to criticism by some investigators. In particular, Trumbor and co-workers found in their analysis of tin solubility in silicon and germanium that the distribution coefficient logarithm near the semiconductor melting temperature rose abruptly, deviating considerably from the dependence in (3.1.3), whereas it was indeed valid at a lower temperature [6]. This is not surprising because there are appreciable positive deviations from the Raoult law near the melting point in the liquid phase [7], which affects the activity coefficient and, hence, the distribution coefficient.

In a study of aluminum solubility in gallium and germanium, Trumbor and co-workers [6] found the respective deviations from Hall's suggestion of a common point in distribution coefficient curves at high temperatures. But this time, the abrupt rise of this coefficient could not be interpreted as being due to the behavior of the activity coefficient, because the liquid Al–Ge system has a negative deviation from the Raoult law and the Ga–Ge liquidus is practically perfect [7]. The slope of the solubility curve was suggested in [8, 9] to result from the impurity ionization.

It was assumed in the derivation of (3.1.3) that the dissolved impurity was in a neutral state in both phases. In the solid phase, however, impurities are partly ionized. The electrical interaction affects the solubility later. This circumstance was taken into account in [9] to obtain an exact expression allowing for interactions in the solution and impurity ionization:

$$K_\alpha^\beta = f_n^{-1} \exp\left[\left(\Delta H_\alpha^f - TS_\alpha^f + \Omega_l\left(1 - x_{\alpha l}\right)^2 - H_\alpha^\beta\right)\Big/ kT\right], \qquad (3.1.7)$$

where f_n is the Fermi–Dirac function and Ω_l is the interaction parameter in a regular solution.

As the temperature increases, the distribution coefficient K given by this formula normally decreases, which fits well the retrograde character of the impurity behavior. The solubility decrease at high temperatures (in the constant temperature approximation of Gibbs partial potential) is associated with a lower impurity activity in the melt, calculated from the equilibrium liquidus curve.

The fact that the Fermi–Dirac function appears in (3.1.7) demonstrates that impurity solubility is affected by all active defects available in a crystal. This permits allowance for double doping effects, in particular, simultaneous introduction of donors and acceptors into a crystal.

Consider, as an illustration, the solubility of a donor impurity. Suppose the solubility is not high and the semiconductor remains nondegenerate. Then, in accordance with the known approximation for the Fermi energy from (3.1.7), one gets [4]

$$K_\alpha^\beta = \frac{\gamma_a}{\alpha_\beta} \exp\left(g_\alpha^\beta / kT\right)\left[R_\alpha^\beta \frac{N_c}{n} \exp\left(\varepsilon_\alpha^\beta / kT\right) + r_\alpha^\beta\right], \qquad (3.1.8)$$

where $R_\alpha^\beta\left(r_\alpha^\beta\right)$ are degeneration factors of the unfilled (filled) electronic state of the impurity. Incorporation of donors increases the free electron concentration. As soon as this concentration reaches a level at which they begin to determine the free carrier concentration, the distribution coefficient begins to decrease. Further introduction of impurities into the melt does not raise appreciably the free electron concentration. If acceptors are introduced simultaneously, the free electron concentration goes down because of the compensation, thereby increasing the donor solubility. Maximum solubility occurs at complete compensation. A similar result was obtained earlier by Glazov and co-workers [10, 11].

It is worth noting that the solubility curve exponent contains the difference $\Delta H_\alpha^{\,f} - \Delta H_\alpha^{\,\beta}$. Therefore, the retrograde character of solubility is associated in both cases with the fact that the solubility enthalpy is higher than the melting heat of the component. The high interaction energy during dissolution decreases the concentration of dissolved impurities, producing a negative solubility region.

3.2 SOLUBILITY OF HYDROGEN-LIKE IMPURITY ATOMS IN GERMANIUM AND SILICON

The authors of [12, 13] took into account the ionization of dissolved impurities in the derivation of expressions (3.1.7) and (3.1.8). The ionization should necessarily be allowed for in the case of hydrogen-like impurities because of the appearance of shallow energy levels.

If the fraction of neutral atoms, F_n, in their total abundance x_c is

$$F_n = \frac{x_{cn}^s}{x_c}, \qquad (3.2.1)$$

one can derive a general expression for the distribution coefficient allowing for the impurity ionization in the solid phase:

$$\ln K_\alpha = -\frac{\Delta H}{RT} - \frac{\Delta S_\alpha^{\mathrm{f}} - \Delta S_\alpha^{\mathrm{s}}}{R} + \ln \gamma_\alpha^{\mathrm{l}} - \ln F_n , \qquad (3.2.2)$$

where ΔH is described by a straight line of (3.1.6), according to Hall.

The necessity to account for the quantity F_n is due to the fact that the equilibrium $C^0 \rightleftarrows C^+ + e^-$ for donors or $C^0 \rightleftarrows C^- + p^+$ for acceptors in the high temperature region can shift appreciably toward the neutral component at high concentrations of intrinsic carriers. The quantity F_n can be easily obtained by considering the equilibrium between neutral and ionized impurities in a crystal. For a hydrogen-like impurity center, it is

$$F_n = \frac{1}{1 + \frac{1}{2}\exp\left(\dfrac{|E_{\mathrm{F}} - E_{\mathrm{i}}|}{RT}\right)}, \qquad (3.2.3)$$

where E_{F} and E_{i} are the positions of the Fermi and local levels of the impurity, counted from the conduction band bottom.

Unfortunately, a straightforward calculation of F_n is complicated by the fact that a solid solution containing over 10^{20} cm^{-3} impurity atoms is degenerate in a wide temperature range [14]. Moreover, the available information on the band structure of germanium and silicon at high temperatures is very limited. A rough value of F_n for germanium can be obtained from Blakemore's plots [15] in the approximation of a simple band model and Boltzmann's statistics. In this approach, F_n varies slightly, within 0.85÷0.90, in the low temperature region (< 800°C), but above 800°C it falls relatively fast to 0.4 at the melting temperature of germanium.

For a degenerate semiconductor, the value of F_n can be found from the expression

$$F_n = 1 + \frac{1}{2} F_{1/2}(\mu/kT), \qquad (3.2.4)$$

where $F_{1/2}(\mu/kT)$ is the known Fermi integral of the "half" index, tabulated in [16], and the Fermi level for this case is calculated from a formula in [15]:

$$\exp\frac{E_F}{kT} = 2N_a \times \left\{ N_v - cN_a + \left[\left(N_v - cN_a \right)^2 + 8N_a N_v \exp\frac{E_i^{Si(Ge)}}{kT} \right]^{1/2} \right\}^{-1} \tag{3.2.5}$$

with the recommended value of $c = 0.27$.

The first detailed treatment of solubility of basic hydrogen-like impurities in germanium and silicon taking account of impurity ionization was carried out by Lehovec [17]. Later, this problem was considered [8, 9], with reference to germanium, in terms of a regular solution approximation, using equations similar to (3.2.2), the only difference being that the dissolution enthalpy was replaced by experimental values of *K*. In one work [9], the authors found the solidus curves from the theoretical Fermi level position, and in the other [8], they solved the inverse problem, namely, estimated the Fermi level position from solubility data for various impurities in germanium. The F_n value was found from Blakemore's plots as a function of temperature and doping impurity concentration.

Comprehensive information on the solubilities of group-III and group-V elements of the periodic table is presented in the book by Glazov and Zemskova [10]. Here, we only list the authors' principal conclusions.

(1) Liquidus curves for the states of group-III and group-V solid solutions can be basically described in terms of the regular solution theory. Solutions with acceptors show a better fit to the theory than those with donors. This may be due to the high donor vapor elasticity, so that a thermodynamic equilibrium in the system is observed only in closed volumes with the counterpressure of saturating vapors at a particular temperature. It is, probably, for this reason that the liquidus curves for these solutions are not quite accurate. In particular, the high temperature liquidus curve for phosphorus in silicon is hypothetical and was extrapolated from the low temperature region of the state diagram.

(2) The impurity solubility is fairly high. Most atoms in a solution are neutral, so electrical measurements cannot provide a total concentration of dissolved atoms. This circumstance should be taken into account when comparing theoretical and experimental values.

Intensive investigations of equilibrium solubility of hydrogen-like impurities were carried out during the 1950–1960s, and their principal results can be found in the reviews [6, 13]. Since the 1960s, the focus has been on silicon. For some impurities (B, C, N, P), experimental data on solubility in germanium are still quite scanty.

Table 3.1. Solubility data for group-III and -V hydrogen-like impurities in germanium.

Impurity	Solubility		K[1]	Dissolution enthalpy ΔH (eV)			
	C_{max}, cm^{-3}	T_{max}, °C		Exper.	Theory [18]	Theory [19]	Theory [22]
B	$\leq 10^{18}$		17.4 [20] 12.2 [21]				
Al	4.3×10^{20}	675	0.073 [14]	0.12 0.31 [23]	–0.02	0.12	–0.04
Ga	4.9×10^{20}	670	0.087 [14]	0.12	–0.01	0.10	0.14
In	4.0×10^{18}	800	7×10^{-4} [6]	0.2 [6]	0.64	0.57	0.69
In	6.0×10^{18}	835		0.85 [24] 0.87[2] [5] 0.70 [25]			
P	2×10^{20}	600	0.01 [6]			0.11	
As	8×10^{19}	800	0.02 [6]	0.54 [27] 0.32[3] [5] 0.40 [28]		0.26	–0.13
Sb	1.2×10^{19}	800	3×10^{-3} [6]	0.70 [27] 0.63 [29] 0.40 [28] 2.19 [3]	0.36	0.45	0.19

[1] K – distribution coefficient at melting temperature.
[2] Theoretical $\Delta H = 0.87 - 3.17\times10^{-4}\,T$ (eV).
[3] Theoretical $\Delta H = 0.32 - 1.13\times10^{-4}\,T$ (eV).

Tables 3.1 and 3.2 give experimental and theoretical results on the solubility and dissolution enthalpy of hydrogen-like impurities in germanium and silicon.

The data in Tables 3.1 and 3.2 sometimes demonstrate a considerable spread of the values of C_{max} and, especially, of ΔH found by different workers. For many impurities, no theoretical calculations are available. A step forward was made by Bulyarsky and co-workers [4, 43, 44], who suggested finding equilibrium impurity solubility by taking into account various interactions between impurities and other point defects and the activities of dissolved atoms. An attempt was made in [43, 44] to calculate Gibbs partial free energies of some hydrogen-like donors and acceptors in silicon. In this treatment [4], the distribution coefficient was represented, in addition to

Table 3.2. Solubility data for group-III and -V hydrogen-like impurities in silicon.

Impurity	Solubility		K[1]	Dissolution enthalpy ΔH (eV)			
	C_{max}, cm^{-3}	T_{max}, °C		Exper.	Theory [18]	Theory [19]	Theory [22]
B	5×10^{20}	1200	0.43 [31]	0.43 [30]			
	8×10^{20}	1200		0.73 [31]			
Al	2×10^{19}	1150	0.002 [6]	0.46 [6]	0.47	0.69	0.17
				0.43 [33]			
				0.66 [32]			
				1.00[2] [5]			
Ga	4.0×10^{19}	1250	0.008 [6]	0.46 [6]	0.44	0.50	0.43
In	1.6×10^{18}	1300	4×10^{-4} [6]	2.51 [34]	1.22	0.93	1.28
P	1.3×10^{21}	1200	0.35 [6]	0.50 [38]		0.03	
	2×10^{20}	1200		0.70 [31]			
	5×10^{21}	1100		0.76[3] [39]			
As	2×10^{21}	1150	0.3 [6]	0.24 [6]		–0.03	–0.06
				0.47 [40]			
	5×10^{19}	1300	0.023 [6]	0.31 [6]	0.34	0.39	0.64
Sb	5.3×10^{19}	1300		0.58 [42]			
				3.49 [41]			

[1] K – distribution coefficient at melting temperature.
[2] Theoretical $\Delta H = 1.00 - 2.56\times10^{-4}\ T$ (eV).
[3] Theoretical $\Delta H = 0.76 - 2.8\times10^{-4}\ T$ (eV).

Formula (3.1.7) derived in the regular solution approximation without allowance for multiply charged states, as

$$K_\alpha^\beta = \frac{\gamma_\alpha}{\gamma_\beta}\exp\left(-\frac{g_\alpha^\beta}{kT}\right)\left[R_\alpha^\beta \exp\frac{E_F - E_\alpha^\beta}{kT} + r_\alpha^\beta\right], \qquad (3.2.6)$$

where the designations are the same as in Section 3.1.

The concentration of a dissolved, say, donor impurity was described in [43, 44] as

$$N_s = \frac{a_s}{a_{si}} N \exp\left(-\frac{g_s}{kT}\right)\left(\frac{1}{2}\exp\frac{E_F - E_s}{kT} + 1\right). \qquad (3.2.7)$$

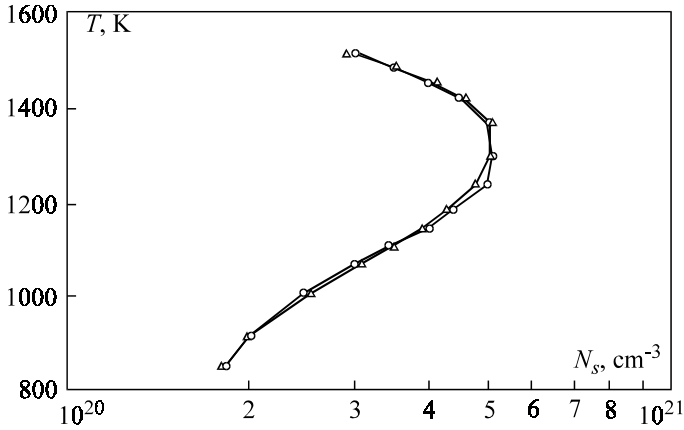

Figure 3.2. Experimental (circles) and calculated (triangles) solidus lines for Al in silicon.

The Fermi level position, E_F, was calculated with (3.2.5), and the activity coefficients were found using the ratios in the regular solution theory [45]:

$$\gamma_\alpha = \exp\frac{\Omega_l\left(1-x_{\alpha l}\right)^2}{kT} \qquad \gamma_\beta = \exp\frac{\Omega_l\left(x_{\alpha l}\right)^2}{kT}, \tag{3.2.8}$$

where the interaction parameter was found from the liquidus curve as

$$\Omega_l = \frac{1}{x_{\alpha l}}\left[TS_\beta^f - \Delta H_\beta^f - kT\lg\left(1-x_{\alpha l}\right)\right]. \tag{3.2.9}$$

Besides, the partial potentials contain the difference between the chemical potentials of mixed components. This difference can be written as

$$\mu_\alpha^0 - \mu_\beta^0 = \Delta H_\alpha^f - \Delta H_\beta^f - T\left(S_\alpha^f - S_\beta^f\right). \tag{3.2.10}$$

After the substitution of (3.2.8) through (3.2.10) into (3.2.6), one gets (3.1.7).

The consistency between (3.1.7) and (3.2.6) is justified since it was shown in Section 1.2 that the approaches based on free energy minimization and on chemical potentials are identical. However, free energy analysis al-

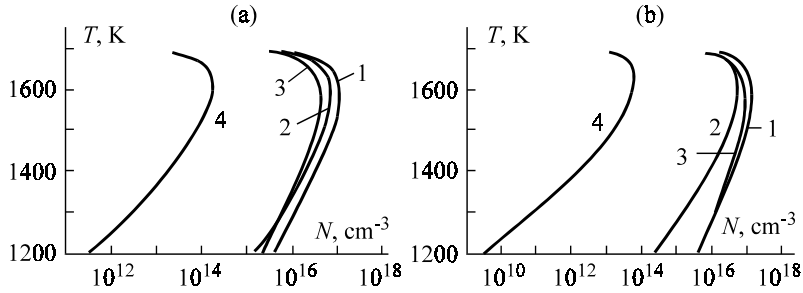

Figure 3.3. Theoretical solidus lines for different Au charge states in undoped silicon (a) and in silicon doped with donors at the concentration $N = 10^{20}$ cm^{-3}: $1 - N_{Au}^{-}$; $2 - N_{Au}^{+}$; $3 - N_{a}^{-}$; $4 - N_{d}^{+}$.

lows a better understanding of the physical mechanisms of interaction and a discussion of more complicated situations, such as multiply charged impurities and their interactions via force fields.

A combined solution of equations (3.2.5) and (3.2.7) with the substitution of (3.1.8), (3.2.6), and (3.2.8) with respect to $\exp(-g_s/kT)$ yielded Gibbs partial potential and temperature dependences of impurity dissolution enthalpy and entropy [43, 44]. As a result, a better agreement was achieved between experimental and calculated solidus curves. This is illustrated in Figures 3.2 and 3.3. The principal result is illustrated in Figure 3.4 showing clearly the nonlinear character of the $\Delta H(T)$ curve. This means that Hall's concept formulated as (3.1.6) does not work, as was suggested by many workers.

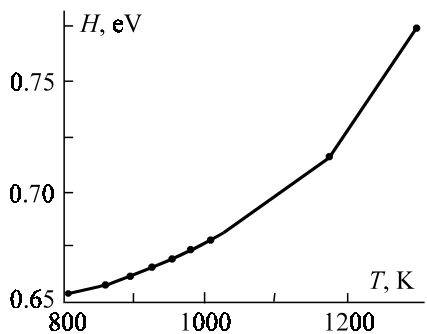

Figure 3.4. The temperature dependence of dissolution enthalpy for Ga in silicon.

3.3 HYDROGEN-LIKE IMPURITY SOLUBILITY IN $A^{III}B^{V}$ COMPOUNDS

Impurity solubility mechanisms in $A^{III}B^{V}$ compounds have been a subject of interest to many investigators [46–50]. Without claiming present all this material, we would like to point out some general mechanisms and factors that distinguish the impurity solubility in these compounds from that in elemental semiconductors.

(1) The process of impurity dissolution is much affected by the formation of complexes, which is stimulated by a high concentration of point defects and their greater diversity than in elemental semiconductors. For details of defect association, the reader is referred to Chapters 5 and 6.

(2) The significance of crystal growth conditions rises. The governing factors in elemental semiconductors are temperature and impurity concentrations in the ambient. An additional factor in $A^{III}B^{V}$ compounds is elasticity of the volatile component vapor. By varying this parameter, we can vary the homogeneity region width and the concentrations of cation and anion vacancies. This, in turn, changes the conditions for impurity dissolution and formation of complexes.

(3) Dislocations and their interaction with impurities also play an important role [47].

It is clear that an adequate description of impurity solubility in such compounds is not a simple task. Nevertheless, it was possible to draw certain conclusions. The concentration of complexes decreases abruptly at high temperatures, and so high temperature data can be treated without considering defects associations. Complexation results in a lower concentration of free charge carriers than the dissolved impurity concentration. However, direct concentration measurements of dissolved material do give an absolute value of dissolved impurity concentration. Therefore, such measurements should be preferred to electrical measurements.

Variations in the external factors, including vapor pressure of volatile components in the ambient, has been accounted for by theory. For this reason, the theoretical considerations and the data treatment presented below may appear useful. As an illustration, let us consider the solubility of tellurium in GaAs, which has been studied in greater detail than other impurities.

The distribution coefficient of tellurium has been discussed by several workers. The experimental results were reviewed in [48] and summarized in Table 3.3 of this book, together with data for other $A^{III}B^{V}$ compounds [52].

Table 3.3. Distribution coefficients of some impurities in $A^{III}B^{V}$ compounds [31, 52].

Impurity	Distribution coefficients						
	AlSb	GaP	GaAs	GaSb	InP	InAs	InSb
S	0.03	0.23	0.17; 0.5–1.0	0.06	0.8	1.0	0.1
Se	0.003	0.29	0.44–0.55	0.18; 0.4	0.6	0.93	0.5; 0.35; 0.17
Te	0.01	0.026	0.3; 0.054–0.016	0.4	–	0.44	3.5; ≈1; 0.54
Zn	0.02	–	0.3–0.9; 0.1	0.3; 0.16; 0.02	–	0.77	2.3; 3.0; 4.9
Cd	0.002	–	< 0.02	–	–	0.13	0.26
Si	0.045	–	0.1; 0.014	1.0	–	0.4	–
Ge	0.026	–	0.02; 0.03	0.02; 0.2; 0.08	0.05	0.07	0.045
Sn	(2÷8)10	–	0.03; <0.02	0.01	0.03	0.09	0.57

The temperature dependence of the distribution coefficient is defined as

$$K = \frac{\gamma_{Te}}{a_{As}} \exp\left(-\frac{g_{Te}^{As}}{kT}\right)\left[\exp\left(\frac{E_F - E_{Te}^{As}}{kT}\right)+1\right]. \qquad (3.3.1)$$

This dependence was used [4] to find the partial free energy of tellurium solubility in GaAs (Figure 3.5). The Fermi energy was calculated on the assumption that the conductivity was determined by tellurium atoms; not all of the atoms, however, are in an active state. The calculations yielded the following thermodynamic parameters of tellurium solubility: ΔH_{Te}^{As} = – 0.52 eV; S_{Te}^{As} = 8.7k at T = 1273 K and $N_d = 1.7\times10^{17}$ cm^{-3}. The enthalpy and entropy were found to be practically constant in the temperature range 970–1500 K and independent of the doping impurity concentration.

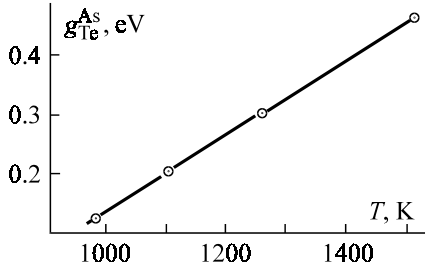

Figure 3.5. The temperature dependence of Gibbs partial potential of Te solubility in GaAs.

3.4 SOLUBILITY OF DEEP IMPURITIES

3.4.1 Solubility in A^{IV} semiconductors

It was shown in Section 2.2 that 3*d*-atoms are primarily dissolved at interstices of elemental semiconductors and always produce deep energy levels in the band spectrum. Here we discuss some general physicochemical and technological problems important to the analysis of deep centers, their distribution in the semiconductor bulk, and behavior during thermal treatment.

Since the formation of every deep center in semiconductors (vacancy, antistructural defect, and, of course, doping impurity) is closely related to the parameters of the ambient, in which the host crystal is formed, the discussion of solubility can be effective only if it considers these parameters. When a thermodynamic equilibrium is established between the semiconductor crystal and the ambient, the conditions are said to be equilibrium. In this case, every type of defect acquires an equilibrium concentration for these particular conditions; the concentration can then be expressed by the respective thermodynamic quantities. In practice, however, we mostly deal with conditions which are close, to a greater or lesser extent, to equilibrium conditions. In that case, we must take into consideration the kinetic factors defining the degree of this closeness.

Let us discuss some aspects of deep center formation by impurities and their behavior in the processes of crystal growth and thermal treatment. The behavior of 3*d*-atoms during crystallization and thermal treatment is primarily determined, as in the case of hydrogen-like and other impurities, by

such impurity parameters as the distribution coefficient K_α, the maximum solubility $C_{\alpha max}$, and the diffusion coefficient D_α at various temperatures. These characteristics of impurity atoms depend on how close the physicochemical properties of the doping element(s) are to those of the host element(s). They also depend on their size ratio and on the electron shell structure of the substituting and the substituted atoms. It is well known that these parameters define the character and specificity of impurity atom incorporation into the crystal lattice and, therefore, the spectrum of the energy levels produced by impurity centers in the semiconductor forbidden gap. Thus, the physical, physicochemical, and technological aspects of impurity behavior in semiconductor materials appear to be intimately interrelated.

For crystallization from a melt, the distribution coefficients of impurities under equilibrium conditions, (K_α^0), can be found from the state diagrams of respective semiconductor–impurity systems as the ratio $C_{\alpha s}^0 / C_{\alpha l}^0$, or they can be calculated from a thermodynamic model for solutions with equal chemical potentials of the impurity in the equilibrium liquid and solid phases, as was demonstrated in Section 3.1. Good results for 3d-impurities can be obtained from the calculation of K_α^0 in the regular solution approximation with formula (3.2.6), which we will borrow from [54] but modify as

$$K_\alpha^0 = \exp\left[\frac{\Delta H_\alpha^f + \Omega_l\left(1 - x_{\alpha l}\right) - \Omega_s\left(1 - x_{\alpha s}\right)^2}{RT} - \frac{\Delta S_\alpha^f}{R}\right]. \qquad (3.4.1)$$

In formula (3.4.1), the parameters of intermolecular interaction in the liquid and solid phases, Ω_l and Ω_s, are taken to be constant (a rigorously regular approximation) or dependent on temperature as, say, $\Omega = a + bT$ (a quasi-regular approximation). The quasichemical solution theory and the assumption of a random particle distribution in the phase of interest can yield an expression for Ω, in the binary system consisting of components *A* and *B*:

$$\Omega = H_{A-B} - 0.5\left(H_{A-A} + H_{B-B}\right), \qquad (3.4.2)$$

where $H_{A\text{-}B}$, $H_{A\text{-}A}$, and $H_{B\text{-}B}$ are interatomic interaction energies of components of different and identical signs.

Regular solutions obey the expressions of (3.2.8). At $\Omega > 0$, $\gamma_\alpha > 1$, and $a_\alpha > x_\alpha$, the interaction in the solution is characterized by repulsion of atoms of components with different signs, while at $\Omega < 0$, $\gamma_\alpha < 1$, and $a_\alpha < x_\alpha$, the attraction of such atoms is dominant. The character of interaction in a liquid

solution affects the equilibrium between the liquid and gas phases: the vapor pressure of the α-component above the solution is higher at $\Omega > 0$ but lower at $\Omega < 0$ than above an ideal solution ($P_\alpha^{id} = x_\alpha P_\alpha^0$) under identical conditions (here, $P_\alpha{}^0$ is the gas pressure of the pure α-component at a given temperature).

Under real crystallization conditions, the time is too short for the equilibrium to be established at the interface because of kinetic limitations. So, crystallization from a melt is usually described using the effective distribution coefficient $K_{\alpha\text{eff}}$ which is, in a simple case, related to the equilibrium coefficient as

$$K_{\alpha\text{eff}} = K_\alpha^0 \Big/ \left[K_\alpha^0 + \left(1 - K_\alpha^0\right) \exp\left(- f\delta / D_\alpha^l\right) \right], \qquad (3.4.3)$$

where f is the growth rate, δ is the diffusion layer thickness in the melt at the crystallization front, and $D_\alpha{}^l$ is the diffusion coefficient of the impurity in the melt.

It is seen from (3.4.3) that the value of $K_{\alpha\text{eff}}$ is affected by the rate ratio of impurity accumulation in the diffusion layer (at $K_\alpha^0 < 1$), by the subsequent equalization of the melt composition defined by f and D_α^l, and by the melt mixing conditions which determine δ.

Generally, the coefficient $K_{\alpha\text{eff}}$ is affected by a number of additional factors, such as the doping level, the interaction between the melt and the ambient, contaminating impurities and their electrical behavior in the semiconductor, the crystallographic orientation of the growth surface, and some others.

Normally, *d*-atoms producing deep centers in semiconductors differ from the host atoms in the electron shell structure and size, due to which reason their building-in into the semiconductor crystal lattice is usually quite difficult in terms of energy. The "resistance" of the host crystal to this building-in leads to relatively large positive values of Ω_s; the values of K_α and $C_{\alpha\ \max}$ are usually much lower than for shallow impurities, as is seen from a comparison of Tables 3.2 and 3.4.

Natural segregation of impurities occurring during oriented crystallization results in their nonuniform distribution along the crystal length. If the flat crystallization front is distorted, the distribution becomes nonuniform in the transverse direction, too. As a rule, the larger the difference between K_α and unity, the more nonuniform is the impurity distribution in both macro- and microvolumes of the growing crystal. The longitudinal impurity distribution in crystals grown by oriented crystallization is described by the equation

Table 3.4. Solubility data for *d*-impurities in silicon.

Impurity	Solubility		K[1]	Dissolution enthalpy ΔH (eV)		
	C_{max}, cm^{-3}	T_{max}, °C		Experim.	Theory ΔH_i	Theory ΔH_s
Sc	2×10^{16}				2.10	
Ti	10^{16}					3.56
V	10^{15}				4.47	
Cr	10^{16}	1250	1×10^{-8}	2.00	1.82	
Mn	10^{16}	1350		2.10	0.56	
Fe	5×10^{16}	1350	8×10^{-6}	2.40	1.59	
Co	2.5×10^{16}	1240		1.45	1.13	
Ni	7×10^{17}	1310		1.40	1.52	
Y					3.56	
Zr	10^{19}				6.02	
Mo	10^{15}				5.78	
Ru	3×10^{16}			3.35	5.27	
Rh	10^{17}			1.68	4.15	
Pd	2.9×10^{16}			1.52	1.16	

[1] K – distribution coefficient at melting temperature; theoretical ΔH values are for the most probable states—interstices (*i*) or sites (*s*).

$$C_j^S = K_{\alpha eff} C_\alpha^{l0} \left[(1-g)^{K_{\alpha eff}} \right]^{-1}, \qquad (3.4.4)$$

where C_α^{l0} is the initial impurity concentration in the melt and g is the crystallized melt fraction. Equation (3.4.4) shows that the accumulation rate of impurities with $K_{\alpha eff}$ in the melt during crystallization and, hence, the steepness of the impurity concentration profile along the crystal increase with decreasing $K_{\alpha eff}$. For this reason, if the initial melt contains, for example, two kinds of background impurities, one of which produces shallow and the other deep levels, the situation in Figure 3.6 may arise. If the initial concentrations of both impurities in the melt are the same, their concentration ratio along the crystal length appears to be different due to the difference in K_α. In the absence of a special doping, this is responsible for the respective change of the Fermi level position and for the difference in the electrophysical parameters along the crystal, for example, an abrupt rise of ρ in the crystal tail.

When deep impurities are introduced into a semiconductor in order to give it the necessary recombinational and insulating properties, the doping

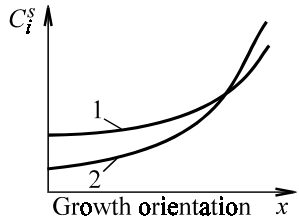

Figure 3.6. A possible distribution of shallow (1) and deep (2) impurities along the single crystal length.

impurity concentration often becomes very high, nearly reaching its solubility limit. An important role then is played by the solidus profile characterizing the temperature dependence of the limit solubility of a particular impurity in a particular semiconductor. The position and configuration of the solidus curve are determined by the physicochemical interaction of the system components and are implicitly present in equation (3.4.1). If we assume, for simplicity, the liquid solutions in the system of interest to be perfect, i.e., $\Omega_l = 0$, it will follow from (3.4.1) with $x_\alpha^s << 1$ that the temperature dependence of K_α^{l0} and the related quantity $x_\alpha{}^s$ are defined by the relation between Ω_s and $\Delta H_\alpha{}^f$. At $\Omega_s < \Delta H_\alpha{}^f$, both quantities increase with decreasing temperature, while at $\Omega_s > \Delta H_\alpha{}^f$ they decrease, giving rise to retrograde solubility. We then observe a certain regularity: the larger the value of Ω_s, the lower the impurity solubility and the greater the retrograde character of the solidus curve. Since deep impurities have large Ω_s values, their specificity reveals itself especially clearly. At least two important conclusions should be drawn from this behavior:

(1) High doping levels require technologies providing crystallization at temperatures with maximum $C_{\alpha max}$;

(2) It is desirable to use less retrograde impurities in order to reduce the probability of decomposition of semiconductor–impurity solid solutions during crystal cooling.

The first condition is important, for example, for the production of single crystals with semi-insulating properties when the concentration of doping impurity must exceed that of shallow background impurities. When low temperature techniques are used (e.g., crystallization from a solution-melt), the solubility of a deep impurity at technological temperatures may prove comparable in value with the concentration of shallow background impurities. A stable compensation of the latter then becomes a problem, and it is difficult to achieve stable properties of crystallized material.

Decomposition of oversaturated impurity solid solutions in a semiconductor, when the impurity changes from an electrically active state to a neutral state as the second phase products, is largely responsible for thermal instability of parameters of semiconductor materials. Decomposition-related problems are quite serious in view of the large number of deep impurities having a well-defined retrograde solubility at high doping levels. In a first approximation, the decomposition probability is determined by the degree of solid solution oversaturation and the seed growth kinetics. The problems of decomposition of oversaturated semiconductor solutions are discussed at length in [55]. Generally, the probability of solution decomposition during its cooling grows with the solidus curve steepness at high temperatures, although the decomposition process is affected by some additional factors, such as defect content in the crystal, the presence and type of contaminating impurities, etc.

The composition macro-inhomogeneities can largely be avoided by using special–purpose techniques, for example, by programming the crystal growth conditions by means of varying the value of $K_{\alpha\text{eff}}$ in accordance with (3.4.3) or by maintaining an approximately constant composition of the liquid phase during the whole crystallization process using various dopants. It is much more difficult to control micro-nonuniformities of impurity distribution in the crystal bulk. They arise from technological conditions, such as instability of equipment performance during crystal pulling from a melt, periodic temperature variations in the melt at the crystallization front because of imperfect thermoregulation, strong convective flows in the melt, distortions in the flat crystallization front, etc. Most of these reasons can be removed. But there are fundamental reasons associated with the crystallization process itself. It is known that crystal growth in directions with small crystallographic indices, which are usually used for the production of most semiconductor single crystals, requires a certain overcooling of the melt in the crystallization front vicinity. Therefore, crystallization represents, to some extent, a self-excited oscillation process. When accumulated overcooling reaches a critical value, it gives rise to a fast layer-by-layer crystallization of material followed by a slowing of the process, because the temperature at the crystallization front begins to increase due to the latent heat release in the phase transition during the crystallization. Then this heat is dissipated, the overcooling is accumulated, and the cycle is repeated. The alternation of these cycles also produces periodic temperature variations at the crystallization front. Variation in the crystallization rate gives rise to a periodic inhomogeneity, in accordance with (3.4.3).

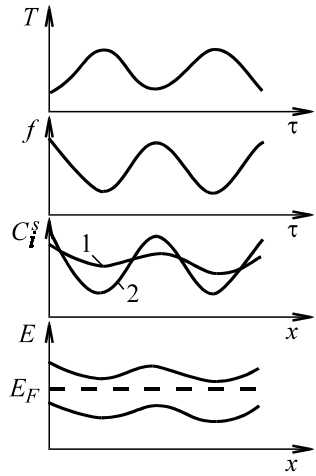

Figure 3.7. A hypothetical origin of periodic micro-nonuniformity in a crystal due to temperature variation at the crystallization front in the presence of shallow donor (1) and deep acceptor (2) impurities in the melt: f – crystallization rate; τ – time; E_F – Fermi level.

Figure 3.7 illustrates a situation which may arise, for example, in simultaneous crystal doping with two impurities, of which one is a deep level impurity. The concentration nonuniformity produces an electrical nonuniformity because of internal electric fields induced by spatial charge separation. It is worth noting that periodic nonuniformities manifest themselves more clearly in dislocation-free single crystals. Growth (high temperature) dislocations are fairly effective sinks for impurity atoms and can essentially smooth out such nonuniformities. Fortunately, periodic nonuniformities at high D_α^s and low cooling rates of the crystal are diffusionally smeared due to their small period.

Among general characteristics of deep impurity solubility in silicon is its correlation, at least for 3*d*-impurities, with the diffusion characteristics, the filling degree of the *d*-shell (outside the crystal), and the energy level (Table 3.5). However, we know little about the solubility of transition 4*d*- and 5*d*-metals in silicon.

The solubilities and energy levels are summarized in Tables 3.6 and 3.7.

Table 3.5. Diffusion and filling characteristics of deep 3*d*-impurity levels in silicon.

Impurity	Sc	Ti	V	Cr	Mn	Fe	Co	Ni
Electronic configuration	$3d^14s^2$	$3d^24s^2$	$3d^34s^1$	$3d^54s^2$	$3d^54s^2$	$3d^64s^2$	$3d^74s^2$	$3d^84s^2$
ΔE, eV	–	1.5	–	1.0	1.5	0.66	0.51	0.47
Type of level	*D*	*D*	*D*	*D*	*D*	*D*	*D*, *A*	*A*
D_0, cm/s	–	2.10	–	10	2.610	6.210	8.510	2.3×10^{-3}

D – donor, *A* – acceptor, $D = D_0 \exp(-\Delta E/kT)$.

Table 3.6. Solubility and energy levels of 4*d*-impurities in silicon.

Impurity	Solubility, cm^{-3}		Energy levels, eV	
	N_{max}	N_a	from *c*-band	from *v*-band
Y		8×10^{13}	0.29*D*; 0.4*D*	0.45*A*
Zr	10^{19}	2×10^{16}	0.17*D*	
Mo	10^{15}	10^{14}	0.33*D*	0.3*D*; 0.34*A*
Ru	3×10^{16}	5×10^{15}	0.24*A*; 0.45*D*	
Rh	10^{17}	5×10^{15}	0.3*A*; 0.55*A*	
Pd	6×10^{16}	5×10^{15}	0.2*A*; 0.28*D*	0.34*A*

D – a donor, *A* – an acceptor, $D = D_0\exp(-\Delta E/kT)$.

Table 3.7. Solubility and energy levels of 5*d*-impurities in silicon.

Impurity	Solubility, cm^{-3}		Energy levels, eV	
	N_{max}	N_a	from *c*-band	from *v*-band
Hf	4×10^{16}	4×10^{16}		0.09*D*
W	–	10^{14}	0.23*A*; 0.3*A*; 0.37*A*	0.34*D*; 0.3*D*
Re	10^{16}	2×10^{15}	0.18*A*; 0.3*A*; 0.53*A*	0.4*D*
Os	6×10^{16}	10^{15}	0.18*A*; 0.53*A*	0.18*D*
Ir	10^{17}	7×10^{15}	0.33*D*; 0.55*A*	
Pt	10^{17}	2×10^{16}	0.25*A*	0.3*A*; 0.3*D*
Au	10^{17}	5×10^{16}	0.54*A*	0.35*D*
Hg	–	10^{14}	0,36*A*; 0.31*A*	0.33*D*; 0.25*D*

It should be mentioned that most of these impurities are multiply charged and amphoteric. They will be discussed later in this chapter but we give their general characteristics here. Besides, one cannot be sure that the energy levels refer to single impurities rather than to their complexes or, in general, to defects.

We know little about the solubility of impurities with empty shells in germanium, except for the group-I atoms—Cu, Au, and Ag. These have found a wide application in fast response semiconductor devices as effective recombination centers in doped germanium and silicon. On the other hand, the application of other impurities with empty *d*-shells is quite limited. The same is true of *f*-atoms in both silicon and germanium. Their doping with *f*-atoms by the conventional methods of pulling from the melt, zone melting, and liquid phase epitaxy, is difficult because of the strong interaction between rare earth elements and oxygen in the liquid phase.

There has been no thermodynamic analysis aimed at determining ΔH_s for many deep level impurities.

3.4.2 Solubility in semiconductor compounds

Impurities with partly filled electron *d*-shells were shown in Section 2.2 to dissolve at the *A*-sublattice sites of $A^{III}B^{V}$ compounds. The solubility data available for some impurities in semiconductors are summarized in Table 3.8. They should be regarded only as preliminary data, because the investigators determined them in samples with different point defect backgrounds (with various degrees of deviation from the stoichiometry), with different dislocation structures, under different technological conditions leading, again, to very different sample stoichiometries, etc. The strong concentration dependence of a doping impurity on the background impurity and on the intrinsic defect content (vacancies in the *A*- and *B*-sublattices) results in a shift from the thermodynamic equilibrium in the semiconductor–impurity system. Besides, the associations of various point defects become in these conditions more important than in elemental semiconductors.

These factors make a rigorous and detailed analysis of solubility of deep impurities in binary and more complex semiconductor compounds quite a difficult task. The reader can find some fragments of such an analysis in the subsequent sections, in particular in Chapter 6.

Table 3.8. Solubility, dissolution enthalpy, and distribution coefficients of deep impurities in $A^{III}B^{V}$ semiconductors.

Impurity	$A^{III}B^{V}$ semiconductors						
	InAs	GaAs		InP	GaP		InSb
	N_{max}	N_{max}	K	N_{max}	N_{max}	K	K
Cu	0.04	0.7	2×10^{-3}	–	6×10^{-4}	–	6.6×10^{-4}
Ag	0.03	0.05	0.1	–	–	–	4.9×10^{-5}
Au	0.02	0.2	–	–	3×10^{-4}	–	1.9×10^{-6}
Cr	–	0.016	6×10^{-4}	–	0.001	–	–
Mn	–	0.01	0.05	0.3	–	0.015	–
Co	–	0.001	8×10^{-5}	–	–	–	–
Ni	–	0.001	0.02	–	–	–	6×10^{-5}
Ti	–	0.033	–	–	–	–	–
Fe	–	0.1	0.03	–	–	10^{-4}	0.04

3.5 SOLUBILITY OF AMPHOTERIC IMPURITIES

3.5.1 A thermodynamic analysis

In the previous section, we mentioned some amphoteric *d*-impurities, but a complete thermodynamic analysis of a semiconductor doped with an amphoteric impurity requires that its total concentration

$$[A] = N_0 + N^- + N^+$$

be related to the parameters of the ambient phase, from which a crystal is being grown, and to the other point defects present in the crystal. To begin with, consider an impurity-free elemental semiconductor consisting of S atoms and then a semiconductor doped with an amphoteric impurity A. The ambient phase is supposed to be the gas phase containing associated S_m-type molecular pairs. Partial gas pressure will be denoted as P_{Sm}. Suppose also that the crystal of interest is a perfect point defect solution. Since the type of solution is taken to be known, let us make a thermodynamic analysis by finding the chemical potentials of intrinsic and impurity point defects (see Section 1.2).

Since an amphoteric impurity can occupy sites and interstices both separately and simultaneously, we will introduce into our consideration a hypothetical defect—an "interstitial vacancy" denoted as V_I. Therefore, we should introduce the concept of an occupied interstice. In a doped crystal, these may be A_I and S_I-type defects; in an impurity-free crystal this is only the S_I defect. In other words, an elemental crystal is considered as a binary SI consisting of two sublattices—S and I.

It is more convenient to begin the analysis with a doped semiconductor. In this case, its composition will be described by the following quantities: S_S, I_I, V_S, and V_I. Vacancies may be charged; therefore,

$$V_S = V_S^0 + V_S^- \quad \text{and} \quad V_I = V_I^0 + V_I^+ . \tag{3.5.1}$$

Here, as usual, the true site vacancies are assumed to have an acceptor character, and hypothetical interstitial vacancies are ascribed a donor character. This is an arbitrary expedient, the more so that the assumption of intrinsic point defects having a charge is of no principal importance, because their true nature must be established experimentally in every particular case. Moreover, if the concentration of charged defects is taken in the first approximation to be much lower than that of un-ionized defects, it will be clear from (3.5.1) that the whole analysis of intrinsic defects can be made, in the first approximation, without indicating their charge. If necessary, the charge of intrinsic defects can be easily introduced with expressions (2.3.6) and (2.3.10).

Therefore, the thermodynamic analysis must yield the equilibrium values of the mole fractions

$$x_S = V_S / S_S , \qquad x_I = V_I / I_I , \qquad \gamma = I_I / S_S . \tag{3.5.2}$$

Of the three mole fractions, it is only the quantity x_S which has a physical meaning for an impurity-free elemental crystal. Indeed, all interstices in such a crystal are free, or

$$V_I = I_I \quad \text{and} \quad x_I = 1. \tag{3.5.3}$$

For such a semiconductor as germanium and silicon, we have

$$I_I = S_S = N_L, \tag{3.5.4}$$

where N_L is the concentration of crystal atoms, which leads to $\gamma = 1$. Nevertheless, it will become clear from further analysis that all the three mole

fractions in a semiconductor compound (or in case of dissociative amphoteric impurities) have a real physical sense. So they remain in (3.5.2) for a further analysis.

The sought for concentration V_S is defined by the equilibrium conditions between the point defects and the external gas phase, as well as the other point defects V_I in the crystal bulk:

$$mV_S + S_m \overrightarrow{\leftarrow} mS_S\,, \tag{3.5.5}$$

$$V_S + V_I \overrightarrow{\leftarrow} 0\,. \tag{3.5.6}$$

These two quasichemical reactions correspond to equations relating partial chemical potentials

$$m\mu\left(V_S^0\right) + \mu\left(S_m\right) = m\mu\left(S_S\right)\,, \tag{3.5.7}$$

$$\mu\left(V_I^0\right) + \mu\left(V_S^0\right) = 0\,. \tag{3.5.8}$$

In these expressions and below, the brackets denote the components with chemical potentials. Using the general expression for the chemical potential of any point defect as a perfect solution component, we can write

$$\mu\left(V_I^0\right) = g\left(V_I^0\right) + T\ln\frac{\left[V_I^0\right]}{\left[I_I\right] + \left[V_I\right]}\,, \tag{3.5.9}$$

$$\mu\left(V_S^0\right) = g\left(V_S^0\right) + T\ln\frac{\left[V_S^0\right]}{\left[S_S\right] + \left[V_S\right]}\,, \tag{3.5.10}$$

where the square brackets stand for the defect concentration. The gas phase is considered as an ideal gas, so that the chemical potential has the form

$$\mu\left(S_m\right) = \Psi_S\left(T\right) + T\ln P_{Sm}\,, \tag{3.5.11}$$

where $\Psi_S(T)$ is a standard potential of the gas phase and T is temperature expressed in energy units.

Expression (3.5.7) also includes the chemical potential $\mu(S_S)$. It is easy to see that the equality $\mu(S_S) = g(S_S)$ is valid, which means that $\mu(S_S)$ represents, physically, the crystal formation energy per atom and is a characteristic of a perfect crystal. It is important for us that this quantity is a function of temperature only.

Substituting the chemical potentials of (2.3.20) and (2.3.21) into (2.3.19), we find

$$\frac{\left[V_I^0\right]\left[V_S^0\right]}{\left(\left[V_I\right]+\left[I_I\right]\right)\left(\left[V_S\right]+\left[S_S\right]\right)} = \exp\left[-\frac{g\left(V_I^0\right)+g\left(V_S^0\right)}{T}\right] = K'_{1S}(T). \qquad (3.5.12)$$

From (3.5.7), (3.5.9), and (3.5.11), we get

$$\frac{\left[V_S^0\right]}{\left[V_S\right]+\left[S_S\right]} P_{Sm}^{1/m} = \exp\left[-\frac{g\left(V_S^0\right)+\Psi_S/m-g\left(S_S\right)}{T}\right] = K'_{2S}. \qquad (3.5.13)$$

There are three unknown quantities in the last two equations: V_S, V_I, and I_I. In order to find them, we should add the condition of site equality in the hypothetical SI crystal:

$$\left[I_I\right]+\left[V_I\right]=\left[S_S\right]+\left[V_S\right]. \qquad (3.5.14)$$

The solution to the set of equations (3.5.12) through (3.5.14) yields general mole fractions (3.5.14). In other words, we can completely define the composition of an impurity-free crystal containing intrinsic atoms and intrinsic defects as constituents. If we remember that the number of site vacancies in an elemental crystal is always smaller than that of occupied sites, $V_S << S_S$, and that all interstices are vacant, or, the other way round, that the number of interstitial vacancies is always larger than that of occupied interstices, $V_I << I_I$, then equation (3.5.14) will transform to

$$\left[V_I\right]=\left[S_S\right]=N_L, \qquad (3.5.15)$$

coinciding with (3.5.4). Consequently, equations (3.5.12) and (3.5.13) will be simplified and together with (3.5.4) will appear as

$$\left[V_S^0\right] = N_L \exp\left[-\frac{g\left(V_I^0\right) + g\left(V_S^0\right)}{T}\right] = N_L K'_{1S}(T), \qquad (3.5.16)$$

$$\left[V_S^0\right] P_{Sm}^{1/m} = N_L \exp\left[-\frac{g\left(V_S^0\right) + \Psi_S/m - g(S_S)}{T}\right] = N_L K_{2S}(T). \qquad (3.5.17)$$

The first of these expressions reflects the equilibrium of intrinsic defects inside the crystal and the second describes the equilibrium between the solid and the ambient gas phases. The concentration of the only type of defect V_S in this crystal, which can be found from (3.5.17), is

$$\left[V_S^0\right] = K_{2S}(T) P_{Sm}^{-1/m}. \qquad (3.5.18)$$

One should note this equality

$$K_{1S}(T) \big/ K_{2S}(T) = P_{Sm}^{-1/m}, \qquad (3.5.19)$$

which is useful because it is sufficient to find one equilibrium constant, K_1 or K_2, experimentally, while the other one can be calculated from (3.5.19), using the known function $P_{Sm}(T)$.

Consider now the same crystal doped with an amphoteric impurity, whose vapor pressure at the doping temperature is denoted as P_{Ak}. In other words, by analogy with S_m, the impurity vapor will be considered as consisting of A_k molecules. In this case, the crystal composition is defined by the following mole fractions, instead of (3.5.2):

$$\left.\begin{aligned} x &= \frac{V_S}{S_S + A_S}, & x_I &= \frac{V_I}{I_I + A_I}, \\ x_{AS} &= \frac{A_S}{S_S + A_S}, & x_{AI} &= \frac{A_I}{I_I + A_I}. \end{aligned}\right\} \qquad (3.5.20)$$

Here, we will assume that the amphoteric impurity has a dissociative nature, namely, A atoms can occupy both sites and interstices in the SI crystal. The equilibrium between the crystal and the gas (vapor) phase will be

characterized by two independent quasichemical reactions (the vapor is an ideal gas)

$$mV_S^0 + S_m \underset{\leftarrow}{\overset{\rightarrow}{}} mS_S, \tag{3.5.21}$$

$$kV_S^0 + A_k \underset{\leftarrow}{\overset{\rightarrow}{}} kA_S^0 \tag{3.5.22}$$

instead of one reaction (3.5.5).

The equilibrium between point defects in the crystal bulk will be expressed, with account taken of the doping impurity, by two reactions

$$V_S^0 + V_I^0 \underset{\leftarrow}{\overset{\rightarrow}{}} 0, \tag{3.5.23}$$

$$A_S^0 + V_I^0 \underset{\leftarrow}{\overset{\rightarrow}{}} A_I^0 + V_S^0. \tag{3.5.24}$$

The equilibrium quasichemical reactions correspond to the relations between the chemical potentials

$$m\mu\left(V_S^0\right) + \mu\left(S_m\right) = m\mu\left(S_S\right), \tag{3.5.25}$$

$$k\mu\left(V_S^0\right) + \mu\left(A_k\right) = k\mu\left(A_S^0\right), \tag{3.5.26}$$

$$\mu\left(V_S^0\right) + \mu\left(V_I^0\right) = 0, \tag{3.5.27}$$

$$\mu\left(A_S^0\right) + \mu\left(V_I^0\right) - \mu\left(A_I^0\right) - \mu\left(V_S^0\right) = 0. \tag{3.5.28}$$

The expressions for μ of the solid state components are taken in the same form as in (3.5.9) and (3.5.10) and for the gas phase as (3.5.11). Then, the solution to the set of equations (3.5.25) through (3.5.28) will provide all unknown concentrations V_S^0, V_I^0, A_S^0, and A_I^0, together with the total concentration of the amphoteric impurity $A^0 = A_I^0 + A_S^0$. The results of this evaluation are given Table 3.9, taking into account the inequalities

$$V_I + A_I << N_L \quad \text{and} \quad V_S + A_S << N_L. \tag{3.5.29}$$

Table 3.9. Intrinsic and impurity point defect concentrations in elemental semiconductors doped with an amphoteric impurity.

Point defect	Concentration vs. temperature and gas pressure
V_S^0	$N_L K_{2S} P_{Sm}^{-1/m}$
V_I^0	$N_L = \dfrac{K_{2S}}{K_{1S}} P_{Sm}^{-1/m} \approx N_L$
V_S^-	$N_L K_{2S} K_{3V} P_{Sm}^{-1/m} \left(n / n_i^2\right)$
A_S^0	$N_L \dfrac{K_{2S}}{K_{2A}} P_{Sm}^{-1/m} P_{Ak}^{1/k} = \dfrac{N_L K_{1S}}{K_{2A}} P_{Ak}^{1/k}$
A_I^0	$\dfrac{N_L K_{2S}}{K_{1S} K_{2A}} P_{Ak}^{1/k} = \dfrac{N_L}{K_{2A}} P_{Sm}^{1/m} P_{Ak}^{1/k}$
A_S^0 / A_I^0	$K_{2S} P_{Sm}^{-2/m} = K_{1S} P_{Sm}^{-1/m}$
A_S^- / A_I^0	$K_{3A} K_{2S} P_{Sm}^{-2/m} \left(n / n_i\right)^2 = K_A \left(n / n_i\right)^2$

The values of temperature functions K in Table 3.9 have the form

$$K_{1S} = \exp\left[-\frac{g\left(V_S^0\right) + g\left(V_I^0\right)}{T}\right], \tag{3.5.30}$$

$$K_{2S} = \exp\left[-\frac{g\left(V_S^0\right) + \Psi_S / m - g\left(S_S\right)}{T}\right], \tag{3.5.31}$$

$$K_{2A} = \exp\left[-\frac{g\left(V_S^0\right) + \Psi_A / k - g\left(A_S^0\right)}{T}\right]. \tag{3.5.32}$$

If an amphoteric impurity dissolves mostly at lattice sites, we have $A_S/A_I >> 1$, but if it is mostly dissolved at interstices, then $A_S/A_I << 1$.

So far, we have discussed un-ionized impurities and intrinsic point defects. But now, let us take their ionization into account. Site vacancies and

amphoteric site atoms will be considered to be acceptors and interstitial amphoteric impurities as donors. The content of interstitial intrinsic atoms will be neglected, as before. The conditions listed above are usually fulfilled in real germanium and silicon crystals.

For simplicity, the amphoteric impurity will be assumed to be singly charged. Then, A_S^- and A_I^+ will be described by formulas (2.3.3) and (2.3.7) which can be used to write the ratio A^-/A_I^+ as

$$\frac{A_S^-}{A_I^+} = \frac{A_S^0}{A_I^0}\frac{g_{dA}}{g_{aA}}\frac{N_v n^2}{N_c n_i^2}\exp\left[\frac{E_{dA} - E_{aA}}{kT}\right], \tag{3.5.33}$$

where E_{dA} and E_{aA} are the energy levels of the amphoteric impurity counted from "their" bands: E_{dA} from the conduction band and E_{aA} from the valence band; g are respective factors of the level degeneracy.

We can also write the concentration of ionized vacancies as

$$V_S^- = \frac{g_{0V}}{g_{aV}}\frac{N_v n}{n_i^2}V_S^0 \exp\left(-\frac{E_{aV}}{kT}\right), \tag{3.5.34}$$

where E_{aV} is the vacancy energy level counted from the valence band top.

By substituting into (3.5.34) the expression for V_S^0 from Table 3.9, we obtain V_S^- in the form shown in the same table. The determining temperature dependent function K_3 is

$$K_{3V} = \frac{g_{0V}}{g_{aV}} N_v \exp\left(-\frac{E_{aV}}{kT}\right). \tag{3.5.35}$$

Let us now turn to equation (3.5.33) and substitute into it the A_S^0/A_I^0 ratio taken from Table 2.13:

$$\frac{A_S^-}{A_I^+} = \frac{g_{dA}}{g_{aA}}\frac{N_v}{N_c}\frac{n^2}{n_i^2}\exp\left(\frac{E_{dA} - E_{aA}}{kT}\right)K_{2S}P_{Sm}^{-2/m} \tag{3.5.36}$$

or

$$\frac{A_S^-}{A_I^+} = K_{3A}K_{2S}P_{Sm}^{-2/m}\left(\frac{n}{n_i}\right)^2 = K_A\left(\frac{n}{n_i}\right)^2, \tag{3.5.37}$$

where

$$K_{3A} = \frac{g_{dA}}{g_{aA}} \frac{N_v}{N_c} \exp\left(\frac{E_{dA} - E_{aA}}{kT} \right) \tag{3.5.38}$$

$$K_A = K_{3A} K_{2S} P_{Sm}^{-2/m} . \tag{3.5.39}$$

For the high temperature region with all centers ionized, we have

$$A_S^- \cong A_S , \qquad A_I^+ \approx A_I . \tag{3.5.40}$$

In this case, the ratio is

$$A_S / A_I = K_A (n/n_i)^2 . \tag{3.5.41}$$

If we assume that the impurity atoms do not have enough time to be redistributed over the crystallochemical positions during the crystal cooling from the doping temperature, the ratio of (3.5.41) will show the amphoteric impurity distribution at sites and interstices in an elemental semiconductor. This ratio, together with the expression for A_S^0 from Table 3.9 makes it possible to evaluate the total concentration of an amphoteric impurity $A = A_S + A_I$, which dissolves in an elemental semiconductor:

$$A = N_L \frac{K_{2S}}{K_{2A}} P_{Sm}^{-1/m} P_{Ak}^{1/k} \left[1 + f + K_A^{-1} \left(\frac{n_i}{n} \right)^2 (1 + \varphi) \right], \tag{3.5.42}$$

where the functions f and φ have the form

$$f = \frac{g_{0A}}{g_{aA}} N_v \frac{n}{n_i^2} \exp\left(-\frac{E_{aA}}{kT} \right), \tag{3.5.43}$$

$$\varphi = \frac{g_{0A}}{g_{dA}} \frac{N_c}{n} \exp\left(-\frac{E_{dA}}{kT} \right). \tag{3.5.44}$$

The solubility of the site component of an amphoteric impurity in an elemental semiconductor is

$$A_S = N_L \frac{K_{2S}}{K_{2A}} (1+f) P_{Sm}^{-1/m} P_{Ak}^{1/k} \tag{3.5.45}$$

and the solubility of the interstitial component of an amphoteric impurity is

$$A_I = N_L \frac{K_{2S}}{K_{2A} K_A} P_{Sm}^{-1/m} P_{Ak}^{1/k} (1+\varphi) \left(\frac{n}{n_i}\right)^2 . \tag{3.5.46}$$

3.5.2 Solubility of dissociative amphoteric impurities

The thermodynamic analysis of the solubility of dissociative amphoteric (multiply charged) impurities will be carried out with reference to their typical representatives—gold and silver in silicon.

The maximum solubilities of these and other dissociative amphoteric impurities in silicon are presented in Table 3.10.

The solidus lines for solid solutions of silicon doped with these impurities were found using the radioactive isotope technique. A comparison with Hall effect measurements shows that all doping atoms are electrically active, so the formulas for nondegenerate semiconductors are applicable in this case.

Since these impurities can have any of the three charged states—neutral, donor, and acceptor state—the formula for a multiply charged impurity is [4]

$$N_P = \frac{a_P}{r_{P0} a_{Si}} N^{Si} \exp\left(-\frac{g_P^{Si}}{kT}\right)\left[1 + r_a \frac{N_v}{P} \exp\left(-\frac{E_a}{kT}\right) + r_d \frac{N_c}{n} \exp\left(-\frac{E_d}{kT}\right)\right], \tag{3.5.47}$$

where $r_a(r_d)$ is the degeneracy factor of the acceptor (donor) state of a multiply charged impurity and $E_a(E_d)$ is the ionization energy of the acceptor

Table 3.10. Maximum total solubilities and distribution coefficients of amphoteric impurities in silicon.

Impurity	N_{max} cm^{-3}	T °C	Distribution coefficient
Ag	2×10^{17}	1350	–
Au	1×10^{17}	1250	3×10^{-5}
Cu	1×10^{18}	1200	4×10^{-3}
Ni	7×10^{17}	1310	–

(donor) state of the same impurity. The concentrations of free electrons and holes are related by the known expression $np = n_i^2$. The analysis of this formula shows that most impurity atoms will be neutral if a pure semiconductor is doped with gold or silver. An additional doping with donors transforms impurity gold to the acceptor state [4].

The conclusion about the redistribution of an amphoteric impurity over crystallochemical positions due to the introduction of other point defects into the semiconductor fully agrees with the analysis made in Section 3.5.1. This conclusion is valid for all associative amphoteric impurities.

Consider now available solubility data for some dissociative impurities in various semiconductors.

Copper impurity in germanium. The state of copper atoms in germanium was discussed in Section 2.3.7 (see Table 2.18). It should be added here that many investigations [58] have shown that copper atoms are dissolved at Cu_s sites and Cu_i interstices in comparable quantities. This means that the Ge<Cu> solid solution is a substitution solution and an incorporation solution simultaneously, with copper concentrations varying with temperature in both cases. Figure 3.8 presents data on C_e measured by electrophysical methods and on total copper concentration C found by the isotope technique. From this, we have

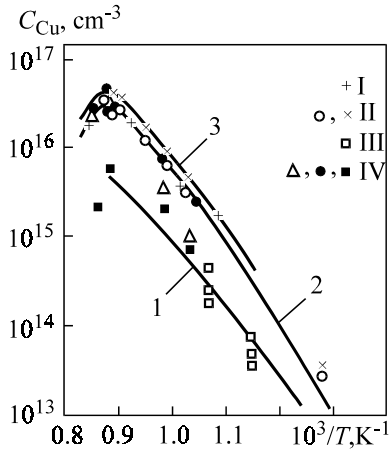

Figure 3.8. Copper solubility in germanium according to the data: I – [59], II – [60], III – [61], IV – [62]; 1 – [60] for C_i, 2 – [63] for C_e, 3 – [64, 65] for C.

$$C_i = \frac{1}{2}(C - C_e), \qquad C_s = C_e + \frac{1}{2}(C - C_e). \tag{3.5.48}$$

In our view, the data of the work [61] are more reliable. The values of ΔH_s and ΔH_i for copper dissolution at sites and interstices are 1.47 and 2.0 eV, respectively.

Copper impurity in silicon. Copper solubility was discussed in detail in [61], and the results of this work are still valid. The solubility was studied in undoped and doped silicon of both types of conductivity separately. For this, the radioactive copper isotope ^{64}Cu was diffused into crystal samples grown by pulling from the melt and by zone melting.

Curve 1 in Figure 3.9 illustrates the results obtained. The copper solubility in *p*-samples doped with boron strongly depended on the boron concentration. The experimental points fit well the calculated dependences

$$N^+ = N_i^+ \exp\left(\frac{\mu_i - \mu}{kT}\right), \tag{3.5.49}$$

$$N^- = N_i^- \exp\left(\frac{\mu_i - \mu}{kT}\right), \tag{3.5.50}$$

when the data of curve 1 in Figure 3.9 are substituted as N_i and the concen-

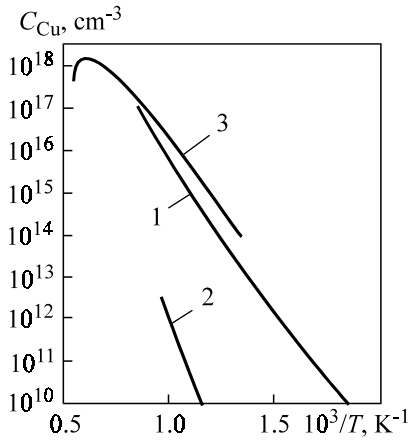

Figure 3.9. Copper solubility in undoped silicon: 1, 3 – interstitial solubility; 2 – site solubility [61].

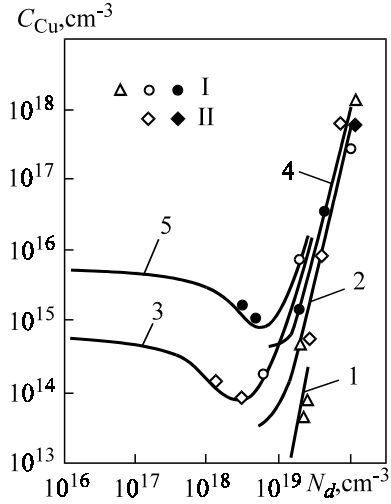

Figure 3.10. Copper solubility in *n*-silicon: I – doped with As; II – doped with P. Temperatures of diffusion doping with Cu, °C: 1 – 500; 2, 3 – 600; 4, 5 – 700.

tration narrowing of the forbidden gap is taken into account. In formulas (3.5.49) and (3.5.50), μ is the Fermi level position in an intrinsic semiconductor.The agreement between the experimental data and equation (3.5.49) indicates unambiguously that curve 1 in Figure 3.9 reflects interstitial solubility, with Cu_i as a single positively charged center.

The results on copper solubility in *n*-silicon doped with phosphorus or antimony [61] are illustrated in Figure 3.10. As the shallow donor concentration increases, the concentration of Cu_i^+ having the same sign decreases, which is indeed observed on the left of the curve maximum. The rapid growth of solubility on the right of the maximum is evidence for the dominance of oppositely charged copper centers Cu^- in this region. These can be only the site centers Cu_s^-. Since C_{Cu_s} rises approximately in proportion to N_d^3, this is more evidence that the copper site in silicon is a triple acceptor [66]. Since the minima in Figure 3.10 are observed at very high concentrations of shallow donors (> 10^{18} cm^{-3}), we have $C_{Cu_s} << C_{Cu_i}$ at the same dissolution temperatures. Taking equations (3.5.49) and (3.5.50) to be valid and relating the first one to interstitial and the second one to site copper atoms, we can describe the total concentration of dissolved copper as

$$C = C_i^i\left(n_i/n\right) + C_i^s\left(n/n_i\right)^3, \qquad (3.5.51)$$

where superscripts *i* and *s* denote interstitial and site solubilities and the subscript indicates intrinsic undoped silicon; the cubic power of the second term reflects triply charged site copper.

Using expression (3.5.51), the authors of [61] found the dependence $C_i^s(T)$ shown by curve 2 in Figure 3.9.

There is eutectic temperature T_{eu} in the diagrams of the Si–Me state; therefore, the copper solubility in the temperature ranges $T < T_{eu}$ and $T > T_{eu}$ should be considered separately. For the Si-Cu system, the eutectic temperature is T_{eu} = 1075 K. Solubility curves 1 and 2 in Figure 3.9 are valid up to this temperature. At higher temperatures, the formation of copper silicide Cu_3Si is possible.

The total copper solubility in silicon from radioactive analysis [61, 67] is shown by curve 3 in Figure 3.9. The dissolution enthalpy in the region up to the eutectic temperature is 1.49 eV.

Gold impurity in silicon. In addition to the information on gold in silicon given in Section 2.3.7 [68], it is necessary to mention that the low distribution coefficients (Table 3.11) do not permit the production of samples heavily doped with gold by pulling from the melt or by zone melting. To do this, large quantities of gold should be introduced into the liquid phase, resulting in strongly nonuniform samples liable to polycrystallization. Besides, impurities concomitant with gold and having higher distribution coefficients will create a heavy foreign impurity background in the samples. For these reasons, conventional growth techniques can produce large quality samples with a low background and a fairly uniform bulk distribution of gold only up to the concentrations $N_{Au} \le 1.5\times10^{15}$ cm^{-3}. Samples with a higher gold content can be produced only by diffusion.

Curve 1 in Figure 3.11 represents data obtained by several authors using diffusion saturation. The high temperature portion of the curve was measured especially carefully in the range of 1000–1380°C [68]. Since the initial samples contained a large number of dislocations, the sink density for Au_i was high and the interstitial gold atoms reached the sink fast. So the data of [68] seem to refer to the site Au_s solubility. This suggestion was made earlier by the authors of [69] who used dislocation-free samples in diffusion experiments and obtained different $C_{Au}(x)$ profiles. They treated them as an interstitial impurity distribution and found curve 2 in Figure 3.11 reflecting, in their view, the interstitial gold solubility in silicon. But at the time those experiments [69] were performed, almost nothing was known about so called swirl defects that may occur in dislocation-free samples. For this reason, the absence of sinks for interstitial gold suggested in [68] seems doubtful. Besides, it is not quite clear how large is the capacity of swirl defects as Au_i sinks.

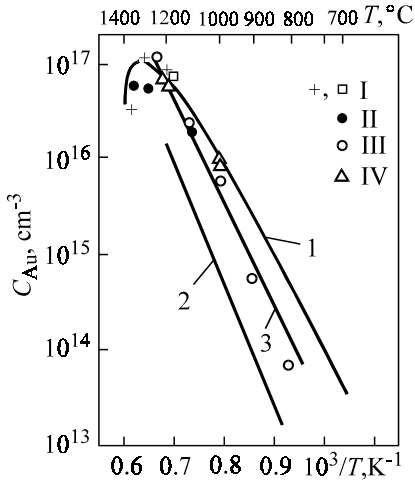

Figure 3.11. Gold solubility at silicon sites (1, 3) and interstices (2), according to the data: I – [68]; II – [70]; III – [71]; IV – [72].

Figure 3.11 also shows the results obtained in [70], which are somewhat lower than those of [68] and are not believed to be quite correct.

Therefore, curve 1 in Figure 3.11 representing site gold solubility in pure silicon can be considered as sufficiently reliable. Curve 2 showing interstitial solubility should, strictly speaking, be checked again in samples with a better controlled density of inner sinks in their up-to-date sense.

Silver impurity in silicon. The solubility of silver reported in [78] is retrograde solubility with a maximum at 1350°C. The solubility below this temperature is satisfactorily described by the exponent law with ΔH = 2.6 eV.

Amphoteric 4d- and 5d-impurities in silicon. Investigations of the diffusion properties and decomposition of 4*d*-impurity atoms of Ru, Rh, and Pd [79, 80] have shown that all of them are amphoteric in silicon. Their basic characteristics are presented in Table 2.20 which also contains data on 5*d*-amphoteric impurities in silicon, namely, W, Re, Os, Ir, Pt, Au, and Ag [81–83]. Note that none of the 4*d*- and 5*d*-impurities have been studied thoroughly. The reliability of the energy level data for some of them is doubtful. Much remains unknown about their nature. Are they single impurity atoms in certain crystal positions or the product of interactions with crystal lattice defects and other impurity atoms?

3.5.3 Solubility of cation–anion impurities in semiconductor compounds

The experimental information on the solubilities of group-IV impurities in $A^{III}B^{V}$ compounds is too scanty to describe its qualitative characteristic—the dissolution enthalpy. Yet, we know that the solubility is fairly high and has a retrograde character. The other parameters of these impurities in $A^{III}B^{V}$ compounds are given in Table 3.11.

The solubility mechanism for some impurities (Ge, Si, and GaAs) differs from that of common doping impurities, for example, Zn, Te, and Se. The difference is due to the fact that the liquid phase equilibrium is established with the crystal surface rather than with its bulk. This follows from the linear dependence of $C_s(C_l)$ described by equation (1.1.23) but not by the square root of (1.1.18). Therefore, one can draw the conclusion about a slow diffusion sinking of group-IV impurities into the crystal bulk (see Section 1.1).

Table 3.11 shows the distribution coefficients, maximum solubilities, and respective temperatures. These data have a considerable dispersion, especially in the distribution coefficients. In addition to occasional reasons, these discrepancies are due to deviations from the stoichiometry, as was predicted theoretically. This can be easily followed by comparing the values of K for

Table 3.11. Some physicochemical parameters of basic $A^{III}B^{V}$ compounds doped with group-IV elements.

Impurity	Distribution coeff. K			Solubility C_m, cm^{-3}			T_m, K		
	GaAs	GaP	InP	GaAs	GaP	InP	GaAs	GaP	InP
C	0.8^{*} 0.2^{*}	10^{-4**}	–	–	–	–	–	–	–
Si	$0.1–0.14^{*}$ 0.062^{**}	$0.1–0.2^{**}$ $0.3–0.5^{**}$	0.55^{*} 30^{**}	1.5×20^{20}	7×10^{20}	3×10^{19}	1353	1323	1335
Ge	0.1– 0.03^{**} $8.3\times10^{-3*}$	$0.1–0.12^{*}$ $0.1–0.1^{**}$	$0.2–0.4^{*}$ $6\times10^{-3**}$	4×10^{20}	2×10^{20}	7×10^{19}	1173	1323	1173
Sn	5×10^{-3}– $8\times10^{-2*}$ $4\times10^{-4**}$	8×10^{-3} 3×10^{-2} $4\times10^{-4**}$	$2.5\times10^{-3*}$	8×10^{19}	2×10^{19}	(2–5)× 10^{18}	1273	1323	1335
Pb	$<0.02^{**}$ $<1\times10^{-5*}$	–	–	–	–	–	–	–	–

* – growth from a stoichiometric melt; ** – growth from a nonstoichiometric solution melt; C_m – maximum solubility, cm^{-3}; T_m – temperature in K, corresponding to C_m.

samples grown from stoichiometric and nonstoichiometric solutions. Therefore, there is a strong influence of the volatile component vapor pressure.

The considerable spread of K values even for the same growth technique may be due to several reasons. A significant role is played by the closeness to the equilibrium conditions, the identification method used, the account or neglect of some group-IV impurities which interact with other point defects, and, finally, the doping level because it also affects the degree of the compound stoichiometry.

Cation–anion impurities in CdSb. Among the few reports available on the solubility of group-IV impurities in CdSb, one should note the work [84] on tin segregation, solubility and diffusion on ingots grown by the Chohkralsky method, and oriented crystallization.

An analysis has shown that the temperature variation of solubility in CdSb has a retrograde character. The maximum solubility at a minimum cadmium gas pressure is achieved at (360 ± 20)°C and is ~1.6×10^{19} cm^{-3}. The maximum solubility increases with the gas pressure.

REFERENCES

3.1. J.L. Mejering, *Physica* **103**: 123–130 (1981).
3.2. R.A. Swalin, *Thermodynamics of Solids*. New York–London: Wiley, 313 p. (1961).
3.3. C.D. Thurmond, J.S Struthers, *J. Electrochem. Soc*. **57**: 831–837 (1953).
3.4. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushchikh defektov v poluprovodnikakh* (Thermodynamics and Kinetics of Interacting Defects in Semiconductors). Moscow: Nauka, 352 p. (1997) (in Russian).
3.5. R.N. Hall, *J. Phys. Chem. Sol*. **3**, No. 1–2: 63–73 (1957).
3.6. F.A. Trumbor, *Bell. System. Techn. J*. **39**, No. 1: 205–223 (1960).
3.7. C.D. Thurmond, M. Kowalchik, *Bell. Syst. Techn. J*. **39**, No. 1: 169–204 (1960).
3.8. V.M. Glazov, L.M. Pavlova, *Doklady AN SSSR* **232**, No. 3: 607–610 (1977) (in Russian).
3.9. V.M. Glazov, L.M. Pavlova, *JFKh* **52**, No. 4: 854–857 (1978) (in Russian).
3.10. V.M. Glazov, V.S. Zemskov, *Fiziko-khimicheskie osnovy legirovaniya poluprovodnikov* (Physical and Chemical Principles of Semiconductor Doping). Moscow: Nauka, 378 p. (1967) (in Russian).
3.11. V.M. Glazov, L.M. Pavlova, *Khimicheskaya termodinamika i fazovye ravnovesiya* (Chemical Thermodynamics and Phase Equilibrium). Moscow: Metallurgia, 559 (1988) (in Russian).
3.12. H. Reiss, *J. Chem. Phys*. **21**, No. 7: 1209–1217 (1953).
3.13. H. Reiss, C.S. Fuller, F.J. Morin, *Bell. Syst. Techn. J*. **35**: 535 (1956).
3.14. F.A. Trumbore, E.M. Porbansky, A.A. Tartaglia, *J. Phys. Chem. Sol*. **11**, No. 3–4: 239–245 (1959).
3.15. J.S. Blakemore, *Proc. Phys. Soc*. **71**, pt 4, No.460: 692–694 (1958).

3.16. V.I. Fistul, *Heavily Doped Semiconductors*. New York: Plenum Press, 418 p. (1969).
3.17. K. Lehovec, *J. Phys. Chem. Sol*. **23**, No. 6: 695–709 (1962).
3.18. V.M. Kozlovslaya, R.N. Rubinshtein, *FTT* **3**, iss. 11: 3354–3362 (1961) (in Russian).
3.19. G.M. Kuznetsov, S.K. Kuznetsova, *Izvestiya AN SSSR, ser. Neorg. mat*. **2**, No. 4: 643–649 (1966) (in Russian).
3.20. H.E. Bridgers, E.D. Kolb, *J. Chem. Phys*. **25**, No.4: 648–650 (1966).
3.21. W.D. Edwards, *J. Electrochem. Soc*. **117**, No. 8: 1062–1064 (1970).
3.22. K. Weisser, *J. Phys. Chem. Sol*. **7**, No. 2: 118–126 (1958).
3.23. V.S. Zemskov, A.D. Suchkova, B.G. Zhurkin *et al*., *Izvestiya AN SSSR, ser. Metall. i topl*., No. 2: 131–134 (1962) (in Russian).
3.24. M.A. Lee, *Sol. St. Electron*. **1**, No. 3: 194–201 (1960).
3.25. F.E. Roberts, *Sol. St. Electron*. **2**, No. 1: 8–13 (1961).
3.26. N.Kh. Abrikosov, V.M. Glazov, Lu Chjen Yuan, *JNKh* **7**, iss. 4: 831–835 (1962) (in Russian).
3.27. F.A. Trumbore, W.G. Spitzer, R.A. Logan *et al*., *J. Electrochem. Soc*. **109**, No. 8: 734–738 (1962).
3.28. C.D. Turmond, F.A. Trumbore, M. Kowalchik, *J. Phys. Chem*. **25**, No. 4: 799–800 (1956).
3.29. B.G. Zhurkin, V.S. Zemskov, D.A. Petrov *et al*., *Izvestiya AN SSSR, ser. Metall. i topl*., No. 6: 86–90 (1959) (in Russian).
3.30. E. Dominigues, M. Jaraiz, *J. Electrochem. Soc*. **133**, No. 9: 1895–1900 (1986).
3.31. O. Madelung, ed., *Numerical Data and Functional Relationships in Science and Technology, New Series* **17**, group III. Berlin–New York: Springer–Verlag (1982).
3.32. D. Navon, V. Chernyshov, *J. Appl. Phys*. **28**, No. 7: 823–824 (1957).
3.33. R.A. Gudmundsen, J. Maserjian, *J. Appl. Phys*. **28**, No. 11: 1308–1316 (1957).
3.34. W. Scott, R.J. Hager, *J. Electr. Mater*. **8**, No. 5: 581–602 (1979).
3.35. I.M. Mackintosh, *J. Electrochem. Soc*. **105**, No. 5:392–401 (1962).
3.36. E. Tannenbaum, *Sol. St. Electron*. **2**, No. 3: 123–132 (1961).
3.37. M. Finetti, P. Negrini, S. Sollmi *et al*., *J. Electrochem. Soc*. **128**, No. 6: 1313–1317 (1981).
3.38. A. Desalvo, G. Ruani, *Phil. Mag., B* **50**, No. 4: 505–516 (1984).
3.39. V.A. Panteleev, N.E. Rudoy, T.S. Gugina, *Izvestiya AN SSSR, ser. Neorg. mat*. **12**, No 10: 1856–1857 (1976) (in Russian).
3.40. R. Angelicci, A. Armigliato, E. Landi *et al*., *Sol. St. Dev., Proceedings 17th ESSDERC Conference*. Bolonga: 405–408 (1988).
3.41. J.J. Rohan, N.E. Pickering, J. Kennedy, *J. Electrochem. Soc*. **106**, No. 8: 705–709 (1959).
3.42. D. Nobili, R. Angelucci, A. Armigliato *et al*., *J. Electrochem. Soc*. **136**, No. 4: 1142–1146 (1989).
3.43. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushchikh primesei v poluprovodnikakh* (Thermodynamics and Kinetics of Interacting Impurities in Semiconductors). Moscow: Nauka, 352 p. (1997).
3.44. S.V. Bulyursky, A.V. Komlev, *Izvestiya RAN, ser. Neorg. mat*. **33**: 134–138 (1997) (in Russian).

3.45. V.M. Glazov, L.M. Pavlova, *Khimicheskaya terodinamika i fazovye ravnovesiya* (Chemical Thermodynamics and Phase Equilibria). Moscow: Metallurgia, 336 p. (1981) (in Russian).
3.46. B.I. Boltaks, *Diffusiya i tochechnye defekty v poluprovodnikakh* (Diffusion and Point Defects in Semiconductors). Leningrad: Nauka, 384 p. (1972) (in Russian).
3.47. M.G. Milvidsky, O.V. Pelevin, B.A. Sakharov, *Fiziko-khimicheskie osnovy polucheniya razlagayushchikhsya poluprovodnikovykh soedinenii* (Physicochemical Principles of Production of Decomposing Semiconductor Compounds). Moscow: Metallurgia, 395 p. (1974) (in Russian).
3.48. H.C. Casey, M.B. Panish, *Heterostructure Lasers*. New York–London: Acad. Press, 364 p. (1978).
3.49. H.C. Casey, M.B. Panish, K.B. Wolfstirn, *J. Phys. Chem. Sol.* **32**: 571–582 (1971).
3.50. C.S. Kang, P.E. Green, *Gallium Arsenide. Symp. Proc.*, Inst. Phys. and Phys. Soc., London: 18–31 (1969).
3.51. M.G. Milvidsky, V. Pelevin, *Izvestiya AN SSSR, ser. Neorg. mat.* **3**: 1159–1163 (1967) (in Russian).
3.52. S.S. Strelchenko, V.V. Lebedev, *Soedineniya $A^{III}B^{V}$* ($A^{III}B^{V}$ Compounds). Moscow: Metallurgia, 144 p. (1984).
3.53. E. Aven, J. Prener, eds., *Physics and Chemistry of A^{II}- B^{VI} Compounds*. Amsterdam: North–Holland Publ. Co. (1967).
3.54. D. Nobel, *Philips Res. Reports* **14**: 361–399 (1959).
3.55. V.I. Fistul, *Raspad peresyshchennykh poluprovodnikovykh tverdykh rastvorov* (Decomposition of Semiconductor Solid Solutions). Moscow: Metallurgia, 240 p. (1977).
3.56 B.I. Boltaks, ed., *Kompensirovannyi kremnii* (Compensated Silicon). Leningrad: Nauka, 122 p. (1972) (in Russian).
3.57. S. Zainabibinov, *Fizicheskie osnovy obrazovaniya glubokih urovnei v kremnii* (The Physics of Deep Level Formation in Silicon). Tashkent: FAN (1984) (in Russian).
3.58. V.I. Fistul, V.N. Tzygankov, A.G. Yakovlev, *FTP* **7**: 2037–2041 (1973) (in Russian).
3.59. C.S. Fuller, J.D. Struthers, *Phys. Rev.* **87**: 526–527 (1952).
3.60. C.S. Fuller, K. Wolfstrin, *J. Phys. Chem. Sol.* **26**: 1463–1473 (1965).
3.61. R.N. Hall, J.H. Racette, *J. Appl. Phys.* **45**: 379–396 (1984).
3.62. C.S. Fuller, J.A. Ditzenberger, *J. Appl. Phys.* **28**: 40–48 (1957).
3.63. H.H. Woodbury, W.W. Tyler, *Phys. Rev.* **105**: 84–92 (1957).
3.64. C.S. Fuller, J.D. Struthers, J.A. Ditzenberger *et al.*, *Phys. Rev.* **93**: 1182–1189 (1954).
3.65. K. Wolfstirn, C.S. Fuller, *J. Phys. Chem. Sol.* **7**: 141–145 (1958).
3.66. A. Milnes, *Deep Impurities in Semiconductors*. New York–London: Wiley, 562 (1973).
3.67. E.R. Weber, *Appl. Phys. A* **30**: 1–22 (1983).
3.68. C.B. Collins, R.O. Carlson, C.J. Galagher, *Phys. Rev.* **105**: 1168–1173 (1957).
3.69. W.R. Wilcox, T.J. La Chapelle, *J. Appl. Phys.* **35**: 240–246 (1964).
3.70. B.I. Boltaks, ed., *Kompensirovannyi kremnii* (Compensated Silicon). Leningrad: Nauka, 121 p. (1972).

3.71. M. Yoshida, K. Saito, *J. Appl. Phys.* **9**: 1217–1228 (1970).
3.72. W.M. Bullis, *Sol. St. Electron.* **9**: 143–168 (1966).
3.73. N. Wiehl, U. Herpers, E. Weber, *J. Radioanal. Chem.* **72**: 69–78 (1982).
3.74. J.H. Aalberts, *J. Appl. Phys. Lett.* **1**: 19–20 (1962).
3.75. M. Yoshida, K. Furusho, *Jap. J. Appl. Phys.* **3**: 521–527 (1964).
3.76. M. Yoshida, K. Saito, *Jap. J. Appl. Phys.* **6**: 573–581 (1967).
3.77. J.W. Chen, A.G. Milnes, *Ann. Mater. Sci.* **10**: 157–228 (1980).
3.78. B.I. Boltaks, Sue Shi Inn, FTT 2: 2677–2684 (1960) (in Russian).
3.79. Yu.A. Zibuts, L.G. Paritsky, S.M. Ryvkin, *FTT* **5**: 3301–3304 (1963).
3.80. J.D. Struthers, *J. Appl. Phys.* **87**: 1560 (1956).
3.81. M.S. Yusupov, Sh. Makhkamov, A. Mirzaev, *FTP* **11**: 592–594 (1977) (in Russian).
3.82. S.A. Azimov, M.S. Yusupov, G. Nurkuziev *et al.*, *FTP* **10**: 2164 (1976) (in Russian).
3.83. S.A. Azimov, A.A. Lebedev, M.S. Yunusov *et al.*, *FTP* **9**: 1828–1829 (1975) (in Russian).
3.84. A.N. Arshavsky, N.Yu. Shmyamov, *FTP* **20**: 104–108 (1986) (in Russian).

Chapter 4

Microscopic Analysis of Impurity Solubility in Semiconductors

4.1 DISSOLUTION ENTHALPY CALCULATION BY WEISSER'S METHOD

In contrast to the macroscopic approach described in the previous chapter, a microscopic analysis is aimed at treating the energy requirements for impurity dissolution in terms of the structure and type of chemical bonding. The relation to the macroscopic theory is through the values of dissolution enthalpy ΔH calculated from the first principles.

A detailed microscopic analysis of impurity solubility in semiconductors was first performed by Weisser [1]. It was based on the consideration of atomic interactions in the solution.

4.1.1 Site solubility

According to Weisser's model, the dissolution of impurity atoms at crystal lattice sites can be represented as a sum of several consecutive stages illustrated in Figure 4.1. The total dissolution enthalpy will be defined by the sum of the energy requirements in each stage. Let us discuss them in some detail.

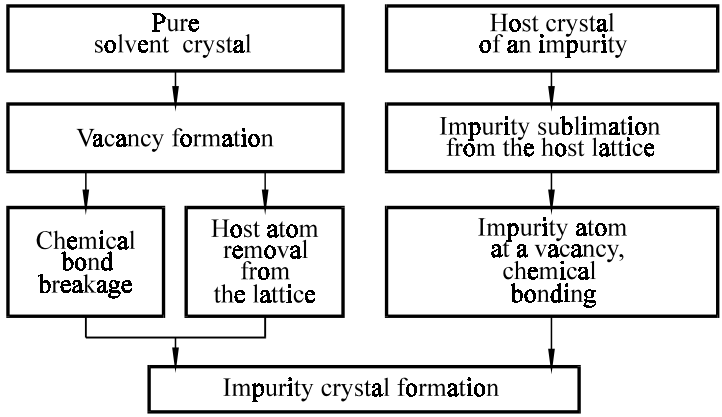

Figure 4.1. Model incorporation of an impurity atom into a semiconductor crystal lattice via substitution.

(1) For the incorporation of impurity atoms into crystal lattice sites, a certain number of sites must become free, i.e., vacancies must be formed in the host lattice. This process requires energy for the breakage of chemical bonds in the crystal and for the removal of liberated host atoms.

(2) Simultaneously, the impurity atoms that were initially in the solid phase in their crystal are sublimated from the lattice. They are transported to the host semiconductor, are adsorbed there, transported by diffusion to vacancies, and condense at them.

(3) The impurity atoms form new chemical bonds with neighboring host atoms at the crystal vacancies. Note that this dissolution pattern does not essentially change even if there is vacancy migration, in addition to impurity diffusion, in the host semiconductor.

The energy balance of the whole dissolution process in Weisser's model can be expressed as

$$\Delta H = 4E_{A0} - \Delta H_{sub}^{A} + \Delta H_{sub}^{B} - \Delta H_{AB}, \qquad (4.1.1)$$

where A refers to host atoms and B to impurity atoms, E_{A0} is chemical binding energy of the host crystal, ΔH_{sub} is sublimation heat, and ΔH_{AB} is the energy released in the condensation of an impurity atom at a host vacancy.

Expression (4.1.1) describes dissolution in an elemental semiconductor, like silicon, consisting of four valent atoms, so it contains the term $4E_{A0}$. The first two terms represent the vacancy formation energy in a host semiconductor. The quantity H_{AB} consists of two components: E_{AB} which is the energy

of chemical bonding between impurity and host atoms, and E_{def}, which is the elastic deformation energy of the lattice, arising from the difference in the sizes of impurity and host atoms. The energy of chemical bonding was described in [1] by the equation

$$E_{AB} = 4\varepsilon_{AB} = \left[\frac{1}{4}\left(\frac{1}{\Delta H_{sub}^{A}} + \frac{1}{\Delta H_{sub}^{B}}\right)\right]^{-1}, \qquad (4.1.2)$$

where ε_{AB} is the energy of a single chemical bond *AB*. Energy E_{def} was calculated in terms of elasticity theory, accounting for the difference Δr in the sizes of the substituting and substituted atoms and for the displacement of the four nearest host atoms by the value Δr_1 smaller than Δr because of the *AB* bond compression. On these two assumptions, E_{def} is

$$E_{def} = 4\tfrac{1}{2} g K_c (\Delta r - \Delta r_1)^2 + 8\pi G r_0 (\Delta r_1)^2, \qquad (4.1.3)$$

where the first term is the energy of chemical bond compression and the second is the tensile energy of the neighboring atoms, g is a dimensionless parameter close to unity, G is the displacement module, r_0 is the initial radius of a substituted host atom, and K_c is the crystal elastic constant defined as

$$K_c = m a_0 (C_{11} + 2C_{12}), \qquad (4.1.4)$$

where $m \approx 1$, a_0 is the host lattice period, C_{11} and C_{12} are elastic constants of the crystal.

After the substitution of numerical values for the constants into (4.1.3) and (4.1.4), we get

$$\begin{aligned} &\text{for Si:} \quad E_{def} = 9.2\times10^{23}\,\Delta r^2 \text{ J/g×atom}, \\ &\text{for Ge:} \quad E_{def} = 7.1\times10^{21}\Delta r^2 \text{ J/g×atom}, \end{aligned} \qquad (4.1.5)$$

where Δr is expressed in centimeters.

Weisser identified the quantity ΔS only with the variation of the oscillation component which was defined by Debye temperatures Θ_A and Θ_B:

$$\Delta S_{cm}^{s} = \Delta S_{osc} = 3k \ln \frac{\Theta_A T_A^f}{\Theta_B T_B^f}. \qquad (4.1.6)$$

The calculations of ΔH from (4.1.1) showed a quantitative difference with experimental values of K, sometimes as large as an order of magnitude or more. This, no doubt, was mostly due to the significant simplifications accepted by Weisser. First of all, this is the calculation of vacancy formation energy as the difference between the breakage energy of the four bonds, E_A, and the sublimation energy of a host atom. Today there are some publications, e.g. [2–4], using semi-empirical approaches and calculations from first principles, which provide more rigorous values of vacancy formation enthalpy ΔH_V in elemental and compound semiconductors. Therefore, it would be more reasonable to replace the first two terms in (4.1.1) by a ΔH value borrowed from one of the above publications.

Further, the E_{AB} values calculated from (4.1.2) cannot be considered correct either, because this formula does not take into account the interactions between an impurity atom and the basic host atoms. These interactions have a complicated character and cannot be described by sublimation energies of impurity and host atoms, required for their removal from the respective lattices. For this reason, a numerical calculation of binding energies from sublimation heats of the respective substances can hardly be regarded as being correct.

The concept of constant covalent radius used by Weisser for finding deformation energy from (4.1.3) cannot be taken for granted either. First, there are many kinds of atomic radii and, second, the atomic size is a variable quantity varying with the particular structure and type of chemical bond. We will consider this problem below.

Another drawback of Weisser's calculations [1] was that the dissolved impurity ionization was ignored. But it must be taken into consideration since an ion located at a lattice site causes a displacement of the electron density of neighboring atoms, inducing polarization. It is pointed out in [5] that the allowance for the ionization decreases the total crystal energy by an appreciable value which may exceed ΔH values found from (4.1.1).

On the other hand, the calculated and experimental values for impurity indium in germanium and silicon, as well as for tin in germanium, more or less coincide. But this is likely to be due to the compensation of one kind of error by another.

4.1.2 Interstitial solubility

Weisser analyzed the energy balance for interstitial dissolution in a way similar to that for site solubility [6, 7]. This process is illustrated schematically in Figure 4.2.

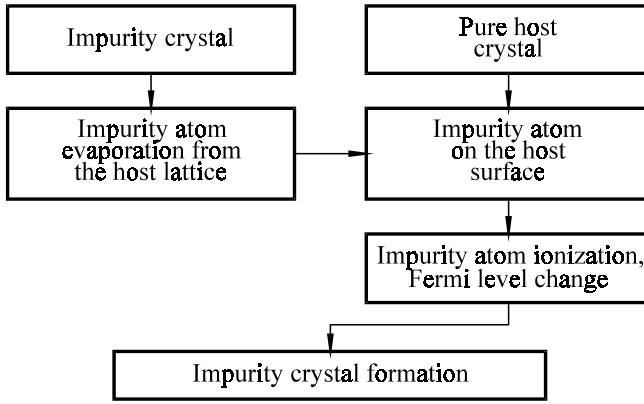

Figure 4.2. Weisser's model of impurity dissolution at crystal interstices [6, 7].

An impurity atom is sublimated from its own lattice by energy ΔH_{sub}. Then it is transported to the semiconductor surface and incorporated into an interstice, thus producing a doped crystal. The atom may become ionized on the semiconductor surface, thereby changing the Fermi level position.

An interstitial atom or ion is affected by repulsive or attractive forces. The expression for ΔH_s for an ionized interstitial impurity is

$$\Delta H_s = \Delta H_{sub} + I = A_B + U_{rep} - U_{im}, \tag{4.1.7}$$

where I is the ionization potential of an impurity atom and A_B is the work function for an electron sublimated from substance B.

For a neutral impurity requiring no energy for ionization, ΔH will be expressed, instead of (4.1.7), as

$$\Delta H_s = \Delta H_{sub} + U_{rep} - U_{im} - A_B. \tag{4.1.8}$$

Since the values of ΔH_{sub}, I, and A_B were tabulated in [8], one has only to find the polarization and repulsion energies to determine ΔH. To calculate the polarization energy U_p, Weisser used the Mott–Littleton method for alkali-halide crystals [9]. This method is as follows.

A positive ion incorporated into a crystal lattice induces polarization of electron shells of the neighboring atoms, and the induced dipoles interact with the impurity atom and with one another. The total energy of the system decreases by U_{im}.

For semiconductors, Weisser suggested the following assumptions:

(1) The dipole moment μ induced in the host atoms varies linearly with an electric field of strength E.

(2) Polarizability of crystal atoms, α_B, is chosen such that it does not depend on the polarization of other atoms, especially, of the nearest neighbors. Polarizability is an isotropic quantity.

(3) The dipole moment is directed along the vector radius away from charge q.

(4) The electric field at the atom center is equal to the vector sum of the field induced by the charge and the fields from all other dipoles induced in the crystal atoms.

Since the dipole moment of an atom, μ_i, depends on the dipole moments of other atoms, the problem of finding a dipole moment includes N simultaneous equations for N lattice atoms. The dipole moments of atoms located in the same sphere are assumed to be identical.

In an n-sphere approximation, the general set of equations for dipole moments is

$$\sum_{i=1}^{n} \mu_i \left(E_{ij} - \delta_{ij} \frac{1}{\alpha_B} \right) = \frac{q}{r_m^2}, \tag{4.1.9}$$

where E_{ij} are matrix elements varying with the structure geometry, r_m is the radius of the respective coordination sphere, and δ_{ij} is the Kronecker delta.

The quantities E_{ij} are the functions of atomic coordinates to be described by the expressions

$$E_{ij} = \sum_{G_j} \frac{1}{R_j (r_k)^5} S_{ij}, \tag{4.1.10}$$

$$\begin{aligned} S_{ij} = & \left[-x_j \left(r_k^2 - 3\xi^2 \right) + y_j (3\xi\eta) + z_j (3\xi\rho) \right] \frac{x_0}{R_j} \\ & + \left[x_j (3\xi\eta) - y_j \left(r_k^2 - 3\eta^2 \right) + z_j (3\eta\rho) \right] \frac{y_0}{R_j} \\ & + \left[x_j (3\xi\rho) + y_j (3\eta\rho) - z_j \left(r_k^2 - 3\rho^2 \right) \right] \frac{z_0}{R_j} \end{aligned} . \tag{4.1.11}$$

The summation is made over all j-atoms of the lattice, located in the j-sphere; x_j, y_j, and z_j are the coordinates of an atom in the j-sphere; x_0, y_0, and z_0

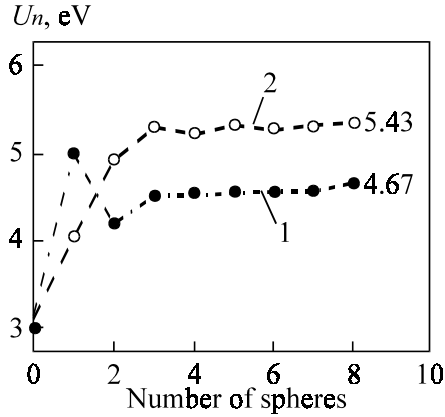

Figure 4.3. Polarization energy of the silicon lattice in different approximations (sphere numbers): 1 – tetrahedral interstice; 2 – hexagonal interstice.

are the coordinates of an atom in the *i*-sphere; ξ, η, ρ, and r_k are the components of interatomic distances of the lattice with the coordinates x_0, y_0, z_0, and x_j, y_j, and z_j; and R_j is the *j*-sphere radius. We also have

$$\begin{aligned} \zeta &= x_j - x_0, \\ \eta &= y_j - y_0, \\ \rho &= z_j - z_0, \\ r_k &= r_j - r_0, \end{aligned} \tag{4.1.12}$$

and r_j, r_0 are the radii of the respective spheres.

Potential φ at the point of charge q location is calculated from

$$\varphi = \sum_j^n \sum_i \frac{\mu_i}{R_j^2} - \frac{q}{R_n + 1}\left(1 - \frac{1}{\varepsilon}\right). \tag{4.1.13}$$

The second term in (4.1.13) takes into account the contribution of all other atoms located outside the *j*-sphere.

As a result, the polarization energy U_{im} is described by the equation

$$U_{im} = -\frac{1}{2} q\varphi\,. \tag{4.1.14}$$

The calculation accuracy is determined by the optimal choice of the number of atoms surrounding an interstitial impurity ion. Weisser [7] used an 8-sphere approximation. The calculated values of U_{im} (Figure 4.3) were constant, beginning with the third order. This result did not change even when the number of spheres was increased to 14 [10].

To find the repulsion energy U_{rep}, Weisser used the Born–Mayer two-parametric potential of the form

$$U_{rep} = A\exp\left[\left(r_i + r_s - r\right)/\rho\right], \tag{4.1.15}$$

where r_i, r_s, and r are the radii of an impurity ion, a host atom, and the distance between impurity atoms, respectively; A and ρ are parameters calculated from crystal compressibility data. The calculation procedure details are described in [11].

We doubt the correctness of this determination, because the structure of silicon and germanium radically differs from the NaCl crystal structure. In particular, the concept of Madelung's constant used in the case of covalent silicon and germanium crystals makes no sense.

It is clear from equation (4.1.15) that the repulsion energy is a function of the host atomic and impurity ion radii, whose choice is ambiguous, as was pointed out above. All systems of radii stem from the assumption that they are constant and additive. The systems of atom–orbital radii are no exception, although they are based on quantum mechanical calculations. Still, they are essentially related to the quasichemical concept of radii being constant and additive.

In reality, the sizes of atoms and ions vary with the kind of host compound, i.e., they change from one nonequivalent bond to another, depending on the structure geometry. Therefore, there is no size of an ion in general, but there is the size of a particular ion. The radius value is always determined for an ionic pair.

Because of the ambiguity of radius and other parameters describing U_{im} and U_{rep}, Weisser's theory failed to yield results fitting the experimental data well. There is a more or less good agreement with experiments only for Li in silicon and germanium.

Millea [12], Hasiguti [13], and Oxenhandler [10] made attempts to improve Weisser's model by choosing different values of atomic radii, the measure of potential drop ρ and the A constant used in the Born–Mayer potential in (4.1.15). In particular, Millea considered the interaction between interstitial ions and the atomic skeleton of the host semiconductor, ignoring valent electrons of Si and Ge atoms but using Si and Ge ionic radii according to Pauling [14]. The values of ρ were taken to be different for different im-

purities, but the calculations were made as in [11] from compressibility data for the respective alkali-halide compounds. They were chosen according to the arrangement of elements in the periodic table—the serial number of an alkali metal increased ($N_{Li} < N_K < N_{Cs}$) with increasing serial number of the impurity in the group ($N_{Li} < N_{Cu} < N_{Au}$). Since $N_{Ge} > N_{Si}$, the serial number of halogen was larger than for Si ($N_{Cl} > N_F$). Naturally, this approach cannot be justified either, although the agreement with experimental data is somewhat better than in Weisser's method.

Hasiguti [13] determined A and ρ from the compressibility values of germanium and silicon, using the Madelung constant for ionic sphalerite crystals, explaining his choice by their structural similarity to the diamond lattice. Obviously, the use of the Madelung constant for the analysis of solubility in germanium and silicon remains questionable.

In contrast to Weisser, Oxenhandler [10] introduced factor F into equation (3.1.17) to account for the degree of the outer shell filling of an impurity atom. He used an original method for the calculation of U_{rep}, representing it as a two-component quantity [15]:

$$U_{rep} = U_{rep}^{cor} + U_{rep}^{ext}, \tag{4.1.16}$$

where U_{rep}^{cor} is the repulsion energy due to the inner electron shells of an impurity atom, defined by the Born–Mayer potential, and U_{rep}^{ext} is the repulsion energy arising from the interaction between the outer *s*-electron of an impurity atom and the chemical bond electrons of a host atom. The latter component was calculated in [17] using Mulliken's method [18] to be discussed below.

The procedure suggested by Oxenhandler makes it possible to extend Weisser's approach to more complex impurities, in particular, to transition metals. This could not be done before because of the restructuring of their electron shells in the dissolution at lattice sites and interstices. However, the use of the Born–Mayer potential together with the ambiguity of its parameters considerably decreased the applicability of the results. In addition to the criticism above, a general disadvantage of the approaches discussed in this section is the neglect of non-zero internal crystal potential U_{cr} in the interstitial space. Of importance is also the lattice relaxation in the vicinity of an interstitial impurity atom.

4.2 DISSOLUTION ENTHALPY IN THE PSEUDO-ALLOY MODEL

An alternative approach to the calculation of ΔH_s was used in [19]. The mixing enthalpy was found from the sum of two terms varying with the mole fraction of the doped compound:

$$\Delta H_s(x) = \Delta H_s^0(x) + \Delta H^i(x), \tag{4.2.1}$$

where $\Delta H_s^0(x)$ was calculated from the difference between the total formation heat of initial binary compounds, say, MeP and GaP as applied to Me-doped GaP, and the formation heats of the pseudobinary alloy $Ga_{1-x}Me_xP$, with account taken of the mole fraction x:

$$\Delta H_s^0(x) = (1-x)\Delta H_{GaP}^0 + x\Delta H_{MeP}^0 - \Delta H_{Ga_{1-x}Me_xP}^0. \tag{4.2.2}$$

The value of ΔH^i in the right-hand side of (4.2.1) was found from a spectroscopic binding model to be

$$\Delta H^i = R\bar{a}^{-3}Df_i,$$

where R is a dimensional factor, $\bar{a}$ is a mean interatomic distance, D is the chemical bonding energy, and f is a factor describing the host crystal ionicity equal to

$$f_i = \frac{c^2}{\left(c^2 + E_H\right)^2}, \tag{4.2.3}$$

where c is the electronegativity of the compound atoms and E_H is their homopolar binding energy.

The change in entropy, ΔS_{cm}, necessary for finding K was calculated in [19], using the well-known formula

$$\Delta S_{cm} = -R(x_1 \ln x_1 + x_2 \ln x_2), \tag{4.2.4}$$

where x are the mole fractions of the components being mixed.

These expressions were further used to calculate ΔH_{cm} and ΔS_{cm} [19], as well as to find the distribution coefficients of the impurities Fe, Mn, and Cr in GaP. The calculated values of K_0 were found to be 1.1×10^{-4}, 1.6×10^{-2}, and

1.05×10^{-2}, respectively. They agree well with the experimental values of K_0 for these impurities: 2×10^{-4}, 2×10^{-2}, and 1.6×10^{-3}.

In spite of this coincidence, this approach has a serious disadvantage because of the ambiguous ionicity value. It was found in [19] in terms of electronegativity. However, there are many electronegativity systems, and the concept itself has no physical sense. So it is hardly justifiable to use this concept in numerical calculations of chemical bonding energies. Moreover, there is a large number of publications on the calculation of ionicity in terms of dielectric theory (see e.g., [20]). Today it is impossible to give a definite value of ionicity for $A^{III}B^{V}$ and $A^{II}B^{VI}$ compounds because of the considerable difference in the values calculated by different workers. On the other hand, the calculations of ΔH^{i}, ΔH_s, and K are very sensitive to the chosen ionicity value.

4.3 WEISSER'S MODIFIED SOLUBILITY THEORY

Weisser's solubility theory was considerably modified for the treatment of impurity atoms with partly filled electron shells [21] and extended to common impurity atoms. In the latter case, the solubility theory must take into account the specific behavior of *d*- and *f*-electrons when impurity atoms are distributed in a host crystal. Since the electron shell configuration is transformed differently at lattice sites and interstices, these two situations should be treated individually.

4.3.1 Interstitial *d*-atom solubility

The dissolution model implies the following assumptions: (a) transition metal atoms in the gas phase are free and have a $d^n s^m$ electronic configuration; (b) impurity atoms do not interact with one another in the gas phase or in the incorporation positions in the crystal lattice, interacting only with host atoms at interstices. Impurity dissolution is presented schematically in Figure 4.4. Here, work is done on an impurity atom to separate it from a free crystal, to transport to the semiconductor surface, and to incorporate into an interstice. In contrast to the scheme in Figure 4.2., there is a change in the electronic configuration from $d^n s^m$ to d^{n+m}.

It follows from general thermodynamics that enthalpy is a function of the system state, i.e., it is independent of the course of the process but is deter-

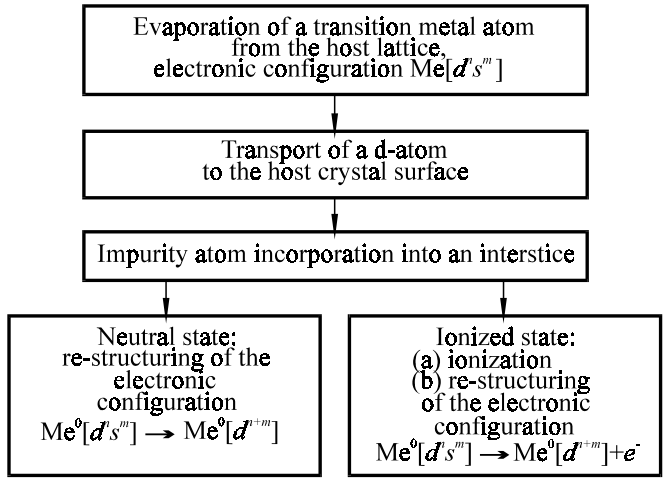

Figure 4.4. Model dissolution of *d*-impurity atoms at silicon interstices.

mined only by the initial and final states of the system of interest. We have assumed the atom state in the gas phase to be the initial state and its state at an interstice as the final state. For convenience, we have accepted a scheme [21], in which a free impurity atom is produced by sublimation from its own lattice. The energy required for this process is equal to atomization energy ΔH_{at} which is the energy characteristic of the initial state of an im-purity atom. In the final (interstitial) state, an impurity atom is affected by repulsive forces because of the electron shell overlap of the impurity and crystal atoms, by the internal crystal field U_{cr}, and by the repulsive forces due to the lattice polarization by an impurity ion, if the impurity center is ionized.

The energy requirements also include the energy necessary for electronic restructuring, or promotion energy ΔP, and so called extra-stabilization energy $\Delta\delta_{ex}$. The extra-stabilization energy, like the repulsion energy, is a function of the electronic configuration of the *d*-impurity at *H*- and *T*-interstices and is defined by the *d*-level splitting by the crystal field. Since the values of $\delta_{ex}^{T,H}$ are fractions of the splitting value D_q [22] and the latter varies between 0.1 and 0.3 eV for different interstitial *d*-impurities, the difference $\Delta\delta_{ex}$ for such small values will lie within the calculation error. So the difference between extra-stabilization energies can be neglected in the first approximation. Dissolution enthalpy ΔH_s^i represents an algebraic sum of energy requirements for the transition from the initial to the final state of the impurity atom. If an interstitial impurity atom is neutral, the polarization energy U_{im} is zero, and ΔH_s^i has the form:

$$\Delta H_s^i\left(\mathrm{Me}^0\right) = \Delta H_{at} + U_{rep} - U_{cr} + \Delta P\,, \qquad (4.3.1)$$

where the superscript *i* indicates impurity dissolution at interstices and the subscript *s* its dissolution in the solid phase.

For an ionized impurity state, one must make allowance for the host lattice polarization energy and the ionization energy of an impurity atom. The latter is evaluated using ionization potential I_s, taken in [21] to be the *s*-electron orbital ionization potential of a free atom. This approach suggests that the electronic configuration of an impurity center becomes d^{n+m-1} due to simultaneous restructuring and ionization.

Dissolution enthalpy is

$$\Delta H_s^i\left(\mathrm{Me}^+\right) = \Delta H_{at} + U_{rep} - U_{cr} - U_{im} + \Delta P + I_s\,. \qquad (4.3.2)$$

The values of ΔH_{at} are tabulated values.

The promotion energy ΔP can be evaluated as the transition energy of an electron between term-averaged levels for the final and initial configurations:

$$\Delta P = E\left(d^{n+m}\right) - E\left(d^n s^m\right), \qquad (4.3.3)$$

for ionized impurities, it is

$$\Delta P = E\left(d^{n+m-1}\right) - E\left(d^n s^m\right). \qquad (4.3.4)$$

The energies of the centers of mass for terms with various valent configurations of atoms were calculated and tabulated in [24].

For common non-transition metal impurities and for impurity atoms with a completely filled *d*-shell, there is no electronic configuration restructuring; hence, $\Delta P = 0$.

Polarization energy U_{im} in expression (4.3.2) was calculated from (4.1.14). Table 4.1 presents the matrix elements of (4.1.10) for *T*- and *H*-interstices and for crystal lattice sites. The numerical values of the matrix elements are dimensionless. They are universal and apply to any semiconductor crystallizing in the diamond- and sphalerite-type of lattice. To go to dimensional values, it is necessary to divide the tabulated values of E_{ij} by a_0^3, where a_0 is the crystal lattice period of a particular semiconductor.

Table 4.1. Matrix elements E_{ij} for polarization energy calculations for the diamond- and sphalerite-types of lattice.

E_{ij}	i	Values of j			
		1	2	3	4
E_{ij}^{T}	1	–14.14	–52.06	–16.17	16.08
	2	–34.71	–18.97	0.96	–10.30
	3	–5.39	0.48	–15.58	–20.54
	4	21.14	–7.73	–16.48	–6.15
E_{ij}^{H}	1	–44.60	24.34	12.82	–7.96
	2	–18.27	–25.57	–10.00	–4.71
	3	12.86	–12.58	–6.10	–27.95
	4	–2.84	–1.47	–7.35	–27.09
E_{ij}^{s}	1	–14.14	–18.45	–16.17	–2.84
	2	–6.15	–21.61	–33.35	–6.47
	3	–5.39	–33.35	–15.58	–10.79
	4	–1.89	–12.94	–21.59	–2.37

The values of U_{im} for silicon were found in [25] to be $U_{\text{im}}^{T} = 4.54$ eV and $U_{\text{im}}^{H} = 5.32$ eV; for site solubility $U_{\text{im}}^{s} = 4.40$ eV.

A limitation of the above ionization energy evaluation is the neglect of the host lattice distortion effect on U_{im}, which seems impossible to calculate rigorously because we do not know the absolute displacements of atoms from their equilibrium positions, necessary in the calculation with (4.1.9) and (4.1.10) involving the distances between impurity centers and host lattice atoms, r_{m} and R_{i}. Lattice distortions were taken into consideration in [21] in the calculation of repulsion energy U_{rep}. Note that we used Mulliken's method [18] in that work instead of the "radius" approach with the Born–Mayer potential (4.1.15). As a result, repulsion energy was expressed as

$$U_{\text{rep}} = \beta\nu \sum_{kl} \bar{I}_{kl} S_{kl}^2 , \qquad (4.3.5)$$

where I_{kl} is the average ionization potential of the k-th electron of a host atom, S_{kl} is the electron overlap integral, β is a calibration parameter ($\beta < 1$) selected from the best agreement between theoretical and experimental values of ΔH of a well-studied model impurity, and ν is a factor accounting for local lattice distortions near an interstitial impurity atom.

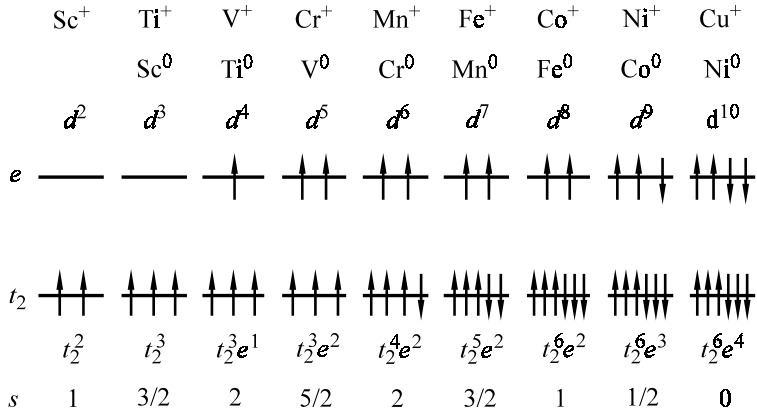

Figure 4.5. The state of 3*d*-impurities at a silicon tetrahedral interstice.

With the hexagonal interstice D_{3d}, the crystal field effect on the *d*-orbitals of an impurity atom is different at the two interstices. The *d*-impurity states at *T*-interstices were described in terms of the model suggested by Ludwig and Woodbury [26], based on the crystal field theory of Roitzin and Firshtein [27] and verified in many experiments. In this model, the internal tetrahedral crystal field partly compensates the 5-fold degeneracy of the *d*-shell. Doubly degenerate *e*-states at a *T*-interstice appear to have a higher energy than triply degenerate *t*-states (Figure 4.5). The outer *s*-electrons go to the *d*-shell under the crystal field action, resulting in the d^{n+m} electronic configuration of the atom.

The electronic state of a *d*-impurity at an *H*-interstice was treated in detail in [21]. The character of the *d*-level splitting is shown in Figures 4.6, 4.7, and 4.8. The filling of the *t*- and *e*-states (*T*-interstice) and of the a_{1g}, e_g', and e_g''-states (H-interstice) is, as usual, described from the Pauli principle and Hund rule. According to the crystal field theory [25], 3*d*-impurities with d^4–d^7 configurations (*T*-interstices) and d^2–d^7 configurations (*H*-interstices) will have high spin states (Figure 4.6), while 4*d*- and 5*d*-impurities will have low spin states (Figure 4.8).

In our calculation of overlap integrals S_{kl} [21], the wave functions of bond electrons in silicon were chosen as a linear combination of atomic orbitals (LCAO), each of which was a *sp*-hybrid built on Slater's wave functions. The wave functions of outer impurity electrons were chosen as Slater's *d*-functions, in accordance with the state of a *d*-atom at a silicon

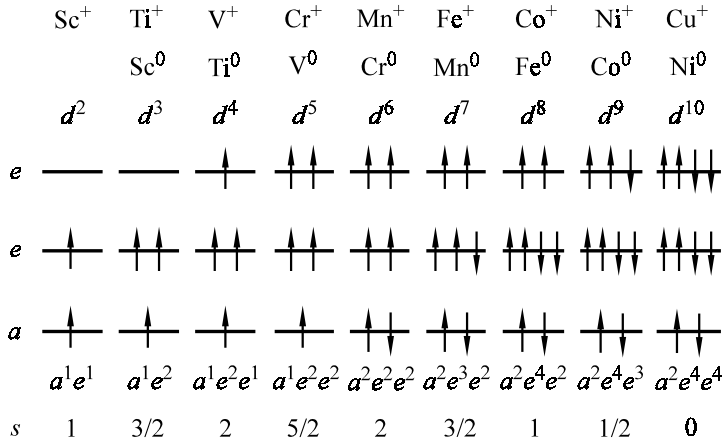

Figure 4.6. The state of 3*d*-impurities at a silicon hexagonal interstice.

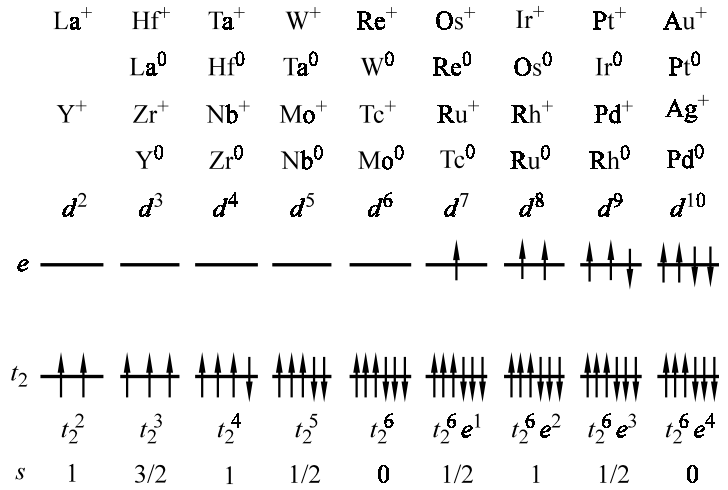

Figure 4.7. The state of 4*d*- and 5*d*-impurities at a silicon tetrahedral interstice.

interstice. The overlap integrals of the *d*-orbitals of an interstitial atom with *sp*-hybrid atomic orbitals were reduced to diatomic ones, with respective coefficients obtained from the transformation matrices of *d*-functions for an impurity atom and of *s*- and *p*-functions for silicon atoms. The summation in

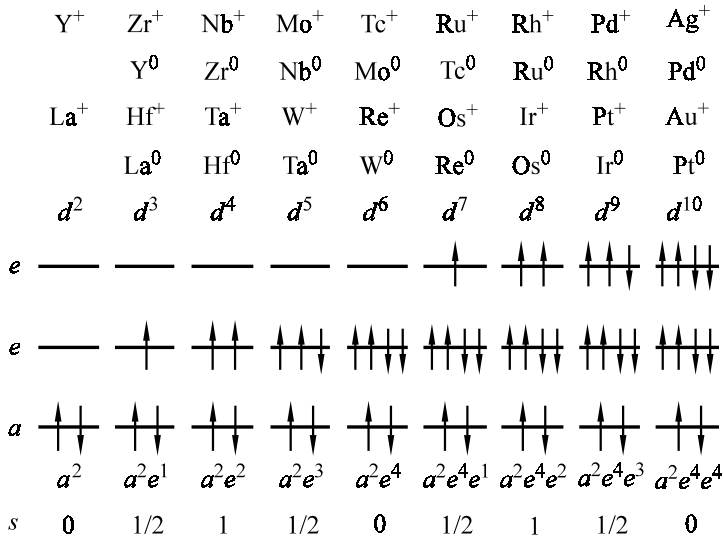

Figure 4.8. The state of 4*d*- and 5*d*-impurities at a silicon hexagonal interstice.

(4.3.5) was performed over all outer electrons of an impurity atom and the bond electrons of silicon atoms within two coordination spheres. The contribution of the third sphere to U_{rep} was estimated to be less than 1%.

Since the calculation of U_{rep} was made for two spheres around an impurity atom, which are subject to the greatest distortions, it appeared important to find coefficient ν in (4.3.5) allowing for local lattice distortions. These may be of two kinds—symmetrical and asymmetrical (Figure 2.11). But we considered only symmetrical distortions [21] which are due to an electron–phonon interaction.

The absolute displacements are described as

$$\Theta = \frac{1}{\Omega}\sqrt{\frac{2\Delta E}{M}}, \qquad (4.3.6)$$

where ΔE is the energy difference in the formation of an occupied and an empty interstice, M is the mass of a cluster consisting of an impurity atom and its neighbors interacting with it, and Ω is the cluster oscillation frequency.

The coefficient ν was expressed in [21] as

$$\nu = \frac{\Theta_x}{\Theta_M} = \sqrt{\frac{\Delta E_x M_M}{\Delta E_M M_x}}\ , \qquad (4.3.7)$$

where the subscript x refers to the impurity under consideration and the subscript M to the model impurity. The model impurities were taken to be $Ni^0(3d^{10})$ and $Cu^+(3d^{10})$, whose charges at silicon interstices were reliably established experimentally.

The β values found from the best agreement between the experimental and calculated values of ΔE were 0.5 for Ni^0 and 0.154 for Cu^+. Lattice distortions for these model impurities were taken into account from the very beginning. For the other impurities, the distortions were allowed for by accepting these β values and introducing the coefficient of (4.3.7) into (4.3.5), but with respect to the model impurity this time.

Expressions (4.3.1) and (4.3.2) for ΔH_s^i also include internal crystal field potential U_{cr} which is difficult to calculate because it depends on the electron density distribution at interstices, whose exact value is unknown. For this reason, there were several assumptions made on the calculation of U_{cr}. It was suggested in [28], for example, that Si^{4+} and Ge^{4+} ions were immersed in a homogeneous valent electron gas. The calculation for silicon in terms of the pseudopotential method yielded U_{cr}^T = 17.5 eV and U_{cr}^H = 21.4 eV. In an alternative model [29], valent electrons were considered as being localized

Table 4.2. Calculated dissolution enthalpies ΔH_s^T (eV) of *d*-impurities at silicon *T*-interstices.

Impurity	$\Delta H_s^T(Me^0)$	ΔH_s^T	Impurity	$\Delta H_s^T(Me^0)$	ΔH_s^T
Sc	6.02	2.10	Ru	5.33	5.27
Ti	8.36	4.21	Rh	4.15	4.17
V	8.03	4.47	Pd	1.16	2.91
Cr	3.12	1.82	Ag	0.42	0.93
Mn	1.81	0.56	La	9.44	1.79
Fe	2.02	1.59	Hf	18.62	8.4
Co	1.13	1.73	Ta	18.11	10.5
Ni	1.52	3.18	W	14.93	9.91
Cu	1.08	1.08	Re	12.87	9.38
Y	9.68	1.13	Os	–	8.9
Zr	11.59	6.02	Ir	7.97	7.6
Nb	10.69	6.37	Pt	3.81	6.12
Mo	7.67	5.78	Au	1.32	3.50
Tc	6.43	4.54			

Table 4.3. The most probable charge states of *d*-impurities at silicon *T*-interstices.

d^2	d^3	d^4	d^5	d^6	d^7	d^8	d^9	d^{10}	$d^{10}s^1$
Sc^+	$^*Ti^+$	V^+	$^*Cr^+$	$^*Mn^+$	$^*Fe^+$	$^*Fe^0$	$^*Co^0$	$^*Ni^0$	
								$^*Cu^+$	$^*Cu^0$
Y^+	Zr^+	Nb^+	Mo^+	Tc^+	Ru^+	Ru^0			
						Rh^+	Rh^0	Pd^0	Ag^0
La^+	Hf^+	Ta^+	W^+	Re^+	Os^+	Ir^+	Ir^0	Pt^0	Au^0

* States confirmed experimentally.

halfway between the lattice atoms. This potential was assumed to be a long-range one. The summation was performed for 16- or 18-atom spheres around an interstice to give U_{cr}^{T} = 5.97 eV and U_{cr}^{H} = 5.67 eV. Both models de--scribe the extreme cases. The true values of U_{cr} must lie within these limits. We calculated ΔH_s^i, using the values of U_{cr}^{T} and U_{cr}^{H} found in [29].

As a result, the interstitial solubility enthalpies were calculated for a large number of transition metal impurities in silicon. *T*-interstitial impurities have the lowest ΔH_s values, i.e., tetrahedral interstices are their ground states. The most probable charge states of the impurities, calculated from the lowest values of $\Delta H_s^T(Me^0)$ and ΔH_s^T, are clear from Tables 4.2 and 4.3.

4.3.2 Site solubility of *d*-atoms

Similarly to the interstitial solubility discussed above, the dissolution enthalpy of transition metals at lattice sites represents an algebraic sum of energy requirements for the dissolution process. For the site component, these are (Figure 4.1): the energy required for the production of elemental transition metal gas, ΔH_{at}, at the initial stage of the process; the energy for vacancy formation, ΔH_{at}^{V}; and the energy D_0 released at the final stage during the chemical bonding of an impurity atom to the nearest host atoms. Therefore, we have

$$\Delta H_s^s = \Delta H_{at} + \Delta H_{at}^{V} - D_0 \,. \qquad (4.3.8)$$

Here, the superscript *s* denotes the site position of an impurity atom. In [21], we accepted for silicon ΔH_f^V = 2.88 eV, as was found in [3].

Let us consider, in some detail, the procedure for the calculation of chemical bonding energy D_0 in expression (4.3.8). For silicon, D_0 can be calculated by Mulliken's quantum chemical method [18], in which

$$D_0 = \sum X_{ij} + \frac{1}{2}\sum K_{mn} - \frac{1}{2}\sum Y_{kl} - \Delta P + E_{\text{i}}, \qquad (4.3.9)$$

where ΣX_{ij} is exchange energy over all binding electron pairs, ΣK_{mn} and ΣY_{kl} are energies of non-binding electron pairs, ΔP is promotion energy, and E_{i} is ion interaction energy. Energy ΣX_{ij} is expressed as

$$\sum X_{ij} = \sum A_i S_{ij} \bar{I}_{ij} \frac{1}{S_{ij}+1}, \qquad (4.3.10)$$

where S_{ij} is the overlap integral of an *i*-electron of silicon and a *j*-electron of the impurity atom, involved in chemical bonding; $\bar{I}_{ij}$ is their average ionization potential; A_i is an empirical parameter equal to 0.65 for *s*-*s* bonds, 1 for *s*-*p* and σ-bonds, and 1.5 for π-bonds.

The electron–electron repulsion energy ΣY_{kl} is

$$\sum Y_{kl} = \beta \sum_{kl} I_{kl} S_{kl}^2, \qquad (4.3.11)$$

coinciding with (4.3.5) for the repulsion energy U_{rep} of interstitial impurities. So, we will further denote ΣY_{kl} as $U_{\text{rep}}^{\text{s}}$.

The exchange energy ΣK_{mn} for non-binding electron pairs can be neglected because it is so low [30].

The ion interaction energy E_{i} was calculated by Mulliken as the square difference of electron electronegativities on Pauling's scale [14]. To avoid ambiguity in the calculation of electronegativities, E_{i} should be identified with the energy of crystal lattice polarization by an impurity ion in the substitution position ($E_{\text{i}} = U_{\text{im}}^{\text{s}}$). The procedure of U_{im} calculation was described in the preceding section, and the numerical value of U_{im}^{s} was found in [21] to be 4.40 eV for singly ionized impurities in silicon.

This approach to finding $E_{\text{i}} = U_{\text{im}}^{\text{s}}$ is justified, because an ionized impurity induces dipole moments in the lattice atoms, causing attraction between the impurity ion and the host atoms, leading, in turn, to an ionic interaction between them.

To calculate the promotion energy ΔP, one must know the electronic configurations of impurity atoms at the crystal sites. The Ludwig–Woodbury model considered above describes well the electronic configurations of interstitial *d*-atoms but seems doubtful when applied to their site positions. There are several reasons for this doubt.

(1) This model predicts the Me^{4+} state, because a d-atom is to give off four electrons for the binding to silicon atoms. But this would require the energy of 30–60 eV [31], which is very unlikely.

(2) Donor states in silicon are to have the transition $d^{n+m-4} \rightarrow d^{n+m-5} + e^-$ and acceptor states must have the transition $d^{n+m-4} + e^- \rightarrow d^{n+m-3}$. To illustrate, the electronic configuration of Cu, Ag, and Au atoms at silicon sites must be d^7, which must change as $d^7 \rightarrow d^6 + e^-$ in the donor-type ionization and as $d^7 + e^- \rightarrow d^8$ in the acceptor ionization, but this would require ~100 eV, which is a very unrealistic value.

(3) The promotion of several electrons, capable of producing tetrahedral bonds, to the excited state on the *p*-shell would require high energies. In particular, the promotion energy for a neutral iron atom $Fe^0(3d^6 4s^2)$ with a consecutive promotion of *d*-electrons to the *p*-shell, producing the $Fe^0(3d^6 4s^1 4p^1)$ states, would require ~3 eV, while for the excitation of the $Fe^0(3d^5 4s^1 4p^2)$ state, the necessary energy would be ~10 eV. Clearly, the promotion of still another 3*d*-electron to the 4*p*-shell to produce $Fe^0(3d^4 4s^1 4p^1)$ would require much more energy than 10 eV (see the valent state energies of *d*-elements in [24]). The same is true of other *d*-impurities.

Therefore, the *sp*-hybrid chemical bonding of *d*-impurities to silicon atoms is very unlikely because of the great energy requirements for the electron shell restructuring of the *d*-atom. More feasible is the formation of tetrahedral bonds between an impurity atom and silicon atoms by involving *d*-electrons into the chemical bonding. With the concept of atomic orbital hybridization, one can suggest the production of d^3s-hybrids geometrically equivalent to a tetrahedron, similar to the sp^3-hybrid. Electronic configurations producing d^3s-hybrids are illustrated in Figure 4.9. It should be noted that *d*-atoms with the $Me^0(d^7 s^2)$ and $Me^0(d^8 s^2)$ configurations in the free state may have the d^2sp bond which represents a distorted tetrahedron. In this case, the lattice distortions around the impurity atom must be appreciable, and the *d*-atom displacement from the lattice site is likely to occur (the Yan–Teller effect).

The *d*-shell electrons which are not involved in tetrahedral bonding are located on the loosened *t*- and *e*-levels, with the *e*-states lying lower than the *t*-states. It is these electrons that should be taken into account in the repulsion energy calculation from (4.3.11).

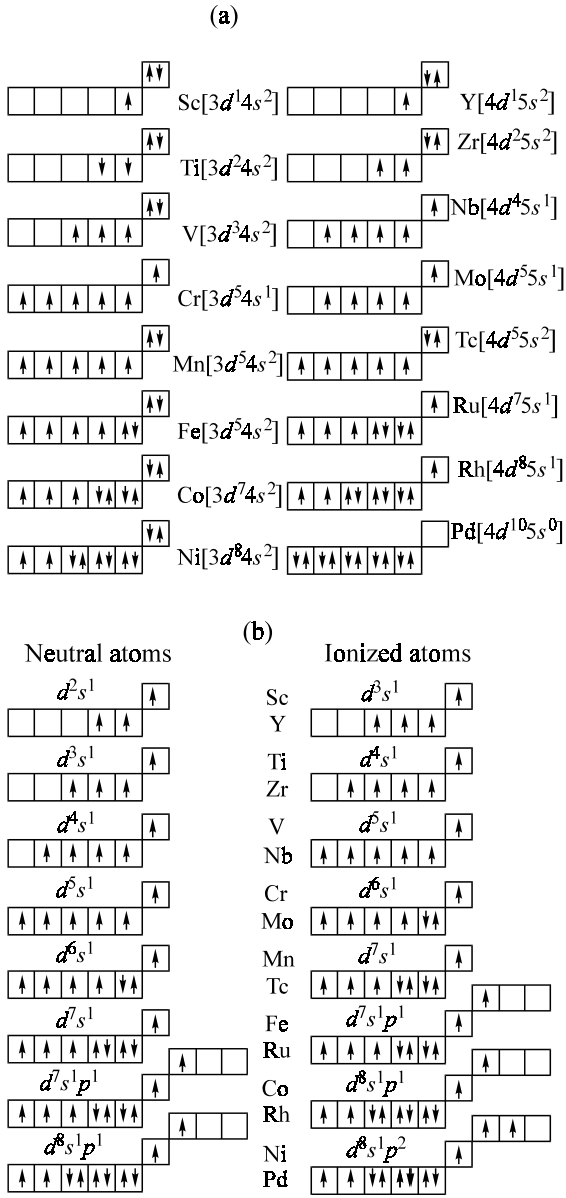

Figure 4.9. Electronic configurations of transition metals in the free state (a) and at silicon lattice sites (b).

The calculation of exchange energy ΣX_{ij} from (4.3.10) requires knowledge of overlap integrals corresponding to the impurity atomic orbital bonding to the sp^3-silicon atomic orbitals. For simplicity, the calculation can be restricted to the atomic orbitals of an impurity and a host atom oriented only toward each other. The details of the calculation can be found in [1].

Expression (4.3.9) for energy D_0 contains the quantity $\Sigma Y_{kl} = U^{\mathrm{s}}_{\mathrm{rep}}$. Its calculation from (4.3.11) similarly to (4.3.5) requires factor ν to allow for the crystal lattice relaxation around an impurity atom. The difference with a site atom is the choice of another model impurity, which may conveniently be a host (silicon) atom. In this case, ν = 1 and the lattice is undistorted. Neglecting $\sqrt{\Delta E_x / \Delta E_{\mathrm{Si}}}$ in (4.3.7), we can define ν^{s} as $\sqrt{M_{\mathrm{Si}} / M_x}$.

The average ionization potential $\bar{I}_{ij}$ from (4.3.10) is

$$\bar{I}_{ij} = \frac{\bar{I}\left(\mathrm{Si}_{sp^3}\right) + \bar{I}\left(Me_{d^3s}\right)}{2}, \tag{4.3.12}$$

where $\bar{I}\left(\mathrm{Si}_{sp^3}\right)$ = 9.5 eV [30],

$$\bar{I}\left(\mathrm{Me}_{d^3s}\right) = \frac{1}{4} I_{\mathrm{s}} + \frac{3}{4} I_{\mathrm{d}} \tag{4.3.13}$$

for the d^3s-hybridization of a *d*-atom, and

$$I\left(\mathrm{Me}_{d^2sp}\right) = \frac{1}{2} I_{\mathrm{d}} + \frac{1}{4} I_{\mathrm{s}} + \frac{3}{4} I_{\mathrm{d}} \tag{4.3.14}$$

for the d^3s-hybridization of an impurity atom. In these expressions, I_{s}, I_{d}, and I_{p} are orbital potentials for the respective electronic configurations of the *d*-atom [32].

The average ionization potential I_{kl} in (4.3.11) is calculated as

$$\bar{I}_{kl} = \frac{\bar{I}_{\mathrm{d}} + \bar{I}_{\mathrm{sp}^3}}{2}, \tag{4.3.15}$$

where I_{d} is the ionization potential of the *d*-electron in the respective valent state of the *d*-atom [32]. The promotion energy in (4.3.9) for neutral site im

Table 4.4. Calculated dissolution enthalpies ΔH_s^s (eV) of site *d*-impurities in silicon.

Impurity	$\Delta H_s^s(Me^0)$	$\Delta H_s^s(Me^-)$	Impurity	$\Delta H_s^s(Me^0)$	$\Delta H_s^s(Me^-)$
Sc		6.55	Y		6.57
Ti	3.56	7.28	Zr	6.69	8.09
V	5.41	7.64	Nb	7.68	9.35
Cr	4.04	no bond	Mo	8.75	no bond
Mn	no bond	no bond	Tc	8.83	9.18
Fe	6.31	3.75	Ru	10.03	8.40
Co	5.08	6.86	Rh	no bond	8.23
Ni	5.06		Pd	no bond	

purities are chosen from the energy difference of the valent states for the corresponding electronic configurations [11]. There are no reports of energy data for the valent states of 5*d*-impurities, whose electronic configurations are Me(d^7s^2) and Me(d^8s^1), and so the dissolution enthalpies of interstitial 5*d*-impurities in silicon have not been calculated.

In contrast to interstitial solubility, the calculation of site dissolution enthalpy does not require internal crystal potential, since the interaction with the nearest neighbors in this case is defined by D_0, while U_{cr} is a long-range potential which cannot affect D_0.

The calculated dissolution enthalpies for transition metal impurities in substitutional positions in silicon are given in Table 4.4.

No bonding occurs between silicon atoms and *d*-atoms at $D_0 < 0$, because the energy requirements for the repulsion and promotion are larger than for the interatomic attraction and exchange. This means that forces pushing an atom out of the site dominate over those confining it to the site. The absence of chemical bonding is designated in Table 4.4 as "no bond", so no ΔH data for such impurities are given. It is still unclear how to calculate interstitial dissolution enthalpy of neutral scandium and yttrium in the silicon lattice, because their configurations in the free state are $Sc^0(3d^14s^2)$ and $Y^0(4d^15s^2)$: the outer shells of these Y and Sc atoms are deficient in one electron and incapable of producing tetrahedral bonds. So the ΔH values have been calculated only for their ionized states $Sc^-(3d^34s^1)$ and $Y^-(3d^35s^1)$. The concepts of d^2sp- and d^3s-hybridization appear to be inapplicable to nickel and palladium atoms in substitutional ionized states in silicon because of a large number of paired *d*-electrons (Figure 4.9). The dsp^2-hybridization of valent electrons is applicable but it corresponds to the square geometry, and strong asymmetric lattice distortions around an impurity atom must be taken into account. This, however, is impossible at the present stage of evolution of the

microscopic solubility theory. Because of the lack of data on the energies of the valent states $Cu^0(3d^74s^14p^3)$, $Ag^0(4d^75s^15p^3)$, and $Au^0(5d^76s^16p^3)$, ΔH_s^s values cannot be analyzed theoretically for the site solubility of these impurities in silicon.

4.4 SOLUBILITY OF INTERSTITIAL *f*-ATOMS IN SILICON

The above treatment was extended to the solubility of interstitial *f*-impurities in silicon [33]. Like for *d*-impurities, it is based on the consideration of interaction potentials of an impurity and a host atom. So the model of *f*-atom incorporation into a silicon crystal will be identical to the one depicted in Figure 4.4, except that the restructuring and ionization will refer to the *f*-electron shell.

Therefore, ΔH_s^i values for neutral and ionized interstitial impurities will be defined by the above equations (4.3.1) and (4.3.2). The total energy balance for *d*-impurities involves the promotion energy ΔP which is the energy of transition of the outer *s*-electrons of an impurity atom to the *d*-shell. It is reasonable to suggest that a similar approach will also be valid for *f*-impurities: $Me(f^ns^m) \rightarrow Me(f^{n+m})$. There are no data on the energy values of the centers of mass of terms for various valent configurations of *f*-elements. For this reason, the promotion energy was ignored in the first approximation [33]. The polarization energy and crystal potential are completely identical to those calculated for *d*-impurities and were found to be 4.54 and 5.97 eV, respectively. The repulsion energy can also be calculated by Mulliken's method using (4.3.5).

The model accepted for *f*-states is basically as follows.

(1) The internal tetrahedral crystal field removes the 7-fold degeneracy of the *f*-shell.

(2) The level splitting of *f*-elements at a tetrahedral interstice corresponds to that for an octahedral symmetry. This is due to a stronger effect of the second sphere around an interstitial atom. There is a singlet state A_{2U} and two triplet states T_{2U} and T_{1U} (Figure 4.10).

(3) The outer *s*-electrons of incorporated impurity atoms are promoted by the crystal field to the *f*-shell, i.e., there is the shell restructuring $Me(f^ns^m) \rightarrow Me(f^{n+m})$.

The filling of the states by *f*-electrons occurs according to the Pauli principle and Hund rule. The states and the level filling by electrons are shown

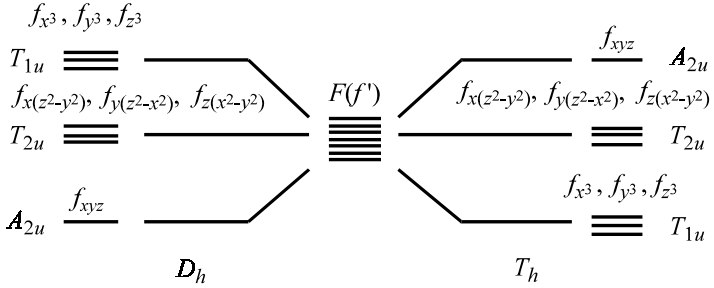

Figure 4.10. Splitting of *f*-levels in a tetrahedral medium.

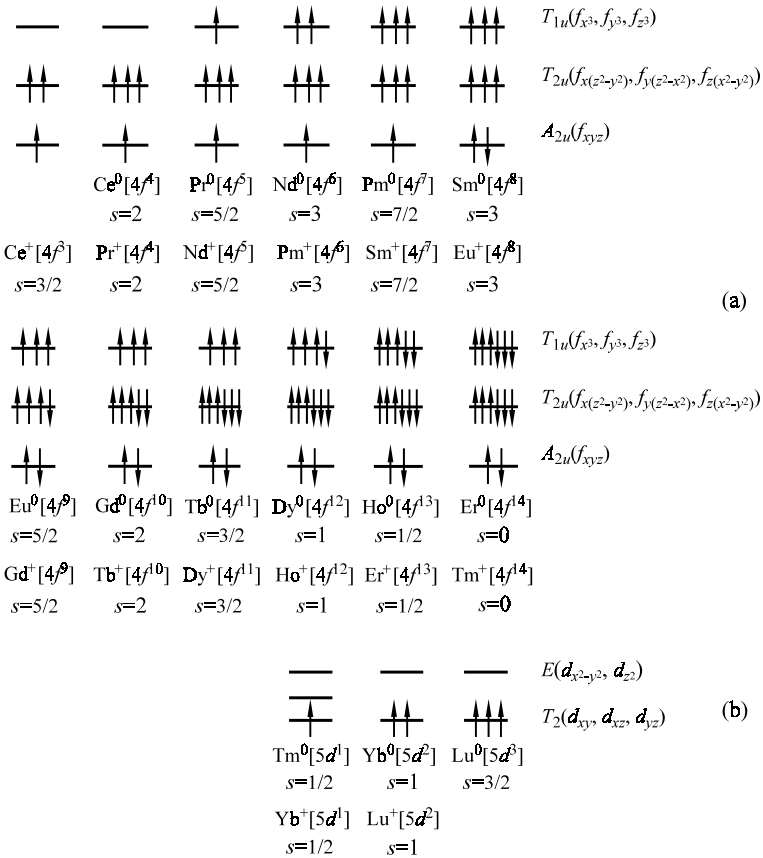

Figure 4.11. Electron filling of splitted *f*-levels at *T*-interstices.

Table 4.5. Dissolution enthalpy of interstitial 4*f*-impurities in silicon.

Configuration	ΔH_{at}, eV	Neutral state		Ionized state	
		U_{rep}, eV	$\Delta H_s(Me^0)$	U_{rep}, eV	ΔH_s
$Ce^0(4f^4)$	4.84	8.30	7.17		
$Ce^+(4f^3)$	4.84			6.98	6.78
$Pr^0(4f^5)$	3.86	10.00	7.89		
$Pr^+(4f^4)$	3.86			9.82	8.59
$Nd^0(4f^6)$	3.40	9.81	7.24		
$Nd^+(4f^5)$	3.40			9.60	7.98
$Pm^0(4f^7)$	2.73	7.74	4.50		
$Pm^+(4f^6)$	2.73			8.79	6.56
$Sm^0(4f^8)$	2.14	7.14	3.31		
$Sm^+(4f^7)$	2.14			6.68	4.14
$Eu^0(4f^9)$	1.83	5.75	1.61		
$Eu^+(4f^8)$	1.83			5.44	2.42
$Cd^0(4f^{10})$	4.15	4.38	2.56		
$Cd^+(4f^9)$	4.15			3.04	2.84
$Tb^0(4f^{11})$	4.03	3.10	1.16		
$Tb^+(4f^{10})$	4.03			3.09	2.46
$Dy^0(4f^{12})$	3.10	2.25	0.62		
$Dy^+(4f^{11})$	3.10			2.29	0.81
$Ho^0(4f^{13})$	3.04	1.71	1.71		
$Ho^+(4f^{12})$	3.04			1.74	0.29
$Er^0(4f^{14})$	2.88	1.20	1.89		
$Er^+(4f^{13})$	2.88			1.30	0.23
$Tm^0(4f^{14}3d^2)$	2.56	4.79	1.38		
$Tm^+(4f^{14}5d^1)$	2.56				
$Yb^0(4f^{14}5d^2)$	1.57	9.20	4.80		
$Yb^+(4f^{14}5d^1)$	1.57			1.44	1.26
$Lu^0(4f^{14}5d^3)$	4.43	11.05	5.51		
$Lu^+(4f^{14}5d^2)$	4.43			2.56	1.91

in Figure 4.11*a*. It should be emphasized that the interstitial *f*-shell of the impurities $Tm(4f^{13}5d^06s^2)$, $Yb(4f^{14}5d^06s^2)$, and $Lu(4f^{14}5d^16s^2)$ appear to be completely filled, whereas the other electrons have to occupy the *d*-shell under the action of the crystal field. In other words, these impurities undergo the following restructuring:

$$Tm(4f^{13}5d^06s^2) \rightarrow Tm(4f^{14}5d^1)$$

$$Yb(4f^{14}5d^06s^2) \rightarrow Yb(4f^{14}5d^2)$$

$$\mathrm{Lu}(4f^{14}5d^{1}6s^{2}) \rightarrow \mathrm{Lu}(4f^{14}5d^{3}).$$

The splitting of levels and their filling by electrons for these impurities are similar to those for *d*-impurities at the silicon *T*-interstice (Figure 4.11b).

The electron states of *f*-atoms at hexagonal interstices were not discussed in [33], because they are more important for the calculation of migration energies than solubility. The overlap integrals for the first two nearest neighbors were calculated by taking the states of *f*-impurities at the *T*-interstice into account.

With the present state of the art, it appears impossible to allow for the crystal lattice distortions for *f*-elements, because the approach discussed above for *d*-impurities implies the use of a model impurity. We, however, failed to choose such an impurity because of the lack of reliable experimental data. Consequently, the coefficient ν in (4.3.5) was taken to be unity. The coefficient β was chosen from the best agreement between the experimental and theoretical values of dissolution enthalpy. Since there are no experimental enthalpy data for *f*-impurities, the coefficient β was intuitively taken to be 0.05. This provides the dissolution enthalpy values of about several electron volts. The coefficient β is quite likely to be refined by further experiments.

The coefficient β for Tm, Yb, and Lu impurities with valent *d*-electrons was taken to be 0.5 for a neutral state and 0.154 for an ionized state, as in the case of *d*-impurities. The calculations of the dissolution enthalpy of neutral and ionized 4*f*-impurities at silicon interstices are given in Table 4.5. The dominant state is that with the lowest enthalpy. It can be concluded, therefore, that interstitial 4*f*-elements in silicon must be mostly neutral, except for Nd, Ho, Er, Yb, and Lu which must be ionized.

4.5 ON SOLUBILITY THEORY FOR SEMICONDUCTOR COMPOUNDS

The above modification of Weisser's theory can be extended to more complex semiconductors, in particular, to $A^{III}B^{V}$ compounds crystallizing in the sphalerite-type lattice. These compounds have some specific features to be taken into consideration. First, it is the existence of two sublattices and, hence, of two types of *T*-interstices. One interstice (T_A) is first surrounded by the nearest A^{III} atoms and then by B^{V} atoms. The other interstice (T_B) is first surrounded by the nearest B^{V} atoms and then by A^{III} atoms. A hexagonal interstice is made up of both A^{III} and B^{V} atoms. This diversity considerably increases the amount of calculations.

A substitutional position requires the treatment of two possible variants—A^{III} and B^{V}. Of importance is the calculation of internal crystal potentials at T_A, $T_{B,}$ and H-interstices of $A^{III}B^{V}$ compounds. The knowledge of these potentials is necessary because they enter expressions like (4.3.2). So far, there have been no calculations of U_{cr} for binary semiconductors.

The calculation of overlap integrals must consider the fact that the bonding between A and B atoms in binary semiconductors is not purely covalent, as in silicon. It can be regarded as a mixed bonding, namely, as covalence involving some ionicity. One way of describing this chemical bonding is by representing the interaction between the nearest A^{III} and B^{V} neighbors as a combination of the sp^3-hybrid wave functions of these atoms.

The calculation of site solubility requires the knowledge of formation energies of vacancies V_A and V_B. The available calculations are contradictory and agree poorly with one another. For this reason, a further development of models permitting a correct calculation of vacancy formation energies is a necessary prerequisite for extending the modified impurity solubility theory to semiconductor compounds. Their specific features are an obstacle to a rigorous treatment of dissolution enthalpies in binary semiconductors.

Nevertheless, a qualitative assessment of ΔH_s^s with reference to group-IV amphoteric impurities in gallium arsenide was made in [34], using the simplified expression (4.3.11). The tabulated values of ΔH_{at} were borrowed from [31], and the vacancy formation energies for gallium and arsenic were taken to be 1.8 and 2.6 eV, respectively. The other terms in (4.3.8) were found from (4.3.9) and (4.3.10). It is noteworthy that all four valent electrons in a group-IV impurity are involved in the production of tetrahedral bonds, which means that all elements are located on the binding electron orbitals. So, the second and third sums in (4.3.9) were neglected [31] because they reflect the interaction of electrons located, according to molecular orbital theory, on loosened orbitals which are free from electrons in this case. The chemical bond between gallium and arsenic atoms was considered as being sp^3-hybridized due to the partial transition of one electron of the As($4s^2 4p^3$) atom to the Ga($4s^2 4p^1$) atom, which, therefore, acquire the electronic configurations As^+($4s^1 4p^3$) and Ga($4s^1 4p^3$). The overlap integrals Si_{ij} in (3.2.1) of the $S(sp^3, sp^3)$ type were reduced to diatomic ones based on Slater's wave functions. The summation in (4.3.10) was performed over all outer electrons of the impurity atom and its nearest neighbors.

The promotion energy ΔP was taken to be the energy of the s-electron transition to the p-orbital, and the differences between the tabulated orbital ionization potentials for the s^2p^2 and s^1p^3 states were equated.

Table 4.6. Calculated dissolution enthalpies (eV) of A^{IV} impurities in GaAs.

Impurity	Substitution		Incorporation	
	Ga-sublattice	As-sublattice	Ga-interstice	As-interstice
C	9.09	9.04	12.20	12.13
Si	5.77	6.22	11.40	12.72
Ge	5.97	6.43	12.07	12.73
Sn	4.88	5.07	10.61	12.40

The quantity E_i in (4.3.9) is ionic interaction energy having the physical meaning of lattice polarization energy. Its calculation by the Mott–Littleton–Weisser method was described in Section 4.1 [see expression (4.1.14)]. The calculations of ΔH_s^s are presented in Table 4.6.

The dissolution enthalpy of interstitial group-IV impurities in both GaAs sublattices was calculated from the simplified expression

$$\Delta H_s^i = \Delta H_{at} + \frac{1}{2}\sum_{ij} Y_{ij} + \Delta P - E_i . \qquad (4.5.1)$$

No chemical bond is formed between an impurity atom and its lattice neighbors. The electrons of the impurity atom are located on loosening orbitals; so, only the repulsion energy was calculated with (4.3.5). The value of ΣK_{mn} was neglected because it is small.

The promotion and polarization energies were found as described above, with the s^1p^3 state taken as the excited state of the impurity atom. The calculated values of ΔH_s^i are also given in Table 4.6.

4.6 COMPARISON WITH EXPERIMENTAL DATA

In spite of the qualitative character of the data presented in Tables 4.2 and 4.4, one can conclude that most d-atoms dissolve at silicon interstices rather than at its sites. Only titanium atoms can occupy sites. The ΔH_s^s values for other *d*-atoms have proved to be too large. Indeed, it is known from experimental data [15] that practically all impurities dissolve at interstices in the silicon lattice.

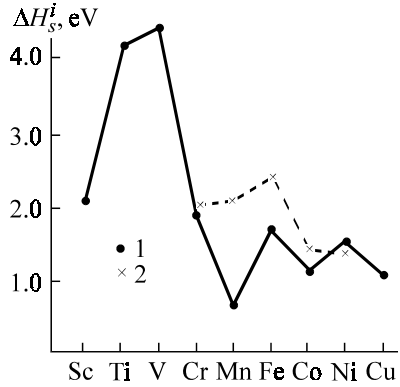

Figure 4.12. Dissolution enthalpy of 3*d*-impurities versus the serial number of the *d*-element: 1 – theory (for high probability charge states); 2 – experiment.

A comparison with absolute experimental values of ΔH_s^s is hard to make for several reasons. The main reason is the absence of experimental temperature dependences of solubility or of the distribution coefficient, from which experimental ΔH_s^s values are found.

But the calculations of ΔH_s^s provided the data of Table 4.3, with the indication of the most probable charge states of interstitial *d*-atoms in silicon. The charge states confirmed by experiments are also indicated there. One can see a good agreement between the experimental and theoretical values for the charge states and electronic configurations of *d*-impurities. A review of many publications on this issue can be found in [16]. Since 4*d*- and 5*d*-impurities are still poorly understood, the theoretical results of Tables 4.2 and 4.3 concerning these impurities can be regarded only as hypothetical.

Indeed, one cannot expect a good agreement with experiment because the modified solubility theory is only qualitative. Still, if we plot ΔH_s^i as a function of the shell filling degree of an impurity atom, as is done in Figure 4.12, the behavior of the experimental and theoretical curves will be identical, and this is a good indication of the model validity. Figure 4.12 shows another specific feature: the solubility curves have bendings at $n = 4$ and 6. Such bendings are typical of many transition metal characteristics—melting and evaporation temperatures, thermal expansion coefficients, compressibility, etc. [35], and this fact also supports the general validity of the dissolution model used. Of much importance for the further development of this model is the change from energy characteristics to concentration dependences. To do this, we need to describe reliably the entropy contribution to solubility, which was totally ignored by the modified Weisser theory.

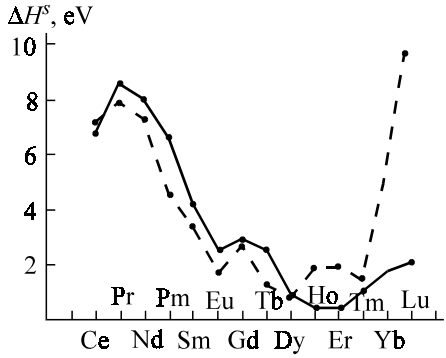

Figure 4.13. Dissolution enthalpy of *f*-impurities in silicon: solid line – ionized impurity; dashed line – neutral impurity.

To check the theoretical conclusions of solubilities of *f*-impurities in silicon is a more difficult task than for *d*-impurities because of the complete absence of experimental ΔH data for *f*-impurities. We would like only to note that the enthalpy dependence on the serial number of the element (Figure 4.13) has characteristic bendings in the middle of the period, namely, at Gd($4f^{10}$) and Eu($4f^{9}$), and an abrupt rise of dissolution enthalpy at the end of the period at Lu($4f^{14}5d^{2}$). These dependences are consistent with changes in the physicochemical properties of 4*f*-elements, such as melting and evaporation temperatures, ionization potentials, and others. The dependences shown in Figure 4.13 also have a bending at Eu and Gd.

The modified Weisser theory makes the treatment of semiconductor compounds all the more qualitative. Indeed, it follows from the minimum ΔH values in Table 4.6 that group-IV impurities in GaAs are to occupy mostly the lattice sites. Besides, the close enthalpy values for the A^{III} and B^{V} lattice sites is a theoretical indication of the amphoteric nature of group-IV atoms in GaAs. These two facts are well known from numerous experimental studies. However, one should pay attention to the excessively large values of ΔH_s^s. If they were correct, these impurities would have a low solubility. On the contrary, the solubility values were shown experimentally to be quite high. This contradiction is due to the neglect of the crystal potential and lattice relaxation effect which are impossible to allow for at present. Were this possible, the absolute values of ΔH_s^s would be lower and close for both sublattices. The conclusion about the amphoteric nature of A^{IV} impurities in GaAs would then be valid.

4.7 QUANTUM CHEMICAL CALCULATION OF DISSOLUTION ENTHALPY

4.7.1 Formulation of the quantum chemical problem

The modified Weisser theory represents a simplified phenomenological approach using the pair potential approximation, which is close, to some extent, to the short-range interaction of an impurity atom with the host neighbors. As a consequence, this theory provides reasonable results for *d*- and *f*-impurities possessing a short-range potential extending to a limited number of coordination spheres (Figure 4.2). It is, however, inapplicable to the solubility treatment of hydrogen-like impurities, whose potential extends much farther from the impurity center.

The delocalized perturbation potential affects such a large crystal region that it cannot be calculated by conventional quantum chemical methods. This problem is complicated by the fact that the energies to be found (ΔH) are to have very low values (< 1 eV) comparable with the error limit for most quantum mechanical methods. So we need to find an approach which could satisfy, at least, two conditions: it must consider the fact that an impurity is actually built into an infinite crystal, and the method accuracy must be high enough to allow calculation of small absolute ΔH values. Moreover, the quantum chemical method must account for the lattice polarization energy released in the displacement of the host electron density by the impurity center, and do this much better than the Weisser theory does.

Following Volkov [36], let us consider the lattice polarization within two coordination spheres of atoms with a high electronegativity, shown schematically in Figure 4.14. For systems with a rigid covalent bonding, this process can naturally be reduced to a displacement of the “center of mass” of electron bond bridges toward the polarizing center, resulting in a consistent redistribution of electron density in the crystal. The total energy of this process can be represented as a sum of terms responsible for the individual bonds.

Therefore, the calculation of polarization for the whole crystal will require an analysis of the tetrahedral crystal structure which can be considered in two ways. The most commonly used concept is that a crystal is simulated by an array of atoms arranged in a special way. For tetrahedral structures (point group symmetry T_d), one atom can be conveniently selected as the central one, the others located on spherical surfaces surrounding the center. The arrangement of atoms on such a surface, termed as a coordination sphere, obeys the symmetry rules and can be found by means of projection

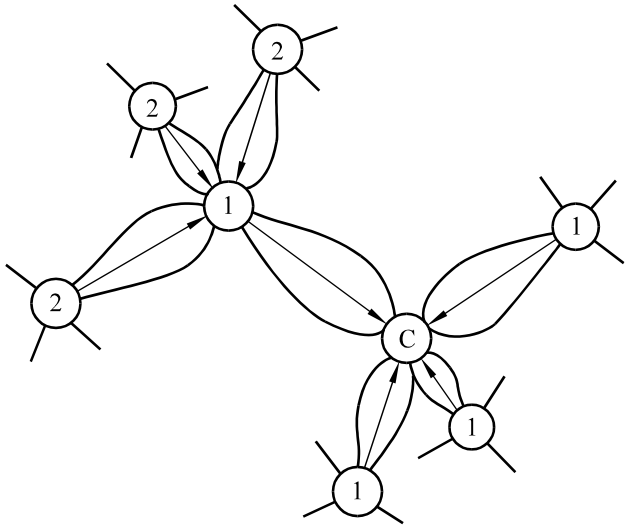

Figure 4.14. Schematic diagram of tetrahedral crystal lattice polarization by the C atom with a higher electrical neutrality within two coordination spheres: 1, 2 – coordination sphere numbers; arrows indicate the direction of electron density displacement of a covalent bond.

operators in the group theory [37], if the coordinates of at least one atom on the sphere are known. The sphere radius R_i is defined by the sphere number i and can be found for the tetrahedral structures of interest as follows:

$$\text{for odd } i: \qquad R_i = d_0\sqrt{\frac{4i-1}{3}}, \tag{4.7.1}$$

$$\text{for even } i: \qquad R_i = d_0\sqrt{\frac{4i}{3}}, \tag{4.7.2}$$

where d_0 is the shortest interatomic distance corresponding to the first coordination sphere radius. Generally, this approach describes fairly well the crystal lattice geometry but leaves aside the problem of the nature and orientation of chemical bonds.

A better concept of the lattice structure for covalent crystals is that based on the graph theory which can be called topological. In this theory, the focus is on the rigid orientation of the interatomic bond. The crystal is simulated

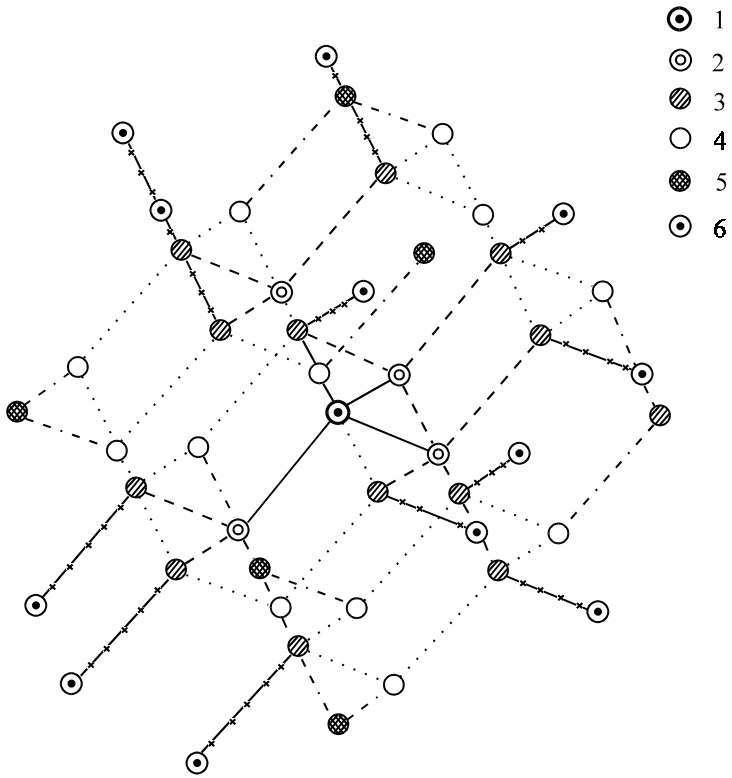

Figure 4.15. A fragment of the tetrahedral structure of a diamond-like semiconductor, including the central atom and atoms of the nearest five coordination spheres (CS): 1 – central atom; 2 – 1st CS atoms; 3 – 2nd CS atoms; 4 – 3rd CS atoms; 5 – 4th CS atoms; 6 – 5th CS atoms.

by a topological $N \times N$ matrix D, symmetrical relative to the principal diagonal (here N is the number of atoms in the system under study). The matrix elements d_{ij} are the topological distances between the atoms, which may be chosen to be different if the choice reflects the additive pattern of localized bonds. Obviously, a crystal is then represented as a system of chemical bonds connecting the atoms, rather than as a system of individual atoms (Figure 4.15). If the minimum bond length d_0 is taken to be unity, d_{ij} can be represented as an integral number corresponding to the minimum number of bonds between the atoms i and j [38, 39]. This approach can sometimes be used for a qualitative evaluation of the system stability by minimizing the

Table 4.7. Numbers of topological spheres $I = (ij)^*$.

CS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
i	0	1	2	2	3	3	4	5	5	6*a*	11*a*	6*b*	6*b*	7*a*	7*b*
j	1	2	3	5	4	6*a*	7*a*	6*b*	8	7*b*	9	7*a*	11	10	12*a*

* *i* and *j* are coordination sphere numbers designating the bond beginning and end; 0 is the central atom; CS – coordination sphere.

half-sum of topological matrix elements or Wiener's number [38–40]. Besides, d_{ij} can be taken to be unity for directly bonded atoms *i* and *j* and zero for the other cases.

For further analysis, it is convenient to unite the atoms of one coordination sphere and transform the matrix in such a way that its dimension would correspond to the total number of coordination spheres N_0. Then the topological matrix element d_{ij} will be expressed as the total number of bonds N_{ij} between the *i*- and *j*-spheres. A fragment of the matrix for the diamond-type crystal is shown in Figure 4.7. It follows from this picture that the atoms of the same coordination sphere may appear to be different in terms of the topological environment. For example, of the 24 atoms belonging to the 6th coordination sphere, 12 are bonded to the 3rd, 7th, and 9th spheres while the other 12 atoms to the 4th, 7th, and 11th spheres. Therefore, it is reasonable to group the atoms of such spheres by the character of binding in the crystal lattice. In the illustration above, the 6th sphere is subdivided into the 6a and 6b spheres. A similar situation is true for the spheres numbered 7, 12, 13, 14, 15, etc.

Since all atoms of a crystal lattice are grouped on the respective coordination spheres, all localized electron pairs of the (*ij*)-type for atoms of the *i*- and *j*-spheres, directly bonded to one another, represent a symmetric array which can be considered as an independent topological (bond) sphere. For example, the topological sphere (23) includes 12 bonds of atoms of the 2nd and 3rd coordination spheres, indicated by a dashed line in Figure 4.15. The double subscript (*ij*), in which *i* is the beginning and *j* the end of the bond, can be replaced by the same index *I*, chosen with the increasing *i*-numbers (Table 4.7) of coordination spheres and the *j*-numbers for spheres with the same *i*.

In this approach, a crystal can be regarded as lying at the base of topological bonds, whose interception points are the lattice atoms. This structural representation of a crystal lattice allows the construction of an inverse topological matrix *N*. The matrix element n_{ijkl} is represented by the total number

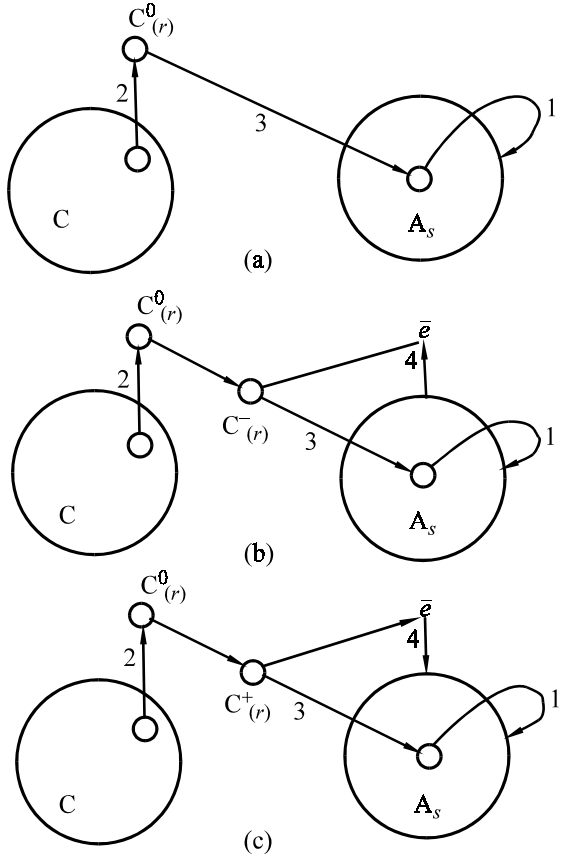

Figure 4.16. Schematic dissolution process of an impurity (*C*) in a semiconductor (*A*): (a) – neutral isovalent impurity; (b) – ionized acceptor impurity; (c) – ionized donor impurity.

of atoms bound simultaneously to the topological spheres (*ij*) and (*kl*). Therefore, the crystal structure can be regarded as the matrix *D* at the base of coordination spheres (Figure 4.15) and as the matrix *N* at the base of topological (bond) spheres. Each of its elements is the number of coordination sphere atoms at the interception of respective bonds (Table 4.7).

Both representations were used in [36] to calculate electron density distributions among lattice atoms in diamond-type tetrahedral semiconductors.

4.7.2 The dissolution model for a substitutional impurity

Quantum chemical theory, like Weisser's theory, treats dissolution enthalpy as the sum of energy contributions to the incorporation of impurity atoms into the crystal solvent to produce a solid solution.

In [36], the initial state was assumed to be a perfect defect-free crystal, and the final state was the impurity solid solution in the crystalline host. It was considered that the solution was infinitely dilute and the impurity was uniformly distributed throughout the crystal. The concentration of impurity atoms was taken to be so small that their interaction with one another could be neglected. In fact, it was an infinitely dilute crystal.

Microscopically, the real process of impurity incorporation into the crystal lattice can be conveniently represented as a sum of several consecutive intermediate states (Figure 4.16) similar to those suggested by Weisser. They can be treated as the system transitions from one virtual state to another, with the total dissolution enthalpy defined by the sum of energy contributions at the individual stages. To calculate these energy contributions (per atom), it is reasonable to single out the following processes leading to the formation of a substitutional impurity center in the host crystal:

– the formation of a vacant site in the host lattice;
– atomization of impurity substance;
– the binding of the impurity to the vacant site;
– ionization of the impurity center (for hydrogen-like impurities).

4.7.3 Quantum chemical calculations of impurity solubility

The details of quantum chemical calculations of impurity solubility in semiconductors at all stages of the process can be found in [1]. Their description would take too much space in this book, so we present only the final results summarized in Table 4.8.

The numerical discrepancy between the calculated and experimental values of dissolution enthalpy ΔH are due to the experimental error (nonequilibrium conditions) and simplifications accepted by the theory. Among the latter are insufficient allowance for lattice relaxation, especially for boron impurity in silicon, and the neglect of the temperature dependence of enthalpy and interimpurity interactions. Nevertheless, there is a satisfactory coincidence in the order of theoretical and experimental ΔH values in all cases, except for boron in silicon. The calculations suggest slightly higher enthalpy values in silicon than in germanium.

Table 4.8. Calculations of dissolution enthalpy (eV) of III-V substitutional impurities in silicon and germanium [43].

Impurity	Silicon		Germanium	
	Calculation	Experiment	Calculation	Experiment
B	2.03	0.43–0.73	1.90	–
Al	0.59	0.43–0.66	0.12	0.12
Ga	0.67	0.46	0.50	0.12
In	0.88	–	0.33	0.20–0.85
C	2.37	2.30	2.62	–
Sn	0.56	0.20	0.33	0.03
N	1.64	–	1.61	–
P	0.73	0.50–0.70	0.40	–
As	0.51	0.47	0.11	–
Sb	0.55	0.24–0.58	0.11	–

Besides, it follows from the quantum chemical calculations that ΔH values decrease with increasing serial number of the impurity element within the same group of elements. This tendency has been observed experimentally for group-IV and group-V elements in silicon.

4.7.4 Perspectives of the quantum chemical method

The improvement of the quantum chemical method for the calculation of impurity dissolution enthalpy in semiconductors primarily depends on refining the atomic parameter system of the CNDO method [41]. This problem will require a detailed account of electron interaction integrals, internuclear repulsion potentials, and bonding parameters of impurity and host atoms.

The authors of [42] took account of the fact that the repulsion energy of atomic skeletons at small internuclear distances R must tend to $1/R$ rather than to the one-center integral of electron interaction. They suggested using additional parameters in the calculation of skeleton interaction energy, which will no doubt complicate the calculation of dissolution enthalpy but will not change quantitatively the general calculation scheme. On the other hand, this may give the theory a more profound physical sense. The idea of re-evaluation of the repulsion energy of atomic skeletons has become quite common in the MINDO and MNDO methods [43, 44, 45]. This approach to the interatomic repulsion calculation may contribute to the solution of the problem of a dimensionality factor for defects considerably distorting the crystal lattice.

Table 4.9. Bonding parameters β_{AB} for $A^{III}B^{V}$ compounds, found from atomization heat [36].

Compounds	AlP	AlAs	AlSb	GaP	GaAs	GaSb	InP	InAs	InSb
$-\beta_{AB}$, eV	3.07	2.66	2.62	3.16	2.79	2.79	3.05	2.67	2.66

The problem of the method parametrization becomes more acute when the quantum chemical method is extended to the calculation of impurity dissolution enthalpies in $A^{III}B^{V}$ compounds. In the first approximation, the bonding parameters β_{AB} for A^{III} and B^{V} atoms can be found from experimental data on atomization heat ΔH_{AB} of $A^{III}B^{V}$ compounds. Enthalpy variation in a decomposition reaction producing simple substances in an elemental gaseous state was considered in [43]:

$$A^{III}B^{V}{}_{gas} = A^{III}{}_{gas} + B^{V}{}_{gas}. \tag{4.7.3}$$

The value of $\Delta H_A{}^{at}$ can be found from tabulated thermodynamic data as

$$\Delta H_{AB} = \Delta H_A + \Delta H_B - \Delta H_{AB}{}^{f}, \tag{4.7.4}$$

where $\Delta H_{AB}{}^{f}$ is the enthalpy of $A^{III}B^{V}$ formation from simple substances in their standard state.

A rather cumbersome procedure of finding the parameters β_{AB} from atomization heat ΔH_{AB} is described in [36]. We give the final results of this work in Table 4.9.

The estimations of dissolution enthalpy for Al, Ga, P, and As in GaAs and GaP, using the bonding parameters from Table 4.9 [36], have shown that this method provides reasonable results ($\Delta H_s{}^{s} \leq 0.2$ eV) and its further improvement is very desirable.

REFERENCES

4.1. K. Weisser, *J. Phys. Chem. Sol.* **15**, No. 2: 118–126 (1958).
4.2. K.H. Benneman, *Phys. Rev.***137**: 1497–1514 (1965).
4.3. G.M. De Munary, L. Gabba, F. Guissiamo *et al.*, *Phys. Stat. Sol., A* **34**: 95–105 (1976).
4.4. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushchikh defektov v poluprovodnikakh* (Thermodynamics and Kinetics of Interacting Defects in Semiconductors). Moscow: Nauka, 352 p. (1997) (in Russian).

4.5. V.I. Fistul, *Amfoternye primesi v poluprovodnikakh* (Amphoteric Impurities in Semiconductors). Moscow: Metallurgia, 240 p. (1992) (in Russian).
4.6. K. Weisser, *J. Phys. Chem. Soc.* **17**: 149–161 (1960).
4.7. K. Weisser, *Phys. Rev.* **126**, No. 4: 1427–1436 (1962).
4.8. R. Hultgreen, P.D. Desai, D.T. Hawkins *et al.*, *Selected Values of Thermodynamic Properties of Elements*. Ohio: Amer. Soc. of Metals, 605 p. (1973).
4.9. N.F. Mott, M.J. Littleton, *Trans. Faraday Soc.* **34**, No. 2: 485–499 (1938).
4.10. Sh.A. Vakhidov, ed., *Radiatsionno-aktiviruemye protsessy v kremnii* (Radiation-Activated Processes in Silicon). Tashkent: FAN, 165 p. (1977) (in Russian).
4.11. G. Leibfried, *Hand. Buch der Physik* **7**, No. 1: 104–324 (1955).
4.12. M.F. Millea, *J. Phys. Chem. Sol.* **27**: 315–325 (1966).
4.13. R.R. Hasiguti, *J. Phys. Sol. Jap.* **21**: 1927–1931 (1966).
4.14. L. Pauling, *The Nature of Chemical Bond in Molecules and Crystals*. New York: Cornell Univ. Press, 644 p. (1960).
4.15. A. Zunger, U. Lindefelt, *Phys. Rev.* **26**: 5989–5992 (1982).
4.16. E.M. Omelyanovsky, V.I. Fistul, *Transition Metal Impurities in Semiconductors*. Bristol–Boston: Adam Hilger Ltd., 243 p. (1986).
4.17. A.Sh. Makhmudov, B.L. Oksengendler, M.C. Yunusov, *FTP* **10**, No. 1: 271–274 (1976) (in Russian).
4.18. R. Mulliken, *J. Phys. Chem.* **56**: 295–311 (1952).
4.19. Yu.S. Dementyev, V.A. Fedorov, N.I. Bletskan, *Izvestiya AN SSSR, ser. Neorg. mat.* **16**: 1164–1167 (1980) (in Russian).
4.20. J.C. Philips, *Rev. Mod. Phys.* **42**: 317–356 (1970).
4.21. V.I. Fistul, V.A. Shmugurov, *FTP* **23**: 677–683 (19890 (in Russian).
4.22. N.B. Bersuker, *Elektronnoe stroenie i svoistva koordinatsionnykh soedinenii* (The Electronic Structure and Properties of Coordination Compounds). Leningrad: Khimiya, 350 p. (1976) (in Russian).
4.23. J.W. Allen, *Deep Centers in Semiconductors*, ed. S.T. Pantelides). New York: Gordon & Breach Publ., 755 p. (1986).
4.24. O.P. Charkin, *Vestnik LGU, ser. Fiz. khim.*, No. 22: 92–102 (1969) (in Russian).
4.25. G.D. Watkins, J.W. Corbett, *Disc. Far. Soc.* **31**: 86–95 (1961).
4.26. G.W. Ludwig, H.H. Woodbury, *Phys. Rev. Lett.*, No. 5: 98–103 (1960).
4.27. A.B. Roitzin, L.A. Firshtein, *Teor. i eksper. khimiya* **2**: 747–766 (1966) (in Russian).
4.28. K.H. Benneman, *Phys. Rev.* **137**: 1497–1514 (1965).
4.29. D. Shaw, ed., *Atomic Diffusion in Semiconductors*. London–New York: Plenum Press, 682 p. (1973).
4.30. S.S. Batzanov, R.A. Zvyagina, *Integraly perekryvaniya in problema effektivnykh zaryadov* (Overlap Integrals and Effective Charges). Novosibirsk: Nauka. 386 p. (1966) (in Russian).
4.31. K. Hellwey, O. Madelung, eds., *Numerical Data and Functional Relationship in Science and Technology, New series* **17**. Berlin: Springer–Verlag, 652 p. (1984).
4.32. O.P. Charkin, M.E. Dyatkina, *Stroenie molekul i kvantovaya khimiya* (Molecular Structure and Quantum Chemistry). Kiev: Naukova dumka, 275 p. (1970) (in Russian).

4.33. V.I. Fistul, V.A. Shmugurov, *FTP* **27**: 1910–1917 (1993) (in Russian).
4.34. V.I. Fistul, V.A. Shnugurov, *FTP* **24**: 1038–1041 (1990) (in Russian).
4.35. S. Sugano, Y. Tanabe, Y. Kamimura, *Multiplets of Transitional Ions in Crystals*. New York: Acad. Press, 360 p. (1970).
4.36. D.A. Volkov, *Kvantovokhimicheski metod pascheta entalpii rastvoreniya primesei III-V grupp v kremnii i germanii* (Quantum Chemical Calculation of Dissolution Enthalpy of III-V Impurities in Si and Ge). Abstr. diss. Moscow: MITKhT, 20 p. (1994) (in Russian).
4.37. J.C. Slater, Electronic Structure of Molecules. New York–London: McGraw–Hill Book Co, 587 p. (1963).
4.38. D. Bonchev, O. Mekenyan, H.G. Fritsche, *Phys. St. Sol., A* **55**, No. 1: 181–187 (1979).
4.39. O. Mekenyan, D. Bonchev, H.G. Fritsche, *Phys. St. Sol., A* **56**, No. 2: 607–614 (1979).
4.40. D.A. Volkov, V.I. Fistul, *FTP* **24**: 475–478 (1990) (in Russian).
4.41. J.M. Sichel, M.A. Whitehead, *Theor. Chim. Acta.* **7**, No. 1: 32–40 (1967).
4.42. R.J. Boyd, M.A. Whitehead, *J. Chem. Soc.* **73**, No. 1: 73–77 (1972).
4.43. V.A. Gubanov, V.P. Zhukov, A.O. Litinsky, Moscow: Nauka, 219 p. (1976) (in Russian).
4.44. Yu.A. Ustynyuk, ed., *Kvantovokhimicheskie medody rascheta molekul* (Quantum Chemical Calculation of Molecules). Moscow: Khimiya, 256 p. (1980).
4.45. A.A. Voityuk, *JSKh* **29**, No. 1: 138–160 (1988) (in Russian).

Chapter 5

Impurity Interactions in Semiconductors

5.1 TYPES OF IMPURITY INTERACTIONS

It follows from the general propositions of phase equilibrium thermodynamics that the concentration of a dissolved impurity is defined from the equality of chemical potentials. One way of macroscopic thermodynamic analysis of solubility C_s is to find particular values of μ_l and μ_s (see Section 1.2). It is simple to do this if both phases are ideal solutions. This assumption was used to obtain the ratios in (1.1.18) and (1.1.19). The principal feature of an ideal solution is the absence of chemical interactions of impurity atoms with one another and with other point defects. In this case, every subsystem consisting of one type of point defects has a partial chemical potential

$$\mu_i = g_i^0 + kT \ln C_i . \qquad (5.1.1)$$

Here, C_i is the concentration of *i*-defects, T is temperature expressed in energy units, and g_i^0 is Gibbs free energy necessary for the incorporation of a single defect into a pure crystal. The total chemical potential of the solid phase represents just an additive value (1.2.16).

The description of impurity solubility in a crystal requires knowledge of defect types "populating" the crystal and the assurance that the defect solutions in the crystal are ideal. Of course, we do not mean just any defects,

but only those whose concentrations are more or less close to the sought for concentration of the doping impurity. Such defects are normally unknown *a priori* in a theoretical treatment. Besides, if a defect has a very large value of g_i^0, its contribution to μ_s may be appreciable even at a low concentration. Both facts make it necessary to take into account as many defects as possible in a thermodynamic analysis of solubility C_s. Consequently, equilibrium is established in a heterogeneous crystal when the chemical potentials of its components in the different phases are identical. These conditions are simple if there are no interactions between the system atoms.

However, a complete absence of interactions is too strong a restriction for the systems under study. It was shown above that impurity ionization shifts the equilibrium processes and may affect the solubility of elements. To take these processes into account, it is necessary to supplement the chemical potential of an atom with that of an electron. This is a consequence of the general change in free energy, both total and partial, due to the interaction of crystal defects.

In Section 1.1, we subdivided all impurity interactions into two groups—external and internal interactions. The former, reflecting the effects of the ambient phases, were discussed in the previous chapters. In this chapter, the focus will be on internal interactions which change the free energy of a crystal representing a closed thermodynamic system. These interactions can be classified as follows.

– *Statistical interactions* are associated with the distribution of structural elements over the crystal lattice positions. They largely affect the configuration entropy of the system.

– *Charge interactions* are electromagnetic and responsible for the fulfillment of the charge conservation law. They, however, do not produce excess potential energy of the crystal.

– *Potential interactions* are a combination of various interactions changing free energy owing to the rise of crystal potential energy. These are all kinds of interactions, in which crystal structural elements do not form quasimolecules (associates) in crystal positions.

– *Associative* (complexation) *interactions* give rise to associates or complexes, i.e., new structural elements of the lattice possessing quasimolecular properties and occupying crystal positions as an entity.

The diversity of impurity interactions makes us consider this problem in detail, as was done in [1].

5.2 STATISTICAL INTERACTION

This kind of interaction is associated with the arrangement of host structural elements and foreign objects (impurities) in crystal lattice positions. The composition of structural elements and the symmetry of objects determine the number of arrangement patterns and, thereby, the *configuration entropy* described by the well-known Boltzmann formula

$$S_{\text{conf}} = k \ln W , \tag{5.2.1}$$

where W is the thermodynamic probability of the system, or the number of arrangement patterns of structural elements and objects which implement its macroscopic thermodynamic state.

An ideal crystal can be produced in just one way—by arranging all atoms in their respective site positions in the lattice, with $W = 1$ and $S_{\text{conf}} = 0$. But such a system is unstable. The tendency for a closed system to increase its entropy leads to disordering processes producing defects.

The origin of a statistical interaction is essentially due to the fact that two objects or structural elements cannot occupy the same position in a crystal. Therefore, they must be arranged in a certain way. The number of arrangement patterns makes its own contribution to the system entropy.

Strictly, Gibbs total energy of a crystal can be described as

$$G = N\mu\left(P, T\right) \tag{5.2.2}$$

only for a one-component system. But it was established in Section 1.1 that a crystal is a multiphase and multi-component system.

If all kinds of interaction, except for the statistical one, are neglected, the Gibbs partial free energy of individual elements and objects can be used to calculate the total Gibbs energy per unit crystal volume:

$$G = \sum_i g_i N_i - TS_{\text{conf}} , \tag{5.2.3}$$

where g_i are respective partial free energies and N_i are concentrations of elements and objects.

Formula (5.2.3) contains partial free energies instead of chemical potentials for the following reasons. These quantities are identical for atoms. For vacancies, the use of chemical potential is incorrect, because a medium consisting of vacancies only is vacuum. There is no vacancy outside the crystal. On the other hand, the contribution of a vacancy to free energy is unques-

tionable. Almost the same is true of complexes (quasimolecules) consisting of several structural elements: of atoms only or of atoms and vacancies. A complex exists only within a crystal, and its free energy is the sum of atomic chemical potentials, the interaction energies of the structural elements of a complex, and the energy of interaction between a complex and the host lattice. In this case, one cannot speak about the chemical potential of a complex, because there is no one-component substance consisting only of complexes, whose free energy could be expressed by formula (5.2.2).

Partial free energy describes interactions occurring within objects and chemical potentials describe interactions of atoms. The latter also includes the heat components of entropy of atoms and complexes and can be expressed as

$$g_i = h_i - TS_{i\mathrm{T}}, \tag{5.2.4}$$

where h_i is partial enthalpy, including internal and interaction energies, and $S_{i\mathrm{T}}$ is oscillation and thermal entropy.

The derivative of total free energy with respect to N_i will contain, in addition to partial free energy, the derivative of configuration entropy

$$\left(\frac{\partial G}{\partial N_i}\right)_{T,\,P} = g_i - T\left(\frac{\partial S_{\mathrm{conf}}}{\partial N_i}\right)_{T,\,P}. \tag{5.2.5}$$

If the derivative is differentiated with respect to one kind of atoms, the chemical potential of the atom can be said to have gained from statistical interaction, and this gain is expressed via the configuration entropy derivative:

$$\mu_i = \mu_i^0 - T\left(\frac{\partial S}{\partial N_i}\right)_{T,\,P}. \tag{5.2.6}$$

This formula is more general than the one above

$$\mu_i = \mu_i^0 + kT \ln x_i\,, \tag{5.2.7}$$

where x_i is the fraction of positions occupied by the i-element. The second term in the right-hand side of (5.2.7) is the consequence of the mixing entropy derivative which is the simplest form of configuration entropy, more exactly, the configuration mixing entropy of two components. Expression

(5.2.6), on the contrary, includes the mixing entropy of many components, with the account of their symmetry and possible arrangement patterns in the crystal lattice.

Various structural elements (atoms and vacancies) having no symmetry and objects (complexes and precipitates) with a more complex structure and their own symmetry will all contribute to the configuration entropy value. The object symmetry may be responsible for its various orientations relative to the crystal axes, thus increasing the configuration entropy. Finally, an appreciable contribution to this entropy component is made by electrons and holes. These can occupy free states in the conduction and valence bands, as well as the energy states of defects with different degrees of degeneracy. All structural elements and objects are involved in statistical interactions.

Consider now various arrangement patterns of defects in a crystal lattice and calculate the configuration entropy for different cases [2, 3].

5.2.1 Configuration entropy of a lattice with N^{β} sites and N_V^{β} vacancies

The number of ways in which N_V^{β} vacancies can be arranged at N sites is equal to the number of positions [3]:

$$A = \frac{N^{\beta}!}{\left(N - N_V^{\beta}\right)!}. \tag{5.2.8}$$

Since vacancies are identical, the thermodynamic states, in which only two vacancies have interchanged positions, are also identical. Therefore, the number of independent arrangement patterns is smaller as many-fold as the number of possible rearrangements of N_V-vacancies. In other words, the number of vacancies is reduced by a factor of $N_V!$ and is equal to the number of combinations

$$W = \frac{N^{\beta}!}{\left(N^{\beta} - N_V^{\beta}\right)! N_V^{\beta}!}. \tag{5.2.9}$$

5.2.2 A lattice with several defect types

Every structural element occupies its crystal position independently. The same is true of host atoms, whose number is

$$N_{\beta}^{\beta} = N^{\beta} - \sum_{\alpha} N_{\alpha}^{\beta} . \tag{5.2.10}$$

With the reasoning above, we have

$$W = \frac{N^{\beta}!}{\left(N^{\beta} - \sum_{\alpha} N_{\alpha}^{\beta}\right)! \prod_{\alpha} N_{\alpha}^{\beta}!} . \tag{5.2.11}$$

5.2.3 A lattice with structural elements in several positions

Structural elements may occupy positions in several sublattices of a binary or multi-component semiconductor, or in different sublattices and types of interstice. Then the general thermodynamic probability is the product of probabilities of defect arrangement in individual subsystems. This is because every arrangement occurs independently. The thermodynamic probability is

$$W = \prod_{\beta} \frac{N^{\beta}!}{\left(N^{\beta} - \sum_{\alpha} N_{\alpha}^{\beta}\right)! \prod_{\alpha} N_{\alpha}^{\beta}} . \tag{5.2.12}$$

5.2.4 The arrangement of complexes

When arranging complexes, one should take into account that every complex may have different orientations in a crystal lattice. Let us select an atom in a complex as the base atom. This atom can be made to occupy sites or interstices as an ordinary structural element. But the complex position may vary with the base atom position. Every turn of the complex relative to the base atom produces a new arrangement which is to be allowed for in the calculation of thermodynamic probability. This concerns all complexes present in a lattice.

Suppose a complex has g_k arrangements in the lattice relative to the fixed base atom. Or, every position of the base atom is g_k-fold degenerate. Then, for one complex, the thermodynamic probability of the system increases g_k times while for all complexes $(g_k)^{Nk}$ times:

$$W = \frac{N^{\beta} g_k^{Nk}}{\left(N^{\beta} - N_k\right)! \, N_k \, !} . \qquad (5.2.13)$$

The degeneracy factor of a complex, g_k, can be calculated from the following considerations. The set of symmetry elements of an ideal lattice forms a point group. The appearance of a complex with its own symmetry reduces the group rank of the ideal lattice. The set of symmetry elements of this new lattice containing the complex is a subgroup of the old lattice. Indeed, a real crystal lattice can only be combined by the same operations as an ideal lattice, but the number of operations will be smaller because of the lower symmetry. All symmetry operations of a particular subgroup make up a complex in the same position, and the thermodynamic states thus produced are identical. The orientation of the complex is changed by operations which are left outside the real lattice subgroup. These operations make up their own subgroup, which is also a subgroup of the ideal lattice. Using the Lagrange theorem, one can conclude that *the degeneracy multiplicity of a complex is equal to the subgroup index in a real lattice*.

5.2.5 The arrangement of electrons and holes

Electrons and holes can occupy free states in the conduction and valence bands and electronic states of defects. If there is no electron degeneracy ($n << N_c$), free electrons have a set of identical levels, whose number per unit volume is equal to the effective density of states in the band [3]. Since these states do not differ in energy, they are N_c-fold degenerate. The thermodynamic probability for n electrons to occupy states within the band is

$$W = \frac{\left(N_c\right)^n}{n!} . \qquad (5.2.14)$$

The factor $(N_c)^n$ gives the number of states related by the N_c-fold degeneracy. The quantity $n!$ in the denominator rules out identical thermodynamic states due to electron rearrangement. Similarly, we have for holes:

$$W = \frac{\left(N_v\right)^p}{p!} . \qquad (5.2.15)$$

To calculate the thermodynamic probability of arrangement of n_α^β electrons in N_α^β defects, one must take into account the spin degeneracy. For the simple case of nondegenerate levels, the following two situations may arise.

(1) An electron occupies an energy level already occupied by another electron. It gets its spin adjusted and occupies the rest of the space. The degeneracy multiplicity is equal to unity.

(2) An electron occupies an empty level and its spin can take any of the two possible orientations. The degeneracy multiplicity is equal to 2.

If the energy level is energy degenerate, the degeneracy multiplicity must be calculated individually.

With the allowance for spin degeneracy, the thermodynamic probability is

$$W = \frac{N_\alpha^\beta ! r_\alpha^{n_\alpha^\beta} R_\alpha^{N_\alpha^\beta - n_\alpha^\beta}}{\left(N_\alpha^\beta - n_\alpha^\beta\right)! n_\alpha^\beta !}, \tag{5.2.16}$$

where r_α^β is the degeneracy multiplicity of an electron-filled state and R_α^β is that of a free state.

Note that the calculation of the probability of an ionized defect state requires that the probabilities of (5.2.11) and (5.2.16) should be multiplied together, because electrons become arranged independent of the arrangement of atoms. The co-factors N_α^β! are reduced by the multiplication. Moreover, it is necessary to account for the arrangement of free electrons in the allowed bands. Thus, we obtain the following thermodynamic probability for ionized atoms:

$$W = \frac{\left(N_\mathrm{c}\right)^n}{n!} \frac{\left(N_\mathrm{v}\right)^p}{p!} \frac{N^\beta ! \left(r_\alpha^\beta\right)^{n_\alpha^\beta} \left(R_\alpha^\beta\right)^{N_\alpha^\beta - n_\alpha^\beta}}{\left(N^\beta - N_\alpha^\beta\right)! \left(N_\alpha^\beta - n_\alpha^\beta\right)! n_\alpha^\beta !}. \tag{5.2.17}$$

We have discussed above all typical situations involving the calculation of thermodynamic probabilities of systems. The configuration entropy is calculated using the Boltzmann formula (5.2.1). Let us now find the derivative [see (5.2.5)] characterizing the statistical interaction of a system with the thermodynamic probability (5.2.17):

$$\frac{\partial S_\mathrm{conf}}{\partial N_\alpha^\beta} = k \ln \frac{N_\alpha^\beta - n_\alpha^\beta}{R_\alpha^\beta \left(N^\beta - N_\alpha^\beta\right)}. \tag{5.2.18}$$

One can see that the contribution to the interaction decreases with decreasing defect concentrations. Note that if the number of defects in the denominator is neglected, the derivative will include positions occupied by defects unfilled by electrons. In contrast with this situation, expression (5.2.7) includes the fraction of positions occupied by a particular kind of atoms, irrespective of whether they are ionized or not. This expression contains no degeneracy multiplicity. This is what we meant when we mentioned that formula (5.2.6) was more rigorous than (5.2.7) and better accounted for the statistical interaction.

5.3 CHARGE INTERACTION

Charge interaction is a manifestation of electromagnetic interaction. In the absence of external electromagnetic fields, a semiconductor crystal tends to preserve its neutrality. Therefore, all charged particles must obey the electrical neutrality law:

The net positive electric charge of all kinds of particles, both free and bound, must be equal to the net charge of negative particles:

$$\sum_i Q_i = 0 . \tag{5.3.1}$$

In this law, if energy states capable of capturing a hole appear in a semiconductor, additional free electrons or energy states that can capture them must be produced. In accordance with this, we have from (5.3.1)

$$n + \sum_i N_{ai}^- = p + \sum_j N_{dj}^+ . \tag{5.3.2}$$

As was mentioned above, charge interactions play an important role in double doping. Doping with additional donors raises acceptor solubility, and vice versa. In this connection, it is necessary to differentiate between donor and acceptor states when writing down Gibbs free energy of a crystal.

5.4 POTENTIAL INTERACTION

The introduction of a defect into a crystal inevitably induces nonuniform force fields of different nature. Defect ionization induces electrostatic fields. When a defect enters a crystal lattice, the latter experiences extension or compression, and the natural fluctuations in defect distribution lead to macroscopic nonuniform mechanical stresses. There is also a gravitational interaction, but it can be ignored in defect formation problems, in contrast to the first two interactions.

The existence of a force field changes the crystal energy, because there appears *potential energy of interacting particles*, in addition to the crystal *internal energy*. This is why this kind of interaction was called a potential interaction [1].

The total crystal energy will then be equal to the sum of free Gibbs energy and potential energy of the force field:

$$E = G + E_{\mathrm{p}} . \tag{5.4.1}$$

Let us find the exact differential of the total energy in a physically small volume, within which the force field value does not change:

$$\mathrm{d}E = \left(\frac{\partial E}{\partial T}\right)_{P,\,N} \mathrm{d}T + \left(\frac{\partial E}{\partial P}\right)_{T,\,N} \mathrm{d}P + \sum_i \left(\frac{\partial E}{\partial N_i}\right)_{T,\,P} \mathrm{d}N_i . \tag{5.4.2}$$

The potential energy of the force field can be assumed to be independent of temperature and pressure; then we have

$$\mathrm{d}E = -S\mathrm{d}T + V\mathrm{d}P + \sum_i \left(\frac{\partial G}{\partial N_i} + \frac{\partial E_P}{\partial N_i}\right)_{T,\,P} \mathrm{d}N_i . \tag{5.4.3}$$

The first bracketed term is the chemical potential of a particle at zero force field. The second term is the potential of the particle interaction with the force field:

$$\varphi_i = \left(\frac{\partial E_P}{\partial N_i}\right)_{T,\,P} . \tag{5.4.4}$$

The total crystal energy is an extensive parameter, and so it can be described by a relation similar to (5.2.2.):

$$E = \sum_i N_i \mu_i \left(P, T, F\right), \tag{5.4.5}$$

where F is the force field strength or another extensive parameter characterizing it.

We get from (5.4.5)

$$\mathrm{d}E = \sum_i \left(N_i \mathrm{d}\mu_i \left(P, T, F\right) + \mu_i \left(P, T, F\right) \mathrm{d}N_i\right). \tag{5.4.6}$$

From the comparison of (5.4.3) and (5.4.6), we find

$$\sum N_i \mathrm{d}\mu_i \left(P, T, F\right) = -S\mathrm{d}T - V\mathrm{d}P, \tag{5.4.7}$$

$$\mu_i \left(P, T, F\right) = \mu_i \left(P, T, \mathrm{O}\right) + \varphi_i. \tag{5.4.8}$$

Thus, *the chemical potential of a particle in a force field increases by the potential of interaction between this particle and the field*.

Lannoo and Bourgoin consider potential interaction [4] as a long-range interaction because its action extends for some distance. The distance, however, is not large: from 0.8 to 80 nm for an elastic field and from 10 to 50 nm for an electrostatic field. So the term "potential interaction" seems preferable.

The charge conservation law can be considered to be strictly valid for most defect formation problems. Therefore, the crystal electrostatic field is zero and does not contribute to chemical potentials. Nevertheless, it is easy to give illustrations to support the validity of (5.4.8) for electrostatic fields. First, it is the well-known Frenkel effect stating that impurity activation energy decreases in a strong electric field. This effect primarily refers to centers possessing a Coulomb potential in interaction with an electron. Figure 5.1 shows that, because of the summation of the Coulomb field attracting an electron to an atom and the external electric field, the activation energy reduces by the value

$$\Delta E_{\mathrm{F}} = \left(\frac{e^3}{\pi\varepsilon_{\mathrm{s}}} F\right)^{1/2}, \tag{5.4.9}$$

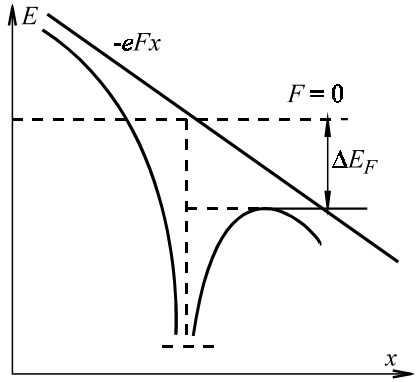

Figure 5.1. Decreasing potential barrier of an attracting center in an electric field.

where F is electric field strength and $\varepsilon_s = \varepsilon\varepsilon_0$ is dielectric permittivity of the semiconductor.

Defect formation in strong electric fields is poorly understood, and there is no information about the electric field effect on defect formation enthalpy.

Electrostatic fields are also due to the ionization of a defect possessing a point charge, whose value is determined by the ion charge state. The ion interaction energy is

$$\Delta E = z_1 z_2 e^2 / 4\pi\varepsilon_s r \,, \qquad (5.4.10)$$

where z_1 and z_2 are the charge states of interacting ions and r is the distance between them.

Formula (5.4.10) is approximate because it accounts only for electrostatic interactions, neglecting other effects. However, it can be used profitably for a qualitative treatment of interactions. If defects have opposite signs, they are attracted to each other and the energy given by formula (5.4.10) is negative. For this reason, Coulomb interaction is one of the causes of the formation of complexes, because mutual attraction stimulates coupling [4–6]. Since an electrostatic interaction is not the only factor leading to complexation, this problem will be discussed individually.

Two situations can prevent coupling in an electrostatic interaction:

(1) when defects of the same sign are repelled from each other and the interaction energy (5.4.10) is positive;

(2) when defects of opposite signs have a low mobility and cannot approach one another to form a stable pair.

The latter often occurs at low temperatures. On the one hand, diffusion coefficients decrease with decreasing temperature, but, on the other, *the electrostatic interaction radius* can be defined as a characteristic distance, at which the interaction energy module (5.4.10) is equal to the most probable particle energy (kT). This radius increases with decreasing temperature, so the situation of interest becomes more probable.

Let us find the Gibbs free energy of a crystal with N sites statistically occupied by charged impurities at concentrations N_α^β and N_γ^β. With formula (5.4.8) and the condition for the absence of stable pairs, we have

$$G = \mu_\beta^\beta N_\beta^\beta + \left(\mu_\alpha^\beta + \varphi_{\alpha\gamma}\right)N_\alpha^\beta + \left(\mu_\gamma^\beta + \varphi_{\gamma\alpha}\right)N_\gamma^\beta - kT\ln W . \qquad (5.4.11)$$

Since no stable pairs are formed, impurities arrange themselves at sites individually. Besides, impurities are completely ionized, so the arrangement of electrons and holes can be ignored:

$$W = \frac{N^\beta !}{\left(N^\beta - N_\alpha^\beta - N_\gamma^\beta\right)!\, N_\alpha^\beta !\, N_\gamma^\beta !} . \qquad (5.4.12)$$

The electrostatic interaction potential is described by (5.4.10):

$$\varphi_{\alpha\gamma} = \varphi_{\gamma\alpha} = \frac{z_\alpha z_\gamma e^2}{4\pi\varepsilon_s r_{\alpha\gamma}} = \frac{Q}{r_{\alpha\gamma}} , \qquad (5.4.13)$$

where $r_{\alpha\gamma}$ is the mean distance between defects, $Q > 0$ is repulsion, and $Q < 0$ is attraction.

In view of a statistical distribution of defects at lattice sites, the mean distance between them can be evaluated from the formula:

$$r_{\alpha\gamma} = \left(N_\alpha^\beta + N_\gamma^\beta\right)^{-1/3} . \qquad (5.4.14)$$

The crystal free energy is expressed as

$$\begin{aligned} G = {} & \mu_\beta^\beta N_\beta^\beta + \mu_\alpha^\beta N_\alpha^\beta + QN_\alpha^\beta\left(N_\alpha^\beta + N_\gamma^\beta\right)^{1/3} \\ & + \mu_\gamma^\beta N_\gamma^\beta + QN_\gamma^\beta\left(N_\alpha^\beta + N_\gamma^\beta\right)^{1/3} - kT\ln W . \end{aligned} \qquad (5.4.15)$$

Hence, we have

$$\frac{\partial G}{\partial N_\alpha^\beta} = \mu_\alpha^\beta + \frac{5}{3} Q \sqrt[3]{N_\alpha^\beta + N_\gamma^\beta} - kT\left[\ln\left(N^\beta - N_\alpha^\beta - N_\gamma^\beta\right) - \ln N_\alpha^\beta\right]. \qquad (5.4.16)$$

One can see that the chemical potential of a non-interacting atom has a gain associated with an electrostatic interaction. This gain affects the impurity concentration. From (5.4.16), we have

$$N_\alpha^\beta = N^\beta \exp\left(-\frac{\mu_\alpha^\beta}{kT} - \frac{5Q}{3kT}\sqrt[3]{N_\alpha^\beta + N_\gamma^\beta}\right). \qquad (5.4.17)$$

Formula (5.4.17) was derived from a simple algorithm and is approximate. In particular, it does not allow for Debye–Huckel corrections for electrostatic screening [1, 4]. Note, however, that the screening effect requires charge carriers with a high mobility. Such carriers do exist in electrolytes, but in solid solutions the impurity mobilities are not high enough to produce screening. For this reason, screening in semiconductors is due to the motion of free charge carriers.

Still, one can conclude from formula (5.4.17) that the concentrations of doping impurities affect one another because of interactions. At $Q > 0$, impurities are repelled from each other and their solubility is lower. The same is true of a single impurity in heavy doping: the mean distance becomes smaller at higher concentrations ($r_\alpha = N_\alpha^{-1/3}$) and the solubility must decrease. But at $Q < 0$, impurities are attracted to one another, and their intermixing increases.

Let us consider the contribution of elastic interaction. When a defect enters a crystal lattice, the latter experiences an extension or compression. Charge–exchange processes cause an additional lattice relaxation and polarization due to electron–phonon interactions. Deformation produces excess energy which can decrease because of pairing or precipitation. Therefore, deformation stimulates the formation of some types of complexes. Elastic strains can disappear owing to dislocation formation.

Of primary interest is how elastic strains influence defect formation. To answer this question, we must find a way of evaluating the potential of elastic forces. We will reproduce the evaluations made above for an electrostatic interaction.

Elastic interaction problems are discussed in detail in the book of Leibfried and Brener [7]. The authors derived the energy of interaction between a defect and an external elastic deformation field and the interaction energy for

two defects creating deformation fields between themselves. The expansion of this interaction energy in terms of the power r^{-1} shows that the principal expansion term varies with distance as r^{-3}. The same result follows from [8]. So, the interaction potential will be

$$\varphi_{\alpha\gamma} = \frac{\sigma}{r^3} = \frac{P_0^{\alpha}}{4\pi e_{11} r^3}\left[S_P\left\{P^b\right\} - 3\left(\hat{R}, P^b, \hat{R}\right)\right]. \qquad (5.4.18)$$

The designations in formula (5.4.18) are the same as in [7].

Using the interaction potential expressed as (5.4.18), we have for free energy in (5.4.11)

$$G = \mu_{\beta}^{\beta} N_{\beta}^{\beta} + \mu_{\alpha}^{\beta} N_{\alpha}^{\beta} + \mu_{\gamma}^{\beta} N_{\gamma}^{\beta} + \sigma\left[N_{\alpha}^{\beta}\left(N_{\alpha}^{\beta} + N_{\gamma}^{\beta}\right) + N_{\gamma}^{\beta}\left(N_{\alpha}^{\beta} + N_{\gamma}^{\beta}\right)\right] - kT \ln W \qquad (5.4.19)$$

Hence,

$$\frac{\partial G}{\partial N_{\alpha}^{\beta}} = \mu_{\alpha}^{\beta} + 2\sigma\left(N_{\alpha}^{\beta} + N_{\gamma}^{\beta}\right) - kT\left[\ln\left(N^{\beta} - N_{\alpha}^{\beta} - N_{\gamma}^{\beta}\right) - \ln N_{\alpha}^{\beta}\right]. \qquad (5.4.20)$$

The second term in (5.4.20) accounts for elastic interactions of defects with one another. Such interactions affect the intermixing of defects. From (5.4.20), we obtain

$$N_{\alpha}^{\beta} = N^{\beta} \exp\left[-\frac{\mu_{\alpha}^{\beta}}{kT} - \frac{2\sigma}{kT}\left(N_{\alpha}^{\beta} + N_{\gamma}^{\beta}\right)\right]. \qquad (5.4.21)$$

One can see that the defect concentration depends on both the interaction potential and the concentration of defects inducing elastic strains.

Thus, external force fields and internal fields created by defect formation processes change crystal energy. Chemical potentials have gains associated with force interactions. As a result, the defect concentration varies with the interaction potential value. One should bear in mind that potentials created by point defects are of the short-range type. For this reason, these effects manifest themselves in heavy doping. Besides, external fields can produce effects similar to the Frenkel effect for Coulomb attracting centers. The defect formation enthalpy decreases and their concentration rises when defect formation occurs in an external force field.

5.5 DEFECT INTERACTION IN A REGULAR APPROXIMATION

Solutions in the liquid and solid phases are classified by the degree of components involvement in the interaction. *An ideal solution* is a solution satisfying the three independent conditions:

(1) the partial internal energy of a component is independent of the solution concentration, and the total internal energy of the solution is the sum of internal energies of the mixing components;

(2) the partial molar volume of a component does not change on mixing;

(3) the partial molar entropy increases by the mixing entropy value

$$\Delta S = -k \ln x_i ,$$

where x_i is the fraction of positions occupied by the i-th component.

The first condition is not always fulfilled because there are different types of interaction (see above).

A regular solution is a solution satisfying the second and third conditions for an ideal solution but it does not satisfy the first condition. There is a certain non-zero mixing enthalpy.

Most solid solutions do not behave as regular solutions but it may be convenient to consider them as such, assuming "regularity" to be a deviation from perfection [6].

The mixing enthalpy of a regular solution is usually found by a quasi-chemical approximation. Suppose two sorts of atoms, *A* and *B*, are miscible. The atoms if each kind individually create ideal solutions. For this reason, only new, additional interactions between the neighboring *A* and *B* atoms can contribute to the mixing enthalpy. Sometimes, chemical bonds are said to be formed, but this is not necessary. Of importance is the fact that their interaction enthalpy is equal to H_{AB}. The mixing enthalpy value is equal to the product of the excess energy produced by this bond and the number of bonds.

Let us denote the coordination number of the crystal lattice as z and find the number of positions in the crystal and the fractions of positions occupied by both sorts of atoms:

$$N = N_A + N_B, \qquad x_A = N_A / N, \qquad x_B = N_B / N. \tag{5.5.1}$$

The probability for atoms to occupy their proper positions is equal to the fraction of positions intended for them. Therefore, the probability for an atom A to occupy a particular position is x_A and for an atom B it is x_B.

In accordance with the probability multiplication rule, the probability for both atoms to occupy their proper positions simultaneously is $x_A x_B$. The probability will not change if the positions of these atoms are interchanged. Therefore, the probability for neighboring positions to be occupied by different atoms is $2x_A x_B$ [16].

The number of A–B bonds is equal to the total number of bonds in a crystal ($zN/2$) multiplied by the formation probability of a bond:

$$N_{AB} = 2x_A x_B \frac{1}{2} zN_0 = zNx_A x_B . \tag{5.5.2}$$

Then the number of A–A and B–B bonds is

$$N_{AA} = \frac{z\left(N_A - N_{AB}\right)}{2}, \qquad N_{BB} = \frac{z\left(N_B - N_{AB}\right)}{2} . \tag{5.5.3}$$

Total crystal enthalpy can be written as

$$\begin{aligned} H &= H_{AA} N_{AA} + H_{BB} N_{BB} + H_{AB} N_{AB} \\ &= \frac{zN_A}{2} H_{AA} + \frac{ZN_B}{2} H_{BB} + zNx_A x_B \left[H_{AB} - \frac{1}{2}\left(H_{AA} - H_{BB}\right)\right]. \end{aligned} \tag{5.5.4}$$

The first and second terms in (5.5.4) describe the enthalpy of miscible components (similar components exist in an ideal solution). The third term is the mixing enthalpy

$$H_{mix} = x_A x_B \Omega , \tag{5.5.5}$$

where the interaction energy is

$$\Omega = zN\left[H_{AB} - \frac{1}{2}\left(H_{AA} + H_{BB}\right)\right]. \tag{5.5.6}$$

It is easy to see that a regular solution transforms to an ideal one under the condition

$$H_{AB} = \frac{1}{2}\left(H_{AA} + H_{BB}\right). \tag{5.5.7}$$

Mixing enthalpy may be positive or negative, depending on the type of interaction. A negative enthalpy value means that atoms are attracted to one another; the physical reason for this attraction is not particularly important. A negative mixing enthalpy value indicates repulsion. Attraction may give rise to a short-range order in the solution, so that the module of mixing enthalpy will increase to become more negative.

With Gibbs free energy of the solution, we find, after differentiation, the well-known result:

$$\mu_B = \mu_B^0 + kT \ln x_B + (1 + x_B)^2 \Omega , \tag{5.5.8}$$

where

$$(1 - x_B)^2 \frac{\Omega}{kT} = \ln \gamma_B , \tag{5.5.9}$$

with γ_B as an activity coefficient. The activity of the component B is

$$a_B = \gamma_B x_B . \tag{5.5.10}$$

Similarly, we can find it for the second component.

These formulas are normally used to describe the behavior of components in a medium, with which a particular crystal is in equilibrium. In case of a liquid phase, a certain short-range order is implied.

Let us consider the contribution of a regular interaction to defect formation processes. Note again that there were only two sorts of atoms in the situation discussed above. One was the solvent and the other the solute. It was between these two kinds of atoms that the interaction took place. From the point of view of defect formation, this means the interaction of a defect with the host lattice, where it was produced. Besides, it was assumed in the derivation of the regular solution formulas that the elements were close to one another and that the interaction was due to short-range forces until chemical bonds were formed. In fact, this is the process of defect formation. Therefore, regular interaction energy must contribute to defect formation enthalpy.

One should keep in mind another circumstance. The fraction of positions occupied by defects is much less than unity, so the value of x_B in the third term of (5.5.8) is neglected. One can see then that, in the first approximation, defect enthalpy has a constant gain due to the regular interaction.

To summarize, a regular interaction of defects in a solid crystal can be neglected. The parameters of this interaction are independent of defect concentrations because these are low. These parameters are summed with thermodynamic parameters of defect formation.

5.6 INTERACTION LEADING TO COMPLEXATION

Complexes, or associates [5], can be formed in crystals under certain conditions, usually at low temperatures. *A complex is a stable structure representing a quasimolecule in the host crystal and possessing specific physical properties*. A complex has its own symmetry different from that of a perfect crystal and, therefore, can occupy several equivalent positions in a lattice with respect to the fixed base atom. It was pointed out in Section 5.2 that the degeneracy multiplicity of the spatial orientation of a complex, r_c, is equal to the subgroup index of the real crystal.

Physically, the nature of forces producing complexes may be different. In particular, an important role in this process is played by an electrostatic interaction. Without going into detail, we would like only to note that every complex has its own formation energy E_c.

Let us see how the process of complexation influences chemical potentials and solubility of the structural constituents of a complex.

We will illustrate this with a simple complex made up of an impurity and a vacancy. Ionization processes will be neglected for simplicity. The crystal free energy in this case can be defined as

$$G = \mu_\beta^\beta N_\beta^\beta + \mu_\alpha^\beta N_\alpha^\beta + g_V^\beta N_V^\beta + g_c^\beta N_c - kT \ln W , \qquad (5.6.1)$$

where

$$W = \frac{N^\beta ! r_c^{N_c}}{\left[N^\beta - \left(N_\alpha^\beta - N_c\right) - \left(N_V^\beta - N_c\right)\right]! \left(N_\alpha^\beta - N_c\right)! \left(N_V^\beta - N_c\right)! N_c !} . \qquad (5.6.2)$$

Taking into account the reasoning concerning the applicability of the chemical potential concept (Section 1.1), we used partial Gibbs free energies for vacancies and complexes in (5.6.1). Expression (5.6.2) took into consideration that it was necessary to arrange free structural elements; so the denominator contains ($N_\alpha{}^\beta - N_c$) for free impurities and ($N_V{}^\beta - N_c$) for vacan-

cies free from complexation. The free energy could be written differently, using the bonding equation, with the minimization performed with Lagrange indeterminate factors. This will be done in Chapter 6.

By differentiating with respect to N_c and equating the derivative to zero, we obtain the familiar result for the concentration of similar complexes [4]

$$N_c = \frac{\left(N_\alpha^\beta - N_c\right)\left(N_V^\beta - N_c\right)}{N^\beta} \exp\left(-\frac{g_c}{kT}\right). \tag{5.6.3}$$

The chemical potential of the impurity involved in complexation is

$$\left(\frac{\partial G}{\partial N_\alpha^\beta}\right)_N = \mu_\alpha^\beta - kT\left[\ln N^\beta - \ln\left(N_\alpha^\beta - N_c\right)\right]. \tag{5.6.4}$$

It follows from (5.6.4) that complexation processes change the defect chemical potential, and this is primarily due to a statistical interaction.

5.7 DEFECT IONIZATION IN SOLIDS

The ionization of defects, like other defect formation processes, can be analyzed in terms of the active mass law.

In accordance with the routine procedure of problem solution with the active mass law, we will write the impurity ionization reaction [see (1.2.30)] of, say, the donor type

$$B \underset{\leftarrow}{\rightarrow} B^+ + e^-. \tag{5.7.1}$$

The respective equilibrium constant is

$$K_{Be} = \frac{\left[B^+\right]}{\left[B\right]} = \exp\frac{\mu_B^+ + \mu_B - \mu_e}{kT}, \tag{5.7.2}$$

where μ_e is chemical potential of an electron, whose value is determined by the Fermi energy, and μ_B^+ is chemical potential of an ionized atom.

The physical meaning of the Fermi energy implies that

$$n = N_c \exp\left(-\mu_e / kT\right) \tag{5.7.3}$$

and that the Fermi level difference between an ionized and an un-ionized state can be taken to be the ionization energy of the defect. Then, with (5.7.3) and (5.7.2), we have

$$K_{Be} = \frac{[B^+]n}{[B]} = N_c \exp\left(-\frac{\varepsilon_B}{kT}\right). \tag{5.7.4}$$

Expression (5.7.4) often used in calculations is not rigorous enough. To demonstrate its inaccuracy, let us turn to quantum transition theory in the solid state, since its details have been described in many fundamental books. Without citing the details here, we will mention only the initial approximations.

The quantum transition of an electron from the ground to an excited state necessarily requires the analysis of the electron–lattice system as a whole. The transition changes the potential energy, or the lattice polarization. The Hamiltonian of the system contains the following terms:

$$\hat{H} = \hat{H}_e(\vec{r}) + \hat{H}_{eL}(\vec{r}, \vec{R}) + \hat{H}_L(\vec{R}), \tag{5.7.5}$$

where $\hat{H}_e$ is the Hamiltonian of electrons (fast subsystem) in the field of atoms (slow subsystem), $\hat{H}_L$ is the Hamiltonian of free oscillations of crystal atoms, $\hat{H}_{eL}$ is an operator of electron–phonon interaction, $\vec{r}$ is the combination of electron coordinates, and $\vec{R}$ is the combination of nuclear coordinates.

An adiabatic approximation implies that the wave function varies slowly with the nuclear coordinates. The Hamiltonian terms in (5.7.5) including the derivatives with respect to the nuclear coordinates (a non-adiabaticity operator) are ignored. With the assumption of the fast and slow subsystems, the wave function can be represented as the product of the wave functions for electrons $\Phi_e(\vec{r}, \vec{R})$ and those for nuclei $\varphi_L(\vec{R})$

$$\Psi = \Phi_e(\vec{r}, \vec{R})\varphi_L(\vec{R}). \tag{5.7.6}$$

The Schrodinger equation is divided into two equations:

$$-\frac{h^2}{8\pi^2 m}\sum_i \Delta_i \Psi_e\left(\vec{r}, \vec{R}\right) + V\left(\vec{r}, \vec{R}\right)\Psi_e\left(\vec{r}, \vec{R}\right) = W\left(\vec{R}\right)\Psi_e\left(\vec{r}, \vec{R}\right), \qquad (5.7.7)$$

$$-\frac{h^2}{8\pi^2}\sum_\alpha M_\alpha^{-1}\Delta_\alpha \Psi_L\left(\vec{R}\right) + W\left(\vec{R}\right)\Psi_L\left(\vec{R}\right) = E\Psi_L\left(\vec{R}\right). \qquad (5.7.8)$$

The first of the two equations describes the state of the electron subsystem in a field of motionless nuclei. The other one describes the motion of nuclei in an averaged electron field. The eigenvalues of electron energy $W(\vec{R})$ are, simultaneously, the potential energy of nuclei, known as *adiabatic potentials*.

A low vibration approximation and the transition to normal coordinates provide adiabatic potentials $W(\vec{R})$ as an expansion with respect to normal vibration modes.

Further calculations are unable to cover the whole vibration spectrum, so a one-coordinate approximation is used. This approximation means that the interaction with one, normally, totally symmetric vibration mode is dominant in a particular system [9]. It sets fairly rigid requirements on the nature of vibrations in the system, requiring, in particular, the dominant role of local vibrations. The conditions under which local and pseudo-local vibrations can arise in the vicinity of a defect are discussed in [9, 10].

Of principal importance is the frequency of intramolecular vibrations in a crystal. If this frequency is in the range of allowed frequencies of the host crystal, the intramolecular vibration energy is generated into the crystal, and the vibration is damped fast. This kind of vibration is unable to make an appreciable contribution to electron–phonon interactions in defect charge–exchange processes. On the contrary, if the frequency of "molecular" vibration is in the range of forbidden lattice frequencies, no wave is generated, and the vibrations are damped slowly. Vibrations involve only atoms located close to a defect. Such vibrations are referred to as local; they can make a considerable contribution to interactions.

However, the local vibration frequency must be beyond the resonance frequencies of the crystal; therefore, the energy of such vibrations must be higher than that of optical phonons in the host lattice. This condition cannot always be fulfilled for intramolecular vibration, but there are situations when local modes lying in the allowed frequency range are only slightly related to the host lattice or to a vibration symmetry not inherent in the crystal. Such vibrations are known as *pseudo-local* and may contribute much to the interactions.

The solution to the Schrodinger equation in the one-coordinate approximation shows that the transition of an electron to an excited state not only raises the potential energy but changes the coordinate of its minimum. This supports the fact that the transition to an excited state due to an electron–phonon interaction is accompanied by displacement of nuclei relative to their initial positions in the system ground state. Two solutions for potential energy in the one-coordinate model are described by simple formulas similar to (2.2.16):

$$E_g = E_1 + \frac{1}{2}\hbar\omega_g Q^2,$$
$$E_e = E_2 + \frac{1}{2}\hbar\omega_e (Q - Q_0)^2, \qquad (5.7.9)$$

where E_g is the potential energy of the ground state and E_e is that of the excited state; $\hbar\omega$ is the energy of the phonon involved in the interaction, the subscripts *g* and *e* indicating that the effective frequencies differ in the ground and excited states; Q is a running value of the configuration coordinate; Q_0 is the coordinate of the potential energy minimum for the excited state.

Formulas (5.7.9) are given in a harmonic approximation. They represent intercepting parabolas known as a configuration coordinate diagram. It should be emphasized that the allowance for anharmonism causes deviations from the parabolic pattern. Besides, the curves do not intercept in the case of resonance interaction (Figure 5.2), in contrast to the case discussed earlier (Figure 2.17).

The electron transition due to an optical excitation involves the Frank–Condon principle. The particle transition from the ground to an excited state occurs too fast for the equilibrium position of nuclei to change at that moment. For this reason, the optical transition with phonon absorption is shown in the diagram by a vertical arrow at point 0; in the case of radiation, the arrow is at point Q_0. Transition energies with radiation and absorption are, respectively,

$$h\nu^{abs} = E_0 + s\hbar\omega_e, \qquad h\nu^{rad} = E_0 - s\hbar\omega_g, \qquad (5.7.10)$$

where *s* is a factor indicating the number of phonons emitted in the thermolization process and $E_0 = E_2 - E_1$ is the difference between minimum energies

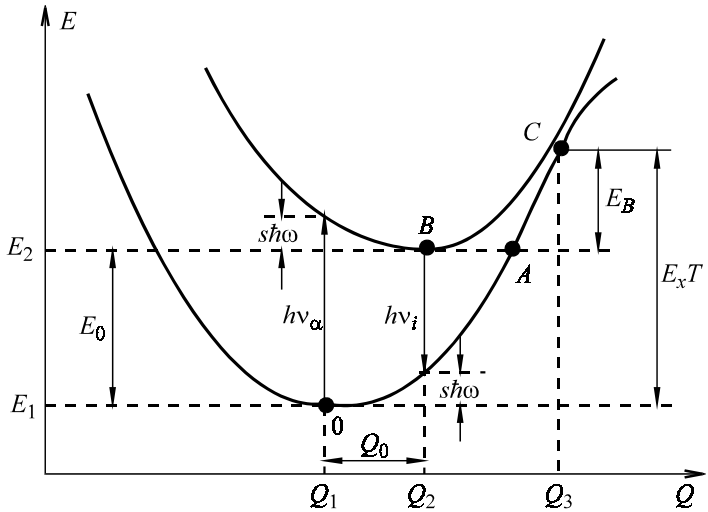

Figure 5.2. The coordination coordinate diagram of the defect ground and excited states in a crystal.

of the excited and ground states, often termed as the energy of purely electron transition. The simple situation described by formula (5.7.10) implies that the phonon energy does not change during the system transition to an excited state.

The quantity $s\hbar\omega$ characterizes the energy released after the optical transition when the system tends to take an equilibrium position. The energy difference between optical absorption and radiation characterizes the lattice polarization energy. It is named the Frank–Condon shift, Stokes losses, or heat release:

$$h\nu^{\text{abs}} - h\nu^{\text{rad}} = \Delta E = 2s\hbar\omega\,, \tag{5.7.11}$$

where ΔE is released heat.

The energy minimum position for the excited state is related to the value of factor s. To find this relation, let us express the optical transition energy through the factor s and Q_0:

$$h\nu^{\text{abs}} = E_0 + \frac{1}{2}\hbar\omega Q_0^2 = E_0 + s\hbar\omega\,. \tag{5.7.12}$$

Hence, we have

$$Q_0 = \sqrt{2s}\,, \qquad s = \frac{1}{2}Q_0^2\,. \tag{5.7.13}$$

Classically, a nonradiative transition from the ground to an excited state must go through the interception point *C* of the potential curves. Absorbing one phonon after another, the system, which was initially at zero point, goes as far as the hump point *C* to join the potential curve for the excited state. In the configuration coordinates model, this is the process of thermal emission. Thermal activation energy is expressed as the energy of a purely electron transition and the heat release is described by a formula which can be derived from (5.7.9) by equating E_g and E_e. At first, we should find the configuration coordinate, at which the transition occurs:

$$E_g = E_e = E_{xt} = \frac{1}{2}\hbar\omega Q_x^2 = E_0 + \frac{1}{2}\hbar\omega\left(Q_x - \sqrt{2s}\right)^2. \tag{5.7.14}$$

From (5.7.14), we have

$$Q_x = \frac{E_0 + s\hbar\omega}{\sqrt{2s} - \hbar\omega}. \tag{5.7.15}$$

Then we find the energy of a classical thermal transition:

$$E_x = \frac{1}{2}\hbar\omega Q_x^2 = \frac{\left(E_0 + s\hbar\omega\right)^2}{4s\hbar\omega}. \tag{5.7.16}$$

The reverse transition from point *B* to the ground equilibrium state 0 must go through point *C*. This transition actually reflects electron capture by a center. Obviously, the capture must have activation energy E_B (Figure 5.2):

$$E_B = \frac{\left(E_0 - s\hbar\omega\right)^2}{4s\hbar\omega}. \tag{5.7.17}$$

From the classical point of view, the transition from point *A* to point *B* is forbidden, because it represents tunneling of nuclei with a large mass and, hence, has a low probability. In practice, such transitions take place at low temperatures. A shift of the transition point leads to the temperature dependence of capture coefficients, observed experimentally. One should bear in mind that the transition energies were derived for a simple situation, when an electron interacted with one type of phonon. This is very unlikely and re-

quires the use of special models of the center. In addition, formulas (5.7.15) and (5.7.16) are valid at temperatures which are hard to obtain experimentally, so it is unreasonable to use them for the calculation of electron–phonon interaction parameters.

Indeed, experiments involving activation energy measurement are carried out at relatively low temperatures, less than 400 K. Experiments using various capacitance spectroscopic techniques are normally made at temperatures below room temperature. The temperature range for activation energy measurements of shallow impurities is near the boiling point for nitrogen or below it. In this case, the crystal lattice does not follow the classical restructuring pattern during the impurity ionization but goes from A to B by tunneling with an activation energy E_0.

The temperature range of defect formation processes occurring at observable rates begins approximately at 600 K. This temperature is sufficient for a system to go to an excited state classically, so that formulas (5.7.15) and (5.7.16) should be valid.

To sum up, experimental activation energies measured at low temperatures cannot always be used in equilibrium constants describing high temperature defect formation, because the contribution of electron–phonon interactions is essential. A configuration diagram shows the energy difference between purely electron, thermal, and optical transitions with absorption and radiation. It is also used to get a qualitative description of the Stokes shift and activation dependences of capture coefficients. Note that equilibrium constants of photochemical reactions differ from those for reactions occurring in the dark. They depend on the spectral composition of the light flow inducing ionization.

The configuration coordinate model accounts for the spectral line broadening. Indeed, absorption may occur not only at the zero point but at any value of Q, since the system can absorb or emit several phonons. The absorbed light energy changes, and transitions at non-zero points have a lower intensity since absorption of several phonons is very unlikely. As a result, the spectral line is broadened to form a bell-shaped band containing information about the parameters of electron–phonon interaction [11]. The analysis of the band shape can yield experimental parameters of this interaction to be used in further investigations.

It is worth making another comment. Thermal emission liberates charge carriers. An electron bound by a level is liberated to go to the conduction band and the hole goes to the valence band. For this reason, many semiconductor researchers relate either the ground or an excited state to band potentials. Strictly, this is not quite the case. The configuration coordinate model

was suggested to describe intra-center transitions. The potential curves described above are the lattice energies in the vicinity of a defect creating a deep level, when an electron is in the ground and an excited state. The model does not allow for the transition from the bound state to the free state. Nonradiative capture in this case accounts for the multiphonon transition mechanism, or the energy transfer from a defect to the lattice. For this to happen, the electron must be captured, or, more exactly, become localized near the defect creating a deep level. So the configuration coordinate diagram does not show band potentials. Rather, one should speak of electron states, when an electron or a hole is liberated from a trap. The energy of such electron states coincides with the conduction band bottom or with the valence band top.

Defect ionization excites the crystal lattice and it becomes polarized. The effective frequency of a phonon involved in electron-vibrational transitions can change during the defect ionization and lattice transition to an excited state [12]. Sometimes, this can be observed experimentally [13].

It follows from the foregoing that a rigorous theory is necessary for the calculation of quantities contained in the equilibrium constant (5.7.2) for the defect ionization reaction (5.7.1). At present, no theory of this kind exists for defect formation processes: the adiabatic approximation is inapplicable and no other approaches have been suggested.

REFERENCES

5.1. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushchikh primesei v poluprovodnikakh* (Thermodynamics and Kinetics of Interacting Impurities in Semiconductors). Moscow: Nauka, 352 p. (1997) (in Russian).

5.2. B.I. Boltaks, *Diffusiya i tochechnye defekty v poluprovodnikakh* (Diffusion and Point Defects in Semiconductors). Leningrad: Nauka, 384 p. (1972) (in Russian).

5.3. V.L. Vinetzky, G.A. Kholodar, *Statisticheskoye vzaimodeistvie elektronov i defektov v poluprovodnikakh* (Statistical Interaction of Electrons and Defects in Semiconductors). Kiev: Naukova dumka, 187 p. (1969) (in Russian).

5.4. M. Nannoo, J. Bourgoin, *Point Defects in Semiconductors* I. Berlin–New York: Springer–Verlag, 263 p. (1981).

5.5. F.A. Kroger, *The Chemistry of Imperfect Crystals*. Amsterdam: North–Holland Publ. Co., 654 p. (1964).

5.6. R.A. Swalin, *Thermodynamics of Solids*. New York–London: Wiley, 313 p. (1961).

5.7. G. Leibfried, N. Brener, *Point Defects in Metals*. Berlin–New York: Springer–Verlag, 429 p. (1978).

5.8. J.D. Eshelby, *Solid State Physics*. **13**. New York: Acad. Press, 376 p. (1956).

5.9. N.N. Kristofel, *Teoriya primesnykh tzentrov malogo radiusa v ionnylh kristallakh* (Theory of Small Radius Centers in Ionic Crystals). Moscow: Nauka, 331 p. (1974) (in Russian).
5.10. K.K. Rebane, A.P. Purga, O.I. Sild *et al*., *Trudy inst. fiziki i astron. AN Est. SSR*, No. 14: 31–75 (1961) (in Russian).
5.11. S.V. Bulyarsky, N.S. Grushko, *Generatzionno-rekombinatzionnye protzessy v aktivnykh elementakh* (Generation and Recombination Processes in Active Elements). Moscow: MGU, 402 p. (1995) (in Russian).
5.12. S.V. Bulyarsky, S.I. Radautzan, V.E. Tezlevan, *Izvestiya AN SSSR, ser. fizika* **40**, No. 9: 1997–1998 (1976) (in Russian).
5.13. K.D. Glinchuk, N.M. Litovchenko, S.I. Skril, *Optoelektronika i poluprovodnikovaya tekhnika* (Optoelectronics and Semiconductor Technology), No. 7: 58–66 (1985) (in Russian).

Chapter 6

Associations of Impurity Atoms

6.1 ION PAIRS

An important type of impurity interactions is ion pairing. The understanding of this phenomenon stems from the theory of electrolytes developed by Debye and Huckel, who considered the electrostatic interaction of oppositely charged ions in a solution.

The Debye–Huckel theory suggests that a dense atmosphere of oppositely charged ions is formed, with time, around every ion in an electrolytic solution. So, the interaction of ions is actually an interaction of ionic atmospheres. The charge of an ionic atmosphere grows with total ion concentration in the solution and decreases with distance from the atmosphere center. In an external electric field, cations and anions move in opposite directions, together with their atmospheres slightly lagging behind, thereby retarding the movement of ions. Ions are also retarded because of the attrac-tion between oppositely charged ionic atmospheres. These retarding effects decrease ion mobility. So the internal energy of an electrolytic solution appears to be the sum of two components: one is U_0 characterizing the inter-nal energy of the uncharged particle subsystem and the other is U_e charac-terizing the subsystem of electrical charges. All thermodynamic functions are thought to consist of two parts corresponding to the uncharged and char-ged components of the solution. The behavior of uncharged particles is des-cribed fairly rigorously by well-known thermodynamic relations. But in order to describe the behavior of charged particles, one has to find the Helmholtz free energy due to the action of charges, or to inter-ion interactions.

Debye and Huckel introduced two assumptions to solve this problem: they replaced the concept of ion point charges by that of a continuous charge distribution of variable density and assumed the field between interacting ions to be a Coulomb field. Both assumptions permitted the use of Poisson's equation for charge distribution. With the allowance for the radial symmetry of the solution, they derived an equation describing the variation of electrostatic potential of any *k*-th ion, φ_k, along the *r*-axis:

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\varphi_k}{\mathrm{d}r}\right)=\frac{4\pi e}{\varepsilon_s}\sum z_j N_j \exp\left(-\frac{qz_j\varphi_k}{kT}\right), \qquad (6.1.1)$$

where ε_s is the dielectric constant of the medium (solution). The exponent of this equation contains the quantity $qZ_j\varphi_k$ representing the energy of the *j*-th ion with charge *Z* in the field of the *k*-th ion.

An exact integration of equation (6.1.1) is impossible, so the Debye–Huckel theory considers several approximations for some particular cases, of which the following two are of the greatest interest.

6.1.1 Point ions

For point ions, the desired potential of the *k*-th ion is

$$\varphi_k=-\frac{qz_k}{\varepsilon_s}\chi\,. \qquad (6.1.2)$$

Comparing this result with the point charge potential known from electrostatics, $\varphi_k = eZ_k/\varepsilon_s r$, one can easily see that the quantity $1/\chi$ having the dimensionality of length plays the same role as the distance *r* in the Coulomb law. Physically, $1/\chi$ is the radius of an ionic atmosphere. Of course, one should remember that this concept is arbitrary, since the same ions cannot compose an ionic atmosphere because of thermal motion. This, in turn, leads to the fact that ions comprising an ionic atmosphere cannot preserve a fixed position in space.

6.1.2 Ions with a fixed radius

For ions with a fixed radius, the desired potential of the *k*-th ion, φ_k, is described as

$$\varphi_k = -\frac{qz_k}{\varepsilon_s}\frac{\chi}{1+b_k\chi}, \tag{6.1.3}$$

where b_k is the final ion radius. This expression differs from (6.1.2) by the factor $1/(1 + \chi b_k)$.

The Debye–Huckel theory was further developed by Byerrum [2] who rejected any attempts to integrate equation (6.1.1) but employed a variational approach to the calculation of ion concentration in the central *k*-th ion field. It was Byerrum who showed that there was a critical interionic distance r_0 characterizing the boundary between completely dissociated ($r > r_0$) and associated ($r < r_0$) ions, the latter producing pair associations. Hence, the conclusion was drawn about the production of ion pairs in concentrated solutions.

The Debye–Huckel–Byerrum theory is, of course, valid only for liquid solutions, in which any spatial positions of ions are admissible and equally probable. In solid crystal solvents, ions occupy fixed positions and are incapable of moving at normal temperatures. Therefore, the concepts of ionic atmosphere and ion pairing are inapplicable directly to impurity ions. However, the basic features of ion pairing do manifest themselves in doped semiconductors because of a larger Bohr orbit of doping impurities. The latter fact is due to a small effective mass of electrons in semiconductors, to a nearly complete ionization of shallow hydrogen-like impurities, and to a low dielectric permittivity of semiconductor crystals.

An approach similar to that suggested by Byerrum can be applied to semiconductors. For this, let us calculate the probability $G(r)\mathrm{d}r$ of location of one-type impurity atom (A) at distance r from an ion of the other type (B). As in the Debye–Huckel theory, a crystal is regarded as a continuous medium, in which possible values of r are not discrete.

Let us bring the reference point of the coordinates into coincidence with the position of an A ion in the lattice. Delineate a sphere of radius r for consideration, such that there are no B ions within its volume. The probability of this event, W_1, will be

$$W_1 = 1 - \int_{r_0}^{r} G(r)\mathrm{d}r, \tag{6.1.4}$$

where r_0 is the shortest distance between ions.

The formation of an ion pair will be considered to occur when a *B* ion is at distance $r + \mathrm{d}r$, i.e., when it is located in a sphere layer between spheres of

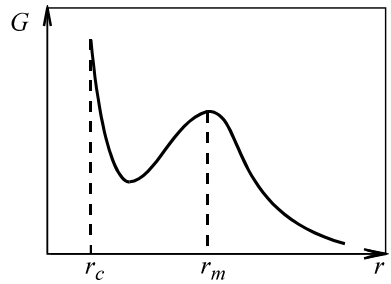

Figure 6.1. The distribution of ionic pairs as a function of interionic distance.

radii r and $(r + \mathrm{d}r)$. This layer volume is approximately equal to $4\pi r^2 \mathrm{d}r$. The probability of this event, W_2, will be

$$W_2 = 4\pi r^2 \mathrm{d}r F(r), \tag{6.1.5}$$

where $F(r)$ is the probability of an elementary pairing event

$$F(r) = N \exp\left(-\frac{\Delta H_r}{kT}\right), \tag{6.1.6}$$

N are the concentrations of A and B ions, (assuming them, for simplicity, to be the same), and ΔH_r is ion interaction energy at distance r.

The probability of a concurrent event is defined as $W_1 \times W_2$. Hence, by equating this product to the probability $G(r)\mathrm{d}r$, we find

$$G(r) = \left[1 - \int_{r_0}^{r} G(r)\mathrm{d}r\right] 4\pi r^2 F(r). \tag{6.1.7}$$

The solution to this integral equation has the form

$$G(r) = 4\pi r^2 F(r) \exp\left[-4\pi \int_{r_0}^{r} r^2 F(r)\mathrm{d}r\right]. \tag{6.1.8}$$

This distribution function is shown in Figure 6.1. Its extremal points are described as

$$r_c = r_0 \frac{\Delta H_r}{2kT} \tag{6.1.9}$$

$$r_m = \sqrt[3]{\frac{\pi N}{2}} . \tag{6.1.10}$$

Therefore, the curve in Figure 6.1 has two maxima. The first peak at r_c, corresponding to the minimum distance between ions, can be treated as the "short-order range" maximum and the peak at r_m as the "long-range order" maximum. It is clear from (6.1.10) that r_m coincides with the mean interionic distance for a random ion distribution. The subdivision of ions into two groups makes it possible to consider all ions separated by distance $r_c \geq r \geq r_0$ as being bound as ion pairs and all ions separated by distance $r > r_c$ as being free.

It follows from (6.1.9) that r_c decreases with increasing temperature, and Figure 6.1 shows that ion pairs in this case will disappear. They will disappear completely at $r_c = r_0$. One can easily derive from this, using (6.1.9), the criterion for ion pairing:

$$\Delta H_r >> 2kT . \tag{6.1.11}$$

In other words, ion pairs can exist until the ion interaction energy remains greater than the energy of ion thermal motion. It should be emphasized that this criterion is totally valid only for a continuous medium, but a crystal doped with impurities is, strictly, not continuous.

Ion pairs can be regarded as new defects, and their concentration can be calculated in terms of the active mass law. For this, ion pairing will be represented as a chemical reaction:

$$\mathrm{A} + \mathrm{B} \underset{\leftarrow}{\rightarrow} (\mathrm{AB})_r + \Delta H_r . \tag{6.1.12}$$

The concentration of ion pairs is defined by the ratio

$$[(\mathrm{AB})_r]/[\mathrm{A}][\mathrm{B}] = K_{(\mathrm{AB})_r} = Z_r f_r \exp(-\Delta H_r / kT), \tag{6.1.13}$$

where Z is the number of ways of ion pairing; in other words, Z reflects the configuration component of entropy variation:

$$Z = \exp(\Delta S_{\mathrm{conf}} / k) . \tag{6.1.14}$$

For an AB ion pair, Z is equal to the number of equivalent positions at the shortest distance from A, which can be occupied by B ions, or, vice versa, to the number of positions occupied by A ions at the shortest distance from B. The quantity f reflects the other component of entropy variation—the vibrational component ΔS:

$$f = \frac{\exp(\Delta S_{\text{vib}})}{k}. \tag{6.1.15}$$

In equation (6.1.13), [A] and [B] are concentrations of free, unpaired ions. If there are pairs with a different interionic distance, or with variable r, the concentration of unpaired ions of one kind, for example, A ions, will be

$$[\text{A}] = N_{\text{A}} - \sum_{r=1}^{r=n} [(\text{AB})_r], \tag{6.1.16}$$

where n is the number of ion pairs with variable r, meeting the criterion of (6.1.11).

Equation (6.1.13) can be re-written as

$$\frac{[(\text{AB})_r]}{\left\{N_{\text{A}} - \sum_{r=1}^{n}[(\text{AB})_r]\right\}\left\{N_{\text{B}} - \sum_{r=1}^{n}[(\text{AB})_r]\right\}} = K_{(\text{AB})_r}. \tag{6.1.17}$$

This equation refers to the equilibrium of ion pairs of one kind (one fixed r value at K) with all other ions. The description of a crystal may need n equations of the (6.1.17) type. For simplicity, we will assume $N_{\text{A}} = N_{\text{B}} = N$, as before. Then, we will have

$$\frac{\sum_{r=1}^{n}[(\text{AB})_r]}{\left\{N - \sum_{r=1}^{n}[(\text{AB})_r]\right\}^2} = K_{\Sigma} = \sum_{r=1}^{n} Z_r f_r \exp\left(\frac{\Delta H_r}{kT}\right). \tag{6.1.18}$$

Since the calculation of K_{Σ} is performed here with respect to discrete values of r, the crystal, therefore, is not considered in this approach as a continuous medium. Such an analysis was carried out in [3]. When the crystal is considered as a continuous medium, the sum in (6.1.18) is replaced by the integral

$$K_\Sigma = 4\pi \int_{r_0}^{r_c} r^2 \exp\left(-\frac{\Delta H_r}{kT}\right) \mathrm{d}r \,. \qquad (6.1.19)$$

The two approaches give close results for one kind of ion pairs. Indeed, in many cases, the concentration of ion pairs with the shortest interionic distance is much higher than that of all other ion pairs. So, these latter can be ignored, and the interaction will be defined by one ΔH value.

A still simpler picture can be obtained if one has a clear understanding of the nature of interaction forces. In principle, variation in ion pairing enthalpy ΔH may be due to various causes: a Coulomb interaction with or without a charge polarization effect, the formation of chemical (e.g., covalent) bonds, elastic interaction, or vibrational effects.

It is generally accepted that a Coulomb interaction is the most important factor, whereas the contribution of other forces is so small that it is usually neglected. In our view, of greater importance, sometimes, is the affinity of one kind of ions to some others, or the formation of stable chemical bonds. Evidence for this was given in [4]. The authors studied the solubility isotherms for A^{III} and B^{V} impurities in germanium and silicon at an equiatomic ratio of the doping impurities. They observed a stable chemical interaction between these impurities, producing chemical complexes.

If we assume that a Coulomb interaction makes the largest contribution to ion pairing, then we have

$$\Delta H = \frac{Z_A Z_B q^2}{\varepsilon r} \,. \qquad (6.1.20)$$

Ignoring, however, the comments above, we can show in the first approximation with (6.1.20) that the integral defining K_Σ in (6.1.19) is

$$K_\Sigma = 4\pi \left(\frac{Z_A Z_B q^2}{\varepsilon}\right)^3 Q(\alpha) \,, \qquad (6.1.21)$$

where $Q(\alpha)$ is the Byerrum function of argument $\alpha = Z_A Z_B q^2/\varepsilon r_0 kT$ tabulated in [2].

6.1.3 Ion pairing manifestation in semiconductor properties

The above simplified approach to the description of ion pairing was sufficient while ion pairs were detected experimentally in the variation of solidus lines, in impurity diffusion, and in electron scattering phenomena. A fairly detailed description of pairing effects can be found in the books [5, 6]. However, the formation of ion pairs proved to be a much more complicated process, as soon as its effect was revealed in charge carrier recombination in semiconductors.

To understand this phenomenon, let us turn to possible electron transitions in the band diagram of a semiconductor containing simple ions and ion pairs. Figure 6.2 shows schematically unpaired ion levels E_d and E_a ($r = \infty$), when interimpurity recombination is impossible. But if two ions produce a pair, the energy levels of each ion in the pair change, approaching the allowed band edges. The level of the positively charged donor goes up because the negative charge of the acceptor partner is now located at a shorter distance r, so that it is difficult for the donor to attach an electron. Similarly, it is difficult for the acceptor to attach a hole, and so its level in the forbidden band goes down.

Quantitatively, the level will change by the value

$$\Delta E = \frac{q^2}{\varepsilon r}, \qquad (6.1.22)$$

and since ion pairs may have a set of r values, a set of paired ion levels will

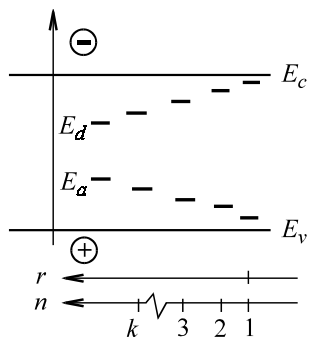

Figure 6.2. Paired ion energy levels as a function of interionic distance: 1,2,3, ... k – pair numbers with decreasing interionic distance r in a pair.

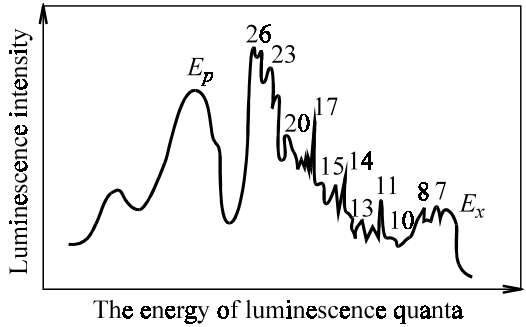

Figure 6.3. Photoluminescence spectrum of Zn-S ion pairs in GaP: the pair number decreases with decreasing interionic distance in a pair.

arise in the crystal. Some levels corresponding to small r can be pushed out of the forbidden band into allowed bands.

Since the ions in a pair are close to each other, the Coulomb interaction occurring between them may produce an electron–hole recombination. This is what is known as interimpurity recombination [7]. Any recombination process is preceded by generation of excess carriers, for which some energy is required. In the reverse process (recombination) this energy is released. Sometimes, it is released as light quanta. This radiative recombination may also occur in a system of paired ion levels, so the crystal will emit the whole spectrum of thin lines with the energies

$$h\nu = \left(E_{\mathrm{d}} - E_{\mathrm{a}}\right)_r . \qquad (6.1.23)$$

Experimentally, a luminescence spectrum of this kind was first observed in ZnS doped with Cu (acceptor) and Ga (donor) [8], as well as in GaP doped with S and Zn. Figure 6.3 illustrates a spectrum observed in that work. One can identify three spectral regions. In the low energy range, there is a wide band with a maximum at E_{s}, where some lines are unresolvable. Then there is a region of well-resolved lines and, finally, a region where this series of lines stops abruptly at ~2.31 eV.

Ion pairs with variable r can be conveniently described by an integral number m (coordination number), such that $m = 1$ corresponds to the most "packed" pair, $m = 2$ to a pair with the next degree of packing, and so on. The coordination numbers are shown in Figure 6.3 near the respective lines. One can see that experiments reveal ion pairs with a fairly large coordination number.

Since the time of that publication [8], the luminescence effect of ion pairing has been observed in many $A^{III}B^{V}$ and $A^{II}B^{VI}$ crystals. But one should bear in mind that interimpurity recombination can often occur in a nonradiative way. Besides, other recombination channels in a crystal may dominate interimpurity recombination, which makes experimental observation of spectra like the one in Figure 6.3 possible only under very strict conditions. This does not mean that, otherwise, ion pairs are absent. This does not mean either that the formation of all *m* ion pairs must always be taken into account. The physical properties of, and processes in, semiconductors have different sensitivities to the presence of ion pairs, so every particular situation should be analyzed to see whether a simple theory is suitable for the treatment of experimental data.

The above consideration of ion pairing implied a fairly high mobility of, at least, one ion in a pair. This is a very important condition because, otherwise, the ion distribution in a crystal will always be only random, without any correlation in their arrangement.

The kinetic feasibility of ion pairing is easy to evaluate by equating *r* to the diffusion pathway from the known relation

$$r = l = \sqrt{Dt}\ , \tag{6.1.24}$$

in which *D* is a diffusion coefficient and *t* is diffusion time.

A trivial conclusion follows from (6.1.24): mobile atoms are more liable to produce ion pairs. For this reason, ion pairing may be essential to the process of solid solution decomposition. Indeed, an excess impurity atom liberated from a lattice site occupies an interstitial position (and it may take an opposite sign). Many impurity ions in interstitial positions possess rather high diffusion coefficients [6]. For this reason, an interstitial ion has a high probability to encounter an oppositely charged ion substituting a lattice site. They interact to produce a stable ion pair. The interaction of impurity atoms with vacancies is a particular case of ion pairing between a vacancy and a donor of opposite signs.

6.2 POLYTROPIC IMPURITIES

In addition to ion pairs, a crystal, especially when doped heavily, may contain associated defects including impurity atoms. Such associations were first observed independently by two groups of researchers [10, 11] during the

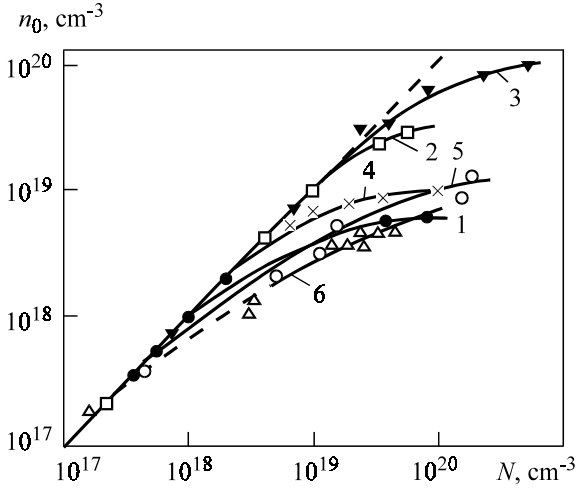

Figure 6.4. A comparison of concentrations of carriers n_0 and impurities N in semiconductors: 1 – Ge<Sb>; 2 – Ge<As>; 3 – Si<As>; 4 – GaAs<Te>; 5 – GaAs<Se>; 6 – GaAs<S>.

investigations of impurity ionization in germanium and silicon. Later, similar experiments were carried out with binary $A^{III}B^{V}$ compounds [12–14]. The experimentally observed effect (Figure 6.4) is that, beginning with some threshold impurity concentrations, there is a regular discrepancy between the chemical concentration of the doping impurity, N, and the concentration of charge carriers (electrons), n, found from Hall coefficient measurements.

These results convincingly indicate that not all of the donor impurity present at a high concentration in the crystal enters a substitution solid solution but exists in several states simultaneously. Since the problem of exact identification of the state of impurity atoms in heavily doped crystals has long been debated, such atoms even acquired a special name—polytropic (multistate) impurities. Some suggestions have been made [15] concerning the factors producing polytropic impurities, among which the main but poorly identifiable factors are the common second-phase inclusions and the formation of impurity associates in the still one-phase solution.

Let us consider the thermodynamics of associate formation, as was done in [16]. Suppose m donor atoms (D) unite with a lattice defect C, or without it, to form a complex Z^q possessing charge q. This process can be represented as the reaction

$$mD^{+} + (m-q)e^{-} + C^{0} \rightleftarrows Z^{q}. \qquad (6.2.1)$$

If equilibrium is established at a certain temperature T, at which electrons obey Boltzmann's statistics, the equilibrium concentrations of the reactants satisfy the active mass law

$$\left[Z^{q}\right] = K(T)\left[D^{+}\right]^{m} n^{m-q}. \qquad (6.2.2)$$

Here, we are making use of the fact that the concentration of intrinsic defects $[C^0]$ depends only on temperature, so it has been included in the equilibrium constant $K(T)$.

Donors are assumed to be completely ionized at all temperatures. The neutrality equation for temperature T is

$$n(T) = \left[D^{+}\right] + q\left[Z^{q}\right] + p. \qquad (6.2.3)$$

At room temperature, the concentration of holes in this equation can be neglected:

$$n_0 = \left[D^{+}\right] + q\left[Z^{q}\right]. \qquad (6.2.4)$$

Substituting (6.2.3) into (6.2.4) and using the relation $p \times n = n_i^2$, we find that the equilibrium electron concentration $n(T)$ at high temperature T is related to the concentration n_0, measured at room temperature T_0, by the expression

$$n = \frac{1}{2}\left[n_0 + \sqrt{n_0^2 + 4n_i^2}\right], \qquad (6.2.5)$$

where n_i is the intrinsic electron concentration at temperature T.

If the condition

$$n_0^2 >> 4n_i^2, \qquad (6.2.6)$$

is fulfilled, expression (6.2.5) will have a simple form:

$$n = n_0. \qquad (6.2.7)$$

Total chemical concentration of donor impurity, N, is a sum of concentrations of free donors and impurity atoms involved in defect associates, often termed as complexes:

$$N = \left[D^{+}\right] + m\left[Z^{q}\right]. \tag{6.2.8}$$

By eliminating $[D^{+}]$, $[Z^{q}]$, and n from equations (5.3.2)–(5.3.5), we get a general relation between n_0 and N:

$$N - n_0 = K(T)\frac{n_0^{m-q}\left(mn_0 - qN\right)^{m}}{\left(m-q\right)^{m-1}}. \tag{6.2.9}$$

We will now consider some particular cases which seem to be most probable.

A neutral complex. By putting $q = 0$ in (6.2.9), we find

$$N = n_0 + mK(T)n_0^{2m}. \tag{6.2.10}$$

It follows from (6.2.10) that $N : N = n_0$ in the low concentration range, while at high concentrations N, we have

$$N = mK(T)n_0^{2m}$$

or

$$\log n_0 = -\frac{1}{2m}\log mK + \frac{1}{2m}\log N\ . \tag{6.2.11}$$

Thus, the function $n_0(N)$ in the logarithmic scale will have a straight asymptote $1/2m$, as is shown by curve 1 in Figure 6.5.

A charged complex. We will consider the simplest charged complex, in which a donor atom unites with a lattice defect. Putting $m = 1$ in (6.2.9), we find

$$N = n_0\frac{1 + K(T)n_0^{1-q}}{1 + qK(T)n_0^{1-q}}. \tag{6.2.12}$$

Positively charged complexes ($q = +1$) can be excluded from consideration at once, because, at best, they do not decrease the electron concentration n.

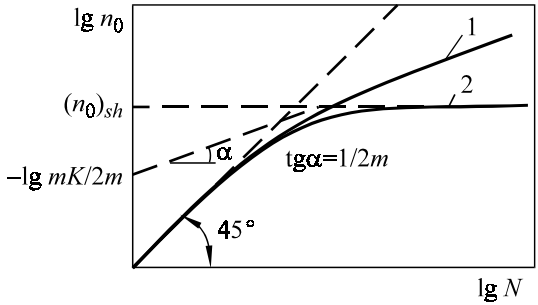

Figure 6.5. The $n_0(N)$ curves for impurity complexation: 1 – formation of neutral complexes; 2 – formation of singly charged complexes.

Generally, their formation at $q > +1$ would lead to an excess of the concentration n over the doping impurity concentration N, but this has never been observed experimentally.

If a complex has a negative charge, then we have

$$q = -|q| \qquad \text{and} \qquad N = n_0 \frac{1 + K(T) n_0^{1+|q|}}{1 - |q| K(T) n_0^{1+|q|}}. \qquad (6.2.13)$$

At low concentrations N, the curve looks as $n = N$. At large N, n_0 reaches saturation (curve 2 in Figure 6.5) to give a shoulder:

$$(n_0)_{\text{sh}} = \sqrt[1+|q|]{\frac{1}{|q| K(T)}}. \qquad (6.2.14)$$

In particular, for $q = -1$, we have

$$(n_0)_{\text{sh}} = \sqrt{\frac{1}{K(T)}}. \qquad (6.2.15)$$

Therefore, the experimental $n(N)$ curves can be accounted for by impurity complexation.

However, the curves will have the same shape when impurity atoms are located in the second-phase inclusions. Therefore, the inequality $n < N$ cannot be considered to indicate unambiguously the formation of complexes.

In reality, the presence of polytropic impurities implies the existence of impurity states which remain, chemically, in the one-phase solution but are

the product of impurity associations. Such impurity complexes manifest themselves in many physical phenomena. For example, their presence in a crystal is to affect the relaxation time of charge carriers, τ, in their scattering, as well as the mobility of electrons and holes, which is directly related to τ: $u = e\tau / m$.

Suppose M atoms have united into Z^q associates. Since the probability of electron scattering is proportional to the number of scattering centers and their square charge, the contribution of M singly ionized free atoms to the mobility will be inversely proportional to M. But M atoms associated into complexes change the mobility in inverse proportion to Mq^2/m, where m is the number of impurity atoms in a complex. The ratio of these quantities, q^2/m, defines the mobility change due to complexation.

There may be three situations concerning the mobility:

– at $q^2/m > 1$, the mobility drops;
– at $q^2/m = 1$, the mobility does not change;
– at $q^2/m < 1$, the mobility rises.

It follows that the formation of neutral complexes ($q = 0$) must increase the electron mobility, while the formation of a complex containing one impurity atom ($m = 1$) must leave it, at best (i.e., at $q = –1$), unchanged; but the mobility generally decreases. If a singly charged complex contains more than one impurity atom ($m > 1$), the electron mobility will increase.

The effects of impurity complexes manifest themselves in many other physical properties of semiconductors, in particular in their heat conductivity [27], mechanical properties [18], diffusion [19], additional optical absorption, for example in GaAs [20]. Being centers of nonradiative recombination, such complexes reduce the quantum yield of photo- and electroluminescence [21].

We will not dwell on the effects of complexes on the physical properties of crystals, because this would take too much space. We will only discuss the available approaches to the explanation of the nature of impurity complexes in heavily doped semiconductors. There are two approaches to this problem. The chemical approach considers the interaction thermodynamics of two impurity atoms. Like in the case of ion pairing, a crystal is regarded as a solution which involves interactions producing AA or BB ion pairs, in addition to AB pairs. Several workers have emphasized the possibility of formation of more complex polymeric groups consisting of more than two atoms [22–24]. The thermodynamic ratios discussed above with reference to ion pairs are, in principle, applicable to complexes. In this case, the enthalpy variation ΔH cannot, of course, be due to a Coulomb attraction. Formally, this approach permits the functional relation $n = f(N)$ to

be found and the theoretical curve to be fitted to the experimental one by choosing the appropriate value of *m* for the number of atoms in a complex, A_m [15, 25].

The formation of impurity complexes in $A^{III}B^{V}$ semiconductors was discussed in [26, 27] in terms of regular solution theory. The authors calculated the Gibbs free energy variation for a large number of reactions in $A^{III}B^{V}$-impurity systems and showed that the decrease in electrical activity in the Te, Se, S series (Figure 6.4) correlated with the growing chemical strength in the A_2Te_3, A_2Se_3, A_2S_3 series. Compounds of Zn and Cd acceptors with B^{V} elements appeared to have a lower strength than $A^{III}B^{V}$ compounds themselves, and this fact correlates with the absence of polytropic *p*-semiconductors.

The chemical approach was able to provide only correlations or, at best, qualitative descriptions of experimental facts. Its principal limitation was that the nature of forces keeping ions of like sign together remained unclear. To explain why donors of like sign could keep together, an electrolytic model was proposed assuming the presence of an oppositely charged ion at the coordination sphere center. It is only a vacancy that may be the central ion in an *n*-crystal, because one can hardly imagine the presence of an acceptor impurity in quantities commensurate with the donor concentration.

Therefore, if a vacancy in Ge and Si is considered to have a four-fold charge, the crystallographic model of a donor complex coherent with the host lattice will be VD_4. This makes it clear why such complexes are absent when Ge and Si are doped with acceptors: vacancies and acceptors then have the same sign.

The suggested model of a complex as a hollow tetrahedron accounts for the fact that a further increase in the donor concentration leads to the formation of second-phase seeds. A comparison between the tetrahedral VAs_4 group and the arsenic structure shows that interatomic distances change but little, and the mutual orientation of atomic layers is preserved. The rearrangement reduces to a by-pair approach of layers, made up of “polymerized” tetrahedra, to each other.

This model also offers only a qualitative explanation; besides, it raises some doubts. First, a vacancy is to be bonded to four donor ions, which seems, intuitively, very unlikely. Second, the state diagrams indicate that the equilibrium phase in Ge and Si doped with arsenic should be GeAs or SiAs, respectively, rather than metallic As. Finally, a four charged vacancy appears quite unlikely. Of course, if a vacancy is doubly charged, this will give rise to the VD_2-type complexes, and the reasoning concerning their polymerization will remain valid. In any case, the phenomenon of polytropy is currently attributed to the formation of associates, in which donor atoms (probably, ions) interact with charged vacancies.

The necessary number of vacancies for complexation to occur seems to be always present in a crystal. These may be vacancies which were in equilibrium at high temperatures, and all processes of their binding to donors occur on crystal cooling from the high growth temperatures (or in a special thermal treatment). Semiconductor compounds of the $A^{III}B^{V}$ group may contain vacancies produced due to deviations from the real crystal stoichiometry. Finally, these may be vacancies formed by sublimation of excess atoms from the lattice sites.

An interesting mechanism of vacancy interaction with impurity ions may take place when the number of vacancies in a crystal is very small. This mechanism may be termed a "vacancy pump." At any given moment of time, the number of vacancies and, hence, of complexes, is quite small, but these complexes migrate to some sinks, say, to dislocations. When an impurity atom reaches a sink, it becomes fixed to it, whereas the vacancy disappears. The disappearance of a vacancy in one place gives rise to a vacancy in another place. Migrating through the crystal, a vacancy has a chance to encounter an impurity atom, to interact with it, and to "tow" it to a sink; then the process is repeated.

This interaction model was suggested in [29] to describe decomposition of metallic solid solutions, but it is quite applicable to semiconductors because a vacancy–impurity complex diffuses much faster than a single impurity atom or a vacancy [30].

6.3 COMPLEXATION THERMODYNAMICS IN A SEMICONDUCTOR COMPOUND

Let us consider a semiconductor compound in equilibrium with the ambient. The probability for point defects to associate and produce secondary complex defects increases with decreasing temperature. This is due to electrostatic and elastic interactions, on the one hand, and to the fairly high mobility of vacancies and interstitial atoms at normal temperatures, on the other. Migrating through a crystal, a vacancy may come to occupy a site next to an impurity atom. They may interact to produce a donor–acceptor complex [31–34]. The diversity of such defects increases with the number of semiconductor constituents.

It is believed that a complex is produced by an electrostatic interaction between oppositely charged defects. But the association of defects decreases the lattice stress and is energetically profitable. Evidently, a gallium vacancy forms more stable complexes with impurities located in the anion sublattice,

because it comes closer to such an impurity. As for IV-group elements producing donor states by substituting cation atoms, a vacancy cannot come up as close to an impurity, so the complex produced is less stable.

High concentrations of doping impurities may stimulate the formation of compounds which can affect the solubility of these impurities. Such processes can occur at heterojunctions and during the formation of natural oxides on the semiconductor surface.

Thermodynamically, the complexation of intrinsic nonstoichiometric defects and molecules statistically distributed in the crystal bulk and not leading to the second phase sublimation can be described by the same theory. However, the details of defect formation theory are not sufficiently elaborated, especially for semiconductor compounds. In this case, it is hard to employ a method based on the active mass law in view of the large number of reactions to be taken into account. The difficulties associated with the choice of reactions and unknown equilibrium constants rise, and so it is preferable to minimize the Gibbs potential of a defect crystal.

This problem was generally solved for simple non-interacting defects [35, 36]. In [36] the problem solution was generalized for a crystal containing simple defects and complexes consisting of several intrinsic defects and impurity atoms. The results obtained allow the treatment of various situations, and this will be demonstrated by several illustrations.

Complexation occurs during crystal cooling from the growth temperature. The established concentrations of host atoms do not change because the exchange of atoms between the crystal and the ambient is limited. The concentration of point defects varies due to their association and interaction with the sinks (dislocations, grain boundaries, etc.).

Consider an idealized system consisting of a crystal AB containing oppositely charged impurity atoms and vacancies. Let us calculate the equilibrium concentration of donor–acceptor complexes, minimizing the Gibbs potential of the system. A real crystal containing simple point defects and complexes can be regarded as an ensemble of particles statistically distributed over definite positions. A particle is understood as a lattice site or interstice, or their combination. Then this particle is a simple defect occupying one position, and the complex occupies several positions simultataneously.

The system of crystal particles is in equilibrium with the ambient, so that the total number of particles remains the same. Concentrations of complexes of various types will be denoted as N_c. Every complex includes several simple particles. The number of simple particles of one kind in a complex will be denoted as m.

Particle concentrations are not independent but are related by the conservation laws for the number of positions in every sublattice

$$\varphi^{\beta} = N^{\beta} - \sum_{\alpha}\left(N_{\alpha}^{\beta} + \sum_{c} N_c m_{\alpha c}^{\beta} \right) = 0\,, \qquad (6.3.1)$$

the number of atoms of each kind

$$\varphi_{\alpha} = N_{\alpha} - \sum_{\beta}\left(N_{\alpha}^{\beta} + \sum_{c} N_c m_{\alpha c}^{\beta} \right) = 0\,, \qquad (6.3.2)$$

the total number of defects of a particular type

$$\varphi_{\alpha}^{\beta} = N_{\alpha\mathrm{tot}}^{\beta} - N_{\alpha}^{\beta} - \sum_{c} N_c m_{\alpha c}^{\beta} = 0\,, \qquad (6.3.3)$$

the total number of particles in the system

$$\varphi^{N} = N - \sum_{\alpha,\beta}\left(N_{\alpha}^{\beta} + \sum_{c} N_c m_{\alpha c}^{\beta} \right), \qquad (6.3.4)$$

and the electroneutrality condition

$$\varphi_{\mathrm{e}} = n - p + \left(\sum_{\alpha,\beta} n_{\alpha}^{\beta} - \sum_{\alpha,\beta} N_{\alpha}^{\beta} d_{\alpha}^{\beta} + \sum_{c} n_c - \sum_{c} N_c d_c \right)\left(1 - \delta_{\alpha}^{\beta}\right). \qquad (6.3.5)$$

These laws permit allowance for all aspects of a statistical interaction in the system. Note also that there are two new conservation laws here, as compared with free defects, which are to be taken into account when treating complexation processes. The two laws are expressed by (6.3.3), which allows for the involvement of structural elements of a real crystal to various complexes, and (6.3.4), which allows for the fact that a complex can occupy positions in different sublattices and types of crystal interstices.

The number of states with the same number of particles of each kind having the same free energy but differing in the spatial arrangement of particles determines the thermodynamic probability. States differing in the rearrangement of one kind of particle are taken to be identical; therefore, such rearrangements must be excluded from thermodynamic probability calcula-

tions. In addition, a complex can occupy a lattice position in several equivalent ways by varying its orientation relative to the crystal lattice directions. The number of such orientations is known as *degeneracy multiplicity*.

Let us discuss a possible method of degeneracy multiplicity calculation. The set of symmetry elements of an ideal lattice forms a point group. The presence of a defect in the crystal lattice reduces the symmetry. However, the set of symmetry elements of this lattice forms a subgroup of the point group.

As in the case of free defects, the lattice sites will be occupied by atoms. The crystal energy states will be occupied by electrons and holes. To calculate the thermodynamic probability for the rearrangement of complexes, let us do the following. We will take one of the atoms as the base one and place it at different lattice sites. The other atoms of this complex will arrange themselves automatically. The various orientations of the complex in the lattice and, hence, the concurrent growth of thermodynamic probability, is taken into account by the degeneracy multiplicity coefficient derived above. The expression for the thermodynamic probability will have the factor $Y_c^{N_c}$.

The account of the degeneracy multiplicity of energy states associated with complexes gives rise to factors R_c and r_c, and the account of electron distribution over complexes gives rise to factors $N_c![(N_c - n_c)!n_c!]^{-1}$. Therefore, we can write the thermodynamic probability for a semiconductor compound containing simple defects and associates as

$$W = \frac{\left(N_{\mathrm{C}}\right)^n \left(N_{\mathrm{V}}\right)^p N!}{n!\,p!} \prod_{\alpha,\beta} \frac{\left(r_\alpha^\beta\right)^{n_\alpha^\beta} \left(R_\alpha^\beta\right)^{N_\alpha^\beta - n_\alpha^\beta}}{n_\alpha^\beta!\left(N_\alpha^\beta - n_\alpha^\beta\right)!\,N_\alpha^\beta!} \times \times \prod_c \frac{\left(Y_c\right)^{N_c} \left(r_c\right)^{n_c} \left(R_c\right)^{N_c - n_c}}{n_c!\left(N_c - n_c\right)!} . \tag{6.3.6}$$

The Gibbs potential of the system under consideration consists of the thermal and configuration components

$$\Delta G = \Delta G_{\mathrm{T}} - \Theta \ln W , \tag{6.3.7}$$

where $\Theta = kT$, ΔG_{T} is the thermal component of Gibbs potential, including defect formation enthalpy and thermal entropy due to the change in the phonon spectral density of an ideal crystal during defect formation.

Indeed, a real crystal lattice can only be matched by the same operations as the ideal lattice, but the number of operations will be smaller because

some symmetry elements will be eliminated by the symmetry reduction. All symmetry operations of a particular subgroup leave the complex *in situ*. The thermodynamic states obtained by these transformations are identical. The orientation of a complex can be changed using the point group symmetry operations not included in the real lattice subgroup. It can be concluded from Lagrange's theorem that the degeneracy multiplicity of defects will be equal to the real lattice subgroup index.

As an illustration, consider a donor–acceptor complex consisting of a vacancy and an impurity atom occupying adjacent sites. Most semiconductors form high symmetry ideal lattices of the O_n or T_d isomorphous groups with the order 24. It is easy to see that a real lattice with this complex possesses the point group C_{3V} with the order 6.

The degeneracy multiplicity factor (Y_c) is equal to the subgroup index which is the quotient of division of the ideal crystal group order by the real crystal subgroup order. In the illustration just given, this factor is 4. For a simple defect, this result can be obtained without resorting to group theory. The appearance of degeneracy multiplicity due to the spatial orientation of the complex as a whole is also a specificity of defect association.

To conclude the discussion of statistical interaction during complexation, let us consider the procedure of finding the thermodynamic probability. The thermal component can be conveniently expressed by introducing what is known as Gibbs partial potentials

$$g_\alpha = H_\alpha^\beta - S_{\alpha T}^\beta T\,,$$

$$g_c = H_c - S_{cT} T\,,$$

where H_c is enthalpy and S_{cT} is thermal entropy of complexation. These parameters take into account interactions and changes in the lattice vibrations during complexation.

Crystal energy increases owing to defect formation and charge carrier excitation. The former is allowed for by summing up all products of Gibbs partial potentials and the concentration of respective defects, while the latter is taken into account by including the products of electron and hole concentrations and respective transition energies. The electron transition from the valence to the conduction band increases free energy by the forbidden gap energy. The number of such transitions is equal to that of free holes, except for the electrons leaving the valence band to occupy acceptor states. The transition of electrons from donor states to the conduction band and from the valence band to acceptor states also increases the system free energy. Considering the foregoing, the thermal component of Gibbs potential is

$$\Delta G_T = \sum_{\alpha,\beta} N_\alpha^\beta g_\alpha^\beta + \sum_c N_c g_c + \\ + \left\{ E_g \left[p - \sum_{\alpha,\beta} n_\alpha^\beta \left(1 - d_\alpha^\beta\right) - \sum_c n_c (1 - d_c) \right] + \sum_{\alpha,\beta} \left(N_\alpha^\beta d_\alpha^\beta + n_\alpha^\beta \right) . \right. \\ \left. + \sum_c E_c \left(N_c d_c + n_c \right) \right\} \left(1 - \delta_\alpha^\beta\right) \qquad (6.3.8)$$

To find equilibrium concentrations of defects, one should minimize the functional

$$\Phi = \Delta G_T - \Theta \ln W + \sum_\beta \lambda^\beta \varphi^\beta + \sum_\alpha \lambda_\alpha \varphi_\alpha + \sum_{\alpha,\beta} \lambda_\alpha^\beta \varphi_\alpha^\beta + \\ + \lambda_e \varphi_e + \lambda^N \varphi_N , \qquad (6.3.9)$$

where λ_α, λ^β, $\lambda_\alpha{}^\beta$, λ_e, and λ^N are Lagrange indeterminate factors. The meaning of λ_α, λ^β, and λ_e was defined in Section 1.2, but only the first and the third factors retain their meaning.

The appearance of factors $\lambda_\alpha{}^\beta$ and λ^N is related to the conservation laws to be taken into account in complexation. If there is a possibility to neglect these processes, $\lambda_\alpha{}^\beta$ can be made to tend to zero. This becomes possible, for example, at high temperatures when the probability of defect association is low.

Let us now substitute expressions (6.3.4) through (6.3.7) into (6.3.9) and write explicitly the functional to be minimized

$$\Phi = \sum N_\alpha^\beta g_\alpha^{\beta'} + \sum N_c g_c' + E_g \left[p - \sum n_\alpha^\beta \left(1 - d_\alpha^\beta\right) - \sum n_c (1 - d_c) \right] \\ + \sum \varepsilon_\alpha^\beta \left[\left(N_\alpha^\beta - n_\alpha^\beta \right) d_\alpha^\beta + n_\alpha^\beta \left(1 - d_\alpha^\beta\right) \right] \left(1 - \delta_\alpha^\beta\right) + \sum \varepsilon_c \left(N_c - n_c \right) d_c \\ + \sum \varepsilon_c n_c (1 - d_c) - \Theta \{ n \log N_c + p \log N_v + N \log N - N - p \log p \\ - p - n \log n + n + \sum \left[n_\alpha^\beta \log r_\alpha^\beta + \left(N_\alpha^\beta - n_\alpha^\beta \right) \log R_\alpha^\beta - n_\alpha^\beta \log n_\alpha^\beta + n_\alpha^\beta \right. \\ \left. - \left(N_\alpha^\beta - n_\alpha^\beta \right) \log \left(N_\alpha^\beta - n_\alpha^\beta \right) + N_\alpha^\beta - n_\alpha^\beta - N_\beta^\beta \log N_\beta^\beta + N_\beta^\beta \right] \\ + \sum \left[N_c \log Y_c + n_c \log r_c + \left(N_c - n_c \right) \log R_c - n_c \log n_c + n_c \right. \\ \left. - \left(N_c - n_c \right) \log \left(N_c - n_c \right) \right] + \left[N_c - n_c \right] \} \\ + \sum \lambda^\beta \left[N^\beta - \sum \left(N_\alpha^\beta + \sum N_c m_{\alpha c}^\beta \right) \right] + \sum \lambda_\alpha \left[N_\alpha - \sum \left(N_\alpha^\beta + \sum N_c m_{\alpha c}^\beta \right) \right] \\ + \sum \lambda_\alpha^\beta \left(N_{\alpha \mathrm{tot}}^\beta - N_\alpha^\beta - \sum N_c m_{\alpha c}^\beta \right)$$

$$+\lambda_e\left[n-p+\sum\left(n_\alpha^\beta-N_\alpha^\beta d_\alpha^\beta\right)+\sum\left(n_c-N_c d_c\right)\right]$$
$$+\lambda^N\left(N-\sum N_\beta^\beta-\sum N_\alpha^\beta-\sum N_c\right) \quad . \tag{6.3.10}$$

To find equilibrium defect concentrations, it is necessary to take partial derivatives of the functional (6.3.10). Some of them make the physical meaning of the Lagrange factors in this equation clear. As a result, we find

$$\lambda_e = E_F, \qquad \lambda_\alpha = \mu_\alpha . \tag{6.3.11}$$

The use of the conservation laws for structural elements in complexation and for the total number of particles changes the factor λ^β. It follows from

$$\frac{\partial\Phi}{\partial N_\beta^\beta}=0 \qquad \text{that} \qquad \lambda^\beta=\mu_\beta-\Theta\log N_\beta^\beta-\lambda^N . \tag{6.3.12}$$

The physical meaning of λ^N is found by calculating the derivative with respect to the total number of particles. It follows from

$$\frac{\partial\Phi}{\partial N}=0 \qquad \text{that} \qquad \lambda^N=\Theta\log N . \tag{6.3.13}$$

By substituting (6.3.12) into (6.3.13), we find

$$\lambda^\beta=-\mu^\beta+\Theta\log\left(N_\beta^\beta/N\right). \tag{6.3.14}$$

Equilibrium defect concentrations are found from the respective derivatives being equal to zero. Begin with point defects:

$$\frac{\partial\Phi}{\partial N_\alpha^\beta}=g_\alpha^\beta+\varepsilon_\alpha^\beta d_\alpha^\beta-\Theta\left[\log R_\alpha^\beta-\log\left(N_\alpha^\beta-n_\alpha^\beta\right)\right]- \\ -\lambda^\beta-\lambda_\alpha-\lambda_\alpha^\beta-\lambda_e d_\alpha^\beta-\lambda^N \quad , \tag{6.3.15}$$

$$\frac{\partial\Phi}{\partial n_\alpha^\beta}=-E_g\left(1-d_\alpha^\beta\right)-\varepsilon_\alpha^\beta d_\alpha^\beta- \\ -\Theta\left[\log r_\alpha^\beta-\log R_\alpha^\beta+\log\left(N_\alpha^\beta-n_\alpha^\beta\right)-\log n_\alpha^\beta\right]+\lambda_e \quad . \tag{6.3.16}$$

Using these expressions and (6.3.10) through (6.3.14), we find

$$N_\alpha^\beta = \frac{a_\alpha}{a_\beta} N_\beta^\beta \exp\left(\frac{g_\alpha^\beta - \lambda_\alpha^\beta}{\Theta}\right)\left[R_\alpha^\beta \exp\left(\frac{\left(E_F - \varepsilon_\alpha^\beta\right)d_\alpha^\beta}{\Theta}\right) + r_\alpha^\beta \exp\left(\frac{\left(E_g - \varepsilon_\alpha^\beta - E_F\right)\left(1 - d_\alpha^\beta\right)}{\Theta}\right)\right], \quad (6.3.17)$$

$$n_\alpha^\beta = r_\alpha^\beta \frac{a_\alpha}{a_\beta} N_\beta^\beta \exp\left(\frac{g_\alpha^\beta - \lambda_\alpha^\beta}{\Theta}\right)\exp\left(\frac{\left(E_g - \varepsilon_\alpha^\beta - E_F\right)\left(1 - d_\alpha^\beta\right)}{\Theta}\right), \quad (6.3.18)$$

where $g_\alpha{}^\beta = g_\alpha^{\beta'} + \mu_\beta{}^0 - \mu_\alpha{}^0$. The account of complexation processes has given another exponential factor indicating the fraction of free point defects. This factor is

$$\exp\left(\lambda_{\alpha c}^\beta\right) = \frac{N_\alpha^\beta}{N_{\alpha tot}}, \quad (6.3.19)$$

where $N_\alpha{}^\beta$ is the number of free point defects and $N_{\alpha tot}{}^\beta$ is the total number of defects, including those bound in complexes.

One should note that the total number of all sorts of particles, N, was reduced and is absent from (6.3.18). This is because we are discussing point defects which occupy positions only in one sublattice or one type of interstice.

Let us find derivatives related to the concentration of complexes:

$$\frac{\partial \Phi}{\partial N_c} = g_c' + E_c d_c - \Theta\left[\log Y_c - \log R_c - \log\left(N_c - n_c\right)\right] - \lambda^N - \sum_\beta \lambda^\beta \sum_\alpha m_{\alpha c}^\beta - \sum_\alpha \lambda_\alpha \sum_\beta m_{\alpha c}^\beta - \sum_\alpha \lambda_\alpha^\beta \sum_\beta m_{\alpha c}^\beta. \quad (6.3.20)$$

As was mentioned above, $m_{\alpha c}{}^\beta$ is the number of one type of structural element in a complex. Consequently, if some constituents of a sublattice or of one type of interstice are not involved in complexation, then $m_{\alpha c}{}^\beta$ are

equal to zero. For this reason, the indices in (6.3.20) acquire values meaningful only for a particular complex:

$$\sum_{\beta}\lambda^{\beta}\sum_{\alpha}m_{\alpha c}^{\beta}+\sum_{\alpha}\lambda_{\alpha}\sum_{\beta}m_{\alpha c}^{\beta}+\lambda^{N}= \\ =\sum_{\alpha,\beta}\left[\mu_{\alpha}-\mu_{\beta}+\log\left(\frac{a_{\beta}}{a_{\alpha}}\right)+\log\left(\frac{N_{\beta}^{\beta}}{N}\right)\right]-\Theta\log N \quad . \qquad (6.3.21)$$

From (6.3.20) and (6.3.21), we have

$$N_c=Y_c N^{p}\prod_{\alpha,\beta}\left(\frac{N_{\alpha}^{\beta}}{N_{\alpha tot}^{\beta}}\right)^{m_{\alpha c}^{\beta}}\exp\left(-\frac{g_c}{\Theta}\right)f_F^{-1}, \qquad (6.3.22)$$

where

$$N^{p}=N\prod_{\alpha,\beta}\left(\frac{a_{\alpha}N_{\beta}^{\beta}}{a_{\beta}N}\right)^{m_{\alpha c}^{\beta}},$$

$$g_c=g_c'-\sum_{\alpha,\beta}\left(\mu_{\alpha}^{0}-\mu_{\beta}^{0}\right)m_{\alpha c}^{\beta},$$

$$f_F^{-1}=R_c\exp\left[\left(E_F-\varepsilon_c\right)\frac{d_c}{\Theta}\right]+r_c\exp\left[\left(E_g-E_F-\varepsilon_c\right)\frac{1-d_c}{\Theta}\right]. \qquad (6.3.23)$$

The concentration of electrons bound by complexes is

$$n_c=Y_c r_c N^{p}\prod_{\alpha,\beta}\left(\frac{N_{\alpha}^{\beta}}{N_{\alpha tot}^{\beta}}\right)^{m_{\alpha c}^{\beta}}\times \\ \times\exp\left(-\frac{g_c}{\Theta}\right)\exp\frac{\left(E_g-E_F-\varepsilon_c\right)\left(1-d_c\right)}{\Theta} \quad . \qquad (6.3.24)$$

Formula (6.3.23) describes the concentration of complexes of any reasonable complexity and the donor and acceptor configurations. The concen-

tration is proportional to the concentration of constituents of a complex to a power equal to the number of these constituents. Formulas (6.3.17) and (6.3.23) permit the analysis of crystal properties, using fairly complicated defect formation models.

6.4 IMPURITY–VACANCY COMPLEXES IN $A^{III}B^{V}$ COMPOUNDS

Cation vacancies in $A^{III}B^{V}$ compounds have a fairly high mobility and are multiply charged acceptors [32, 37]. Consequently, the probability of interaction between these vacancies and donor impurities is quite high. Chokralsky-grown *n*-GaAs samples possessing a high concentration of vacancies at the growth temperature have a luminescence spectrum with a maximum at 1.17–1.20 eV at 77 K. In crystals doped with Te, this band is attributed to the complex of a Ga vacancy and a Te donor atom substituting As at the adjacent lattice site [32]. Similar centers were identified in GaP [38, 39]. The authors of [40] noticed that the symmetry of the complex contradicted the model of two defects at neighboring sites in the anion and cation sublattices. The symmetry was found to be lower and multiwedge, rather than three-wedge. A detailed study of polarized luminescence in uniaxial deformation led to the conclusion that the symmetry reduction was due to additional distortion by the Yan–Teller interaction, rather than to the attachment of a third constituent [41, 42]. The same authors observed an alignment of complexes along a certain crystal direction under the deformation action and studied the tree-particle complexes $V_{As}V_{Ga}Sn_{As}$ and $V_{As}V_{Ga}Te_{As}$. In spite of the Yan–Teller interaction reducing the symmetry, the first type complexes are thermodynamically simple two-particle complexes.

Let us discuss the complexation of charged vacancies and impurities of opposite signs. These structural elements are attracted to each other by Coulomb forces to produce a neutral donor–acceptor complex. Suppose that a particular vacancy interacts with a particular impurity. The numbers of free vacancies, complexes, and impurities are related as

$$N_V + N_c = N_{tot}, \qquad N_{im} + N_c = N_{im\,tot}, \tag{6.4.1}$$

where N_{tot} is the total number of vacancies after the crystal growth and $N_{im\,tot}$ is the total number of impurities. From formula (6.3.23), we have for a donor-type complex:

$$N_c = 4N \frac{N_V N_{im}}{N_{V\,tot} N_{im\,tot}} \exp\left(-\frac{g_c}{\Theta}\right)\left[\frac{2N_c}{n}\exp\left(-\frac{\varepsilon_c}{\Theta}\right)+1\right], \quad (6.4.2)$$

where N is the total number of positions and factor 4 is a degeneracy factor.

Suppose that the concentration of impurity atoms is larger than that of vacancies. The condition $N_{im\ tot} >> N_{V\ tot}$ automatically yields $N_{im\ tot} \approx N_{im}$. From the conservation law for the number of vacancies, we have

$$N_c = \frac{N' N_{V\,tot}}{N' + N_{V\,tot}}, \quad (6.4.3)$$

where

$$N' = 4N \exp\left(-\frac{g_c}{\Theta}\right)\left[\frac{2N_c}{n}\exp\left(-\frac{\varepsilon_c}{\Theta}\right)+1\right].$$

The interaction of partners in a donor–acceptor complex is mainly of the Coulomb type. If an *n*-semiconductor has a shallow donor and a deep acceptor, the former will be charged positively and the latter negatively; as a result, they will be attracted to each other. However, the interaction enthalpy is not high [43]. Besides, it is negative because the interaction results in attraction. For this case, the following inequalities are satisfied: $N' >> N_{V\ tot}$ and $N_c \approx N_{V\ tot}$. In other words, all vacancies produced during crystallization are bound into complexes.

Therefore, when the concentration of shallow donors in an *n*-semiconductor is higher than that of vacancies, there should be no free vacancies. Indeed, electron paramagnetic resonance studies show that an EPR signal which could be attributed to vacancies is only observed in a *p*-semiconductor [32], which agrees with the above hypothesis.

Another fact is noteworthy. The main contribution to the formation of a neutral complex is made by acceptor vacancies. Their concentration can be written as

$$N_V^A = \frac{1}{a_A} N_A^A \exp\left(-\frac{g_V^A}{\Theta}\right)\left[\frac{2n}{N_c}\exp\left(\frac{E_g - \varepsilon_V^A}{\Theta}\right)+1\right]. \quad (6.4.4)$$

In the first approximation for material doped with one impurity type, the following equality is valid at high temperature:

$$n = N_{\text{im}}\,. \tag{6.4.5}$$

Then, substituting (6.4.4) and (6.4.5) into (6.4.2), we have

$$N_{\text{c}} = 4N^2 \frac{N_{\text{im}}}{a_{\text{A}} N_{\text{V tot}} N_{\text{tot}}} \exp\left(-\frac{g_{\text{V}}^{\text{A}}}{\Theta}\right)\left[\frac{2N_{\text{c}}}{N_{\text{im}}}\exp\left(-\frac{\varepsilon_{\text{c}}}{\Theta}\right)+1\right]\times$$
$$\times\left[\frac{2N_{\text{im}}}{N_{\text{c}}}\exp\left(\frac{\varepsilon_{\text{g}}-\varepsilon_{\text{V}}^{\text{A}}}{\Theta}\right)+1\right]. \tag{6.4.6}$$

At different proportions of the terms between the brackets of the last co-factor, the concentration of complexes is proportional either to that of free impurities or to their square concentration. The quantity $\varepsilon_{\text{V}}^{\text{A}}$ can be evaluated from the point of transition between the linear and quadratic dependences. This concentration behavior of complexes agrees with the data of [44].

Of the great diversity of impurity–vacancy complexes in $A^{III}B^{V}$ compounds, it is worth mentioning one of the Cr impurity states in GaAs. In addition to the states discussed in Section 2.2, an associate containing a deep donor of the type $Cr_{Ga}^{-}D_{As}^{+}$ was observed, with Cr_{Ga}^{-} having the d^4-configuration [53]. This associate has the cubic (not trigonal) symmetry. The configuration diagram of this associate defect, borrowed from [53], is shown in Figure 6.6*a*. The authors of this model attributed the role of D_{As}^{+} to the anion vacancy *V*. Both components of this associate create deep t_2-symmetry levels in the GaAs forbidden gap (Figure 6.6*b*). The Cr^{-} level, free from electrons

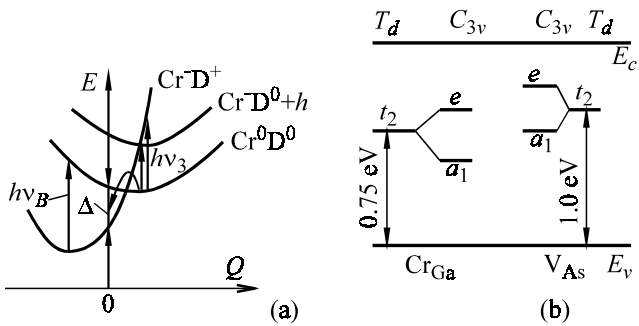

Figure 6.6. The configuration diagram (a) and energy levels of [CrD] associated components (b) in a tetrahedral (T_{d}) and trigonal ($C_{3\text{V}}$) field: arrows (a) – light-induced optical transitions [199].

in the ground state, is lower than the filled V^0 level (Figure 6.6*b*). When the components approach each other, an electron transition is possible from the vacancy to the Cr impurity, according to the reaction

$$V^0 + Cr^0(d^3) \rightarrow [Cr^- V^+]. \qquad (6.4.7)$$

Because of the strong localization of the wave function near every partner in the associate, the electron "senses" a double-well asymmetric potential and is transported by the tunneling effect at $T \rightarrow 0$.

The theory of this mechanism [45] explains successfully the observable details of optically induced light absorption modulation spectra which allowed the authors of [45] to identify the associate $Cr_{Ga}^- D_{As}^+$ state. Such tunneling states appear to be typical of associate defects, whose components occupy neighboring crystal positions and represent deep level centers.

6.5 IMPURITY–VACANCY COMPLEXES IN SILICON

Silicon crystals heated up to 1000–1250°C actively generate vacancies and host interstitial atoms. Obviously, impurities will interact with both to produce associate defects. Complexes with host interstices are likely to accumulate at sinks, because most intrinsic interstices return to their positions due to recombination with vacancies or to the substitution of other impurities at lattice sites. They may also migrate to other crystal areas to produce various complexes like oxides, silicides, etc. Therefore, there are vacancies left, which can form complexes with impurity atoms. Impurities can produce an electron bond, although only dangling, with the silicon lattice.

A dynamic equilibrium concentration of such complexes will be established at a given temperature. On fast cooling from a high temperature, some of the complexes will decay but others will be frozen as impurity–vacancy associates. In silicon, they often reveal themselves as deep level impurity centers in the forbidden gap.

Evidence for the existence of impurity–vacancy complexes in silicon was obtained in experiments on irradiation of doped crystals with γ-quanta and fast electrons. This technique appeared very convenient for the investigation of this problem, since the energy position of impurity levels could be reliably determined in samples prior to their irradiation. In the control (undoped) samples, it was possible to detect the positions of the remnant, stable vacancies. Monitoring of transformations of both and of their concentration

variations induced by irradiation allows a reasonable conclusion to be made concerning their states, including the processes of impurity–vacancy interactions.

The principal result of many investigations of this kind was that the interactions of primary radiational defects, mostly vacancies, with impurity atoms give rise to deep level complexes. A typical example is the formation of *E*-centers representing *V*-P complexes in silicon doped with phosphorus [46]. This complex is an acceptor with the energy level at $E_c - 0.43$ eV, which anneals at ~400 K. Similarly, a *V*-Sn complex is observed in silicon with the level $E + 0.35$ eV.

The authors of [47] identified the Au-*V* interaction in silicon, and the energy levels of *V*-Fe complexes were found in [48] to be $E_c - 0.36$ eV and $E_v + 0.22$ eV. The doping impurities Ni, Co, and Mn also interact with radiational defects in silicon [49, 50]. In the latter case, the structure of complexes proves to be much more complex due to the interstitial state of these impurities in the silicon lattice. For this reason, the impurity atoms have a high mobility and a good chance to encounter not only vacancies but other point defects, primarily impurity oxygen and carbon. This complicates identification of various types of complexes and makes their structure quite sophisticated. This effect manifests itself clearly in silicon samples doped with another interstitial fast diffusing impurity—lithium—producing complexes of the Li–O–V and Li–V_2 types with the levels $E_c - 0.27$ eV and $E_v + 0.48$ eV [51, 52].

Thus, the general tendency for the formation of impurity–vacancy complexes is related to a high concentration of vacancies and/or heavy doping with active impurities. For this reason, complexation is stimulated not only by radiational but also by thermal doping of nonequilibrium vacancies [53].

6.6 IMPURITY SYNERESIS

Impurity and impurity–vacancy interactions can also reveal themselves in the redistribution of impurity atoms over crystallochemical positions. This effect is very clearly observed in the distribution of amphoteric impurities. It was shown in Section 2.3 that some impurities, for example Cu atoms, can occupy both sites and interstices in elemental semiconductors, such as Ge and Si. The former have acceptor properties and the latter donor properties. In semiconductor compounds, such as $A^{III}B^{V}$, amphoteric properties are exhibited by group-IV atoms which are donors in the A^{III} positions and accep-

tors at the B^V sublattice sites. Since an amphoteric impurity has an opposite sign in various crystallochemical positions, its redistribution can be easily followed by determining charge carrier concentrations equal to the difference between the concentrations of atoms in these positions: $n_0 = N_D - N_A$ or $p_0 = N_A - N_D$. The values of n and p are found from Hall coefficient measurements.

At first sight, the distribution of an amphoteric impurity over possible positions in a crystal (or, the function $n_0(A)$, where A is the total amphoteric impurity concentration) can be found using a macroscopic approach and standard thermodynamic methods. This attempt was made in [54] in the study of interactions between vacancies and isovalent impurities. The general nature of this approach permits its application for the study of amphoteric and, more generally, of any impurities. Similarly, vacancies can be replaced by any other point defects—intrinsic or impurity defects, including a self-identical amphoteric A atom. Following [55], the authors of [54] took into account the interaction of defects in both sublattices, D_{A1} and D_{B1}, with atoms A_{A1} and A_{B1} and with host atoms A_1 and B_1 by preserving the first non-zero cross terms of the second order of smallness in the expansion of crystal thermodynamic potential Φ in terms of small (as compared to N_{A1} and N_{B1}) concentrations of all other components of the A_1B_1 compound. Below, the concentrations N_{A1}, N_{B1}, and so on will just be replaced by A_1 and B_1, etc. These cross terms have the form [55]:

$$g_{A_{A1}} D_{A_1} \left(A_{A1} D_{A1} / N_L \right), \tag{6.6.1}$$

where $g_{A_{A1}}$ and D_{A1} are the Gibbs energies of a single associate defect consisting of an amphoteric atom A in the A_1 position and of defect D in the same sublattice A_1; N_L is the concentration of atoms in each sublattice of the A_1B_1 crystal.

With the cross terms, the thermodynamic potential will be

$$\begin{aligned}\Phi = {} & g^0_{A1} A_1 + g^0_{B1} B_1 + g^0_{A_{A1}} A_{A1} + g^0_{A_{B1}} A_{B1} + g^0_{D_{A1}} D_{A1} + g^0_{D_{B1}} D_{B1} \\ & + g_{A_{A1}} D_{A_1} \frac{A_{A_1} D_{A_1}}{N_I} + g_{A_{A1} D_{B1}} \frac{A_{A1} D_{B1}}{N_I} + g_{A_{B1}} D_{A1} \frac{A_{B1} D_{A1}}{N_I} \\ & + g_{A_{B1} D_{B1}} \frac{A_{B1} D_{B1}}{N_I} - T \ln \frac{A_1 + A_{A1} + D_{A1}}{A_1! A_{A1}! D_{A1}!} - T \frac{B_1 + A_{B1} + D_{B1}}{B_1! A_{B1}! D_{B1}!}\end{aligned} \tag{6.6.2}$$

To find partial chemical potentials of the partners, expression (6.6.2) is differentiated, neglecting the empty sites:

$$\mu_{D_{A1}} = \frac{\partial \Phi}{\partial D_{A1}} = g^0_{DA_1} + g_{A_{A1}D_{A1}} x_{A_{A1}} + g_{A_{B1}D_{A1}} x_{AB_1} + T \ln x_{DA_1} \qquad (6.6.3)$$

$$\mu_{D_{B1}} = \frac{\partial \Phi}{\partial D_{B1}} = g^0_{DB1} + g_{A_{A1}D_{B1}} x_{A_{A1}} + g_{A_{B1}D_{B1}} x_{A_{B1}} + T \ln x_{DB_1} \qquad (6.6.4)$$

$$\mu_{B1} = \frac{\partial \Phi}{\partial B_1} = g^0_{B_1} + T \ln x_{B1} = g^0_{B1} + T \ln(1 - x_{AB1}). \qquad (6.6.5)$$

It is seen from these equations that defect formation energies are not constant but vary with the crystal–solid solution composition:

$$g_{D_{A1}} = g^0_{DA1} + g^0_{A_{A1}D_{A1}} x_{A_{A1}} + g_{A_{B1}D_{A1}} x_{A_{B1}} \qquad (6.6.6)$$

$$g_{D_{B1}} = g^0_{D_{B1}} + g_{A_{B1}D_{B1}} x_{A_{B1}} + g_{A_{A1}D_{B1}} x_{A_{A1}}. \qquad (6.6.7)$$

For the chemical potential of the B_1 component, we accept the relation well known from thermodynamics:

$$(\mu_{B1})_m = \Psi(T) + T \ln P_{(B1)_m}, \qquad (6.6.8)$$

which makes allowance for the volatility of the B_1 partner representing $(B_1)_m$ molecular structures, when under pressure P in the gas phase. The function $\Psi(T)$ represents the standard potential of the gas phase at temperature T expressed in energy units. In other words, equation (6.6.8) expresses the assumption of the gas phase as being an ideal gas.

The relations between the atomic fractions of the partners are described by the equations

$$x_{A1} + x_{A_{A1}} + x_{D_{A1}} = 1 \qquad (6.6.9)$$

$$x_{B1} + x_{A_{B1}} + x_{D_{B1}} = 1, \qquad (6.6.10)$$

which reflect the site balance in each sublattice of the A_1B_1 compound. In addition, there are two equations reflecting the relation between chemical potentials: one for the defect formation–annihilation process

$$\mu_{D_{A1}} + \mu_{D_{B1}} = 0 \tag{6.6.11}$$

and the other for evaporation–condensation of the volatile B_1 component

$$\mu_{B1} = \mu_{D_{B1}} + \frac{1}{m}\mu_{(B1)_m}. \tag{6.6.12}$$

By substituting expressions (6.6.3) through (6.6.5) and (6.6.8) into expressions (6.6.9) through (6.6.12), we obtain the mole fraction ratios of defects containing an amphoteric impurity:

$$\frac{x_{D_{A1}}}{x^0_{D_{A1}}} = \frac{1}{1 - x_{A_{B1}}} \exp\left[- \frac{g_{A_{A1}D_{A1}} x_{A_{A1}} + g_{A_{B1}D_{A1}} x_{A_{B1}}}{T} \right], \tag{6.6.13}$$

$$\frac{x_{D_{B1}}}{x^0_{D_{B1}}} = \left(1 - x_{A_{B1}}\right) \exp\left[- \frac{g_{A_{A1}D_{B1}} x_{A_{A1}} + g_{A_{B1}} x_{A_{B1}}}{T} \right]. \tag{6.6.14}$$

In these equations, $x^0_{D_{A1}}$ and $x^0_{D_{B1}}$ are equilibrium atomic fractions of defects in a pure, impurity-free crystal. Therefore, if D are understood as vacancies, these equations describe the change in the homogeneity region of the A_1B_1 compound doped with an A_1 impurity (isovalent doping) interacting directly with intrinsic point defects—vacancies. For a compound semiconductor, only equation (6.6.13) is necessary.

Let us analyze equations (6.6.13) and (6.6.14). The simplest case is the *interaction between the nearest neighbors*—an impurity atom and a defect located in different sublattices of a A_1B_1 compound. Then, $g_{A_{A1}D_{A1}} = 0$ and $g_{A_{A1}D_{B1}} = 0$. Hence, an amphoteric impurity can change the vacancy concentration only in the adjacent sublattice and cannot do this in its own sublattice. For an appreciable change in $x_{D_{A1}}$ or $x_{D_{B1}}$, the quantitative evaluation requires that the following condition be fulfilled:

$$\frac{gx}{T} \sim 1. \tag{6.6.15}$$

By putting $x \sim 0.01$, which is an extreme value for many impurities in semiconductors, and equating T to the crystallization temperature of the semiconductor ($T \approx 1400$–1500 K), the condition of (6.6.15) will be satisfied

at $g > 12$ eV. This energy is too large if the binding between an impurity atom and a defect (vacancy) in the other sublattice is regarded as resulting from an elastic or Coulomb interaction of the nearest neighbors. The energy of such interactions does not exceed 1–5 eV. Therefore, an additional generation of vacancies by the doping impurity is unlikely in an elastic or purely Coulomb interaction. This may become possible, say, for InSb grown from a melt at $T = 470$°C and for $A^{III}B^{V}$ crystals grown by liquid-phase epitaxy at $T = 700$–800 K.

The next case is the *interaction in the same sublattice*. Then $g_{A_{A1}D_{A1}} \neq 0$ and $g_{A_{B1}D_{B1}} \neq 0$, but these values may have the same or different signs. If the impurity repels the defect, $g > 0$. It is quite obvious that the g values in the exponents of equations (6.6.13) and (6.6.14) may, in practice, have different signs. Then the absolute exponent may become small, resulting in a considerable change in the vacancy concentration in amphoteric doping. A simple illustration of this situation is the case when an amphoteric impurity occupies lattice sites of a semiconductor compound and is an acceptor [A_{A1}^{-}], while interstices are occupied by a donor [A_{B1}^{+}]; vacancies are then acceptors [D_{A1}^{-}]. Here, interstices are treated as the B_1 "sublattice" with $g_{A_{A1}D_{A1}} > 0$ and $g_{A_{B1}D_{B1}} < 0$. Thus, the solubility, i.e., concentration, of impurities in semiconductors is determined by the vacancy content in the crystal, but sometimes the concentration of compensating defects (vacancies) themselves depends on the amphoteric impurity concentration. This is the *effect of impurity syneresis*.

In the study of group-IV impurity behavior in GaAs, the authors of [56] suggested a possible transition of impurity atoms, for example silicon, from the charged to the neutral state, rather than from one sublattice to the other. This transition depends on the silicon concentration and the crystal growth conditions determining the ensemble of intrinsic point defects. The latter determine the transition of a group-IV impurity to the neutral state either by being involved in neutral associates or by accelerating the loss of electrical activity of the impurity. The principal feature of this model is its self-consistency: the concentration of the electrically active fraction of a group-IV impurity depends on the defect content of the host crystal, while this content itself varies with the impurity content.

This kind of syneresis has been confirmed by many studies of doping impurity behavior in various semiconductor compounds, irrespective of the model used.

In practice, a semiconductor is often doped with two impurities simultaneously. Suppose one impurity is amphoteric and the other is just a common impurity M. The question is whether the distribution of the amphoteric impurity over sublattices of a semiconductor compound will vary with the con-

centration of impurity M due to their interaction. Generally, the above approach allowing for the interaction in thermodynamic potential Φ is totally applicable to this problem. What is to be done is to replace D_{A1} and D_{B1} defects by M_{A1} and M_{B1} concentrations and to solve equations (6.6.13) and (6.6.14) with respect to the mole fractions $x_{A_{A1}}$ or $x_{A_{B1}}$. The resulting general solution is too cumbersome. A simpler approach, without losing the generality, was suggested in [57] and refined in [58]; we will follow it below.

It will be assumed that non-amphoteric impurity M is introduced into a crystal without an external action. The crystal already contains amphoteric impurity A in two possible positions A_1 and A_2. The role of M impurity will be played by an intrinsic point defect D. In the problem discussed above, the D concentration was a dynamic variable, but now M (or D) is a constant value varying only with the intensity of the external action (diffusion, ion implantation, etc.) for introducing M into the crystal.

The energies required for the incorporation of amphoteric atoms to both positions will be denoted as E_1 and E_2 and the interaction potentials of all amphoteric atoms as V_{11}, V_{22}, V_{12}, and V_{21}. Obviously, we will have $V_{12} = V_{21}$. The potentials of M impurity interaction with amphoteric atoms will be V_{1M} and V_{2M}.

The results of a theoretical treatment in terms of concentrations [58] are:

$$A_1 = N_0\left\{\exp\frac{1}{kT}\left[E_1 + V_{11}\frac{A_1}{N_0} + V_{12}\frac{A_2}{N_0} + V_{1M}\frac{M}{N_0}\right] + 1\right\}^{-1} \qquad (6.6.16)$$

$$A_2 = N_0\left\{\exp\frac{1}{kT}\left[E_2 + V_{21}\frac{A_1}{N_0} + V_{22}\frac{A_2}{N_0} + V_{2M}\frac{M}{N_0}\right] + 1\right\}^{-1}, \qquad (6.6.17)$$

where N_0 is total amphoteric impurity concentration. The quantities A_1 and A_2 can be found from the following equation [57]:

$$\ln\frac{A-Q}{A+Q} = \frac{1}{kT}\left[E_1 - E_2 + \frac{M}{N_0}(V_{1M} - V_{2M}) + \frac{A}{2}(V_{11} - V_{22}) - \frac{Q}{2}(V_{11} - 2V_{12})\right], \qquad (6.6.18)$$

where

$$A = \frac{A_1 + A_2}{N_0}, \qquad (6.6.19)$$

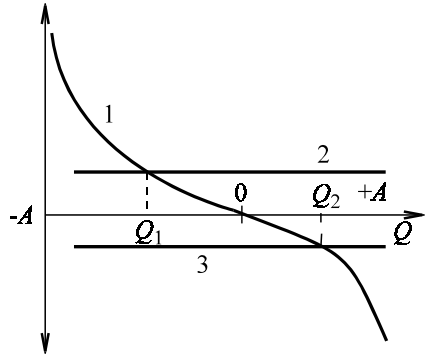

Figure 6.7. A plot of the solution to equation (6.6.21): 1 – for the left-hand side of the equation; 2, 3 – for the right-hand side of the equation at $E_1 > E_2$, $V_{1M} > V_{2M}$ (2) and $E_1 < E_2$, $V_{1M} < V_{2M}$ (3).

$$Q = \frac{A_1 - A_2}{N_0} . \qquad (6.6.20)$$

For simplification, only potentials V_{1M} and V_{2M} in (6.6.18) are taken to be non-zero. Then, the solution is $Q_1 < 0$, or $Q_2 > 0$. Let us assume that amphoteric atoms do not interact. Then, (6.6.18) will transform to

$$\ln\frac{A-Q}{A+Q} = \frac{1}{kT}\left[E_1 - E_2 + \frac{M}{N_0}\left(V_{1M} - V_{2M}\right)\right], \qquad (6.6.21)$$

i.e., M impurity is capable of redistributing A impurity atoms over positions 1 and 2 in (6.6.20).

It is easy to see that this conclusion will hold even without simplifying equation (6.6.18). Its right-hand side will not be parallel to the Q-axis in the coordinates of Figure 6.7 but approach it at a certain angle. It is seen that, owing to the interaction with A_1 and A_2, M impurity can redistribute the amphoteric impurity in such a way that even the crystal conductivity type will change, provided that the A_1 and A_2 states have different signs. Suppose the A impurity is a charged acceptor at a site and a charged donor at an interstice. The transition from one type of conductivity to the other will occur abruptly at a certain critical M concentration, at which

$$\ln\left[\frac{A-Q}{A+Q}\right]=0\,. \tag{6.6.22}$$

Hence, we find from (6.6.21)

$$M=N_0\frac{E_1-E_2}{V_{2M}-V_{1M}}\,. \tag{6.6.23}$$

In a more general case, with all amphoteric atoms at any position interacting with one another, equation (6.6.18) equalized to zero at point $Q = 0$ will yield

$$M=N_0\frac{E_1-E_2+(A/2)(V_{11}-V_{22})}{V_{2M}-V}\,. \tag{6.6.24}$$

Here again, an impurity syneresis manifests itself: M impurity redistributes A impurity, with the value of M depending on the A concentration. Due to the impurity syneresis, the curves $n_0(A)$, where n_0 is electron concentration in an n-type crystal, can have fanciful shapes very different from curves for zero interaction of M impurity with amphoteric atoms at constant M (or D) concentration. In practice, one can find numerous examples of complicated $n_0(A)$ curves. This is good evidence for the existence of impurity interactions and for the manifestation of impurity syneresis. On the other hand, the lack of knowledge of E_1 and E_2, as well as of interaction potentials V_{11}, V_{12}, and V_{22}, make the processing of experimental $n_0(A)$ curves nearly impossible.

6.7 COMBINED COMPLEXATION

Both types of interaction—impurity–vacancy and ion–impurity pairing—often occur together. A typical example is the precipitation of excess impurity from an oversaturated semiconductor solution, discussed in [59, 60] with reference to germanium doped with two impurities simultaneously—copper and antimony.

Figure 6.8 shows the kinetic curves for solid solution decomposition in Sb-free samples (curve 5) and Sb-doped samples (curves 1–4) at different annealing temperatures. The differences in these curves were interpreted in

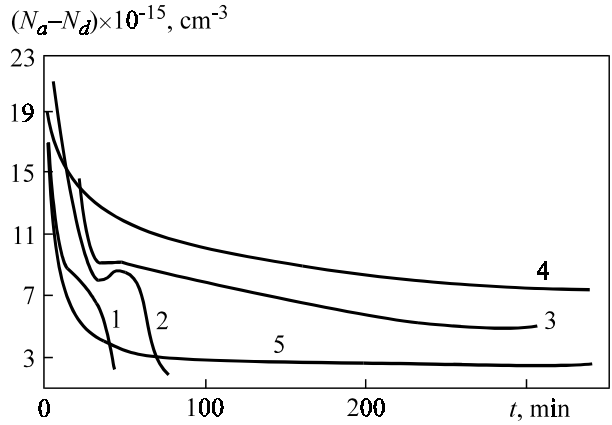

Figure 6.8. Variation in the Hall concentration $N_a - N_d$ with annealing time t in Ge<Cu> samples (curve 5) and in Ge<Cu,Sb> samples (curves 1–4) at temperatures, °C: 1 – 500; 2 – 485; 3 – 450; 4 – 425; 5 – 485.

[59, 60], with the allowance for doubly charged vacancy, in terms of the following family of reactions between various defects:

$$\mathrm{Cu_s^-} \underset{K_2}{\overset{K_1}{\rightleftarrows}} \mathrm{Cu_i^+} + V^{=}, \tag{6.7.1}$$

$$\mathrm{Cu_s^-} \underset{K_4}{\overset{K_3}{\rightleftarrows}} \mathrm{Cu_i^+} + V^{-} + e^{-}, \tag{6.7.2}$$

$$D^{+} + nV^{=} \underset{K_6}{\overset{K_5}{\rightleftarrows}} Z^{-(2n-1)}, \tag{6.7.3}$$

$$m\mathrm{Cu_i^+} + Z^{-(2n-1)} \underset{K_8}{\overset{K_7}{\rightleftarrows}} Z', \tag{6.7.4}$$

$$\mathrm{Cu}_i^+ \xrightarrow{K_9} \text{sinks}, \tag{6.7.5}$$

$$V^{=} \xrightarrow{K_{10}} \text{sinks}. \tag{6.7.6}$$

The charge states of interacting point defects are designated with account taken of the Fermi level position at the forbidden band center at annealing temperatures. That vacancies are doubly charged was demonstrated in [61].

The key reaction linking all other reactions is (6.7.4); its product is a complex with a zero charge structure $(D^+ - V^{=} - \mathrm{Cu}_i^+)^0$. This process is preceded by the formation of another complex $(D^+ - V^{=})^-$ having an acceptor character. The appearance of these new acceptors corresponds to a "lower" decomposition rate represented as a kink in the kinetic curves in Figure 6.8.

The formation of charged associates was supported by charge carrier mobility measurements. As soon as a slower decomposition region appeared in the kinetic curves, the mobility dropped, indicating the production of new charged centers. Different behavior was observed in Sb-free samples. As copper atoms sublimated from the lattice sites, the number of scattering centers decreased, leading to a higher mobility. Moreover, the mobility rise fitted well the theoretical curve calculated on the assumption that the scattering centers were copper ions.

Finally, the formation of new acceptors (Z associates) in the slower kinetic curve region was detected by direct measurements of the low temperature dependence of the Hall coefficient. The energy level position of Z-centers was found from the curve slope to be $E_v + 0.08 \pm 0.02$ eV.

Similar complexes were observed in the Ge–Sb–Ni systems [62]. It was shown in experiments with nonstoichiometric GaAs crystals [63, 64] that gallium vacancies make a considerable contribution to complexation. Therefore, complicated interactions of impurity atoms, ions, and vacancies are common phenomena in doped semiconductors.

To conclude, we present an analytical expression taking into account both interaction types—donor–acceptor ion pairing and complexation:

$$N_A = \frac{N_D}{1+\left[1+\left(2n_i/N_A^0\right)^2\right]^{1/2}} + \sqrt{\left\{\frac{N_D}{1+\left[1+\left(2n_i/N_A^0\right)^2\right]^{1/2}}\right\}^2 + \left(N_A^0\right)^2}$$

$$+\frac{KN_{\mathrm{D}}\left(N_{\mathrm{A}}^{0}\right)^{2}\left\{1+\left[1+\left(2n_{\mathrm{i}}/N_{\mathrm{A}}^{0}\right)^{2}\right]^{1/2}\right\}}{2+K\left(N_{\mathrm{A}}^{0}\right)^{2}\left\{1+\left[1+\left(2n_{\mathrm{i}}/N_{\mathrm{D}}^{0}\right)^{2}\right]^{1/2}\right\}}, \tag{6.7.7}$$

where N_{A} is the extreme acceptor concentration in the presence of a donor, $N_{\mathrm{A}}{}^{0}$ is the acceptor concentration in the host semiconductor, $N_{\mathrm{D}}{}^{0}$ is the donor concentration, and K is a complexation constant. The first two terms in (6.7.7) allow for ion pairing (donor–acceptor interaction) and the third one describes complexation.

The value of K appears to be always indeterminate in experimental data processing. It must be found from a particular case at $n_{\mathrm{i}}/N_{\mathrm{A}}{}^{0} >> 1$; then, (6.7.7) transforms to

$$N_{\mathrm{A}} = N_{\mathrm{A}}^{0} + N_{\mathrm{D}} \Bigg/ \left(1+\frac{1}{KN_{\mathrm{A}}^{0}n_{\mathrm{i}}}\right). \tag{6.7.8}$$

For example, Ge<Cu> samples have $N_{\mathrm{Cu_s}} = 2.3\times10^{16}$ cm^{-3} at $T = 850°$C, while Ge<Sb,Cu> samples have $N_{\mathrm{Cu_s}} = 4.6\times10^{16}$ cm^{-3} at $N_{\mathrm{Sb}} = 5\times10^{16}$ cm^{-3} and the same temperature. The value of K calculated for these conditions from (6.7.8) is 3.7×10^{-36} cm^{-6}.

6.8 INDIRECT ION–ION INTERACTION

Many experimental studies have demonstrated an increased solubility of one impurity in the presence of another impurity, which could not be interpreted as being due to ion pairing or complexation. The former was unlikely because of the low impurity concentration, since the mean distance between impurity atoms

$$r = \frac{0.7}{\sqrt[3]{N_{\mathrm{im}}}} \tag{6.8.1}$$

was 20.0 μm at $N_{im} = N_{Sb} + N_{Cu} = 10^{17}$ cm^{-3}, and the Coulomb interaction energy at such distances was 0.06 eV, which was much less than $2kT$ varying from 0.15 to 0.19 eV. Elastic interaction at these distances is also negligible.

Experiments showed [60] that vacancy concentration variation in Ge<Cu,Sb> samples did not affect the order of magnitude of copper solubility. For this reason, the complexation via the reaction

$$Cu_i + V + Sb \rightleftarrows (Cu_s^q\, Sb) \tag{6.8.2}$$

was discarded.

A similar situation was observed in Ge<Sb,Ni> samples [65]. Complexation due to a direct interaction of antimony and copper atoms could not contribute much to the experimentally established fact of increased Cu_s solubility in germanium in the presence of Sb impurity.

This effect can be accounted for by an indirect interaction of impurity atoms occurring via the electron subsystem of the crystal. In this case, the Fermi level is at the forbidden gap center, and the process of copper dissolution can be represented as:

in undoped germanium

$$\left[Cu^0\right] \rightleftarrows Cu_s^{=} + Cu_i^{+} + 2p + e_1 + \Delta H_0, \tag{6.8.3}$$

in donor-doped germanium

$$\left[Cu^0\right] \rightleftarrows Cu_s^{=} + Cu_i^{+} + D^{+} + 2p + e_1 + e_2 + \Delta H_1, \tag{6.8.4}$$

where $[Cu^0]$ is copper concentration in the ambient phase; p, e_1, and e_2 are concentrations of holes and electrons determined, respectively, by the concentrations $C_s^{=}$, C_i^{+} and D^{+}; ΔH_0 and ΔH_1 are copper dissolution enthalpies in undoped and doped germanium, respectively. It follows from these reactions that electrons, excited to the donor level, charge Cu^0 atoms without activation up to $Cu^{=}$ in the presence of a donor compensating impurity (in contrast to undoped germanium). This is supported by the fact that the effect of high copper solubility in the concentration range of interest is only determined by the donor concentration and is independent of the kind of donor impurity (Sb, P, As) [60, 61].

There is no doubt that direct complexation makes an increasingly greater contribution with increasing donor content. This process is accompanied by the generation of new energy centers in the semiconductor forbidden gap.

6.9 APPLIED ASPECTS OF COMPLEXATION

6.9.1 Deep center content versus growth temperature and free electron concentration

Complexation has a profound effect on defect formation in semiconductor compounds. Most point defects are isolated at high growth temperatures. Hardening freezes the defect concentration without affecting complexation. The most mobile defects seek for sinks and find them in associates being formed. These processes were discussed in the previous section. Note that their investigation allows determination of some thermodynamic characteristics of defects. This is considered with reference to GaP in [58].

Epitaxial *n*-type GaP layers were grown by liquid phase epitaxy. Two sets of samples differing in the growth conditions were studied. The temperature range of cooling was the same in both cases: minus 1020–950°C. The first group of samples consisted of epitaxial structures grown in graphite piston cassettes in an open system with a continuous hydrogen flow. However, the concentration of free electrons in epitaxial layers varied from 5×10^{15} to 5×10^{17} cm^{-3} because of the addition of sulphur to the melt. The second group of samples was grown in a quasi-closed volume. No additional impurities were added. The growth temperature range was varied in order to find the temperature dependence of deep center content. The content of centers was measured by the thermally stimulated capacity method. Normally, three deep levels with the activation energies 0.35, 0.53, and 0.65 eV were observed in the temperature range of 150–280 K.

Since the levels of interest lie in the upper half of the band and their activation energies are smaller than the forbidden gap halfwidth, the thermal emission rate of electrons, e_n^+, exceeds that of holes, e_p^+. The levels are completely ionized. For these conditions, the concentration of centers can be described as

$$N_t = \frac{2(N_d - N_a)\Delta C}{C},$$

where $N_d - N_a$ is shallow donor concentration obtained from the CVC data and ΔC is the capacity variation due to ionization of a deep center.

The measurements are presented in Figures 6.9 and 6.10. The concentration of centers rises linearly with that of uncompensated donors and with the growth temperature of *n*-GaP epitaxial layers. It follows from formula

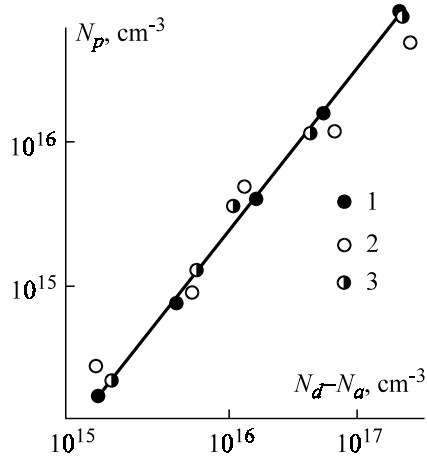

Figure 6.9. Deep center concentration versus uncompensated donor concentration in *n*-GaP for various levels of deep centers: 1 – 0.35 eV; 2 – 0.53 eV; 3 – 0.63 eV.

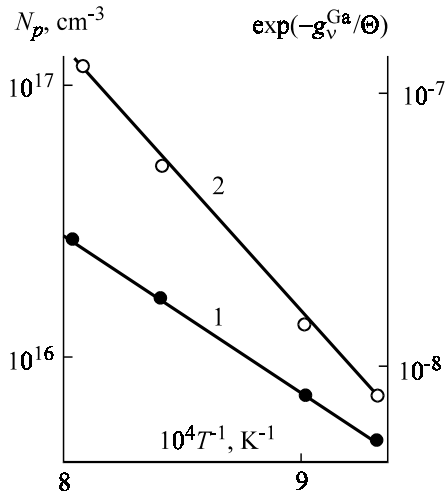

Figure 6.10. The deep center concentration (1) and the parameter $\exp(-g_V^{Ga}/\theta)$ (2) versus GaP annealing temperature.

(6.4.4) that an increase in deep level concentration with increasing free electron content is characteristic of acceptor centers; their energy levels, however, lie in the upper half of the forbidden gap.

The epitaxial layers discussed were grown from high purity materials by different growth techniques at different laboratories. So it is very unlikely that the formation of the centers was associated with impurities. More probable is the suggestion that the centers were associated with intrinsic defects. A gallium vacancy creates three levels in the lower half of the band. Antistructural defects can create levels in the upper half. So far, only P_{Ga} defect has been identified in GaP [64, 66]. The energy levels of this defect were identified by the DLTS technique in [67]. Mechanical stress applied at a current of 5 A×cm^{-2} gives rise to a dislocation network and to a deep donor level in the forbidden gap located 0.8 eV lower than the conduction band. The concentration of such deep centers is proportional to the applied stress and time. Removal of external stress by passing current through a diode produces two levels—at 0.71 and 1.18 eV below the conduction band bottom. The activation energies of these centers correspond to the theoretical values of the first and second ionization levels of the antistructural defect P_{Ga}. Optical data consistent with these results are presented in [68]. However, the P_{Ga} defect is a donor, so there should be no dependence on free electron concentration (because the center creates deep levels and must be neutral at growth temperatures) or its concentration should decrease with increasing *n*.

A semi-empirical calculation method was suggested [69], in which a cation antisite defect, Ga_P, creates levels at 0.84, 1.14, and 1.44 eV, counted from the valence band. No one has ever observed this center. The calculations made in this work [69] are so different from its experimental data that one can hardly identify these defects as being antistructural, even with the theoretical error.

An alternative suggestion is that these centers have a complex structure. We mentioned earlier that a reasonable model is that of a complex consisting of a cation vacancy and a shallow donor, in particular, the V_{Ga}S-type of complexes. Such complexes may be more stable than those with silicon because of the defect localization at neighboring sites.

It was shown in Section 6.4 that the concentration of complexes is proportional to vacancy concentration which, in turn, is proportional to free electron concentration. Vacancy concentrations rise with growth temperature.

6.9.2 Homogeneity region width in $A^{III}B^{V}$ compounds

The formation of point defects, both intrinsic and due to impurities and complexes, results in the chemical composition of the crystal being different

from its chemical formula. The crystal composition is then said to deviate from the stoichiometric pattern.

The deviation from crystal stoichiometry can be described by the mole fraction of the excess component. Let us assume that the excess A component in a binary compound corresponds to a positive stoichiometric deviation and the excess B component to a negative deviation. The summation will be made with respect to this characteristic. The stoichiometric deviation of a binary compound is

$$\Delta = \sum_{\alpha} N_{\alpha}^{\mathrm{B}} \Big/ N^{\mathrm{B}} - \sum_{\alpha} N_{\alpha}^{\mathrm{A}} \Big/ N^{\mathrm{A}} + \sum_{j} \left(N_j^{\mathrm{A}} \Big/ N_j - N_j^{\mathrm{B}} \right), \qquad (6.9.1)$$

where j is a running index for various types of interstices.

The first term in (6.9.1) unites all structural elements substituting B lattice atoms, thereby promoting the excess A component. The second term does exactly the opposite. The third term allows for the mole fractions of atoms at different interstitial vacancies. In summation, the subscript α runs through all types of substitutional defects, except for the values of A and B corresponding to antistructural defects. Antistructural disordering does not affect the homogeneity region width. Indeed, the transition of, say, a gallium atom from its site to an anion site does not change the total number of gallium atoms in the system. Therefore, there is no direct influence. However, the antistructural defect that will appear will compensate *n*-type material, leading to a change in the homogeneity region width.

The same is true of complexation processes. Without changing the total number of defects, these processes change the Fermi level position, thereby changing the stoichiometric deviations. We will illustrate this with the following [58]. The stoichiometric deviation through point defect concentrations will be expressed as

$$\Delta = \sum (-1)^{\delta_{\mathrm{A}}^{\beta}} \frac{a_{\alpha}}{a_{\beta}} \exp\left(-\frac{g_{\alpha}^{\beta}}{\Theta} \right) f_{\alpha\beta}^{-1} + \sum (-1)^{\delta_{\mathrm{B}}^{j}} a_{\alpha} \exp\left(\frac{g_{\alpha}^{j}}{\Theta} \right) f_{\alpha j}^{-1}, \qquad (6.9.2)$$

where

$$f_{\alpha\beta}^{-1} = R_{\alpha}^{\beta} \exp\left[\frac{\left(E_{\mathrm{F}} - \varepsilon_{\alpha}^{\beta} \right) d_{\alpha}^{\beta}}{\Theta} \right] + r_{\alpha}^{\beta} \exp\left[\frac{\left(E_{\mathrm{g}} - E_{\mathrm{F}} - \varepsilon_{\alpha}^{\beta} \right) \left(1 - d_{\alpha}^{\beta} \right)}{\Theta} \right],$$

$$f_{\alpha j}^{-1} = R_\alpha^j \exp\left[\frac{\left(E_F - \varepsilon_\alpha^j\right)d_\alpha^j}{\Theta}\right] + r_\alpha^j \exp\left[\frac{\left(E_g - E_F - \varepsilon_\alpha^j\right)\left(1 - d_\alpha^j\right)}{\Theta}\right].$$

Index β in formula (6.9.2) runs through two values—*A* and *B*, while index *j* runs through values related to interstices.

Therefore, the calculated deviation values are affected by the factors:

(1) the compound growth conditions (via temperature and activity coefficients of various ligands);

(2) partial free energies of defect formation and ionization energies;

(3) doping levels and compensation degrees.

Gallium arsenide has been studied much better than other $A^{III}B^{V}$ compounds. In spite of this, it is useful to evaluate the variation limits of the homogeneity region width in additional donor doping of GaAs. Such calculations were performed in [58], using the available parameters of defect formation and ionization [71, 72]. The results of the calculation are shown in Figure 6.11.

Doping changes the homogeneity region width so much that this should not be ignored in designing various technological processes. Similar calculations for GaP [58] provided results illustrated in Figures 6.12 and 6.13. The reader can find in this work other illustrations of applied aspects of impurity interaction.

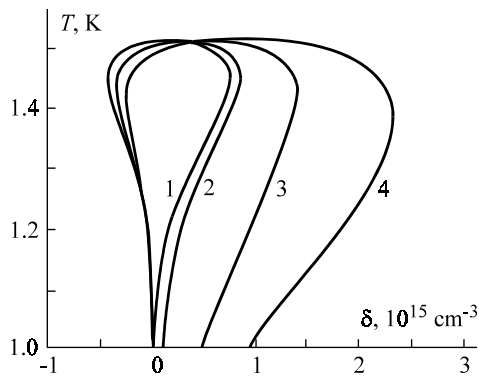

Figure 6.11. The homogeneity region width δ in GaAs with account taken of doping and complexation (calculation) at various impurity concentrations, cm^{-3}: 1 – 0; 2 – 1×10^{17}; 3 – 5×10^{17}; 4 – 1×10^{18}.

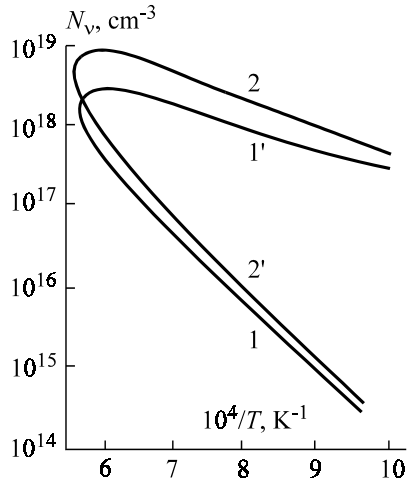

Figure 6.12. The temperature dependence of the vacancy concentration for Ga (1) and P (2) in GaP: 1, 2 – Ga-enriched; 1', 2' – P-enriched.

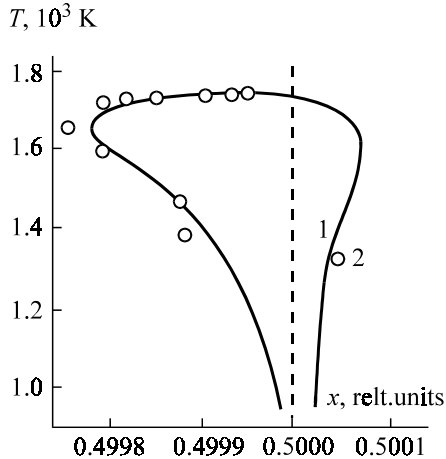

Figure 6.13. The homogeneity region width for GaP: solid curve – calculation; circles – experiment [73, 74].

REFERENCES

6.1. V.M. Glazov, *Osnovy fizicheskoi khimii* (The Fundamentals of Physical Chemistry). Moscow: Vysshaya shkola, 456 p. (1981) (in Russian).

6.2. J. Bjerrum, *Metal Amine Formation in Aqueous Solutions*. Copenhagen, 308 p. (1957).

6.3. J.S. Prenner, F.E. Williams, *Phys. Rev*. **101**: 1427–1435 (1956).

6.4. N.Kh. Abrikosov, V.M. Glazov, Lu Chjen Yuan, *J. Neorg. Khimii* **7**: 831–836 (1962) (in Russian).

6.5. V.M. Glazov, V.S. Zemskov, *Fiziko-khimicheskie osnovy legirovaniya poluprovodnikov* (Physico-Chemical Principles of Semiconductor Doping). Moscow: Nauka, 371 p. (1967) (in Russian).

6.6. B.I. Boltaks, *Diffuziya i tochechnye defekty v poluprovodnikakh* (Diffusion and Point Defects in Semiconductors). Leningrad: Nauka, 384 p. (1972) (in Russian).

6.7. J.S. Blakemore, *Semiconductor Statistics*. London: Pergamon Press, 302 p. (1962).

6.8. E.A. Apple, F.E. Williams, *J. Electrochem. Soc*. **106**: 224–227 (1959).

6.9. M.P. Lorenz, T.H. Morgan, H.D. Petti, *Trudy IX mezhdun. konfer. po fizike poluprov*. (Proc. IXth Intern. Conf. on Semicond. Physics). Leningrad: Nauka, 524–529 (1969).

6.10. W.G. Spitzer, F.A.Trumbore, R.A. Logan, *J. Appl. Phys*. **32**: 1822–1829 (1961).

6.11. V.I. Fistul, M.G. Milvidsky, E.M. Omelyanovsky *et al*., *DAN SSSR* **149**: 1119–1122 (1963) (in Russian).

6.12. L.J. Vieland, I. Kudman, *J. Phys. Chem. Sol*. **24**: 437–440 (1965).

6.13. V.I. Fistul, E.M. Omelyanovsky, O.V. Pelevin *et al*., *Izvestiya AN SSSR, ser. Neorg. Mat*. **2**: 657–658 (1966) (in Russian).

6.14. M.Ya. Dashevsky, L.S. Okun, P.Z. Plotkina, *Izvestiya AN SSSR, ser. Neorg. mat*. **4**: 684–687 (1968) (in Russian).

6.15. V.I. Fistul, *Heavily Doped Semiconductors*. New York: Premium Press, 418 p. (1969).

6.16. V.I. Fistul, P.M. Grinshtein, N.S. Rytova, *FTP* **4**: 84–88 (1970) (in Russian).

6.17. D.G. Arasly, M.I. Aliev, V.I. Fistul, *Izvestiya AN AzSSR*, No. 5: 103–108 (1965) (in Russian).

6.18. M.G. Milvidsky, O.G. Stolyarov, A.V. Berkova, *FTT* **6**: 3170–3173 (1964) (in Russian).

6.19. R.B. Fair, C.R. Weber, *J. Appl. Phys*. **44**: 273–277 (1973).

6.20. E.P. Rashevskaya, V.I. Fistul. *FTT* **9**: 3618–3620 (1967) (in Russian).

6.21. M.G. Milvidsky, V.B. Osvensky, G.P. Proshko, *FTP* **6**: 224–228 (1972) (in Russian).

6.22. C.D. Thurmond, J.D. Struthers, *J. Phys. Chem.* **57**: 831–837 (1953).
6.23. J. Teltow, *Ann. Phys.* **5**: 63–65 (1949).
6.24. C.S. Fuller, W. Kaiser, C.D. Thurmond, *J. Phys. Chem. Sol.* **16**: 161–167 (1961).
6.25. C.S. Fuller, K.B. Wolfstirn, *J. Appl. Phys.* **34**: 2287–2291 (1963).
6.26. V.B. Ufimtzev, A.N. Krestovnikov, *Izvestiya AN SSSR, ser. Neorg. mat.* **4**: 1578–1583 (1968) (in Russian).
6.27. V.B. Ufimtzev, F.A. Gimelfarb, *Izvestiya AN SSSR, ser. Neorg. mat.* **9**: 2073–2077 (1973) (in Russian).
6.28. N.A. Bulyenkov, V.I. Fistul, *DAN SSSR* **180**: 1415–1418 (1968) (in Russian).
6.29. L.A. Cirifalco, H. Herman, *Acta Met* **13**: 583–597 (1965).
6.30. A.S. Damask, G.D. Dines, *Point Defects in Metals*. New York–London: Gordon & Breach Publ., 291 p. (1963).
6.31. C.J. Hwang, *J. Appl. Phys.* **40**: 4584–4592 (1969).
6.32. E.W. Williams, *Phys. Rev.* **168**: 922–932 (1968).
6.33. M.G. Milvidsky, V.B. Osvensky *et al.*, *FTP* **6**: 224–228 (1972) (in Russian).
6.34. V.A. Telezhkin, K.B. Tolpygo, *FTP* **15**: 1084–1089 (1983) (in Russian).
6.35. M.G. Milvidsky, V.B. Osvensky, *Styrukturnye defekty v monokristallakh poluprovodnikov* (Structural Defects in Semiconductor Single Crystals). Moscow: Metallurgia, 324 p. (1984) (in Russian).
6.36. S.V. Bulyarsky, V.P. Oleinikov, *Phys. Stat. Sol., B* **141**: K7–K10 (1987).
6.37. G.C. Osbourn, *Phys. Rev.* **122**: 2898–2902 (1980).
6.38. I.A. Buyanova, O.S. Ostapenko, M.K. Sheikman, *FTT* **27**, No. 3: 748–755 (1985) (in Russian).
6.39. I.A. Buyanova, O.S. Ostapenko, M.K. Sheikman, *FTP* **20**: 1791–1800 (1986) (in Russian).
6.40. N.S. Averkiev, A.A. Gutkin, E.B. Osipov *et al.*, *FTP* **25**, No. 1: 50–57 (1991) (in Russian).
6.41. N.S. Averkiev, A.A. Gutkin, E.B. Osipov, *FTP* **26**, No. 7: 1269–1281 (1992) (in Russian).
6.42. A.A. Gutkin, M.A. Reshchikov, V.R. Sosnovsky, *Semiconductors*. Amer. Inst. of Physics (1993).
6.43. F.A. Kroger, *The Chemistry of Imperfect Crystals*. Amsterdam: North–Holland Publ. Press, 654 p. (1964).
6.44. P. Krispin, *Phys. Stat. Sol., A* **69**: 193–200 (1982).
6.45. R.A. Vanem, K.A. Kikoin, P.A. Lyuk *et al.*, *Pis'ma v JETF* **39**: 416–419 (1984) (in Russian).
6.46. V.S. Vavilov, A.E. Kiv, O.P. Niyazova, *Mekhanismy obrazovainya i migratzii defektov v poluprovodnikakh* (Formation and Migration of Defects in Semiconductors). Moscow: Nauka, 368 p. (1981) (in Russian).
6.47. F.T. Taguchi, K. Makashima, *J. Phys. Soc. Japan* **31**: 955 (1971).

6.48. T. Nakano, Y. Inuishi, *J. Phys. Soc. Japan* **19**: 851–853 (1964).
6.49. M.K. Bakhadyrkhanov, S. Zainabidinov, A.T. Teshabaev, *FTP* **11**: 285–290 (1977) (in Russian).
6.50. M.K. Bakhadyrkhanov, S. Zainabidinov, *FTP* **11**: 2051–2053 (1977) (in Russian).
6.51. V.S. Vavilov, I.V. Smirnova, V.A. Chapin, *FTT* **4**: 1128–1131 (1962) (in Russian).
6.52. V.S. Vavilov, I.V. Kryukova, V.A. Chapin, *FTT* **6**: 2634–2637 (1964) (in Russian).
6.53. K.D. Glinchuk, N.M. Litovchenko, *UFKh* **10**: 177 (1965) (in Russian).
6.54. N.S. Rytova, *FTP* **20**: 1514–1517 (1986) (in Russian).
6.55. L.D. Landau, E.M. Lifshitz, *Statisticheskaya fizika* (Statistical Physics). Moscow: Nauka, 566 p. (1964) (in Russian).
6.56. S.P. Grishina, M.N. Kamalov, L.I. Kolesnik, *FTP* **15**: 1449–1454 (1981) (in Russian).
6.57. O.I. Danilevich, *K teorii ravnovesnykh sostoyanii primesei v poluprovodnikah* (On the Theory of Equilibrium Impurity States in Semiconductors). Kiev: AN USSR, Inst. Theor. Physics, Preprint ITF 79–6P, 11 p. (1979) (in Russian).
6.58. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushchikh defektov v poluprovodnikakh* (Thermodynamics and Kinetics of Interacting Defects in Semiconductors). Moscow: Nauka, 352 p. (1997) (in Russian).
6.59. V.I. Fistul, V.N. Tzygankov, A.G. Yakovenko, *FTP* **6**: 1396–1398 (1972) (in Russian).
6.60. V.I. Fistul, V.N. Tzygankov, A.G. Yakovenko, *FTP* **7**: 2037–2041 (1973) (in Russian).
6.61. A. Hizaki, *J. Phys. Sol. Japan* **21**: 34–40 (1966).

Chapter 7

Impurity Kinetics in Semiconductors

7.1 IMPURITY MIGRATION ENERGY

The impurity migration model to be discussed in this chapter is based on a modified Weisser theory concerning the transition of an impurity atom from the equilibrium interstitial position to a new equilibrium position through a saddle point. The theory of equilibrium positions, or the type of interstice (tetrahedral or hexagonal), discussed in Section 4.3, can be easily extended to impurity migration.

The migration energy ΔE can be represented as

$$\Delta E_{\mathrm{m}} = \Delta \mathrm{U}_{\mathrm{rep}} - \Delta U_{\mathrm{im}} - \Delta U_{\mathrm{cr}} - \Delta \delta_{\mathrm{ex}}, \tag{7.1.1}$$

where

$$\Delta U_{\mathrm{rep}} = U_{\mathrm{rep}}^{T} - U_{\mathrm{rep}}^{H} \tag{7.1.2}$$

$$\Delta U_{\mathrm{im}} = U_{\mathrm{im}}^{H} - U_{\mathrm{im}}^{T} \tag{7.1.3}$$

$$\Delta U_{\mathrm{cr}} = U_{\mathrm{cr}}^{H} - U_{\mathrm{cr}}^{T} \tag{7.1.4}$$

$$\Delta \delta_{\mathrm{ex}} = \delta_{\mathrm{ex}}^{H} - \delta_{\mathrm{ex}}^{T}. \tag{7.1.5}$$

The first three expressions representing, respectively, the energy differences of repulsion, polarization, and crystal field at *H*- and *T*-interstices were defined in Section 4.3; $\Delta\delta_{ex}$ is the difference in extrastabilization energies of these interstices. The small values of $\Delta\delta_{ex}$ will be neglected, as was done for the solubility calculations in Section 4.3. If the migrating impurity is ionized, expression (7.1.1) should be replaced by

$$\Delta E_m = \Delta U_{rep} - \Delta U_{im} - \Delta U_{cr} - \Delta\delta_{ex} + \Delta I_s, \tag{7.1.6}$$

where

$$\Delta I_s = I_s^T - I_s^H \tag{7.1.7}$$

represents the ionization potential difference of the impurity atom at both interstices.

For a particular semiconductor, ΔU_{im} and ΔU_{cr} are constant values independent of the kind of impurity, so ΔE_m for different impurities will be defined by ΔU_{rep} and ΔI_s. The numerical values of U_{im} and U_{cr} used in Section 4.3 are ΔU_{im} = 0.78 eV and ΔU_{cr} = 0.3 eV. The calculations of ΔU_{rep} with crystal lattice relaxation for transition metal impurities in silicon allowed the determination of ΔE_m values (Table 7.1). The comparison with experimental data suggests that Fe, Co, and Ni impurities in silicon diffuse over interstices in the neutral state, Cu diffuses in the Cu^+ state, and Ag and Au diffuse in a more complicated way, because their experimental values for ΔE_m are far from the calculated values for diffusion as Me^0 and Me^+. Below, we will show that they may migrate via dissociative diffusion.

Table 7.1. Migration energies (eV) of amphoteric transition metal impurities in silicon [1].

Impurity	ΔE_m (Me^0)	ΔE_m (Me^+)	ΔE_m (exp.)
Fe	0.61	0.38	0.69 [75]
Co	0.57	0.31	0.51 [76]
Ni	0.48	0.39	0.47 [72]
Cu	0.34	0.42	0.43 [77]
Pd	0.47	0.29	–
Ag	0.44	0.24	1.6 [78]
Pt	0.34	1.43	–
Au	0.53	0.53	0.38 [79]

A theoretical analysis of migration energies ΔE_m is possible, at present, only for elemental semiconductors of the silicon-type. The difficulties associated with $A^{III}B^{V}$ and $A^{II}B^{IV}$ semiconductor compounds are similar to those discussed in Section 4.3.

7.2 MICROSCOPIC THEORY OF IMPURITY KINETICS

Impurities exhibit various migration effects at high temperatures. One is *dissociative diffusion* which represents a combined migration of impurity atoms over lattice interstices and vacancies with an exchange of positions, i.e., the capture of an impurity atom by a vacancy to produce a free interstice. A more general migration process is *decomposition of an oversaturated solid solution*, involving diffusion as a stage in a more complicated process [2]. There are several models to describe decomposition of a semiconductor–impurity solid solution. These have been discussed in detail in the book [2]. Irrespective of which model applies better to which case, a common feature is that the interaction of impurity atoms with other point defects is involved in any migration process.

A consistent account of these interactions in kinetic theory requires the introduction of a variable diffusion coefficient, but this requires the knowledge of its functional dependence on the concentrations of all "participants" of the kinetic process. One should bear in mind that these concentrations are interdependent and continuously vary in time, and that impurity atoms, especially amphoteric atoms, change their charge when changing their position in the crystal. Finally, if one takes into account the ability of impurity atoms to form associate defects, the problem of direct introduction of a variable diffusion coefficient into theory will appear meaningless.

More fruitful is the approach based on the equivalence of diffusional and quasichemical descriptions of concentration variation of impurity atoms and defects. In this approach, the processes involving impurity atoms are treated as chemical reactions, namely, the reactions of impurity transition from a site (s) to an interstice (i) to produce a vacancy, and vice versa:

$$\begin{array}{c} A_s^{Z_s} + Zp^+ \underset{\leftarrow}{\overset{\rightarrow}{}} A_i^{Z_i} + V^{Z_V}, \\ \qquad\qquad \Big| K_3 \quad \Big| K_4 \\ \qquad\qquad \text{sinks} \quad \text{sinks} \end{array} \tag{7.2.1}$$

where the superscript in brackets denotes the charge value of every reagent. This reaction also involves positive holes. Here, a *p*-semiconductor is discussed as an illustration; an *n*-semiconductor was analyzed in [3]. The quantities K_1 and K_2 are the rate constants of direct and reverse reactions. For an amphoteric impurity changing its charge in the transition from a site to an interstice, this change is allowed for by the inequality $Z_s \neq Z_i$ for both the value and sign. If an amphoteric impurity has several charge states simultaneously but is in the same position in the crystal, this can be allowed for by introducing into (7.2.1) the sum $A^{Z_{s1}} + A_s^{Z_{s2}} + \ldots$ or $A_i^{Z_{i1}} + A_i^{Z_{i2}} + \ldots$ We will omit this cumbersome series. The possible changes in the charge state are determined by the change in the Fermi level which can always be taken into account by using its relation to the carrier concentration.

There are two other kinetic processes taken into account by (7.2.1): the migration of interstitial impurity atoms and vacancies to sinks. These can also be treated as quasichemical reactions with their own rate constants K_3 and K_4. This assumption is permissible if the conditions formulated in [4] are fulfilled: the sink identity for all defects, a uniform distribution of sinks throughout the crystal, their constant number and unlimited capacity, as well as the absence of elastic stress fields.

Another important quasichemical reaction is the interaction between a site impurity and a vacancy to produce an associate defect X^{Z_X}:

$$A_s^{Z_s} + V^{Z_V} \underset{K_6}{\overset{K_5}{\rightleftarrows}} X^{Z_X} + Z' + p^+ . \qquad (7.2.2)$$

$$\Big| \; K_Z$$

sinks

The number of electrons involved in each of the above reactions is defined by the charge conservation condition:

$$Z = Z_s - Z_V - Z_i , \qquad Z' = Z_X - Z_V - Z_s . \qquad (7.2.3)$$

The time variation of volume average concentrations are written as kinetic equations

$$\frac{dC_s}{dt} = -K_1 C_s p^Z + K_2 C_i C_V - K_6 C_s C_V + K_5 C_X p^{Z'}$$

$$\frac{dC_i}{dt} = K_1 C_s p^Z - K_2 C_i C_V - K_3\left(C_i - C_i^0\right)$$

$$\frac{dC_V}{dt} = K_1 C_s p^Z - K_2 C_i C_V - K_6 C_s C_V + K_5 C_X p^{Z'} - K_4\left(C_V - C_V^0\right)$$

$$\frac{dC_X}{dt} = -K_5 C_X p^{Z'} + K_6 C_s C_V - K_7\left(C_X - C_X^0\right), \qquad (7.2.4)$$

where

$$K_2 = 4\pi r_0 (D_i + D_V), \qquad K_3 = \gamma\rho D_i,$$

$$K_4 = \gamma\rho D_V, \qquad K_6 = 4\pi\rho_0 D_V, \qquad (7.2.5)$$

$$K_7 = \gamma\rho D_X.$$

Here, D_i, D_V, and D_X are diffusion coefficients of respective defects; ρ is the sink density per unit area; γ is a factor determined by sink geometry, approximately equal to unity; r and ρ are effective capture radii for an interstitial atom–vacancy pair and a substitutional atom–vacancy pair, respectively.

The rate constants for direct and reverse reactions are not independent. Their relationship can be easily found from equilibrium conditions with $dC_s/dt = 0$ and $dC_X/dt = 0$ and, respectively, $C_s = C_s^0$, $C_i = C_i^0$, and $C_X = C_X^0$. With these conditions, we obtain from (7.2.6)

$$K_1 = K_2 \frac{C_i^0 C_V^0}{C_s^0} p_0^{-Z}, \qquad K_5 = K_6 \frac{C_V^0 C_s^0}{C_X^0} p_0^{-Z'}. \qquad (7.2.6)$$

The set of equations (7.2.4) must be supplemented by the crystal neutrality condition which permits finding the concentration of the fifth component—free holes, p:

$$p = Z_s C_s + Z_i C_i + Z_V C_V + Z_X C_X + n_i^2 / p. \qquad (7.2.7)$$

Here, the last term represents the concentration of negatively charged electrons.

All Z values in (7.2.7) are taken with their own signs corresponding to the character of ionization of the reactants. The set of equations (7.2.4) and (7.2.7) contains concentration products, so they are nonlinear. To solve this

set of equations, one should accept additional assumptions leading to its linearization.

There are two assumptions to be made. One is the approximation of equilibrium defect concentration (EDC) [5], which implies that the concentration of the most mobile defects is equal, at any moment of time, to the thermodynamically equilibrium concentration at a given temperature. The EDC approximation is fairly rough, because point defects are in quasi-equilibrium with one another at every moment of time, and the presence of a nonequilibrium fraction of one defect type (say, impurity atoms at lattice sites) entails an excess content of other defects.

The other approximation [6,7] is based on the assumption that excess concentrations of all defects are much lower than their equilibrium concentrations. This approximation was termed in [7] nonequilibrium defect concentration (NDC) [7]. We believe that the latter approach is preferable because it better fits experimental conditions.

The set of equations (7.2.4) was most completely solved in the work [7], where the NDC approximation was mathematically represented as the condition

$$\frac{\delta C}{C^0} = \frac{C - C^0}{C^0}. \tag{7.2.8}$$

It follows from (7.2.7) that in the absence of equilibrium, the neutrality equation is

$$\left(1 + \frac{n_i^2}{p}\right)\delta p = Z_s \delta C_s + Z_j \delta C_j + Z_V \delta C_V + Z_X \delta C_X. \tag{7.2.9}$$

Normally, the concentration of one component, say, site impurity, is larger than the other concentrations. Then, since we have $p >> p_i$, expression (7.2.9) should be replaced by

$$\delta p = Z_s \delta C_s. \tag{7.2.10}$$

To simplify the theoretical treatment, we will assume one type of defect to be dominant. The further presentation is an illustrative application of the above reasoning to the general process of solid solution decomposition, assuming at the beginning that an impurity atom leaves the substitution solution to be incorporated and then diffuses to sinks. This means that the

rates of reactions (7.2.2) are too low, and K_5 and K_6 can be taken to be zero. The kinetic equations in this case are

$$\left.\begin{aligned}
\frac{dC_s}{dt} &= -K_1 C_s p^Z + K_2 C_i C_V \\
\frac{dC_i}{dt} &= K_1 C_s p^Z - K_2 C_i C_V - K_3\left(C_i - C_i^0\right) \\
\frac{dC_V}{dt} &= K_1 C_s p^Z - K_2 C_i C_V - K_4\left(C_V - C_V^0\right)
\end{aligned}\right\}. \qquad (7.2.11)$$

In the linear approximation allowing for the neutrality condition (7.2.7), we have

$$\begin{aligned}
\frac{d}{dt}\delta C_s &= -K_2 \frac{C_i^0 C_V^0}{C_s^0}\left(1 + ZZ_s \frac{C_s^0}{p_0 + n_0}\right)\delta C_s \\
&\quad + K_2 C_V^0\left(1 - ZZ_V \frac{C_i^0}{p_0 + n_0}\right)\delta C_i + K_2 C_i^0\left(1 - ZZ_V \frac{C_V^0}{p_0 + n_0}\right)\delta C_V \\
\frac{d}{dt}\delta C_i &= K_2 \frac{C_s^0 C_V^0}{C_s^0}\left(1 + ZZ_s \frac{C_s^0}{p_0 + n_0}\right)\delta C_s \\
&\quad - \left[K_2 C_V^0\left(1 - ZZ_i \frac{C_i^0}{p_0 + n_0}\right) + K_3\right]\delta C_i - K_2 C_i^0\left(1 - ZZ_V \frac{C_V^0}{p_0 + n_0}\right)\delta C_V \\
\frac{d}{dt}\delta C_V &= K_2 \frac{C_i^0 C_V^0}{C_s^0}\left(1 - ZZ_s \frac{C_s^0}{p_0 + n_0}\right)\delta C_s \\
&\quad - K_2 C_V^0\left(1 - ZZ_i \frac{C_i^0}{p_0 + n_0}\right)\delta C_i - \left[K_2 C_i^0\left(1 - ZZ_V \frac{C_V^0}{p_0 + n_0}\right) + K_4\right]\delta C_V
\end{aligned} \qquad (7.2.12)$$

The characteristic equation of this set

$$\begin{vmatrix} \frac{1}{\tau} - K_2 \frac{C_i^0 C_V^0}{C_s^0}\left(1 + \right. & K_2 C_V^0 \left(1 - \right. & K_2 C_i^0 \left(1 - \right. \\ \left. + ZZ_s \frac{C_s^0}{p_0 + n_0}\right); & \left. - ZZ_s \frac{C_i^0}{p_0 + n_0}\right); & \left. - ZZ_V \frac{C_V^0}{p_0 + n_0}\right); \\ K_2 \frac{C_i^0 C_V^0}{C_s^0}\left(1 + \right. & \frac{1}{\tau} - K_2 C_V^0 \left(1 - \right. & - K_2 C_i^0 \left(1 - \right. \\ \left. + ZZ_s \frac{C_s^0}{p_0 + n_0}\right); & \left. - ZZ_i \frac{C_i^0}{p_0 + n_0}\right); & \left. - ZZ_V \frac{C_V^0}{p_0 + n_0}\right); \\ K_2 \frac{C_i^0 C_V^0}{C_s^0}\left(1 + \right. & - K_2 C_V^0 \left(1 - \right. & \frac{1}{\tau} - K_2 C_i^0 \left(1 - \right. \\ \left. + ZZ_s \frac{C_s^0}{p_0 + n_0}\right); & \left. - ZZ_i \frac{C_i^0}{p_0 + n_0}\right); & \left. - ZZ_V \frac{C_V^0}{p_0 + n_0}\right) \end{vmatrix} = 0 \quad (7.2.13)$$

has three roots. The smallest of them determines the time constant of the decomposition process.

To avoid cumbersome formulas, let us make some simplifying assumptions. We will put $Z_s = -1$, which is valid for the acceptor state of a site impurity but will not consider its state with $Z_s = -2$. Of course, the latter assumption simplifies the situation but oversimplifies the theoretical results. Besides, the site state will be assumed to be dominant and this means that the sample conductivity type remains unchanged during the kinetic process.

To solve the set of equations (7.2.12), let us introduce the variable x, defined as

$$x = \frac{(1/\tau)}{\left[\gamma\rho\left(D_i + D_V\right)\right]}, \quad (7.2.14)$$

and the designations

$$\alpha_i = \frac{4\pi r_0 C_i^0}{\gamma\rho}, \qquad \frac{C_i^0}{C_s^0} = \beta_i, \quad (7.2.15)$$

$$\alpha_V = \frac{4\pi r_0 C_V^0}{\gamma\rho}, \qquad \frac{C_V^0}{C_s^0} = \beta_V . \tag{7.2.16}$$

The characteristic equation (7.2.13) takes the form:

$$\begin{vmatrix} x-(1+Z)\alpha_i\beta_V & \alpha_V & \alpha_i \\ (1+Z)\alpha_i\beta_V & x-\alpha_V-\dfrac{D_i}{D_i+D_V} & -\alpha_i \\ (1+Z)\alpha_i\beta_V & -\alpha_V & x-\alpha_i-\dfrac{D_V}{D_i+D_V} \end{vmatrix} = 0 , \tag{7.2.17}$$

where Z should be substituted with its own sign: minus for an acceptor and plus for a donor.

Equation (7.2.17) is nothing else but the cubic equation

$$x^3 - \left[1+\alpha_i+\alpha_V+(1+Z)\alpha_i\beta_V\right]x^2 + \left[\frac{D_iD_V}{(D_i+D_V)^2} + \right.$$
$$\left. +(1+Z)\alpha_i\beta_V + \frac{\alpha_iD_i+\alpha_VD_V}{D_i+D_V}\right]x - \tag{7.2.18}$$
$$-(1+Z)\alpha_i\beta_V\frac{D_iD_V}{(D_i+D_V)^2} = 0$$

Since $x << 1$, we neglect the cubic term and, considering C_V to be a small value, put

$$\alpha_V << 1 \quad \text{and} \quad \beta_V << 1. \tag{7.2.19}$$

Then, (7.2.18) transforms to the quadratic equation $ax^2 + bx + c = 0$ with such a structure of coefficients that the inequality $b^2/(4ac) >> 1$ is always valid for $\alpha_i \geq 1$. Hence, the smallest root will be

$$x = \frac{c}{b}, \tag{7.2.20}$$

namely,

$$x = (1+Z)\frac{\alpha_i\beta_V \dfrac{D_i D_V}{(D_i + D_V)^2}}{\dfrac{D_i D_V}{(D_i + D_V)^2} + (1+Z)\alpha_i\beta_V + \dfrac{\alpha_i D_i + \alpha_V D_V}{D_i + D_V}}. \qquad (7.2.21)$$

The sink density ρ can be considered to be high during solid solution decomposition at any impurity concentration. Even if the sink density in the sample is small, its surface acting as a sink provides large ρ values. Then, it follows from (7.2.15) and (7.2.16):

$$\alpha_i << 1, \qquad \alpha_V << 1 \qquad (7.2.22)$$

$$\beta_i << 1, \qquad \beta_V << 1 \qquad (7.2.23)$$

Note that an amphoteric impurity dominant at lattice sites will be described by the relations $\beta_i \cong 1$ and $\beta_V \cong 1$, instead of (7.2.23). They will also be valid for deep inhomogeneous impurities. With the reverse substitution of (7.2.14) and using (7.2.22) and (7.2.23), we obtain the final dimensional solutions for two possible relations between coefficients D_i and D_V:

$$\frac{1}{\tau} = (1+Z)\gamma\rho\frac{\alpha_i\beta_V D_V}{(D_V/D_i) + \alpha_i} \quad \text{at} \quad D_i >> D_V \qquad (7.2.24)$$

$$\frac{1}{\tau} = (1+Z)\gamma\rho\frac{\alpha_i\beta_V D_i}{(D_i/D_V) + \alpha_V} \quad \text{at} \quad D_V >> D_i. \qquad (7.2.25)$$

These expressions are valid for shallow hydrogen-like impurities. For deep impurities, the solutions have a different form:

$$\frac{1}{\tau} = (1+Z)\gamma\rho\frac{\alpha_i\beta_V D_V}{(D_V/D_i)\alpha_V + (1+Z)\beta_V\alpha_i + \alpha_i} \quad \text{at } D_i >> D_V \qquad (7.2.26)$$

$$\frac{1}{\tau} = (1+Z)\gamma\rho\frac{\alpha_i\beta_V D_i}{(D_i/D_V)\alpha_i + (1+Z)\beta_V\alpha_i + \alpha_V} \quad \text{at } D_V >> D_i. \qquad (7.2.27)$$

The solutions obtained as various values of $1/\tau$ determine the experimental time constants of solution decomposition:

Table 7.2. Effective activation energies Q_{eff} for interstitial decomposition of amphoteric impurities at $\alpha_i >> 1$, $\alpha_V << 1$, $\beta_i \cong 1$, $\beta_V \cong 1$.

Kinetic conditions	Activation energies
	$D_V/D_i >> 1$
$(Z+1)\beta_V > D_i/D_V > \beta_V/\beta_i$	$-E_i^M$
$(Z+1)\beta_V < D_i/D_V > \beta_V/\beta_i$	$-E_V + E_s - E_V^M$
$(Z+1)\beta_V > D_i/D_V < \beta_V/\beta_i$	$-E_i^M$
$(Z+1)\beta_V < D_i/D_V < \beta_V/\beta_i$	$-E_i + E_s - E_i^M$
	$D_V/D_i << 1$
$(Z+1)\beta_V > D_V/D_i < \beta_i/\beta_V$	$-E_V$
$(Z+1)\beta_V < D_V/D_i < \beta_i/\beta_V$	$-E_V + E_s - E_V^M$
$(Z+1)\beta_V > D_V/D_i > \beta_i/\beta_V$	$-E_V^M$
$(Z+1)\beta_V < D_V/D_i > \beta_i/\beta_V$	$-E_i + E_s - E_i^M$

Table 7.3. Effective activation energies Q_{eff} for interstitial decomposition of amphoteric impurities at $\alpha_i << 1$, $\alpha_V << 1$, $\beta_i \cong 1$, $\beta_V \cong 1$.

Kinetic conditions	Activation energies
$1 > D_i/D_V < \alpha_V$	$-E_i + E_s - E_i^M$
$1 > D_i/D_V > \alpha_V$	$-E_i - E_V + E_s - E_V^M$
$1 < D_i/D_V < \alpha_i$	$-E_V + E_s - E_V^M$
$1 > D_i/D_V > \alpha_i$	$-E_i - E_V + E_s - E_i^M$

$$\tau = \tau_0 \exp(Q/kT), \tag{7.2.28}$$

where Q is the effective activation energy representing a combination of the activation energies of diffusion in $D_i(T)$ and $D_V(T)$ and the energy defining the temperature dependences of solubility $C_i^0(T)$, $C_V^0(T)$, and $C_s(T)$ present in the quantities α_i, α_V, β_i, and β_V. From a comparison of experimental values of Q and Q_{eff} (Tables 7.2 and 7.3), one can understand the decomposition mechanism in the crystals under study.

For compiling Tables 7.2 and 7.3, we used the identity $\alpha_V/\alpha_i = \beta_V/\beta_i$ which follows from (7.2.15) and (7.2.16), because the values of $\gamma\rho$ present in α were very difficult to determine experimentally. The above solutions for $1/\tau$ were obtained in the NDC approximation. The solutions in the EDC ap-

proximation will be similar but having no factor ($Z + 1$). They are simpler and easier to use in experimental data processing. Let us see what physical picture corresponds to each approximation.

The determinant in (7.2.17) is symmetrical relative to interstitial atoms and vacancies. So, for simplicity, consider only the case with $D_V >> D_i$. Let us expand (7.2.17) into the third column elements, taking account of the inequality $x << 1$:

$$\alpha_i \begin{vmatrix} (Z+1)\alpha_i\beta_V & \alpha_V \\ (Z+1)\alpha_i\beta_V & x - \alpha_V - (D_i/D_V) \end{vmatrix} + \alpha_i \begin{vmatrix} x - (Z+1)\alpha_i\beta_V & \alpha_V \\ (Z+1)\alpha_i\beta_V & -\alpha_V \end{vmatrix} -$$

$$-(\alpha_i + 1) \begin{vmatrix} x - (Z+1)\alpha_i\beta_V & \alpha_V \\ (Z+1)\alpha_i\beta_V & x - \alpha_V - (D_i/D_V) \end{vmatrix} = 0. \qquad (7.2.29)$$

The diagonal element of the third column of (7.2.17) contains two terms. The first one, α_i, characterizes the elimination rate of free vacancies from the crystal volume due to recombination with interstitial impurity atoms, according to reaction (7.2.1). The second term, $D_V/(D_i + D_V) \approx 1$, describes the migration of free vacancies to sinks. If vacancy diffusion to sinks dominates, i.e., $\alpha_i << 1$, the expansion of (7.2.29) will contain only the third term:

$$\begin{vmatrix} x - (Z+1)\alpha_i\beta_V & \alpha_V \\ (Z+1)\alpha_i\beta_V & x - \alpha_V - (D_i/D_V) \end{vmatrix} = 0, \qquad (7.2.30)$$

which corresponds to the EDC approximation.

Indeed, after a fast elimination of nonequilibrium vacancies from a crystal, their concentration comes to equilibrium in a very short time equal to the diffusion time. If a vacancy recombines with an interstitial atom sooner than it diffuses to a sink, i.e., $\alpha_i >> 1$, the expansion of (7.2.29) will contain all the terms. For the set of equations (7.2.13), this will correspond to the condition $(d/dt)\delta C_V = 0$, or to the NDC approximation.

Thus, the physical sense of the NDC approximation is that the concentration of the most mobile defect is "adjusted" fast to the concentrations of the other reactants and comes to a quasi-equilibrium state with slower reactants.

A similar analysis of the set of kinetic equations (7.2.4) can be carried out for a more complicated decomposition process occurring via reactional diffusion and a mixed mechanism. Reactional diffusion implies the formation of associate defects (substitutional impurity–vacancy complex) and

their migration to the sinks. The mixed mechanism represents a simultaneous migration of associate defects and interstitial atoms to the sinks. For reactional diffusion, the set of equations (7.2.4) takes the form:

$$\frac{dC_s}{dt} = -K_6 C_s C_V + K_5 C_X n^{Z'},$$

$$\frac{dC_V}{dt} = -K_6 C_s C_V + K_5 C_X n^{Z'} - K_4 \delta C_V, \qquad (7.2.31)$$

$$\frac{dC_X}{dt} = K_6 C_s C_V - K_5 C_X n^{Z'} - K_7 \delta C_X,$$

where C_X is the concentration of associate defects and $Z' = Z_s - Z_V - Z_X$.

Let us ignore, for the time being, a possible multiple charge of the site amphoteric impurity by taking $Z = 1$. Then we can write, in the linear approximation (7.2.8) with the neutrality equation:

$$\frac{d}{dt}\delta C_s = -(Z'+1)K_6 C_V^0 \delta C_s - K_6 C_s^0 \delta C_V + K_6 \frac{C_V^0 C_s^0}{C_X^0}\delta C_X,$$

$$\frac{d}{dt}\delta C_V = -(Z'+1)K_6 C_V^0 \delta C_s - \left(K_6 C_s^0 + K_4\right)\delta C_V + K_6 \frac{C_V^0 C_s^0}{C_X^0}\delta C_X,$$

$$\frac{d}{dt}\delta C_X = (Z'+1)K_6 C_V^0 \delta C_X + K_6 C_s^0 \delta C_V - \left(K_6 \frac{C_V^0 C_s^0}{C_X^0} + K_7\right)\delta C_X. \qquad (7.2.32)$$

The characteristic cubic equation of this set for dimensionless time constants

$$x = \frac{1}{\tau}\frac{1}{\gamma \rho D_V} \qquad (7.2.33)$$

has the form

$$x^3 - x^2\left[(Z'+1)\alpha_V + \alpha_s + \alpha_V\frac{C_s^0}{C_X^0} + 1 + \frac{D_X}{D_V}\right] + x\left[(Z'+1)\left(1+\frac{D_X}{D_V}\right)\alpha_V + \right.$$
$$\left. + (1+\alpha_s)\frac{D_X}{D_V} + \alpha_V\frac{C_s^0}{C_X^0}\right] - (Z'+1)\alpha_V\frac{D_X}{D_V} = 0 \qquad (7.2.34)$$

where we included, in addition to (7.2.15),

$$\alpha_s = \frac{4\pi r_0 C_s^0}{\gamma\rho}. \qquad (7.2.35)$$

As in the case of interstitial decomposition mechanism, the cubic term will be neglected. Assuming $D_X/D_V << 1$ and $\alpha_s << 1$, we obtain an equation of the type $ax^2 + bx + c = 0$:

$$x^2\left(1+\frac{C_V^0}{C_X^0}\right) - x\left(\frac{D_X}{D_V} + \frac{C_V^0}{C_X^0}\right) + (Z'+1)\frac{D_X C_V^0}{D_V C_s^0}. \qquad (7.2.36)$$

After the reverse substitution, the smallest root of this equation, $x = c/b$, is

$$\frac{1}{\tau} = (Z'+1)\frac{\gamma\rho D_X\, C_V^0/C_s^0}{(D_X/D_V) + \left(C_V^0/C_X^0\right)}. \qquad (7.2.37)$$

In contrast to the interstitial decomposition mechanism, there are only two limit kinetic conditions here:

$$\frac{D_X}{D_V} << \frac{C_V^0}{C_X^0}: \qquad \frac{1}{\tau} = (Z'+1)\gamma\rho D_X\left(C_X^0/C_s^0\right), \qquad (7.2.38)$$

$$\frac{D_X}{D_V} >> \frac{C_V^0}{C_X^0}: \qquad \frac{1}{\tau} = (Z'+1)\gamma\rho D_V\left(C_V^0/C_s^0\right). \qquad (7.2.39)$$

Inequality (7.2.38) takes into account the identical relation

Table 7.4. Effective activation energies Q_{eff} for reactional and mixed decomposition.

Decomposition mechanism	Kinetic condition	Activation energy
Migration of defect associates	$D_X/D_V << C_V^0/C_X^0$	$-E_X^M - E_X + E_s \equiv E_s^M$
Mixed:	$D_X/D_V >> C_V^0/C_X^0$	$-E_V^M - E_V + E_s$
Simultaneous migration of associates and interstitial atoms	$D_i/D_X << 1$	$-E_i^M$
	$D_i/D_X >> 1$	$-E_s^M - E_s + E_X = -E_X^M$

$$D_s \equiv D_X\left(C_X^0 / C_s^0\right) \qquad \text{at} \qquad C_X^0 << C_s^0 . \tag{7.2.40}$$

The values of Q_{eff} for this decomposition mechanism (migration) are given in Table 7.4.

For the mixed mechanism occurring via migration of associates and interstitial atoms, the linearization of the set of equations (7.2.4) gives four linear equations, making it difficult to derive an analytical solution for $1/\tau$. To reduce the number of equations, we will assume, as in [2], that vacancies are the most mobile defects and that $C_V = C_V^0$ at every moment of time. In other words, vacancies will be treated in the EDC approximation and other defects in the NDC approximation.

The linearized set of equations has the form:

$$\begin{aligned}
\delta\frac{dC_s}{dt} &= K_2 C_V^0 \delta C_i - (Z+1)K_1\left(C_s^0\right)^Z \delta C_s + K_5 \delta C_X - K_6 C_V^0 \delta C_s , \\
\delta\frac{dC_i}{dt} &= -K_2 C_V^0 \delta C_i + (Z+1)K_1\left(C_s^0\right)^Z \delta C_s - K_3 \delta C_i , \\
\delta\frac{dC_X}{dt} &= K_6 C_V^0 \delta C_s - K_3 \delta C_X - K_7 \delta C_X .
\end{aligned} \tag{7.2.41}$$

It should now be found what kind of diffusion limits the process. Assume at first $D_X >> D_i$, i.e., the limiting factor is interstitial diffusion. Then, in the first approximation, we have $\delta dC_X/dt = 0$, and δC_X is found from the last equation of (7.2.41) as

$$\delta C_X = \frac{K_6 C_V^0 \delta C_s}{K_5 + K_7}. \tag{7.2.42}$$

By substituting (7.2.42) into the first two equations of (7.2.41) and equating the characteristic equation to zero with respect to $1/\tau$, we find

$$\left(\frac{1}{\tau}\right)^2 - \frac{1}{\tau}\left[(Z+1)K_1\left(C_s^0\right)^Z + K_6 C_V^0 + K_2 C_V^0 + K_3\right] + \\ + K_3\left[(Z+1)\left(C_s^0\right)^Z + K_6 C_V^0\right] + K_6 C_V^0 K_2 C_V^0 = 0. \tag{7.2.43}$$

This equation was derived for $K_7 >> K_5$, i.e., under the condition that associates reach sinks before they dissociate.

Let us take $K_3 >> K_2 C_V^0$; otherwise, interstitial atoms would annihilate rather than reach sinks. If interstitial diffusion is the limiting factor, then

$$(Z+1)K_1\left(C_s^0\right)^Z + K_6 C_V^0 + K_2 C_V^0 >> K_3. \tag{7.2.44}$$

Using the above assumptions and simple transformations, we obtain the minimum time constant of decomposition:

$$\frac{1}{\tau} = K_3 = \gamma\rho D_i. \tag{7.2.45}$$

Similarly, the other limiting case will be described as

$$\frac{1}{\tau} = K_7 = \gamma\rho D_X \equiv \frac{\gamma\rho D_s C_s^0}{C_X^0}. \tag{7.2.46}$$

The respective expressions for Q_{eff} can be found in Table 7.4.

7.3 DISSOCIATIVE DIFFUSION OF IMPURITIES

Dissociative diffusion represents impurity migration involving a change of the crystallographic position. It is described by quasichemical reactions (7.2.1). Its difference from decomposition is that the diffusant enters a sample from the outside because diffusion is a technological process. In decomposition, it comes from the inside, i.e., from the lattice sites. For this reason, the theoretical treatment of dissociative diffusion is basically similar to that of decomposition carried out in Section 7.2. However, the dissociative diffusion analysis requires the consideration of surface concentration C_{sur}, which decreases with depth in time and along the coordinate. The measure of its decrease in time is the time constant $1/\tau$ determined by the migration mechanism in the sample bulk. Its coordinate dependence is determined by the sample shape and the boundary conditions of the diffusion problem.

The treatment of diffusion involves two problems. The first problem deals with the macroscopic diffusion coefficient D, while the second reveals its microscopic nature, because D is a combination of partial diffusion coefficients of the components D_s, D_i, D_V, and D_X and of their concentrations C_s, C_i, C_V, and C_X. It does not seem worthwhile discussing the first problem, since it is of interest mostly to particular situations which will be described in Section 7.5. The general analysis of the second problem (Section 7.2) is too cumbersome to be compared with experiments. Also, a simplified treatment is often sufficient for a qualitative and semi-quantitative interpretation of experimental data. This simplified theory will be presented in this section.

The principal simplification is the use of the EDC approximation, but it will be necessary to find which defect is in equilibrium. Below, the criteria for its identification will be formulated. Further, the charge states of defects will be neglected in the first approximation. The charge value can be introduced into the derived formulas any moment by multiplying the result by $(Z + 1)$, as was done in Section 7.2. Finally, this treatment will be restricted to two defects—impurity atoms, occupying sites and interstices, and vacancies. In other words, the interaction of defects producing associates, or reactional diffusion, will be ignored.

With these assumptions, the set of kinetic equations (7.2.4) becomes simplified:

$$\frac{dC_s}{dt} = -K_1 C_s + K_2 C_i C_V, \tag{7.3.1}$$

$$\frac{dC_i}{dt} = K_1 C_s - K_2 C_i C_V - K_3\left(C_i - C_i^0\right), \tag{7.3.2}$$

$$\frac{dC_V}{dt} = K_1 C_s - K_2 C_i C_V - K_4\left(C_V - C_V^0\right). \qquad (7.3.3)$$

The relation between K_1 and K_2 is defined, as before, by expression (7.2.6) but with $Z = 0$.

The nature of sinks was not discussed in Section 7.2, so we will specify it now.

Sinks for interstitial impurity atoms and vacancies may be dislocations, the sample surface, and defects of the impurity associate type. If the latter are neglected (this can be done for the usual samples containing dislocations), the rate constants K_3 and K_4, according to [8] will be:

$$K_{3,4} = \left(\pi/l\right)^2 D_{i,V} + \frac{2\pi n_D D_{i,V}}{\ln\left(r_D/r_0'\right)}. \qquad (7.3.4)$$

The first term in this expression reflects recombination on the surface of a sample of thickness l. The second term corresponds to recombination of impurities or vacancies on dislocations with density n_D, with the effective capture radius r_0'. The value of r_D is equal to the half mean distance between dislocations, which is, in turn, defined by their density:

$$r_D = \frac{1}{\sqrt{\pi n_D}}. \qquad (7.3.5)$$

The linearization of equations (7.3.1) through (7.3.3) can be carried out in the EDC approximation with $C_i = C_i^0$ or $C_V = C_V^0$. Which of these conditions should be chosen for experimental data processing? The answer to this question was found in [8] by analyzing the distribution "tails" of concentration C_s, C_i, and C_V established at the end of the diffusion process when relation (7.2.8) becomes valid. Substituting it into (7.3.1) through (7.3.3) for every concentration and neglecting the terms of the second order smallness $\delta C_i \delta C_V$, we obtain a set of differential equations similar to (7.2.12) but much simpler because of the above assumptions:

$$\frac{d\delta C_s}{dt} = -K_1 \delta C_s + K_2\left(C_i^0 \delta C_V + C_V^0 \delta C_i\right),$$
$$\frac{d\delta C_i}{dt} = K_1 \delta C_s - K_2\left(C_i^0 \delta C_V + C_V^0 \delta C_i\right) - K_3 \delta C_i, \qquad (7.3.6)$$

$$\frac{d\delta C_V}{dt} = K_1\delta C_s - K_2\left(C_i^0\delta C_V + C_V^0\delta C_i\right) - K_4\delta C_V .$$

The sets of equations (7.2.12) and (7.3.6) will drop out when $Z = 0$ and (7.2.6) are substituted into (7.2.12).

The left-hand sides of equations (7.3.6) will vanish in the range of large time intervals when the concentrations are close to equilibrium values. Then, the solutions for C_s, C_i, and C_V have the exponential form $\exp(\alpha t)$, where α is the minimum root of the characteristic equation

$$x^3 + ax^2 + bx + c = 0, \tag{7.3.7}$$

in which

$$\begin{aligned} a &= K_1 + K_4 + K_3 + K_2C_V^{\,0} + K_2C_i^{\,0}, \\ b &= K_1K_4 + K_1K_3 + K_2K_4C_V^{\,0} + K_2K_3C_i^{\,0} + K_3K_4, \\ c &= K_1K_3K_4. \end{aligned} \tag{7.3.8}$$

Similarly to the solution to equation (7.2.13), the minimum root of equation (7.3.7) is $x = c/b$. It is this root that determines the value of α

$$\alpha = \frac{K_1K_3K_4}{K_1K_4 + K_1K_3 + K_2K_4C_V^0 + K_2K_3C_i^0 + K_3K_4} \tag{7.3.9}$$

and essentially coincides with (7.2.21) in terms of diffusion coefficients.

To linearize equations (7.3.1) through (7.3.3), let us put, at first, $C_i = C_i^0$. Then the equation of the set (7.3.2) turns to zero and the other two equations transform to

$$\begin{aligned} \frac{dC_s}{dt} &= -K_1C_s + K_2C_i^0C_V \\ \frac{dC_V}{dt} &= K_1C_s - K_2C_i^0C_V - K_4\left(C_V - C_V^0\right). \end{aligned} \tag{7.3.10}$$

The respective characteristic equation defining α in this particular case is

$$x^2 + x\left(K_1 + K_2C_i^0 + K_4\right) + K_1K_4 . \tag{7.3.11}$$

The minimum root α is defined as

$$\alpha = \frac{K_1 K_4}{K_1 + K_2 C_i^0 + K_4}. \qquad (7.3.12)$$

Similarly, the author of [8] derived the expression

$$\alpha = \frac{K_1 K_3}{K_1 + K_2 C_V^0 + K_3}, \qquad (7.3.13)$$

which is valid for $C_V = C_V{}^0$, i.e., another linearization condition for equations (7.3.1) through (7.3.3). It is easy to see, from the comparison of (7.3.12) and (7.3.13) with the general result in (7.3.9), that the criteria for this or that linearization condition are the relations:

$$K_1 K_3 + K_2 K_3 C_i^0 + K_3 K_4 > K_1 K_4 + K_2 K_4 C_V{}^0 \quad \text{at} \quad C_i = C_i^0, \qquad (7.3.14)$$

$$K_1 K_4 + K_2 K_4 C_V{}^0 + K_3 K_4 > K_1 K_3 + K_2 K_3 C_i^0 \quad \text{at} \quad C_V = C_V{}^0. \qquad (7.3.15)$$

If the reaction constants, the sink concentrations, and the sources of vacancies and interstitial atoms are known, one can find from (7.3.14) and (7.3.15) which point defects—vacancies or interstitial atoms—are closer to equilibrium and, thereby, to know which equation—(7.3.2) or (7.3.3)—should be used together with (7.3.1) for the interpretation of experimental diffusion data on dissociative and cation–anion amphoteric impurities.

If (7.3.14) is applicable, dissociative diffusion will be described by the set of equations (7.3.10), whose solution at $K_2 C_i^0 > K_1 K_4$ with the initial conditions $t = 0$ and $C_s = C_V = 0$ is [9]

$$C_s = C_s^0 - C_s\left(1 + \frac{C_V K_4}{C_s^0 K_2 C_i^0}\right)\exp\left(-\frac{C_V^0}{C_s^0} K_4 t\right) + \frac{C_V^0 K_4}{K_2 C_i^0}\exp\left(-K_2 C_i^0 t\right). \qquad (7.3.16)$$

At large t, the ratio $C_V{}^0 K_4/(C_s{}^0 K_2 C_i{}^0 < 1)$ and expression (7.3.16) transform to

$$C_s = C_s^0\left[1 - \exp\left(-\frac{C_V^0}{C_s^0} K_4 t\right)\right]. \qquad (7.3.17)$$

If the other criterion, i.e., (7.3.15), is applicable, only the first two equations remain from the sets of (7.2.1) through (7.3.3); their solution derived under similar conditions is

$$C_s = C_s^0 - C_s^0\left(1+\frac{C_i K_3}{C_s^0 K_1 C_V}\right)\exp\left(-\frac{C_i}{C_s^0}K_3 t\right)+\frac{C_i^0 K_3}{C_V^0 K_1}\exp\left(-K_1 C_V^0 t\right) \tag{7.3.18}$$

and for large t, we have

$$C_s = C_s^0\left[1-\exp\left(-\frac{C_i^0}{C_s^0}K_3 t\right)\right]. \tag{7.3.19}$$

The foregoing referred to fixed sources and sinks of vacancies and interstitial impurity atoms. But in real crystals, there are also unfixed sources and sinks, which change their capacity with the degree of generation or absorption of vacancies or atoms. An example of an unfixed source is a Frank dislocation loop. Such loops act as vacancy sources, and as the vacancies are generated, they become larger, changing their ability to generate vacancies.

The mathematical description of diffusion, taking unfixed sources or sinks into account, is similar to that of diffusion with a movable boundary. It is, however, difficult to apply this theory to the mixed case of simultaneous action of fixed and unfixed sinks (or sources), although it is this situation that is so typical of dissociative diffusion of impurities in semiconductors.

An alternative approach to the analysis of vacancy diffusion [9] implies the presence of independent and non-interacting fixed (C_{Vf}) and unfixed (C_{Vv}) sinks and sources. The reader can find this cumbersome analysis in the original work [9]. Here, we will give only the final expression for $C_s(t)$ in dissociative diffusion when fixed and unfixed vacancy sinks (sources) act simultaneously:

$$C_s = C_s^0\left[f(t)+\frac{\beta^2 C_s^0}{4}\left[f(t)\right]^2\right]. \tag{7.3.20}$$

The following designations are used here: $\beta = \pi\sqrt{b/n_f}$, where b is the Burgers vector of a dislocation loop and n_f is the concentration of fixed sinks and sources; $f(t)$ is the function [9]

$$f(t) = \frac{2\left[1 - \exp\left(-\sqrt{\beta^2 C_s^0 + 1} K_{Vf} \frac{C_V^0}{C_s^0} t\right)\right]}{1 + \sqrt{\beta^2 C_s^0 + 1} - \left(1 - \sqrt{\beta^2 C_s^0 + 1}\right) \exp\left(-\sqrt{\beta^2 C_s^0 + 1} K_{Vf} \frac{C_V^0}{C_s^0} t\right)}. \quad (7.3.21)$$

in which K_{Vf} is the rate constant of vacancy migration to fixed sinks.

In the particular case of unfixed sources (sinks), or at $\beta = \infty$, equation (7.3.20) transforms to

$$C_s = C_s^0 \text{th}^2 \varphi, \quad (7.3.22)$$

where φ is the function of time defined in [9]. So, instead of (7.3.20), we will have

$$C_s = C_s^0 \text{th}^2 \left(\frac{\pi^2}{12} a D_V \sqrt{n_V C_s^0} \frac{C_V^0}{C_s^0} t\right). \quad (7.3.23)$$

To conclude this section, Table 7.5 gives expressions for $C_s(t)$ and respective conditions, which should be used for the analysis of experimental data on dissociative diffusion.

If the criterion of (7.3.14) or (7.3.15) is not fulfilled, the equations in Table 7.5 derived in the EDC approximation become invalid. Then, experimental data are to be analyzed in terms of the more general theory developed in Section 7.2.

Table 7.5. Dissociative diffusion equations and conditions for data analysis.

Type of sink (source)	Equilibrium concentration	Applicable formula	Condition	Formula $C_s(t)$	Formula $C_s(t \to \infty)$
Fixed	$C_i = C_i^0$	(7.3.14)	$C_s\|_{t=0} = 0$ $C_V\|_{t=0} = 0$	(7.3.16)	(7.3.17)
	$C_V = C_V^0$	(7.3.15)	$K_2 C_i^0 > K_1 K_4$ $C_s\|_{t=0} = 0$ $C_i\|_{t=0} = 0$	(7.3.18)	(7.3.19)
Unfixed	$C_i = C_i^0$	(7.3.14)	–	(7.3.23)	–
Mixed	$C_i = C_i^0$	(7.3.14)	–	(7.3.20)	–

7.4 KINETIC EFFECTS IN SUBSURFACE LAYERS

The impurity diffusion is characterized by high rates, because it is stimulated by an interstitial present as a single diffusant or a component of dissociative diffusion. This circumstance leads to important features of impurity profiles in heated semiconductor samples. The principal mechanism determining the impurity profiles at a semiconductor surface involves the interaction between impurity atoms and vacancies, which are always present in excessive quantities near the surface. Initially, this mechanism was attributed to the injection of vacancies from the surface into the crystal bulk. But it has recently been established [10] that an essential contribution is made by the process, in which vacancies unite to produce pores with their subsequent decoration by the diffusing impurity. The production of excess nonequilibrium vacancies is of no importance. For example, the nonequilibrium processing of a GaAs sample in [10] was performed by arsenic evaporation and in [11] by chemical etching of silicon. So, the theoretical analysis of near-surface impurity kinetics was carried out in [12] in a general form, irrespective of the vacancy injection technique used.

Suppose there are excess nonequilibrium vacancies, N_{V0} in a subsurface crystal layer of thickness δ (Figure 7.1). What processes occur in it when the crystal is heated stage-by-stage? In the first stage, the oversaturated vacancy solution is decomposed and vacancies diffuse. They produce the "second vacancy phase"—pores. After most vacancies have been utilized for the production of pores and the nonequilibrium vacancy concentration becomes low, their diffusion to the surface and into the bulk will be accompanied by coalescence of pores (the second stage). At the end of this stage, all excess vacancies will come up to the crystal surface.

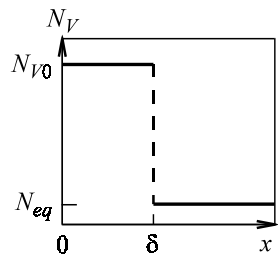

Figure 7.1. The vacancy profile at the initial moment of time prior to thermal treatment of the sample; N_{eq} – equilibrium vacancy concentration in the semiconductor.

The system evolution in the first stage is described by equations [12]

$$\frac{\partial N_V}{\partial t} = D\frac{\partial^2 N_V}{\partial r^2} - 4\pi N D r N_V, \tag{7.4.1}$$

$$\frac{\partial r^2}{\partial t} = \frac{2N_V}{N_L} D, \tag{7.4.2}$$

where N_V is the concentration of nonequilibrium vacancies, D is their diffusion coefficient, r is the pore radius, N_L is the vacancy concentration in a pore or reciprocal volume $1/\Omega$ per vacancy, and N is the concentration of pores.

The first right-hand term in (7.4.1) describes nonequilibrium vacancy diffusion from the surface into the sample bulk and the second term describes the vacancy fraction extracted for pore formation. The initial and boundary conditions of the problem, according to Figure 7.1, are

$$N_V\big|_{x=0} = 0, \qquad N_V\big|_{t=0} = \begin{cases} N_{V0} & \text{at} \quad x \le \delta \\ 0 & \text{at} \quad x > \delta \end{cases}, \qquad r\big|_{t=0} = 0. \tag{7.4.3}$$

The first condition implies a fast absorption of vacancies on the surface, i.e., the characteristic time for excess vacancy absorption at $x = 0$ is much smaller than the other characteristic times in this problem.

Consider the solution to the problem for the first stage, as was done in [12]. For this, substitute (7.4.2) into (7.4.1), integrate with respect to time and introduce new variables according to the equalities:

$$x = \mu\xi, \qquad t = \nu\tau, \qquad \tau^2 = \lambda q, \tag{7.4.4}$$

where

$$\mu^6 = \frac{N_L}{2N_{V0}} \frac{1}{\left(8/3\pi N\right)^2}, \tag{7.4.5}$$

$$\nu = \mu^2 D, \tag{7.4.6}$$

$$\lambda = \mu^{-4}\left(8/3\pi N\right)^2. \tag{7.4.7}$$

As a result, expressions (7.4.1) and (7.4.2) take the dimensionless form:

$$\frac{\partial q}{\partial \tau} = \frac{\partial^2 q}{\partial \xi^2} - q^{3/2} + f(\xi), \tag{7.4.8}$$

$$N_V = N_{V0} \frac{\partial q}{\partial \tau}, \tag{7.4.9}$$

$$f(\xi) = \begin{cases} 1 & \text{at} \quad \xi \le \Delta \\ 0 & \text{at} \quad \xi > \Delta \end{cases} \qquad \Delta = \frac{\delta}{\mu} \tag{7.4.10}$$

with the boundary and initial conditions:

$$|q|_{\tau=0} = 0, \qquad |q|_{\xi=0} = 0, \qquad |q|_{\xi=\infty} = 0. \tag{7.4.11}$$

The function $f(\xi)$ is found by substituting the initial conditions into equation (7.4.8). The first stage duration τ varies with the elimination time of most vacancies, which can be found from (7.4.8) by excluding the diffusion term $\partial^2 q/\partial \xi^2$ (in the order of magnitude, ≈ 1). The necessary condition for pore formation at the surface is the restriction $\tau_i \le \tau_s$, which means that the pore formation time should not be larger than the time of vacancy migration to the surface. At the end of the first stage, $\partial q/\partial \tau = 0$, and the established pore profile can be found from (7.4.8) by equating this expression to zero with the boundary conditions of (7.4.11). The qualitative view of the solution obtained is shown in Figure 7.2. The parameter determining the profile shape is Δ having the sense of the characteristic time ratio of the vacancy migration to the surface and the formation of a pore.

Every curve in Figure 7.2 can be subdivided into three regions of ξ coordinate variation. The respective solutions for the regions are [12]:

$$0 \le \xi \le \xi_{\max}: \qquad \int_0^q \frac{dq}{\sqrt{2\left(\frac{2}{5} q^{5/2} - q + q_\Delta\right)}}, \tag{7.4.12}$$

$$\xi_{\max} \le \xi < \Delta: \qquad \xi_{\max} + \int_0^{q_{\max}} \frac{dq}{\sqrt{2\left(\frac{2}{5} q^{5/2} - q + q_\Delta\right)}} = \xi, \tag{7.4.13}$$

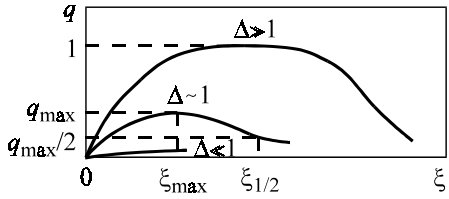

Figure 7.2. The square pore radius R_p^2 as a function of distance to the semiconductor surface (in dimensionless units) at various values of the Δ parameter; $\xi_{1/2}$ – the coordinate defined by the equation $q(\xi_{1/2}) = q_{max}/2$.

$$\Delta \le \xi < \infty : \qquad q = \frac{400}{\left(\xi + \xi_0\right)^4}, \tag{7.4.14}$$

where

$$\xi_{max} = \int_0^{q_{max}} \frac{dq}{2\left(\frac{2}{5}q^{5/2} - q + q_\Delta\right)}, \quad \xi_0 = \left(\frac{400}{q_\Delta}\right)^{1/4} - \Delta, \tag{7.4.15}$$

q_Δ and q are defined by the equations

$$\frac{2}{5}q_{max}^{5/2} - q_{max} + q_\Delta = 0, \tag{7.4.16}$$

$$\int_{q_\Delta}^{q_{max}} \frac{dq}{\sqrt{2\left(\frac{2}{5}q^{5/2} - q + q_\Delta\right)}} = \Delta - \xi_{max}. \tag{7.4.17}$$

In the limit cases, we have

$$\Delta << 1: \quad q_{max} \approx q_\Delta \approx \frac{\Delta^2}{2}, \quad \xi_{1/2} \approx 2.2\Delta + \frac{6.2}{\sqrt{\Delta}}, \tag{7.4.18}$$

$$\Delta >> 1: \quad q_{max} \approx 1, \qquad \xi_{1/2} \approx \Delta + 2.5, \tag{7.4.19}$$

It is seen from Figure 7.2. that the profile has a stepwise shape at $\Delta >> 1$; at $\Delta << 1$, the profile halfwidth increases and the profile becomes slightly smeared.

It should be noted that the pore profile at $\xi \geq \Delta$ coincides with that defined as a particular case in [13]:

$$r \sim \left(1 + \frac{x}{x_0}\right)^{-2} \tag{7.4.20}$$

with the characteristic value of x_0

$$x_0 = \left(\frac{N_V D N^2 \Omega}{v}\right)^{-1/5}, \tag{7.4.21}$$

where v is the surface motion rate during crystal evaporation or etching.

The duration of the second stage when the pores become larger, or the "lifetime" of pores, can be evaluated from the coalescence characteristic time [14]:

$$t_{II} = \frac{r^3 kT}{D\sigma\Omega^2 \left(N_V\right)_{eq}}, \tag{7.4.22}$$

where σ is the surface tension at the crystal–vacuum interface.

The dimensional characteristic times of the first stage are

$$t_I = v\tau_1 = \frac{\mu^2}{D}, \tag{7.4.23}$$

$$t_s = v\tau_s = \frac{\delta^2}{D}. \tag{7.4.24}$$

The quantitative evaluation of the times [12] shows that the relations

$$t_I \leq t_s << t_{II} \tag{7.4.25}$$

are valid in a wide temperature range for Ge, Si, and GaAs. Therefore, we can draw the conclusion that a quasi-stationary pore profile is formed in the

subsurface layer of these crystals. Its formation requires the fulfillment of the condition $\Delta \geq 1$:

$$\delta\left(\frac{128\pi^2}{9}\frac{N_{V0}N^2}{N_L}\right) \geq 1. \qquad (7.4.26)$$

The typical characteristic time values for GaAs at 1000 *K* lie within the following ranges: t_I = 0.2–20 s, t_s = 1–10 s, $t_{II} = 4\times10^5$–4×10^7 s. If an impurity is diffused into a crystal with the subsurface vacancy profile, it will decorate the profile if it can cover a distance ~δ for a time t_A shorter than t_{II}. Therefore, the condition for the formation of a subsurface impurity profile is

$$t_A = \frac{\delta^2}{D_A} \leq t_{II}, \qquad (7.4.27)$$

where D_A is the impurity diffusion coefficient. The analytical expression to describe the impurity profile derived in [13] is a power function

$$N_A \sim \left(1+\frac{x}{x_0}\right)^{-5}, \qquad (7.4.28)$$

in which x_0 is defined by (7.4.21).

The above analysis and evaluation of the characteristic times of the stages in the interaction between impurities and vacancies, which are quite abundant in the subsurface layer of any semiconductor sample, have shown that the impurity profile cannot be described by the conventional diffusion equation. Therefore, it cannot be used to find impurity diffusion coefficients. Moreover, expression (7.4.28) is very close to the exponent in the range of not very large *x* [13], which can be easily taken for $\exp(D_A t/l^2)$ used in diffusion theory.

It is necessary to emphasize that the probability of impurity decoration of a pore profile decreases with increasing temperature. This is because the pore coalescence time t_{II} is too short for the impurity profile to be established and vacancies do not come up to the crystal surface. Most amphoteric impurities have high diffusion coefficients, and this circumstance facilitates the fulfillment of inequality (7.4.27) practically at any reasonable temperature.

Normally, shallow hydrogen-like impurities have low diffusion coefficients, so they rarely decorate pores. Impurities with partly filled *d*-shells (Ni, Fe, Cu, and others) and amphoteric impurities containing an interstitial

component and diffusing via the dissociative mechanism possess high diffusion coefficients. This facilitates the fulfillment of (7.4.27) at various temperatures. Vacancy porosity in solids is a well-known fact frequently observed in metals in normal diffusion and radiational swelling [16]. The mechanism of these processes is related to the Kirkendale–Frenkel effect, in which porosity arises at the interface of two solids because of the different diffusion coefficients of their constituent atoms.

Vacancy porosity in semiconductors is less familiar. It was first observed in the study of selenium diffusion in GaAs [17]. Interestingly, the semiconductor crystal contacted, in this experiment, selenium vapor, i.e., the gas-phase, but not a solid. The authors also interpreted their results as being due to the Kirkendale–Frenkel effect.

It follows from the above theory that the necessary condition for the formation of pores is oversaturation of the sample subsurface layer by vacancies. These conditions were created in [14] when a GaAs crystal was presaturated uniformly by copper and subjected to thermal treatment during continuous pumping of arsenic vapor. Evaporation of arsenic atoms provided the initial stepwise vacancy profile corresponding to Figure 7.1. Copper atoms decorated the pore profile, which was registered experimentally. The profile of copper atoms obtained by secondary ion mass-spectrometry is shown in Figure 7.3, together with the initial fairly uniform distribution of copper atoms in the sample prior to thermal treatment. Besides, the copper concentration after diffusion saturation was 2×10^{23} m^{-3}, which is close to copper solubility in GaAs at the saturation temperature of 1073 K.

The samples thus prepared were annealed in a quartz ampoule continuously pumped out at 950 K. The annealing time of 6 hours was longer than the characteristic time of copper migration to the sample surface due to decomposition of the Cu–As solid solution.

Figure 7.3 also shows the calculated profiles. Curve 1 was plotted in accordance with the equation suggested in [10]:

$$N_V = N_{V0} \exp\left(-\tilde{x}/L_V\right), \qquad (7.4.29)$$

where $\tilde{x}$ is the coordinate equal to $\tilde{x} = x - vt$ at the boundary motion rate v; L_V is the characteristic length of the vacancy profile.

Curve 2 in Figure 7.3 was plotted using formula (7.4.28). One can see that only the beginning of the experimental copper profile fits equation (7.4.29). There is an interesting feature looking like a subsurface maximum. To explain this maximum, consider the problem of a stationary pore and va-

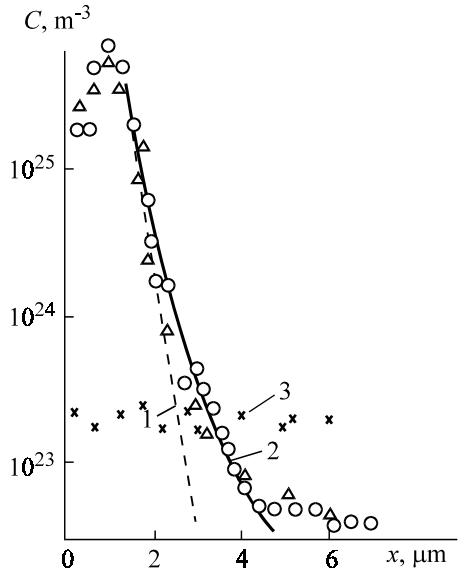

Figure 7.3. Theoretical and experimental Cu profiles in GaAs after vacuum treatment at 950 K [14, 16]: 1 – from formula (6.5.29); 2 – from formula (6.5.28); 3 – initial Cu distribution.

cancy profile, taking into account the interaction between vacancies and impurity atoms (for example, copper) diffusing via the dissociative mechanism. We have the set of equations

$$\left.\begin{aligned}
\frac{\partial N_{\mathrm{V}}}{\partial t} &= D_{\mathrm{V}}\frac{\partial^2 N_{\mathrm{V}}}{\partial x^2} - \frac{N}{\Omega}\frac{\partial R^3}{\partial t} - K_1 N_{\mathrm{i}} N_{\mathrm{V}} + K_2 N_{\mathrm{s}} \\
\frac{\partial N_{\mathrm{i}}}{\partial t} &= D_{\mathrm{i}}\frac{\partial^2 N_{\mathrm{i}}}{\partial x^2} - K_1 N_{\mathrm{i}} N_{\mathrm{V}} + K_2 N_{\mathrm{s}} \\
\frac{\partial N_{\mathrm{s}}}{\partial t} &= D_{\mathrm{s}}\frac{\partial^2 N_{\mathrm{s}}}{\partial x^2} - K_1 N_{\mathrm{i}} N_{\mathrm{V}} + K_2 N_{\mathrm{s}} \\
\frac{\partial R}{\partial t} &= \frac{\Omega}{R}\left\{D_{\mathrm{V}}\Delta N_{\mathrm{V}} + D_{\mathrm{s}}\Delta N_{\mathrm{s}}\right\}
\end{aligned}\right\}, \qquad (7.4.30)$$

where

$$\Delta N_{\mathrm{V}} \equiv N_{\mathrm{V}}\big|_{r=l} - N_{\mathrm{V}}\big|_{r=R},$$

$$\Delta N_s \equiv N_s\big|_{r=l} - N_s\big|_{r=R}\,.$$

Assuming $D_i >> L^2/t$, where L is the sample size and t is diffusion time, we obtain $N_i = N_{i0}$, with N_{i0} as a constant. Going over to the moving coordinate system and assuming a stationary distribution of impurity atoms, vacancies, and pores, i.e., $(\partial N/\partial t)_{\tilde{x}=0} = 0$ and $(\partial R^2/\partial t)_{\tilde{x}=0} = 0$, we obtain the following set of equations

$$\left.\begin{aligned} D_V\frac{\partial^2 N_V}{\partial \tilde{x}^2} + \frac{\partial N_V}{\partial \tilde{x}} + \frac{N}{\Omega}V\frac{\partial R^3}{\partial \tilde{x}} - K_1 N_{i0} N_V + K_2 N_s = 0 \\ D_s\frac{\partial^2 N_s}{\partial \tilde{x}^2} + V\frac{\partial N_s}{\partial \tilde{x}} + K_1 N_{i0} N_V - K_2 N_s = 0 \\ \frac{D_V \Delta N_V + D_s \Delta N_s}{R} + V\frac{1}{\Omega}\frac{\partial R}{\partial \tilde{x}} = 0 \end{aligned}\right\}. \qquad (7.4.31)$$

The boundary conditions are

$$\begin{aligned} N_V\big|_{\tilde{x}=0} &= N_{V0}\,, \qquad N_V\big|_{\tilde{x}\to\infty} = N_{V\infty} << N_{V0} \\ N_s\big|_{\tilde{x}=0} &= N_{s0}\,, \qquad N_s\big|_{\tilde{x}\to\infty} = \frac{K_1}{K_2} N_{i0} N_{V\infty}\,. \end{aligned} \qquad (7.4.32)$$

The distribution of components, described by the set of equations (7.4.31), possesses three characteristic lengths D/v, $(D\tau)^{1/2}$, and L_{pore}, where D is any of the diffusion coefficients D_i, D_V, or D_s. All characteristic lengths are related to crystal surface motion, the interactions between vacancies and impurities and those between diffusing components and pores, respectively.

Of special interest is the case of low sublimation rates of arsenic atoms from the sample, or $D/v >> (D\tau)^{1/2}$, L_{pore}. From here, we have the restrictions $v << (D\tau)^{1/2}$, D/L_{pore}. Under the conditions

$$\sqrt{D\tau} << L_{pore} \qquad \text{or} \qquad \sqrt{D\tau} >> L_{pore} \qquad (7.4.33)$$

the set of equations is simplified for small and large distances from the crystal surface.

Consider the case with $(D\tau)^{1/2} << L_{pore}$. At small $x << L_{pore}$, equations (7.4.31) transform to

$$\left.\begin{aligned} D_{\mathrm{V}}\frac{\partial^2 N_{\mathrm{V}}}{\partial \tilde{x}^2} - K_1 N_{\mathrm{i}0} N_{\mathrm{V}} + K_2 N_{\mathrm{s}} \\ D_{\mathrm{s}}\frac{\partial^2 N_{\mathrm{s}}}{\partial \tilde{x}^2} + K_1 N_{\mathrm{i}0} N_{\mathrm{V}} - K_2 N_{\mathrm{s}} \end{aligned}\right\}. \tag{7.4.34}$$

The solution of this set of equations yields

$$\left.\begin{aligned} N_{\mathrm{s}} &= \left(N_{\mathrm{s}0} - N_{\mathrm{s}}^{\infty}\right)\exp\left(-\tilde{x}/L_1\right) + N_{\mathrm{s}}^{\infty} \\ N_{\mathrm{V}} &= \left(N_{\mathrm{V}0} - N_{\mathrm{V}}^{\infty}\right)\exp\left(-\tilde{x}/L_1\right) + N_{\mathrm{V}}^{\infty} \end{aligned}\right\}, \tag{7.4.35}$$

where

$$\frac{1}{L_1} = \sqrt{\frac{K_1 N_{\mathrm{i}0}}{D_{\mathrm{V}}} + \frac{K_2}{D_{\mathrm{s}}}},$$

$$N_{\mathrm{V}}^{\infty} = \frac{K_2\left(D_{\mathrm{V}} N_{\mathrm{V}0} + D_{\mathrm{s}} N_{\mathrm{s}0}\right)}{D_{\mathrm{V}} K_2 + D_{\mathrm{s}} K_1 N_{\mathrm{i}0}},$$

$$N_{\mathrm{s}}^{\infty} = \frac{K_1 N_{\mathrm{i}0}\left(D_{\mathrm{V}} N_{\mathrm{V}0} + D_{\mathrm{s}} N_{\mathrm{s}0}\right)}{D_{\mathrm{V}} K_2 + D_{\mathrm{s}} K_1 N_{\mathrm{i}0}}.$$

It is easy to get the solution to the set of equations (7.4.31) for large $x >> (D\tau)^{1/2}$ as well. It coincides with (7.4.27):

$$\left.\begin{aligned} N_{\mathrm{V}} &= \frac{N_{\mathrm{V}0} - N_{\mathrm{V}}^{\infty}}{\left(1 + \tilde{x}/x_0\right)^5} + N_{\mathrm{V}}^{0} \\ N_{\mathrm{s}} &= \frac{K_1}{K_2} N_{\mathrm{i}0} N_{\mathrm{V}} \end{aligned}\right\}, \tag{7.4.36}$$

where

$$x_0 = \left(\frac{900\nu}{8\pi^2 N^2 N_{\mathrm{V}}^{\infty} \Omega \tilde{D}}\right)^{1/5}. \tag{7.4.37}$$

Thus, formulas (7.4.35) and (7.4.36) are the basis for the analysis of experimental pore profiles produced by the interaction of impurity atoms with vacancies and pores.

Let us return to the analysis of the profile in Figure 7.3 to find L_1 from (7.4.35). For this, we take for copper atoms $D_s = 10^{-14}$ m^2/s, for vacancies $D_V = 10^{-12}$–10^{-13} m^2/s [10], and the characteristic time $\tau \leq 1$ necessary for equilibrium to be established in the reaction $Cu_i + V_{As} \rightleftarrows C_{As}$. Then, $L_1 = 1\mu m$, in agreement with the experimental distance from the sample surface to the copper distribution maximum. Therefore, this maximum is due to the copper surface concentration being lower than its quasi-equilibrium concentration in the initial region of the pore profile.

7.5 DIFFUSION PROFILES OF INTERACTING IMPURITIES

7.5.1 General principles

In most situations of practical importance, diffusion occurs in multicomponent systems. However, the research into the behavior of impurity profiles in multicomponent diffusion has been quite limited [19–28]. Our consideration of diffusion profiles will mostly follow M. I. Sinder's model [18].

The interpretation of diffusion in a multicomponent system is difficult for two reasons. One is that the experiment providing reliable information is very complicated technically. The other reason why diffusion is difficult to

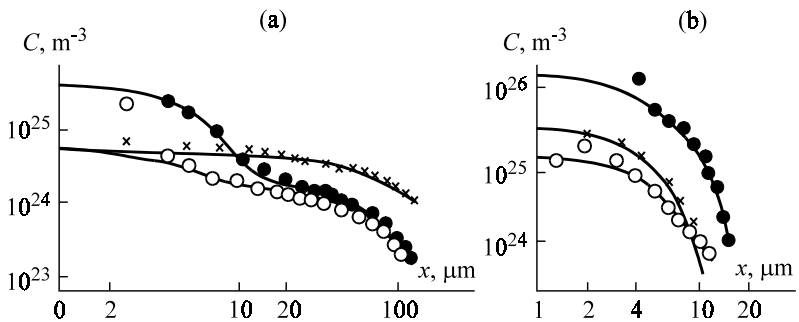

Figure 7.4. The distributions of Sb (a) and In (b) in different Ge samples [23].

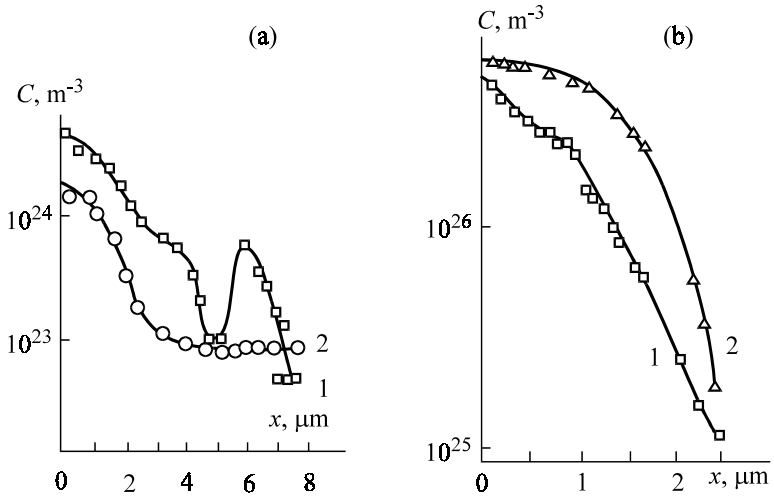

Figure 7.5. The distributions in silicon: (a) – of iron in the presence of P (1) and in the absence of P (2); (b) – of phosphorus in the presence of Fe (1) and in the absence of Fe (2) [15].

interpret is that the interaction of impurity atoms must be taken into account, and this circumstance makes theoretical models quite sophisticated. But we can derive simple analytical expressions permitting a comparison with experiments only in certain approximations [19]. We will resort to numerical methods of equation solution with a preliminary qualitative analysis of experimental data, because it will provide the basis for accepting or discarding certain factors from the mathematical consideration.

A theoretical interpretation of diffusion profiles is also difficult because of the necessity to draw a distinction between the effects of interacting and non-interacting impurities. Moreover, even if only one impurity diffuses, the diffusion process itself may involve many components. We demonstrated this above when discussing dissociative diffusion (Section 7.3) and the strong effect of interaction with vacancies at the sample surface (Section 7.4) and in its bulk (Section 6.5). Figures 7.4–7.7 illustrate impurity profiles for simultaneous and stepwise impurity diffusion in semiconductors.

A simple mathematical model of sequential diffusion, widely used in practice, is possible only if the diffusion coefficient of the pre-doping impurity is much higher than that of the subsequent impurity. This condition permits the consideration of this problem as that of impurity diffusion into a uniformly doped sample [23, 25, 26]. Otherwise, it would be necessary to

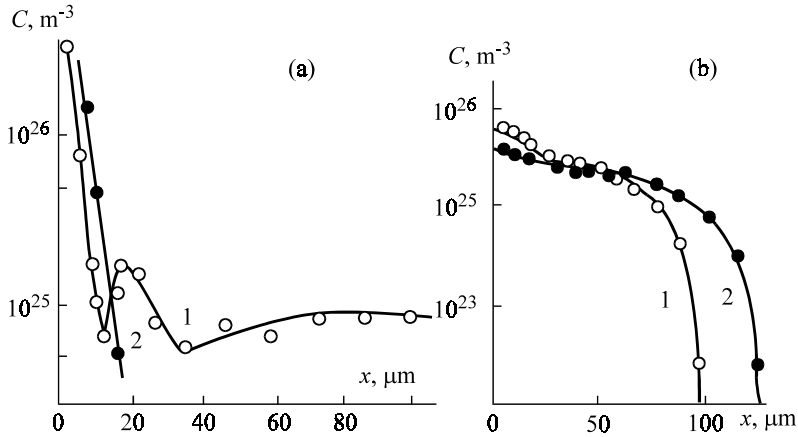

Figure 7.6. Concentration profiles for Cu and Zn in GaAs in sequential diffusion: (a) – of Cu into a Zn-doped sample (1); (b) – of Zn into a Cu-doped sample (1); curves 2 – control distributions of Cu (a) and Zn (b) [3].

consider diffusion into a nonuniformly doped sample, which would involve sophisticated computations.

Figures 7.4–7.7 show complicated profiles which cannot be described by a simple diffusion equation

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2}, \tag{7.5.1}$$

whose solution is represented by the well-known expressions

$$C = C_0 \operatorname{erfc}\frac{x}{2\sqrt{Dt}} \quad \text{at} \quad \sqrt{Dt} \ll L, \tag{7.5.2}$$

$$C = C_0\left[1 - \frac{4}{\pi}\exp\left(-\frac{\pi^2 Dt}{4L^2}\right)\operatorname{Sin}\frac{\pi x}{2L}\right] \quad \text{at} \quad \sqrt{Dt} \sim L. \tag{7.5.3}$$

Many experimental profiles have extrema under certain conditions.

Complicated profiles can be analyzed within two approaches. One is based on the thermodynamics of irreversible processes. It employs the interaction models of diffusing components—the complexation model and

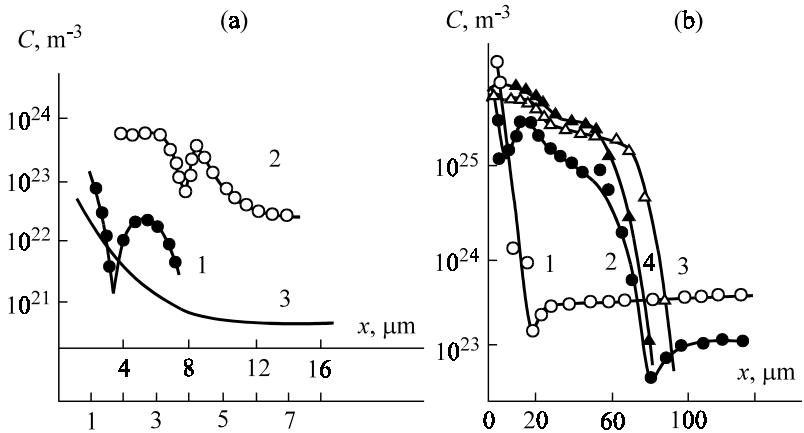

Figure 7.7. Concentration profiles (a) of Fe in silicon after B diffusion at T = 1273 K: 1 – $t = 3\tau$; 2 – $t = 5\tau$; 3 – control, no B diffusion in sample annealing at $t = 5\tau$; (b) of Cu (2) and Zn (4) in simultaneous diffusion into GaAs; 1, 3 – control distributions of Cu and Zn in their separate diffusion.

the model of interaction via internal fields: an electric field [19] and an elastic field [27, 28].

Note that the impurity profiles for multicomponent diffusion in semiconductors have a common pattern for various solids. For example, the profiles of simultaneous diffusion of silicon, chromium, and carbon into iron have the same pattern. We should remember, however, that the study of impurity profiles in metals and alloys is complicated by a high density of dislocations, second-phase products, small-angle boundaries, and structural defects. Of course, their interactions in semiconductors are easier to avoid.

7.5.2 Impurity interactions in terms of thermodynamics of irreversible processes

Let us analyze diffusion kinetics in terms of the linear thermodynamics of irreversible processes [29, 30]. The equation for component flows j_i expressed through chemical potential gradients is

$$j_i = -\sum_{k=1}^{n} L_{ik} \frac{\partial \mu_{ik}}{\partial x}, \qquad (7.5.4)$$

where L_{ik} are Onzager's coefficients and μ_{ik} is the chemical potential of the k-th component.

Going from the chemical potentials to the concentrations of components, C_k, we can write

$$j_i = \sum_{k=1}^{n} D_{ik} \frac{\partial C_k}{\partial x}, \tag{7.5.5}$$

where

$$D_{ik} = \sum_{k=1}^{n} L_{im} \frac{\partial \mu_{ik}}{\partial C_k}$$

are partial diffusion coefficients.

From the continuity theorem for a one-dimensional case, we have

$$\frac{\partial C_i}{\partial t} = \sum_{k=1}^{n} \frac{\partial}{\partial x}\left(D_{ik} \frac{\partial C_k}{\partial x} \right). \tag{7.5.6}$$

Therefore, diffusion in an n-component system is described by n^2 diffusion coefficients, each of which is generally a variable depending on the concentration of components and other state parameters. Solutions to the set of equations (7.5.6) were derived in [30, 36] for various boundary conditions at constant diffusion coefficients D_{ik} in simultaneous and sequential diffusion of two components into a semi-infinite body.

Before discussing the details of these solutions, we should specify the concepts of sequential and simultaneous diffusion. Usually, these concepts are associated with various initial conditions and with the presence or absence of components in the sample bulk. To make the classification of solutions to diffusion equations convenient, we will define them in terms of various boundary conditions. If *constant concentrations of components* are assigned *at the boundary*, such a process will be termed *simultaneous diffusion*. If the boundary condition assigns *the absence of one component* in the sample bulk at the initial moment of time, such a process will be referred to as *sequential diffusion*.

Equation (7.5.6) for two components is written as

$$\left.\begin{aligned} \frac{\partial C_1}{\partial t} &= D_{11} \frac{\partial^2 C_1}{\partial x^2} + D_{12} \frac{\partial^2 C_2}{\partial x^2} \\ \frac{\partial C_2}{\partial t} &= D_{21} \frac{\partial^2 C_1}{\partial x^2} + D_{22} \frac{\partial^2 C_2}{\partial x^2} \end{aligned}\right\}. \tag{7.5.7}$$

The initial and boundary conditions for sequential diffusion are

$$C_1(x, 0) = C_1^{\,0}, \qquad C(\infty, t) = C_1^{\,0}, \qquad C_2(0, t) = C_2^{\,0},$$

$$C_2(\infty, t) = 0, \qquad C_2(x, 0) = 0,$$

$$D_{11}\frac{\partial C_1}{\partial x} + D_{12}\frac{\partial C_2}{\partial x}\bigg|_{x=0} = 0\,.$$

For simultaneous diffusion, they are

$$C_1(x, 0) = C_2(0, t) = 0 \quad \text{at} \quad x > 0,$$

$$C_1(\infty, t) = C_2(\infty, \mathrm{t}) = 0, \quad C_1(0, t) = C_1^{\,1}, \quad C_2(0, t) = C_2^{\,1},$$

where $C_1^{\,0}$, $C_2^{\,0}$, $C_1^{\,1}$, and $C_2^{\,1}$ are constants.

The solutions to the set of equations (7.5.7) for sequential diffusion (Figure 7.8) are [28]:

$$\left.\begin{aligned} C_1 &= C_1^0 + C_2^0 D_{12}\frac{\operatorname{verfc}\frac{x}{2u\sqrt{t}} - \operatorname{uerfc}\frac{x}{2v\sqrt{t}}}{v\left(u^2 - D_{11}\right) - u\left(v^2 - D_{11}\right)} \\ C_2 &= \frac{C_2^0}{1 - \frac{u}{v}\frac{v^2 - D_{11}}{u^2 - D_{11}}}\operatorname{erfc}\frac{x}{2u\sqrt{t}} + \frac{C_2^0}{1 - \frac{v}{u}\frac{u^2 - D_{11}}{v^2 - D_{11}}}\operatorname{erfc}\frac{x}{2v\sqrt{t}} \end{aligned}\right\} \tag{7.5.8}$$

and for simultaneous diffusion (Figure 7.9), they are [28]:

$$\left.\begin{aligned} C_1 &= -\frac{1}{D}\left[C_1^1\left(u^2 - D_{22}\right) + C_2^1 D_{12}\right]\operatorname{erfc}\frac{x}{2u\sqrt{t}} + \\ &\quad + \frac{1}{D}\left[C_1^1\left(v^2 - D_{22}\right) + C_2^1\right]\operatorname{erfc}\frac{x}{2u\sqrt{t}} \\ C_2 &= -\frac{1}{D}\left[C_1^1 D_2^1 - C_2^1\left(v^2 - D_{22}\right)\right]\operatorname{erfc}\frac{x}{2u\sqrt{t}} + \\ &\quad + \frac{1}{D}\left[C_1^1 D_{21} - C_2^1\left(u^2 - D_{22}\right)\right]\operatorname{erfc}\frac{x}{2u\sqrt{t}} \end{aligned}\right\}, \tag{7.5.9}$$

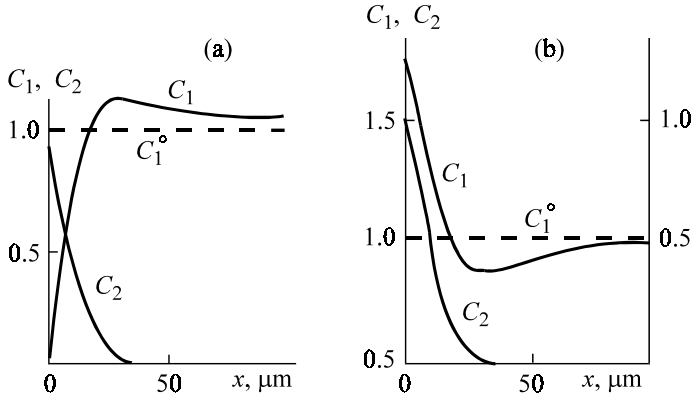

Figure 7.8. Distributions of concentrations C_1 and C_2 in sequential diffusion: (a) – at $D_{12} > 0$; (b) – at $D_{12} < 0$ [3].

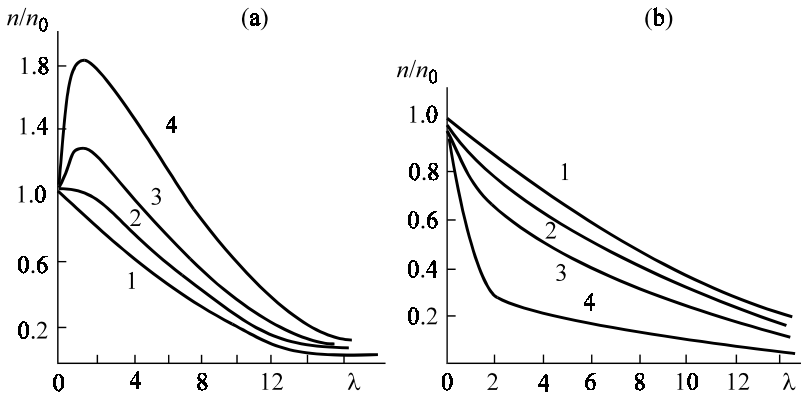

Figure 7.9. Distributions of the active component in simultaneous diffusion [3] of (a) impurities of different signs and (b) impurities of the same sign ($\lambda \equiv x/2t^{1/2}$ is the Boltzmann variable): (a) – $n_0 = 1$; $p_0 = 0.1$ (1), 0.5 (2), 2 (3), 5 (4); (b) – $n_0 = 10$; $p_0 = 0.2$ (1), 3 (2), 5 (3), 10 (4).

where

$$u^2 = \frac{1}{2}\left(D_{11} + D_{22} - D\right), \qquad v^2 = \frac{1}{2}\left(D_{11} + D_{22} + D\right),$$

$$D = \sqrt{\left(D_{11} - D_{22}\right)^2 + 4D_{12}D_{21}}\ .$$

It is seen from Figure 7.8 that in sequential diffusion, the distribution of component 1 pre-doped into the sample contains an extremum: a maximum at $D_{11} > 0$ and a minimum at $D_{12} < 0$. In simultaneous diffusion of two components, their concentrations decrease monotonically into the sample bulk due to the initial and boundary conditions assigned at a fairly large distance from the surface. At the surface, however, the distribution of the components may show a maximum (Figure 7.9). The condition for the appearance of the maximum can be easily derived from (7.5.9).

Therefore, the distribution of components can be described in terms of phenomenological diffusion coefficients. An advantage of this approach is its universal character permitting the description of experimental impurity profiles without going into the interaction details. It follows from this that a non-monotonic character of impurity distributions is one of the characteristics of the interaction between the components. On the other hand, the assumption of coefficients D_{ik} being independent of concentration is a fairly rough approximation.

The phenomenological character of coefficients D_{ik} does not permit their numerical evaluation or the understanding of interimpurity interaction. This is a common property of all phenomenological models. For this reason, model approaches are more often used to analyze impurity profiles.

7.5.3 Impurity interactions in terms of a model approach

The electrostatic interaction model for impurity atoms analyzes internal field effects on impurity diffusion. It is based on a combined solution of diffusion equations and Poisson's solution for electrostatic potential [18, 23, 32]. We will discuss briefly some qualitative features arising in electrical interactions [19].

In sequential diffusion, the distribution of a pre-diffused impurity shows an extremum: a maximum in donor–donor and acceptor–acceptor interactions and a minimum in donor–acceptor interactions. In simultaneous diffusion of two impurities (Figure 7.9), the impurity concentrations in

donor–acceptor interactions decrease monotonically with distance from the sample surface.

When the impurities have the same sign, the distribution pattern of the rapidly diffusing impurity may have a maximum if the slowly diffusing impurity has a high surface concentration (Figure 7.9*a*). The distribution of the latter remains monotonic.

Of interest is another feature. If the diffusion coefficients of impurities differ considerably, the rapidly diffusing impurity has little effect on the concentration distribution of the other impurity, irrespective of the initial and boundary conditions. The distribution pattern of the former impurity has two regions: a region of an abrupt change of concentration with depth, associated with the electric field effect of the slowly diffusing impurity, and a region of free diffusion.

The comparison of solutions derived from the set of equations (7.5.7) and from the field-affected interaction model shows their agreement. An advantage of the "field" model is the relation between the characteristic features of impurity profiles and physically clear reasons, as well as diffusion coefficients.

It might seem from the analogy with an electric field that impurity elastic field should have a strong effect on diffusion. This, however, is not the case. Let us discuss this problem in some detail, as was done in [33–35].

The study of elastic interactions of point defect pairs has a rather long history [35]. It has recently been established that the solution to the problem of strong interaction greatly depends on the boundary conditions. For example, the authors of [36] consider two elastically interacting point defects in the form of two spheres made from foreign material, built into a cavity of a somewhat different radius. The interaction energy E_{int} of the spheres is

$$E_{\text{int}} \cong AR^{-6}, \tag{7.5.10}$$

where R is the interdefect distance, $A = (G/G_1 - 1)$, G and G_1 are the modules of the matrix shift and inclusion.

Hence, the energy of elastic interaction for dilatation centers ($G = G_1$) is zero. This seems to be the reason why most authors take into account only the stress field arising from a nonuniform distribution of defects, although elastic interaction of defect pairs also exists in an unstressed crystal, i.e., at $\sigma = 0$ [35, 37, 38]. Any real solid is finite. When a dilatation center is introduced into it, the body surface is compressed or dilated, unless it is fixed rigidly. The induced surface forces deform the lattice, creating what is known as an imaginary deformation field. The relative change in the body volume is described as

$$\varepsilon = \operatorname{div}\vec{u} + \operatorname{div}\sum_{\alpha}\left\langle U_{\alpha}^{\infty}\right\rangle = \sum_{\alpha}\frac{\Omega_{\alpha}C_{\alpha}}{\nu}, \qquad (7.5.11)$$

where U is intrinsic energy of a defect; $\Omega = \vartheta_1 - \vartheta$ is the volume difference between the sphere (ϑ_1) and the cavity (ϑ), into which the sphere was inserted; C is the defect concentration; ν is Poisson's coefficient; summation is made over all defect kinds α.

The first term in (7.5.11) is the body lattice deformation and the second term represents deformation concentrated on defects (dilatation centers). The total energy of pair interaction of centers α and β, or the elastic contribution to the free energy of a homogeneous defect solid solution for bodies with unfixed boundaries was found to be [35]

$$E_{\text{int}} = -\frac{1}{2}N_s\sum_{\alpha}\sum_{\beta}K\frac{\Omega_{\alpha}\Omega_{\beta}}{\vartheta}C_{\alpha}C_{\beta}\frac{2}{3}\frac{1-2\nu}{1-\nu}, \qquad (7.5.12)$$

where K is the hydrostatic compression module and N_s is the number of sites in a defect-free crystal.

Normal diffusion conditions are close to this case of free crystal boundary. The respective expression for a diffusional defect flow was found in [35]:

$$\vec{j}_{\alpha} = -D_{\alpha}\vec{\nabla}C_{\alpha} + \frac{2}{3}\frac{D_{\alpha}}{kT}K\frac{1-2\nu}{1-\nu}\Omega_{\alpha}C_{\alpha}\sum_{\beta}\vec{\nabla}\frac{\Omega_{\beta}C_{\beta}}{\vartheta} + \frac{1}{3}\frac{D_{\alpha}}{kT}C_{\alpha}\Omega_{\alpha}\vec{\nabla}\sigma_{kk}, \qquad (7.5.13)$$

where σ_{kk} is a stress tensor.

If diffusion into a thick plate is uniform, we have

$$\sigma_{xx} = \sigma_{yy} = -\frac{\Omega}{3}\frac{E}{1-\nu}C,$$
$$\sigma_{xy} = \sigma_{yz} = \sigma_{xz} = \sigma_{zz} = 0, \qquad (7.5.14)$$

where E is Young's module.

The lattice is compressed ($\sigma_{kk} < 0$) if the covalent radius of an impurity atom is larger than that of the host atom ($\Omega > 0$). By substituting (7.5.14) into (7.5.13) with $K = E/3(1 - 2\nu)$, it is easy to obtain a zero effect of elastic

fields on one-dimensional diffusion of dilatation centers in an isotropic medium, which is consistent with the conclusion of [38]. It is interesting that the authors of [39, 40] thought it necessary to add a drift term to the expression for impurity atom flow to describe the diffusion.

Finally, most semiconductors are known to be anisotropic. This means that the elastic interaction value and the stress concentration vary with the direction $\vec{n}$ of the sample cleavage. According to [40], we have

$$\vec{j} = -D_{\text{el}}\vec{\nabla}C, \tag{7.5.15}$$

$$D_{\text{el}}(\vec{n}) = D\left\{1 - \frac{2}{9}\frac{\Omega^2 C}{\nu kT}\left[\frac{E(\vec{n})}{1-\nu(\vec{n})} - Y(\vec{n})\right]\right\}, \tag{7.5.16}$$

where D is a diffusion coefficient of atoms uninvolved in the lattice deformation, $Y(\vec{n})$ is a complex function of elastic constants C_{11} and C_{12}, of components n_x, n_y, n_z, and an anisotropic factor

$$\xi = \frac{C_{12} + 2C_{44} + C_{11}}{C_{44}} \neq 0. \tag{7.5.17}$$

The pair interaction of dilatation centers does not completely compensate the stress field, and the diffusion coefficients must differ in different crystallographic directions, namely, [35]:

$$D_{111} > D_{110} > D_{100}. \tag{7.5.18}$$

These relations were observed experimentally in silicon in diffusion of phosphorus with a high surface concentration [41].

Diffusion conditions, in which the boundary of a solid can be considered to be fixed rigidly, are very rare in semiconductor practice. But if this situation does take place, one should use, instead of (7.5.16), the solution derived by Krivoglazov [42], showing that concentration elastic stresses affect diffusion even at $\zeta = 0$ and the relations of (7.5.18) are also fulfilled. Those interested in the mathematical results concerning this situation can turn to the original work [42].

The next interaction model which describes impurity diffusion in semiconductors is the complexation model. Complexation mechanisms are diverse [43] but any model implies that the range of forces producing complexes must be small. It is only under this condition that a complex can be

regarded as a point defect. Otherwise, say, in electrostatic interaction, the complexation model will allow only for short-range forces, with the long-range component replaced by an average field induced by the nearest neighbors. This is characteristic of the internal electric field model. Therefore, the two models—the complexation model and the electric field model—supplement each other and describe, respectively, the short- and long-range interactions between diffusing components. It is easy to notice, however, that the interaction in the field model may be both attraction and repulsion, whereas the complexation model describes only attraction. Short-range repulsion is usually taken into account by a factor in the diffusion coefficient.

7.5.4 Diffusion theory for immobile complexes

The equations for the diffusion of two components A and B with the instantaneous formation of the complex $Q \rightleftarrows A + B$ are

$$\left.\begin{aligned} \frac{\partial N_A}{\partial t} &= D_A \frac{\partial^2 (N_A - Q)}{\partial x^2} + D_Q \frac{\partial^2 Q}{\partial x^2} \\ \frac{\partial N_B}{\partial t} &= D_B \frac{\partial^2 (N_B - Q)}{\partial x^2} + D_Q \frac{\partial^2 Q}{\partial x^2} \\ \frac{(N_A - Q)(N_B - Q)}{Q} &= K \end{aligned}\right\}, \qquad (7.5.19)$$

where N_A, N_B, and Q are the total concentrations of the components; D_A, D_B, and D_Q are their diffusion coefficients; K is a complexation constant.

Here, we will discuss the solution to the set of equations (7.5.19) for small $D_{Q,}$ as compared to D_A and D_B, or for practically immobile complexes. The authors of [44, 45] considered the problem of sequential diffusion of component A into a sample uniformly doped with component B with the initial boundary conditions:

$$\left.\begin{aligned} N_A(x,0) &= 0 & N_B(x,0) &= N_B(\infty) \\ N_A(0,t) - Q(0,t) &= H_A(0) & \frac{\partial}{\partial x}\left[N_B(x,t) - Q(x,t)\right]\Big|_{x=0} &= 0 \end{aligned}\right\}. \qquad (7.5.20)$$

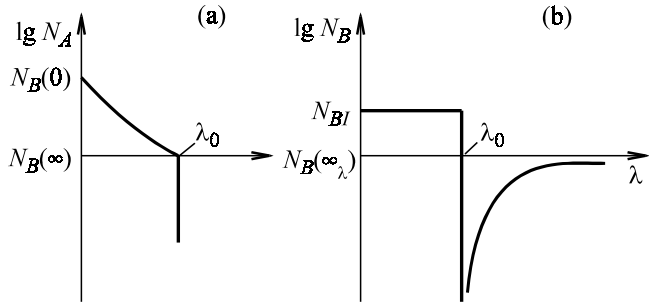

Figure 7.10. Distributions of (a) *A* atoms and (b) *B* atoms in a strong complexation after sequential diffusion [45].

The solution to equations (7.5.19) with the boundary conditions of (7.5.20) was obtained in the extreme cases of strong complexation ($K = 0$) and weak complexation ($K >> N_A, N_B$).

(1) A strong complexation is shown in Figure 7.10 at K = 0:

$$N_A = \begin{cases} N_{B1} + H_A(0)\left(1 - \dfrac{\operatorname{erf}(\lambda/d_A)}{\operatorname{erf}(\lambda_0/d_A)}\right) & \lambda < \lambda_0 \\ 0 & \lambda > \lambda_0 \end{cases} \tag{7.5.21}$$

$$N_B = \begin{cases} H_B & \lambda < \lambda_0 \\ N_B(\infty)\left(1 - \dfrac{\operatorname{erfc}(\lambda/d_B)}{\operatorname{erfc}(\lambda_0/d_B)}\right) & \lambda > \lambda_0 \end{cases}, \tag{7.5.22}$$

where $\lambda = x/2t^{1/2}$ is Boltzmann's variable, $d_A^2 = D_A$, $d_B^2 = D_B$, and

$$N_{B1} = \frac{N_B(\infty) d_B \exp\left(-\lambda_0^2/d_B^2\right)}{\sqrt{\pi}\lambda_0 \operatorname{erfc}(\lambda_0/d_B)}.$$

The characteristic value of λ is defined by the equation

$$\frac{N_B(\infty) d_B \exp\left(-\lambda_0^2/d_B^2\right)}{\lambda_0 \operatorname{erfc}(\lambda_0/d_B)} = \frac{H_A(0) d_A \exp\left(-\lambda_0^2/d_A^2\right)}{\lambda_0 \operatorname{erfc}(\lambda_0/d_A)}. \tag{7.5.23}$$

(2) A weak complexation is shown in Figure 7.11 at $K >> N_A, N_B$:

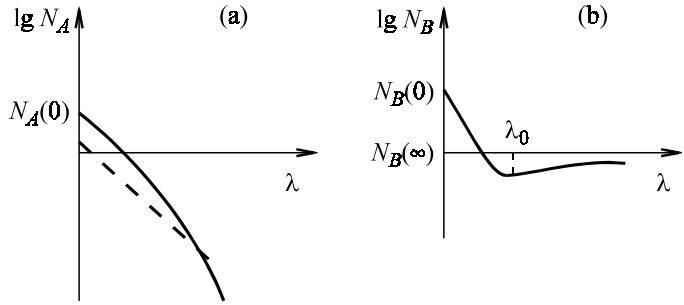

Figure 7.11. Distributions of (a) *A* atoms and (b) *B* atoms in a weak complexation after sequential diffusion [45].

$$N_{\mathrm{A}} = H(0)\left\{1+\frac{N_{\mathrm{B}}(\infty)}{K}\operatorname{erfc}\left(\lambda\sqrt{1+\frac{N_{\mathrm{B}}(\infty)}{K}}\Big/ d_{\mathrm{A}}\right)\right\}, \qquad (7.5.24)$$

$$N_{\mathrm{B}} = N_{\mathrm{B}}(\infty)\left[1+\frac{H_{\mathrm{A}}(0)}{K}\frac{d_{\mathrm{B}}^2}{d_{\mathrm{B}}^2-d_{\mathrm{A}}^2}\left(\operatorname{erfc}(\lambda/d_{\mathrm{A}})-\frac{d_{\mathrm{A}}}{d_{\mathrm{B}}}\operatorname{erfc}(\lambda/d_{\mathrm{B}})\right)\right]. \qquad (7.5.25)$$

It is seen from Figures 7.10 and 7.11 that the surface concentration of component *A* in both complexation types increases by the concentration value of the complexes. Simultaneously, there is an increase in the subsurface concentration of component *B* due to its inflow into the complexation region. The distribution of component B has a minimum, which is very distinct in a strong complexation but is hardly visible in a weak complexation.

The set of equations (7.5.19) was solved in [46] for simultaneous diffusion without restrictions on the complexation constant, i.e., without drawing a distinction between the types of complexation. It was suggested only that the diffusion coefficient of the component uniformly distributed in the sample was much larger than that with zero concentration.

The boundary and initial conditions for this problem are

$$\left.\begin{aligned} \left|N_A - Q\right|_{x=0} &= H_A(0), \quad \left|N_A\right|_{t=0} = 0 \quad x>0 \\ \left|N_B - Q\right|_{x=0} &= H_B(0), \quad \left|N_B\right|_{t=0} = H(\infty) \quad x>0 \end{aligned}\right\}. \qquad (7.5.26)$$

The solution to this problem can be facilitated by using Boltzmann's variable which transforms the set of equations (7.5.19) to

$$\left.\begin{aligned} d_{\mathrm{A}}^2 H_{\mathrm{A}}'' + 2\lambda\left(H_{\mathrm{A}} + Q\right)' = 0 \\ d_{\mathrm{B}}^2 H_{\mathrm{B}}'' + 2\lambda\left(H_{\mathrm{B}} + Q\right)' = 0 \end{aligned}\right\}, \tag{7.5.27}$$

where

$$H_{\mathrm{A}} \equiv N_{\mathrm{A}} - Q, \qquad H_{\mathrm{B}} \equiv N_{\mathrm{B}} - Q.$$

By putting into (7.5.27), alternately, $d_{\mathrm{A}} = 0$ at $\lambda \to \infty$ and $d_{\mathrm{B}} \to \infty$ at $x = 0$, we find the external and internal expansions for H_{A} and H_{B}:

$$H_{\mathrm{A}}^{\mathrm{ex}} = 0\,, \qquad H_{\mathrm{B}}^{\mathrm{ex}} = C_\lambda \operatorname{erfc}\lambda_{\mathrm{B}} + H_{\mathrm{B}}(\infty) \tag{7.5.28}$$

$$H_{\mathrm{B}}^{\mathrm{in}} = 0\,, \qquad H_{\mathrm{A}}^{\mathrm{in}} = \left[H_{\mathrm{A}}(0) - C_2\right]\operatorname{erfc}\lambda_{\mathrm{A}}\sqrt{1 + \frac{H_{\mathrm{B}}(0)}{K}} + C_2\,, \tag{7.5.29}$$

where C_1 and C_2 are constants derived from the "sowing" conditions of the internal and external expansions:

$$N_{\mathrm{A}} = \left[H_{\mathrm{A}}(0) + \frac{H_{\mathrm{A}}(0)H_{\mathrm{B}}(0)}{K}\right]\operatorname{erfc}\lambda\sqrt{1 + \frac{H_{\mathrm{B}}(0)}{K}}\Bigg/ d_{\mathrm{A}} \tag{7.5.30}$$

$$N_{\mathrm{B}} = H_{\mathrm{B}}(\infty) + \left[H_{\mathrm{B}}(0) - H_{\mathrm{B}}(\infty)\right]\operatorname{erfc}\frac{\lambda}{d_{\mathrm{B}}} + \frac{H_{\mathrm{A}}(0)H_{\mathrm{B}}(0)}{K}\operatorname{erfc}\lambda\sqrt{1 + \frac{H_{\mathrm{B}}(0)}{K}}\Bigg/ d_{\mathrm{A}}\,.$$

Expressions (7.5.30) somewhat differ from those derived in the original work [46]. They are more simple and contain no constants with an ambiguous physical sense. The N_{B} distribution minimum appears under the condition $H_{\mathrm{B}}(0) < H_{\mathrm{B}}(\infty)$, which means that the boundary concentration of the free component *B* must be smaller than in the sample bulk. The minimum arises from two opposite *B* flows—one consisting of *B* atoms bound in complexes and migrating into the sample bulk; the other consisting of free *B* atoms migrating outward.

The solutions to the set of equations (7.5.9) under the condition $D_{\mathrm{Q}} << D_{\mathrm{A}}$, D_{B} show that complexation has a greater effect on the profile of the

higher mobility component. As in the field model, the distribution of the rapidly diffusing component has two regions. In one region, the concentration distribution is determined by the bound component and in the other by the free-diffusing component.

In sequential diffusion, the distribution pattern of the initially uniform component contains only a minimum, but if the field mechanism is involved, it contains both a minimum and a maximum. In simultaneous "field" diffusion into an undoped sample, the distribution pattern of the higher mobility component may have a maximum, which is absent if the process involves low mobility complexes ($D_Q << D_A, D_B$) or nearly immobile complexes.

The kinetic models of impurity interaction discussed in this chapter have a strong restriction on the mobility of complexes. These models have another limitation—the neglect of complexation kinetics, which means that "the time for equilibrium to be established between complexes and free impurities is much shorter than the characteristic times of diffusion" [45]. In other words, the reaction producing complexes Q is instantaneous, as was pointed out at the beginning of this section.

REFERENCES

7.1. V.I. Fistul, *Amfoternye primesi v poluprovodnikakh* (Amphoteric Impurities in Semiconductors). Moscow: Metallurgia, 240 p. (1995) (in Russian).
7.2. V.I. Fistul, *Raspad peresyshchennykh poluprovodnikovykh tverdykh rastvorov* (Decomposition of Oversaturated Semiconductor Solid Solutions). Moscow: Metallurgia, 240 p. (1977) (in Russian).
7.3. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushchikh defektov v poluprovodnikakh* (Thermodynamics and Kinetics of Interacting Defects in Semiconductors). Moscow: Nauka, 352 p. (1977) (in Russian).
7.4. A.S. Damask, G.D. Dines, *Point Defects in Metals*. New York–London: Gordon & Breach Publ., 291 p. (1963).
7.5. A. Seeger, K.P. Chik, *Phys. St. Sol.* **29**: 455–542 (1968).
7.6. M. Yoshida, *J. Appl. Phys. Japan* **8**: 1211–1213 (1969).
7.7. N.S. Rytova, V.I. Fistul, *FTP* **5**: 1961–1968 (1971) (in Russian).
7.8. M. Yoshida, *J. Appl. Phys. Japan* **8**: 1211–1213 (1969).
7.9. M. Yoshida, K. Saito, *Jap. Appl. Phys.* **9**: 1217–1228 (1970).
7.10. V.I. Fistul, V.I. Pokrovsky, *FTP* **13**: 1402–1406 (1979) (in Russian).
7.11. V.I. Fistul, A.I. Chernova, *FTP* **14**: 990–994 (1980) (in Russian).
7.12. V.I. Fistul, M.I. Sinder, *FTP* **15**: 1182–1186 (1981) (in Russian).
7.13. M.I. Sinder, V.I. Fistul, *FTP* **14**: 1953–1957 (1980) (in Russian).

7.14. Ya.E. Geguzin, *Diffuzionnaya zona* (Diffusion Zone). Moscow: Nauka, 344 p. (1979) (in Russian).
7.15. V.A. Uskov, E.V. Kurilchik, *Uch. zap. Gorkovskogo univ*. Voprosy fiziki tverdogo tela, 132–136, Gorki (1973) (in Russian).
7.16. I.A. Akhiezer, L.N. Davydov, *Metalofizika* **3**, No. 3: 3–15 (1981) (in Russian).
7.17. S.S. Khludkov, T.T. Lavrishchev, In: *Arsenid Galliya* (Gallium Arsenide): 290–295, Tomsk (1969) (in Russian).
7.18. V.I. Fistul, M.I. Sinder, *FTP* **17**, No. 11: 1195–2008 (1982) (in Russian).
7.19. V.A. Uskov, In: *Svoistva legirovannykh poluprovodnikov* (Properties of Doped Semiconductors). Moscow: Nauka. 129–134 (1977) (in Russian).
7.20. B.I. Boltaks, *Diffuzitya i pochechnye defekty v poluprovodnikakh* (Diffusion of Point Defects in Semiconductors). Leningrad: Nauka, 384 p. (1972) (in Russian).
7.21. A.F. Willoughby, *Rep. Progr. Phys*. **41**: 1662–1705 (1978).
7.22. V.A. Uskov, E.V. Kurilchik, V.V. Vaskin, Izvestiya AN SSSR, ser. Neorg. mat. 7, No. 10: 1689–1693 (1973) (in Russian).
7.23. V.V. Vaskin, V.S. Metrikin, V.A. Uskov *et al*., *FTT* **8**, No. 12: 3467–3473 (1966) (in Russian).
7.24. B.I. Boltaks, R.Sh. Malkovich, V.A. Pokoeva *et al*., *FTT* **10**, No. 4: 755–757 (1976) (in Russian).
7.25. R.Sh. Malkovich, V.A. Pokoeva, *FTT* **19**, No. 9: 1731–1736 (1977) (in Russian).
7.26. R.Sh. Malkovich, V.A. Pokoeva, *FTT* **18**, No. 9: 2606–2610 (1976) (in Russian).
7.27. V.A. Panteleev, T.S. Gugina, *FTT* **19**, No. 1: 181–184 (1977) (in Russian).
7.28. V.A. Panteleev, T.S. Gugina, V.N. Leonov, *Izvestiya vuzov, ser. Fizika* **45**, No. 6: 124–127 (1977) (in Russian).
7.29. S.R. De Groot, P. Mazur, *Non-Equilibrium Thermodynamics*. Amsterdam: North–Holland Publ. Co., 454 p. (1962).
7.30. G.V. Shcherbedinsky, *Izuchenie vzaimodeistviya ugleroda s legiruuyshimi elementami v troinoi sisteme* (Carbon Interaction with Doping Elements in Ternary Systems). Avtoref. dokt. diss. Moscow: TZNI–CHERMET, 33 p. (1973) (in Russian).
7.31. G.V. Shcherbedinsky, L.A. Kondrachenko, *FMM* **29**, No. 4: 788–795 (1970) (in Russian).
7.32. V.V. Vaskin, *FTP* **2**, No. 1: 102–109 (1968) (in Russian).
7.33. V.A. Panteleev, V.A. Muravyev, *FTT* **19**, No. 2: 682–686 (1977) (in Russian).
7.34. V.A. Panteleev, M.I. Vasilevsky, Yu.L. Kalinkin, *FTT* **25**, No. 10: 2930–2931 (1983) (in Russian).
7.35. M.I. Vasilevsky, G.M. Golemshtok, V.A. Panteleev, *FTT* **27**, No 1: 187–193 (1985) (in Russian).
7.36. I.M. Lifshitz, L.V. Tanatarov, *FMM* **12**, No. 3: 331–337 (1961) (in Russian).
7.37. V.S. Eremeev, *FMM* **45**, No. 3: 599–607 (1978) (in Russian).
7.38. G.N. Gaidukov, B. Ya. Lyubov, *FTT* **21**, No. 6: 1701–1709 (1979) (in Russian).
7.39. A.M. Kosevich, *Fizicheskaya mekhanika realnykh kristallov* (Physical Mechanics of Real Crystals). Kiev: Naukova dumka, 328 p. (1981) (in Russian).
7.40. F.C. Larche, J.W. Cahn, *Acta Metalurgica* **30**, No. 10: 1835–1845 (1982).

7.41. O.V. Aleksandrov, I.P. Matkhanova, *Izvestiya AN SSSR, ser. Neorg. mat.* **19**, No. 4: 517–520 (1982) (in Russian).
7.42. M.A. Kriviglaz, *FMM* **17**, No. 2: 161–167 (1962) (in Russian).
7.43. H. Reiss, C.S. Fuller, *Bell Syst. Techn. J.* **35**, No.3: 535–636 (1956).
7.44. V.V. Vaskin, V.A. Uskov, *FTT* **10**, No. 4: 1239–1241 (1968) (in Russian).
7.45. V.V. Vaskin, V.A. Uskov, *FTT* **11**, No. 7: 1763–1769 (1969) (in Russian).
7.46. V.A. Uskov, *Izvestiya AN SSSR, ser. Neorg. mat.* **9**, No. 6: 1949–1050 (1973) (in Russian).
7.47. G.V. Shcherbedinsky, V.I. Shaidurov, *DAN SSSR* **181**, No. 5: 1193–1196 (1968) (in Russian).

Chapter 8

Impurity Migration in the Formation of Mobile Complexes

8.1 DIATOMIC COMPLEXES: FORMATION AND DECOMPOSITION

Complexes are often formed by particles possessing electrical activity. This is likely to be associated with long-range potential interaction. Thermal treatment inevitable in semiconductor technology or device heating during its application accelerates particle association.

Consider a set of kinetic equations for the formation of two-particle complexes. Suppose that the association occurs in an elemental semiconductor and involves two constituents according to the reaction $A + B \rightleftarrows K$. This process is described by the set of equations:

$$\begin{aligned}
\frac{dN_c}{dt} &= -e_c N_c + C_c N_A N_B \\
\frac{dN_A}{dt} &= e_c N_c - C_c N_A N_B \qquad (8.1.1) \\
\frac{dN_B}{dt} &= -e_c N_c + C_c N_A N_B,
\end{aligned}$$

where e_c is the thermal decay rate of a complex; C_c is the capture coefficient of complexation; N_c is the concentration of complexes; N_A and N_B are concentrations of A and B particles, respectively.

Equations (8.1.1) are written in accordance with the theory presented above. The superscripts have been omitted, because the process occurs in an elemental semiconductor and the particles are substitutional atoms. However, this set of equations can describe a semiconductor of any complexity. Moreover, the role of a particle can be played by another complex; in that case (8.1.1) describes its enlargement and transformation to a more complex associate. This set of equations will be solved under the following boundary conditions:

$$N_A(t = 0) = N_A^0, \qquad N_B(t = 0) = N_B^0, \qquad N_c(t = 0) = 0. \tag{8.1.2}$$

This suggests that there are no complexes at the initial moment of time and that the concentration of partners in a complex is minimal.

The variables in (8.1.1) are not independent. At every moment of time, they are related by the conservation law for the number of particles of sort A and sort B:

$$N_A^0 = N_A(t) + N_c(t), \qquad N_B^0 = N_B(t) + N_c(t). \tag{8.1.3}$$

With these equations, the complexation kinetics will be

$$\frac{dN_c}{dt} = aN_c^2 + bN_c + d\ , \tag{8.1.4}$$

where

$$a = C_c, \qquad b = -e_c - C_c(N_A^0 + N_B^0), \qquad d = C_c N_A^0 N_B^0.$$

Equation (8.1.4) can be solved using the following substitutions.
(1) $N_c(t) = Y(t) + N_c^e$, where N_c^e is the equilibrium solution to be found from the equation

$$aN_c^2 + bN_c + d = 0, \tag{8.1.5}$$

and the equation for $Y(t)$ will be

$$\frac{dY}{dt} = aY^2 + \left(b + 2N_c^e\right)Y\ . \tag{8.1.6}$$

(2) Equation (8.1.6) reduces to a linear equation after the substitution of expression $Y(t) = 1/Z(t)$:

$$\frac{dZ}{dt} = \left(b + 2N_c^e\right)Z - a\,. \tag{8.1.7}$$

With the solution to linear equation (8.1.7) and the initial conditions (8.1.2), the time dependence of the concentration of complexes can be found from the expression

$$N_c(t) = N_c^e + \left\{\frac{c}{\lambda}\left[\exp(-\lambda t) - 1\right] - \frac{1}{N_c^e}\exp(-\lambda t)\right\}^{-1}, \tag{8.1.8}$$

where $\lambda = -e_c - C_c(N_A^0 + N_B^0 - 2N_c^e)$.

The equilibrium solution is found from the quadratic equation (8.1.5):

$$N_c^e = \frac{1}{2}\left(\frac{e_c}{C_c} + N_A^0 + N_B^0 - \sqrt{D}\right)$$
$$D = \left(\frac{e_c}{C_c} + N_A^0 + N_B^0\right)^2 - 4N_A^0 N_B^0\,. \tag{8.1.9}$$

One equation root is discarded in view of the condition $N_c^{ec} \le N_A^0, N_B^0$.

The function $N_c^e = f(T)$ for low temperatures ($T \to 0$) is equal to the constant ($\min\{N_A^0, N_B^0\}$); as the temperature rises, it decreases as $\exp(-g_c/kT)$, ($g_c < 0$).

Figure 8.1 presents the results of complexation kinetics modeling. The curves have a characteristic shape, while the time constant is not strictly constant. An important feature is the temperature dependence of the steady state value. At low temperatures, the process is shifted toward the larger number of complexes. Their maximum concentration is equal to the concentration of those partners in an associate which are less abundant in the semiconductor. Some complexes decompose at higher temperatures and the amplitude of the kinetic process goes down.

The temperature dependence of the steady state concentration is shown in Figure 8.2, together with isochronous annealing data for *p*-silicon [1]. One should note that the curves converge in the high temperature region and

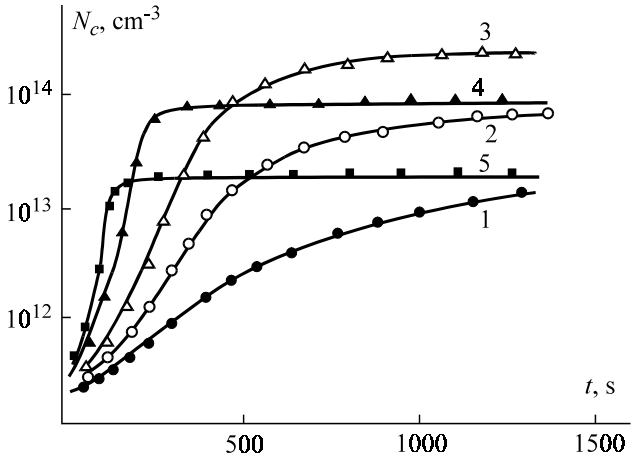

Figure 8.1. Complexation kinetics at different temperatures, °C: 1 – 500; 2 – 550; 3 – 600; 4 – 650; 5 – 700.

significantly differ at low temperatures. If the experimental data coincide with the bell-shaped curve, the steady state concentration of complexes is a kink. This is because complexation at low temperatures is confined to diffusion processes and the time is too short for the concentration to reach the steady state level during isochronous annealing. More exactly, the restriction is due to the capture of one partner by another, and the capture coefficient

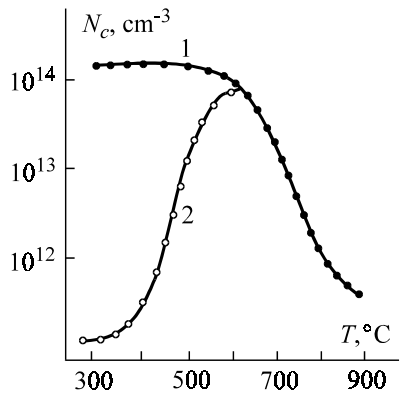

Figure 8.2. The temperature dependence of (1) the steady state concentration of complexes and (2) their concentration after isochronous annealing.

determining this process contains the diffusion coefficient [2, 3]. The difference between the steady state and experimental curves allows determination of the capture coefficient. The parameter fitting of the steady state process in the high temperature region yields the defect concentration and complexation energy.

Complexation kinetic coefficients can be found from the concentration kinetics of complexes at various temperatures. This can be done using equation (8.1.8). Complexation parameters can also be derived from the decrease in the number of *A* and *B* particles. But this is not always possible because of experimental difficulties. A very common experiment involves an isochronous semiconductor annealing. This experiment can be described by equations (8.1.8) and (8.1.9). For this, the annealing time in (8.1.8) must be taken to be fixed ($t = t_0$) and the concentration of complexes must be considered as a function of temperature. Modeling shows that the experimental plot in the concentration–temperature coordinates represents a bell-shaped curve. The calculations show that complexation is very effective in a certain temperature range. As the annealing time becomes longer, the concentration maximum shifts toward lower temperatures.

Formulas (8.1.8) and (8.1.9) are difficult to use for an experimental data description, but they can be simplified on the following assumption. Suppose the initial concentration of one type of particles is much higher than that of the other (for definiteness, $N_A^0 >> N_B^0, N_c^e$); then (8.1.8) and (8.1.9) will transform to

$$N_c^e \approx \frac{N_A^0 N_B^0}{N_A^0 + N_B^0 + g/C_c} \approx \left(\frac{1}{N_B^0} + \frac{1}{N_A^0} + \frac{g}{C_c N_A^0 N_B^0} \right)^{-1} \qquad (8.1.10)$$

$$N_c(t) = N_c^e\left[1 - \exp(\lambda t)\right], \quad \lambda(T) = -e_c - C_c N_A^0 . \qquad (8.1.11)$$

This approximation corresponds to a fairly common situation when, say, the impurity concentration exceeds the vacancy concentration. At non-zero initial concentration of complexes, the following solution can be obtained easily:

$$N_c(t) = N_c^e + \left(N_0 - N_c^e\right)\exp(\lambda t) , \qquad (8.1.12)$$

where $N_c(t = 0) = N_0$. The annealing curve will be described by formula (8.1.12) if the time in it is taken to be fixed ($t = t_0$ is annealing time) and N_c is considered as a function of temperature.

The derived formulas can describe the formation of electrically active complexes in a large number of situations.

The temperature effect on the concentration of electrically active centers was studied in [1] using thermally stimulated currents. The temperature dependence of the concentration of centers with E_v + 0.46 eV in silicon during isochronous annealing was shown to be a bell-shaped curve with their maximum concentration of 7×10^{13} cm^{-3}. The authors suggested that this kind of center was a complex produced by point defects; but there were no suggestions concerning the nature of this center. Some of the center parameters were found in [4] with formula (8.1.12). The calculation procedure for finding complexation parameters was as follows.

(1) The tail of the bell-shaped curve lying in the high temperature region is described by formula (8.1.10). This is because the concentration rapidly reaches its equilibrium value in this region. After the parameter fitting, the equilibrium defect concentration can be written as

$$N_c^e = \frac{1}{6.2\times10^{-15} + 7.97\times10^{-5}\exp(-1.79/kT)}. \qquad (8.1.13)$$

The concentration of one partner in the center is 1.6×10^{14} cm^{-3}, according to (8.1.10). The energy of the center formation is approximately equal to 1.8 eV.

(2) The expression for equilibrium concentration (8.1.10) and formula (8.1.12) were used to derive the dependence

$$\lambda(T) = \frac{1}{t_0}\ln\frac{N_c - N_c^e}{N_0 - N_c^e}.$$

The initial concentration of the centers in question was found to be 1.78×10^{11} cm^{-3}.

(3) The activation energy of complexation can be found from the slope of the curve in the low temperature region if one plots the logarithmic function $\lambda(T)$ versus $1/T$. The high temperature slope can yield the activation energy of decomposition of the complex.

For this case, the activation energy of complexation was found to be 2.5 ± 0.1 eV, which coincides with that of oxygen diffusion (the latter was found to be 2.52 eV [5]). This suggests that an electrically active complex is produced by attachment of an oxygen atom to some particles. The concentration of such particles is ~1.6×10^{14} cm^{-3}. These may be background

Fe, Cu, Mn, and C atoms. No activation energy of decomposition has been traced in this temperature range; therefore, its probability is quite low. In accordance with (8.1.12) and (8.1.13), the concentration of such centers is described as

$$N_c(T) = N_c^e(T) + \left(N_0 - N_c^e(T)\right)\exp\left[-8.3\times10^{11}\exp\left(-\frac{2.5}{kT}\right)t_0\right]. \qquad (8.1.14)$$

Thus, isochronous annealing curves can be used to find the energy characteristics of the complex. These characteristics, in turn, can help to identify the center composition. Its nature may be clarified, with account taken of the energy level position, by quantum chemical methods or by detailed investigations of its complexation kinetics.

8.2 DIFFUSION MODEL FOR MOBILE COMPLEXES[1]

In contrast to the kinetics of immobile complexes described by equation (7.5.24), the set of equations for the diffusion of components in the reaction $A + B \rightleftarrows Q$ has the general form:

$$\frac{\partial H_A}{\partial t} = D_A\frac{\partial^2 H_A}{\partial x^2} - K_1 H_A H_B + K_2 Q,$$

$$\frac{\partial H_B}{\partial t} = D_B\frac{\partial^2 H_B}{\partial x^2} - K_1 H_A H_B + K_2 Q, \qquad (8.2.1)$$

$$\frac{\partial Q}{\partial t} = D_Q\frac{\partial^2 Q}{\partial x^2} + K_1 H_A H_B - K_2 Q,$$

where H_A is the concentration of free component *A*, H_B is the concentration of free component *B*, Q is the concentration of complexes, K_1 is the rate constant of complexation, and K_2 is the rate constant of decomposition of a complex.

[1] Here, we follow M. I. Sinder's dissertation *A Theoretical Treatment of Sub-surface Profiles in Semiconductors* (Moscow, 1982) done under the author's supervision, and the publications [6–8].

The set of equations (8.2.1) has no analytical solution. For this reason, it is usually solved by computational techniques to obtain the dependences $H_A(x, t)$, $H_B(x, t)$, $Q(x, t)$.

Let us analyze qualitatively the processes occurring during *A*, *B*, and *Q* diffusion into a sample uniformly doped with *A*, *B*, and *Q* impurities. We will first evaluate the order of magnitude of the terms of equations (8.2.1):

$$\frac{\partial H_A}{\partial t} \sim \frac{H_A}{t}, \qquad D_A \frac{\partial^2 H_A}{\partial x^2} \sim \frac{H_A}{x^2}, \qquad K_1 H_A H_B - K_2 Q \sim \frac{H_A}{\tau} \quad \text{and so on,}$$

where τ is the characteristic time of the reaction A + B $\rightleftarrows$ Q, or the characteristic time for chemical equilibrium to be established in a uniform closed system.

By substituting these evaluations into (8.2.1), we get

$$\frac{1}{t} \approx \frac{D_A}{x^2} + \frac{1}{\tau}$$

$$\frac{1}{t} \approx \frac{D_B}{x^2} + \frac{1}{\tau} \qquad (8.2.2)$$

$$\frac{1}{t} \approx \frac{D_Q}{x^2} + \frac{1}{\tau}.$$

It follows from (8.2.2) at short times $t << \tau$ that the terms $\sim 1/\tau$ describing the interaction of diffusing particles can be neglected; (8.2.1) transforms to

$$\frac{\partial H_A}{\partial t} = D_A \frac{\partial^2 H_A}{\partial x^2},$$

$$\frac{\partial H_B}{\partial t} = D_B \frac{\partial^2 H_B}{\partial x^2}, \qquad (8.2.3)$$

$$\frac{\partial Q}{\partial t} = D_Q \frac{\partial^2 Q}{\partial x^2}.$$

This means that the three components diffuse independently into the sample bulk at small time values.

There may be three situations at large times, $t >> \tau$.

At the sample surface with $x << (Dt)^{1/2}$ (D is an arbitrary coefficient of diffusion D_A, D_B, or D_Q), it follows from (8.2.2) that ~1/t terms describing the time variation of concentration can be neglected; (8.2.1) transforms to

$$D_A \frac{\partial^2 H_A}{\partial x^2} - K_1 H_A H_B + K_2 Q = 0,$$

$$D_B \frac{\partial^2 H_B}{\partial x^2} - K_1 H_A H_B + K_2 Q = 0, \qquad (8.2.4)$$

$$D_Q \frac{\partial^2 Q}{\partial x^2} + K_1 H_A H_B - K_2 Q = 0.$$

This set of equations describes steady state profiles of particles at the sample surface.

In the sample bulk with $x >> (Dt)^{1/2}$, the terms describing the interaction of particles become much larger than the other terms, and it immediately follows from (8.2.1) and (8.2.2) that

$$K_1 H_A H_B = K_2 Q. \qquad (8.2.5)$$

The two missing equations can be obtained by taking into account the other terms in equations (8.2.1). To have only commensurable terms in the equations, we add the first two equations of (8.2.1) to the third term to get the expressions

$$\frac{\partial H_A}{\partial t} + \frac{\partial Q}{\partial t} = D_A \frac{\partial^2 H_A}{\partial x^2} + D_Q \frac{\partial^2 Q}{\partial x^2}, \qquad (8.2.6)$$

which, together with (8.2.5), form a full set of equations. It coincides with the set of equations (8.2.1) suggested in [9] and describes a consistent migration of particles into the sample bulk. For the finite sample length L, we get the following conditions for the transformation of equations (8.2.1) to (8.2.5) and (8.2.6):

$$\sqrt{D\tau} << x << L, \qquad \tau << t << L^2/D. \qquad (8.2.7)$$

Note that if the sample is sufficiently long, the diffusion term for the region $x >> (Dt)^{1/2}$, $(D\tau)^{1/2}$ in the sample bulk can be neglected, and the set of equations (8.2.1) transforms to the conventional chemical kinetics equations:

$$\frac{\partial H_A}{\partial t} = -K_1 H_A H_B + K_2 Q\,,$$

$$\frac{\partial H_\mathrm{B}}{\partial t} = -K_1 H_\mathrm{A} H_\mathrm{B} + K_2 Q\,, \qquad (8.2.8)$$

$$\frac{\partial Q}{\partial t} = K_1 H_A H_B - K_2 Q\,.$$

Let us consider in more detail the sets of equations (8.2.4), (8.2.5), and (8.2.6) describing stationary and nonstationary distributions of particles in a sample at large times $t >> \tau$.

The boundary conditions at $x = 0$ for equations (8.2.4) coincide with those for (8.2.1):

$$H_\mathrm{A}\big|_{x=0} = H_{\mathrm{A}0}\,, \qquad H_\mathrm{B}\big|_{x=0} = H_{\mathrm{B}0}\,, \qquad Q\big|_{x=0} = Q_0\,. \qquad (8.2.9)$$

The solution to the set of equations (8.2.4) with the boundary conditions of (8.29) and a limited concentration at $x \to \infty$ is

$$H_\mathrm{A} = H_{\mathrm{A}0} + \frac{D_\mathrm{Q}}{D_\mathrm{A}}\big[Q_0 - Q(x)\big],$$

$$H_\mathrm{B} = H_{\mathrm{B}0} + \frac{D_\mathrm{Q}}{D_\mathrm{B}}\big[Q_0 - Q(x)\big], \qquad (8.2.10)$$

$$Q = \frac{1}{D_\mathrm{Q}} Y(x)\,.$$

The dependences $H_\mathrm{A}(x)$, $H_\mathrm{B}(x)$, and $Q(x)$ have a monotonic character. For $K_1 H_{\mathrm{A}0} H_{\mathrm{B}0} = K_2 Q_0$, the solutions to (8.2.10) transform to identical constants. Therefore, the necessary condition for a steady state subsurface profile is $K_1 H_{\mathrm{A}0} H_{\mathrm{B}0} \neq K_2 Q_0$, which means that the concentration of particles at the sample boundary must be the equilibrium concentration for the crystal bulk. Physically, there must be a subsurface layer of finite thickness, for which

these conditions are fulfilled. Therefore, a steady state profile is produced by "screening" of nonequilibrium concentrations on the crystal surface. This situation is similar to that for a Debye screening layer under a charged surface [10]. If there is an electric double layer, the electrical neutrality in the bulk is established due to electrical interactions of diffusing particles. In the case under consideration, equilibrium is established due to a chemical interaction of diffusing particles.

We can show that (8.2.10) at $x \to \infty$, $H_A(x) \to H_A(0)$, $H_B(x) \to H_B(x)$, and $Q(x) \to Q(x)$ satisfies the relations

$$D_A H_A(0) + D_Q Q(0) = D_A H_{A0} + D_Q Q_0$$

$$D_B H_B(0) + D_Q Q(0) = D_B H_{B0} + D_Q Q_0 \qquad (8.2.11)$$

$$K_1 H_A(0) H_B(0) = K_2 Q(0),$$

i.e., a chemical equilibrium is established in the sample bulk, or, more exactly, the particle concentrations satisfy the active mass law (8.2.5).

To solve the set of equations (8.2.4) and (8.2.6), it is necessary to define the initial and boundary conditions. The boundary concentrations at $x << (Dt)^{1/2}$ ($x \to 0$) are equal to $H_A(0)$, $H_B(0)$, and $Q(0)$, respectively. This is because there is a variation region $(D\tau)^{1/2} << x << (Dt)^{1/2}$, in which both sets of equations—(8.2.4) and (8.2.6)—are valid. They are smoothly sewn together in the indicated range of x, permitting the use of concentrations $H_A(0)$, $H_B(0)$, and $Q(0)$ as the boundary conditions for these sets of equations.

If the initial concentrations H_A^0, H_B^0, and Q^0 are equilibrium concentrations ($K_1 H_A^0 H_B^0 = K_2 Q^0$), they act as the initial conditions for equations (8.2.4) and (8.2.6). If the initial concentrations are nonequilibrium ones ($K_1 H_A^0 H_B^0 \neq K_2 Q^0$), equilibrium bulk concentrations $H_A(\infty)$, $H_B(\infty)$, and $Q(\infty)$ are established at large times $t >> \tau$, whose values can be obtained by solving the set of equations (8.2.8):

$$H_A(\infty) = \frac{1}{2}\left\{H_B^0 - H_A^0 - K + \sqrt{\left(H_A^0 - H_B^0 + K\right)^2 + 4K\left(H_B^0 + Q^0\right)}\right\}$$

$$H_B(\infty) = \frac{1}{2}\left\{H_A^0 - H_B^0 - K + \sqrt{\left(H_A^0 - H_B^0 + K\right)^2 + 4K\left(H_B^0 + Q^0\right)}\right\} \qquad (8.2.12)$$

$$Q(\infty) = \frac{1}{2}\left\{H_A^0 + H_B^0 + K + 2Q^0 - \sqrt{\left(H_A^0 - H_B^0 + K\right)^2 + 4K\left(H_B^0 + Q^0\right)}\right\},$$

where $K = K_1/K_2$.

The concentration distribution in the sample bulk remains uniform, and the values of $H_A(\infty)$, $H_B(\infty)$, and $Q(\infty)$ will act as the initial conditions for the set of equations (8.2.4) and (8.2.6).

Thus, the qualitative analysis of diffusion of particles *A*, *B*, and *Q* into a plane sample uniformly doped with the same particles *A*, *B*, and *Q* involved in the reaction $A + B \rightleftarrows Q$ provides the following picture of the process.

At short times, $t << \tau$, there is an independent diffusion of *A*, *B*, and *Q* into the sample. Then, at $t \sim \tau$, there is an interaction of particles because of the reaction $A + B \rightleftarrows Q$, resulting, first, in a change of the particle motion away from the sample surface and, second, in a chemical equilibrium in the crystal bulk ($x >> (Dt)^{1/2}$, $(D\tau)^{1/2}$). At times $\tau << t << L^2/D$, a steady state distribution is established at the sample surface. A chemical equilibrium is established in the bulk at $x >> (Dt)^{1/2}$, $(D\tau)^{1/2}$. A diffusion-like wave propagates through the region of $(D\tau)^{1/2} << x << L$, representing a consistent movement of the particles from the surface into the bulk. Thus, a wave propagates through the bulk, "absorbing" the equilibrium distribution in the bulk $H_A(\infty)$, $H_B(\infty)$, and $Q(\infty)$ and prescribing a new equilibrium distribution $H_A(0)$, $H_B(0)$, and $Q(0)$, which has been established at the steady state profile boundary. At $t >> L^2/D$, a uniform equilibrium concentration distribution is established in the sample: $H_A(0)$, $H_B(0)$, and $Q(0)$. At the surface, there are steady state distributions extending as far as the length $(D\tau)^{1/2}$.

When particles *A*, *B*, and *Q* diffuse into a plane sample, the absence of one of them, say, H_B, is prescribed at its boundary:

$$\left.\frac{\partial H_B}{\partial x}\right|_{x=0} = 0, \quad H_A\big|_{x=0} = H_{A0}, \qquad Q\big|_{x=0} = Q_0.$$

Then at times $t >> \tau$, an equilibrium will be established on the sample surface, and the concentration H_B will have the value $H_B|_{x=0} = KQ_0/H_{A0}$. Since the surface concentrations will satisfy relation (8.2.5), no subsurface steady state profile will be formed but the rest of the process will remain the same as in the previous case.

If the absence of two kinds of particles, for example H_B and *Q*, is prescribed at the sample boundary

$$\left.\frac{\partial H_B}{\partial x}\right|_{x=0} = 0, \qquad \left.\frac{\partial Q}{\partial x}\right|_{x=0} = 0, \qquad H_A\big|_{x=0} = H_{A0},$$

then the equilibrium will be established on the surface at times $t >> \tau$, exactly as above, and there will be no steady state surface profiles. However, in contrast to the case above with the completely defined boundary conditions for (8.2.4) and (8.2.6) at $x = 0$ ($H_A|_{x=0} = H_{A0}$, $H_B|_{x=0} = KQ_0/H_{A0}$, $Q|_{x=0} = Q_0$), we have now only one boundary condition $H_A|_{x=0} = H_{A0}$. The missing boundary condition is

$$D_B \frac{\partial H_B}{\partial x} + D_Q \left.\frac{\partial Q}{\partial x}\right|_{x=0} = 0$$

reflecting the conservation of the total amount of B impurity during diffusion. Indeed, equations (8.2.6) are valid for any moments of time. By integrating the second equation with respect to x, we will get

$$\frac{\partial}{\partial t}\int_0^\infty H_B dx + \int_0^\infty Q dx = D_B \frac{\partial H_B}{\partial x}\bigg|_0^\infty + D_Q \frac{\partial Q}{\partial x}\bigg|_0^\infty = 0;$$

hence,

$$\int_0^\infty (H_B + Q) dx = 0.$$

The above boundary condition unambiguously defines the solution to the sets of equations (8.2.4) and (8.3.6).

Similarly, one can analyze diffusion processes occurring in a plane sample, when its one half is uniformly doped with, say, A particles, at the initial moment of time and the other half with B particles. As in the above case, we can show that the concentration distribution is described by (8.2.4) and (8.2.6) at large times, $t >> \tau$. Note that no steady state profiles are formed in this case. Steady state profiles at the boundary separating the crystal regions doped with different impurities A and B can arise only if the reaction rate constants K_1 and K_2 in the region doped with A are not equal to those in the region doped with B.

Of special interest for the analysis of experimental diffusion profiles in semiconductors is the nonstationary impurity diffusion from the sample surface into the bulk, described by equations (8.2.4) and (8.2.6). Since we consider problems only with initial uniform distribution, we can make

Boltzmann's substitution $\lambda \equiv x/2t^{1/2}$ simplifying significantly the sets of equations (8.2.4) and (8.2.6):

$$\begin{aligned} d_A^2 H_A'' + d_Q^2 Q'' + 2\lambda\left(H_A + Q\right)' &= 0 \\ d_B^2 H_B'' + d_Q^2 Q'' + 2\lambda\left(H_B + Q\right)' &= 0 \\ H_A H_B &= KQ\,, \end{aligned} \tag{8.2.13}$$

where $d_A{}^2 \equiv D_A$, $d_B{}^2 \equiv D_B$, $d_Q{}^2 \equiv D_Q$, and the primes designate differentiation with respect to λ.

The set of equations (8.2.13) can also be written in terms of total concentrations $N_A \equiv H_A + Q$ and $N_B \equiv H_B + Q$:

$$\begin{aligned} d_A^2\left(N_A - Q\right)'' + d_Q^2 Q'' + 2\lambda N_A' &= 0 \\ d_B^2\left(N_B - Q\right)'' + d_Q^2 Q'' + 2\lambda N_B' &= 0 \\ \left(N_A - Q\right)\left(N_B - Q\right) &= KQ\,. \end{aligned} \tag{8.2.14}$$

The solution to the set of equations (8.2.13) will be considered for the three most common situations: sequential diffusion, simultaneous diffusion, and interdiffusion. The boundary conditions are as follows.

(1) For sequential diffusion:

$$\begin{aligned} H_A\big|_{\lambda=0} &= H_A(0) \qquad & H_A\big|_{\lambda=\infty} &= H_A(\infty) \\ d_B^2 H_B' + d_Q^2 Q'\big|_{\lambda=0} &= 0 \qquad & H_B\big|_{\lambda=\infty} &= H_B(\infty)\,. \end{aligned} \tag{8.2.15}$$

(2) For simultaneous diffusion:

$$\begin{aligned} H_A\big|_{\lambda=0} &= H_A(0) \qquad & H_A\big|_{\lambda=\infty} &= H_A(\infty) \\ H_B\big|_{\lambda=0} &= H_B(0) \qquad & H_B\big|_{\lambda=\infty} &= H_B(\infty)\,. \end{aligned} \tag{8.2.16}$$

(3) For interdiffusion

$$H_A\big|_{\lambda=-\infty} = H_A(-\infty) \qquad H_A\big|_{\lambda=+\infty} = 0$$
$$H_B\big|_{\lambda=-\infty} = 0 \qquad\qquad H_B\big|_{\lambda=+\infty} = H_B(+\infty)\,. \tag{8.2.17}$$

Sequential diffusion corresponds to the absence of two particle flows at the sample boundary

$$\left.\frac{\partial H_B}{\partial x}\right|_{x=0} = 0 \qquad\qquad \left.\frac{\partial Q}{\partial x}\right|_{x=0} = 0$$

in the solution of equations (8.2.1). Simultaneous diffusion corresponds to the prescription of, at least, two concentrations at the sample boundary. Interdiffusion is for diffusional homogenization of the sample, whose one half is doped with *A* particles and the other half with *B* particles It would be easy to re-write the boundary conditions (8.2.15) through (8.2.17) in terms of total concentrations N_A and N_B.

The set of equations (8.2.13) with the boundary conditions (8.2.15)–(8.2.17) generally has no analytical solutions. So we will consider some limiting cases, following mainly the work [11].

The strong complexation approximation ($K = 0$). Physically, the strong complexation approximation means that the reaction $A + B \rightleftarrows Q$ is sharply shifted toward the formation of a complex. Formally, it follows from equation (8.2.5) at $K = 0$ that the concentration of, at least, one free component is zero, i.e., $H_A = 0$ or $H_B = 0$. This generally leads to two regions on the λ-axis, separated by a special point λ_0. An independent diffusion of the *Q* complex and a free component (H_A or H_B) occurs within each region. From the continuous concentration condition, H_A and H_B at λ_0 are equal to zero, i.e., particles H_A and H_B migrate toward point λ_0, which physically means the chemical reaction front in the formation of the *Q* complex.

The weak complexation approximation ($K >> H_A, H_B$). The weak complexation approximation corresponds to the shift of the reaction $A + B \rightleftarrows Q$ toward the formation of free components. From (8.2.13) in the zero approximation ($K = \infty$), we have $Q^0 = 0$, and the components *A* and *B* will diffuse independently. In the next approximation, we have

$$Q^{(1)} = \frac{H_A^0 H_B^0}{K} << H_A, H_B$$

permitting a trivial perturbation theory to be built.

The linear approximation ($\Delta N_A << N_A$, $\Delta N_B << N_B$). Here, the change in the concentrations of all particles during their diffusion is much smaller than the concentration values. These conditions permit linearization of the set of equations (8.2.14)

$$\begin{aligned} D_{AA}(\Delta N_A)'' + D_{AB}(\Delta N_B)'' + 2\lambda(\Delta N_A)' = 0 \\ D_{BB}(\Delta N_B)'' + D_{BA}(\Delta N_A)'' + 2\lambda(\Delta N_B)' = 0 . \end{aligned} \tag{8.2.18}$$

This set of equations coincides with the one considered in [12], but the coefficients D_{AA}, D_{BB}, D_{AB}, and D_{BA} were introduced formally in [12]. In our case, they have a clear physical meaning and are expressed via the diffusion coefficients D_A, D_B, and D_Q, the rate constant of the reaction $A + B \rightleftarrows Q$, and the initial concentrations $H_A(\infty)$, $H_B(\infty)$, and $Q(\infty)$:

$$\begin{aligned} D_{AA} &= \frac{D_A\left[K + H_A(\infty)\right] + D_Q H_B(\infty)}{K + H_A(\infty) + H_B(\infty)} \\ D_{BB} &= \frac{D_B\left[K + H_B(\infty)\right] + D_Q H_A(\infty)}{K + H_A(\infty) + H_B(\infty)} \\ D_{AB} &= (D_Q - D_A)\frac{H_A(\infty)}{K + H_A(\infty) + H_B(\infty)} . \end{aligned} \tag{8.2.19}$$

The approximation of close diffusion coefficients. In the case of $D_A = D_B + D_Q$, the set of equations (8.2.14) has a simple analytical solution. It is of much interest because one can follow the variation in the distributions H_A, H_B, and Q as a function of the reaction rate constant K.

8.3 SOLUTION OF DIFFUSION EQUATIONS FOR VARIOUS BOUNDARY CONDITIONS

Here we will present the solutions to equations (8.2.14) with the boundary conditions (8.2.15)–(8.2.17). The boundary conditions are typical for the situation when a semiconductor is doped with one sort of impurity.

8.3.1 Sequential diffusion

Consider the problem of diffusion of component *A* into a semi-infinite sample uniformly doped with component *B* in the absence of *B* sublimation from the sample. The boundary conditions for this problem are

$$N_{\mathrm{A}}\big|_{\lambda=0} = N_{\mathrm{A}}(0) \qquad N_A\big|_{\lambda=\infty} = 0$$
$$d_{\mathrm{B}}^2\left(N_{\mathrm{B}} - Q\right)' + d_{\mathrm{Q}}^2 Q'\big|_{\lambda=0} = 0 \qquad N_{\mathrm{B}}\big|_{\lambda=\infty} = N_{\mathrm{B}}(\infty). \tag{8.3.1}$$

These conditions correspond to the following boundary conditions of equations (8.2.1):

$$H_{\mathrm{A}}\big|_{t=0} = 0, \qquad Q\big|_{t=0}, \qquad H_{\mathrm{B}}\big|_{t=0} = H_{\mathrm{B}}^0, \quad x > 0$$
$$H_A\big|_{x=0} = H_{x0}, \quad \left.\frac{\partial Q}{\partial x}\right|_{x=0} = 0, \qquad \left.\frac{\partial H_{\mathrm{B}}}{\partial x}\right|_{x=0} = 0 \tag{8.3.2}$$

or

$$H_{\mathrm{A}}\big|_{t=0} = 0, \qquad Q\big|_{t=0}, \qquad H_{\mathrm{B}}\big|_{t=0} = H_{\mathrm{B}}^0, \quad x > 0$$
$$Q\big|_{x=0} = Q_0, \qquad \left.\frac{\partial Q}{\partial x}\right|_{x=0} = 0, \qquad \left.\frac{\partial H_{\mathrm{B}}}{\partial x}\right|_{x=0} = 0. \tag{8.3.3}$$

Conditions (8.3.3) are more general than (8.3.2) because if we prescribe the *Q* content at the boundary, the content of the free component H_{A} at the boundary may be zero, which is ruled out for conditions (8.3.2).

(1) *The strong complexation approximation* ($K = 0$). Suppose $N_{\mathrm{B}}(\infty) > (d_{\mathrm{Q}}/d_{\mathrm{B}})N_A(0)$, corresponding to the absence of a free particles at the boundary. Then, the solution to (8.2.14) has the form (Figure 8.3):

$$N_{\mathrm{A}} = Q = N_{\mathrm{A}}(0)\operatorname{erfc}\left(\lambda/d_{\mathrm{Q}}\right)$$
$$N_{\mathrm{B}} = N_{\mathrm{B}}(\infty) - \frac{d_{\mathrm{Q}}}{d_{\mathrm{B}}} N_{\mathrm{A}}^0 \operatorname{erfc}\left(\lambda/d_{\mathrm{B}}\right) \tag{8.3.4}$$
$$N_{\mathrm{B}} = H_{\mathrm{B}} + Q.$$

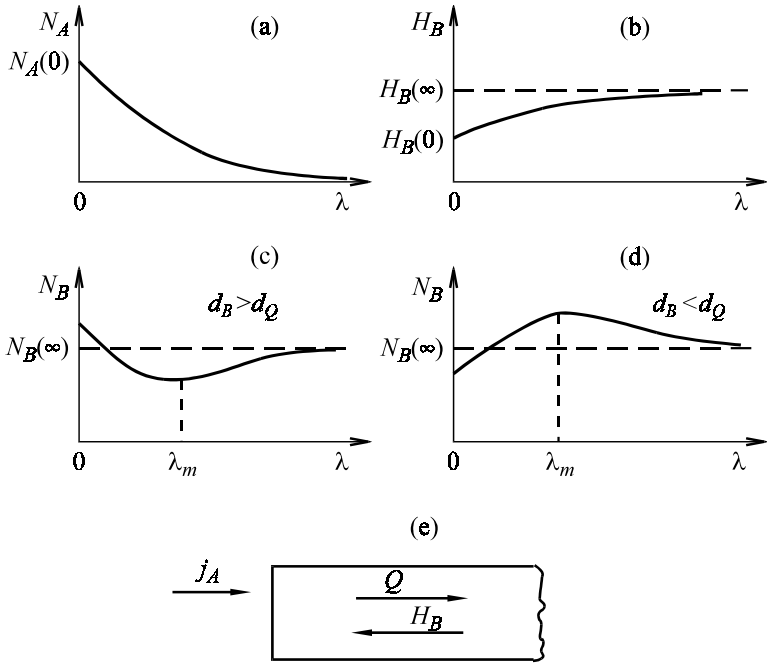

Figure 8.3. Distributions of components and directions of flows in sequential diffusion in the strong complexation approximation at $N_B(\infty) > d_Q N_A(0)/d_B$.

One can see from Figure 8.3 *c* and *d* that the N_B curve has an extremum with the coordinate λ_m:

$$\lambda_m^2 = \frac{d_B^2 d_Q^2}{d_B^2 - D_Q^2} \ln \frac{d_B^2}{d_Q^2} .$$

It is a minimum at $d_B > d_Q$ (Figure 8.3*c*) and a maximum at $d_B < d_Q$ (Figure 8.3*d*). The extremum appears in the distribution of N_B and of complexes Q if the particle flows are those shown in Figure 8.3*e*. Complexes Q migrate inward where their concentration is zero. The formation of complexes at the boundary, where they bind some free particles H_B, leads to an outward flow of the component H_B.

Suppose now $N_B(\infty) < (d_Q/d_B)N_A(0)$, corresponding to the presence of the free component H_A at the sample boundary. The solution to (8.2.14), represented in Figure 8.4, is

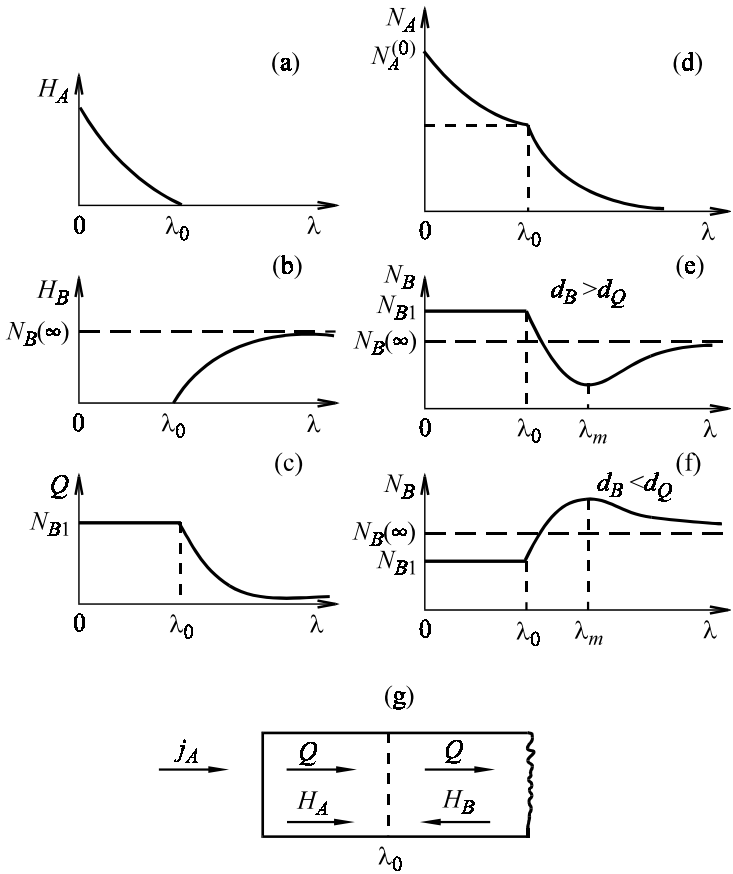

Figure 8.4. Distributions of components and directions of flows in sequential diffusion in the strong complexation approximation at $N_B(\infty) < d_Q N_A(0)/d_B$.

$$H_A \equiv N_A - Q = \begin{cases} \left(N_A(0) - N_{B1}\right)\left(1 - \dfrac{\operatorname{erfc}\frac{\lambda}{d_A}}{\operatorname{erfc}\frac{\lambda_0}{d_A}}\right) & \lambda \le \lambda_0 \\ 0 & \lambda > \lambda_0 \end{cases}$$

$$H_{\mathrm{B}} \equiv N_{\mathrm{B}} - Q = \begin{cases} 0 & \lambda \le \lambda_0 \\ N_{\mathrm{B}}(\infty)\left(1 - \dfrac{\operatorname{erfc}\frac{\lambda}{d_{\mathrm{B}}}}{\operatorname{erfc}\frac{\lambda_0}{d_{\mathrm{B}}}}\right) & \lambda > \lambda_0 \end{cases} \tag{8.3.5}$$

$$Q \equiv \begin{cases} N_{\mathrm{B1}} & \lambda \ge \lambda_0 \\ N_{\mathrm{B1}} \dfrac{\operatorname{erfc}\frac{\lambda}{d_{\mathrm{Q}}}}{\operatorname{erfc}\frac{\lambda_0}{d_{\mathrm{Q}}}} & \lambda < \lambda_0 \end{cases}.$$

Constants λ_0 and N_{B1} are found from the balance equation for N_{A} and N_{B}:

$$\begin{gathered} \int_0^\infty \left[N_{\mathrm{B}}(\lambda) - N_{\mathrm{B}}(\infty)\right] \mathrm{d}\lambda = 0 \\ 2\int_0^\infty N_{\mathrm{A}}(\lambda)\,\mathrm{d}\lambda = -d_{\mathrm{A}}^2\left(N_{\mathrm{A}} - Q\right)' - d_{\mathrm{Q}} Q'\big|_{\lambda=0}. \end{gathered} \tag{8.3.6}$$

The equalities of (8.3.6) are derived by integrating the first two equations in (8.2.14) with respect to λ from 0 to ∞. They are the integral form of the particle conservation law in this problem. By substituting the solutions to (8.3.5) into (8.3.6), we eventually get the equations describing λ_0 and N_{B1}:

$$\begin{gathered} N_{\mathrm{B1}} T(\lambda_0/d_{\mathrm{Q}}) = N_{\mathrm{B}}(\infty) T(\lambda_0/d_{\mathrm{B}}) \\ N_{\mathrm{A1}} T(\lambda_0/d_{\mathrm{Q}}) = \left(N_{\mathrm{A}}(0) - N_{\mathrm{B1}}\right) S(\lambda_0/d_{\mathrm{A}}) \\ T(x) \equiv \frac{\exp(-x^2)}{\sqrt{\pi}x\operatorname{erfc}(x)} \quad S(x) \equiv \frac{\exp(-x^2)}{\sqrt{\pi}x\operatorname{erfc}(x)}. \end{gathered} \tag{8.3.7}$$

The same equations can be derived from the condition at the reaction front:

$$d_{\mathrm{A}}^2 H_{\mathrm{A}}'\big|_{\lambda=\lambda_0} + d_{\mathrm{Q}}^2 Q'\big|_{\lambda=\lambda_0+0} - d_{\mathrm{Q}}^2 Q'\big|_{\lambda=\lambda_0-0} + 2\lambda_0 Q\big|_{\lambda=\lambda_0} = 0 \tag{8.3.8}$$

$$d_B^2 H_B'\big|_{\lambda=\lambda_0} + d_Q^2 Q'\big|_{\lambda=\lambda_0+0} - d_Q^2 Q'\big|_{\lambda=\lambda_0-0} + 2\lambda_0 Q\big|_{\lambda=\lambda_0} = 0,$$

which are obtained by integrating equations (8.2.13) in the vicinity of the reaction front. Note that $N_A(0) - N_{B1} \equiv H_A(0)$.

It is clear from Figure 8.4 *e*, and *f* that an extremum appears in the distribution N_B with the coordinate λ_m, which is defined by the equality

$$\lambda_m^2 = \lambda_0^2 + \frac{d_B^2 d_Q^2}{d_B^2 - d_Q^2} \ln \frac{d_B^2}{d_Q^2}.$$

The presence of the free component H_A at the boundary changes qualitatively the distribution pattern. Two regions with $\lambda < \lambda_0$ and $\lambda > \lambda_0$ appear. The free component H_B is absent from the first regions. The free component H_A migrates inward to the complexation front λ_0. The free component H_A is absent from the $\lambda > \lambda_0$ region. There is a H_B flow outward to the complexation front, while a flow of Q complexes moves inward from this front.

The existence of counterpropagating flows H_B and Q gives rise to an extremum in the $\lambda > \lambda_0$ region. The absence of Q flow from the $\lambda < \lambda_0$ region is associated with the boundary condition

$$d_B^2 (N_B - Q)' + d_Q^2 Q'\big|_{x=0} = 0$$

and with the absence of free H_B in this region.

In the limit $d_Q \to 0$, the solutions to (8.3.5) transform to the solutions to (7.5.21) and (7.5.22).

(2) *The weak complexation approximation* ($K >> N_A$, N_B). In the zero approximation ($K = \infty$), we have $Q = 0$; so it follows from (8.2.14) that the migration of N_A and N_B is described by the common diffusion equations. Therefore, we have

$$N_A^0 = N_A(0)\operatorname{erfc}\frac{1}{d_A}, \qquad N_B = N_B(\infty).$$

In the first approximation with respect to ${N_A}^0/K$, ${N_B}^0/K$, we get $Q^{(1)} = ({N_A}^0 {N_B}^0)/K$. Considering $Q^{(1)}$ as a perturbation, we derive, to an accuracy of N_A/K or N_B/K, the following expressions for N_A and N_B (Figure 8.5):

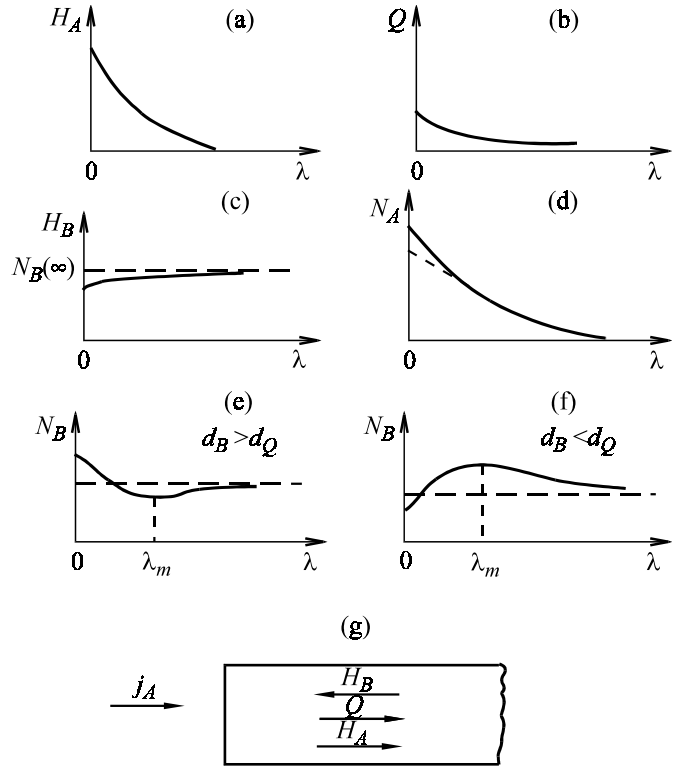

Figure 8.5. Distributions of components and sequential diffusion flows in the weak complexation approximation.

$$N_{\mathrm{A}} = N_{\mathrm{A}}(0)\operatorname{erfc}\frac{\lambda}{\sqrt{d_{\mathrm{A}}^2 + \dfrac{N_{\mathrm{B}}(\infty)}{K}\left(d_{\mathrm{Q}}^2 - d_{\mathrm{A}}^2\right)}}$$

$$N_{\mathrm{B}} = N_{\mathrm{B}}(\infty)\left\{1 + \frac{N_{\mathrm{A}}(0)}{K}\frac{d_{\mathrm{Q}}^2 - d_{\mathrm{B}}^2}{d_{\mathrm{A}}^2 - d_{\mathrm{B}}^2}\left(\operatorname{erfc}\lambda/d_{\mathrm{A}} - \frac{d_{\mathrm{A}}}{d_{\mathrm{B}}}\operatorname{erfc}\lambda/d_{\mathrm{B}}\right)\right\}. \tag{8.3.9}$$

It is seen from Figure 8.5*f* that the N_{B} distribution has an extremum, like in a strong complexation, associated with the counterpropagating Q and H_{B}

flows: a minimum at $d_B > d_Q$ and a maximum at $d_B < d_Q$. The extremum coordinated is defined by the equality

$$\lambda_m^2 = \frac{d_A^2 d_B^2}{d_B^2 - d_A^2} \ln \frac{d_B^2}{d_A^2}.$$

Note that there is no abrupt reaction front in a weak complexation.

The applicability of expressions (8.3.9) imposes additional restrictions on the complexation constant:

$$\frac{K}{N_B(\infty)} >> \left| 1 - \frac{d_Q^2}{d_B^2} \right|$$

$$\frac{K}{N_A(0)} >> \left| 1 - \frac{d_Q^2}{d_B^2} \right|, \tag{8.3.10}$$

which is equivalent to the condition that a complex is not to migrate too fast, as compared with H_A and H_B. The above restrictions follow from the requirement of the smallness of perturbation $Q^{(1)}$ in equation (8.2.14). At $d_Q \to 0$, the solutions to (8.3.9) transform to (7.5.24) and (7.5.25).

(3) *The linear approximation* ($\Delta N_A << N_A$; $\Delta N_B << N_B$). With (8.2.19), we have the following expressions for the coefficients of equations (8.2.18):

$$D_{AA} = D_A \frac{K}{K + N_B(\infty)} + D_Q \frac{K}{K + N_B(\infty)}$$

$$D_{BB} = D_B \equiv d_{BB}^2, \qquad D_{AB} = 0 \tag{8.3.11}$$

$$D_{BA} = (D_Q - D_B) \frac{N_B(\infty)}{K + N_B(\infty)}.$$

Using formulas (7.5.7), we get

$$N_A = N_A(0) \operatorname{erfc} \lambda / d_{AA} \tag{8.3.12}$$

$$N_B = N_B(\infty) + \frac{N_A(0) D_{BA}}{d_{BB}^2 - d_{AA}^2}\left[\operatorname{erfc}\lambda/d_{BB} - \frac{d_{BB}}{d_{AA}}\operatorname{erfc}\lambda/d_{AA}\right].$$

The N_B distribution has an extremum: a minimum at $d_Q < d_B$ and a maximum at $d_Q > d_B$. The extremum coordinates are defined by the equation

$$\lambda_m^2 = \frac{d_{AA}^2 d_{BB}^2}{d_{AA}^2 - d_{BB}^2} \ln \frac{d_{AA}^2}{d_{BB}^2}.$$

The condition for the applicability of this approximation is

$$N_A(0) << \frac{d_{AA}(d_{AA} + d_{BB})}{|D_{BA}|} N_B(\infty), \qquad (8.3.13)$$

indicating the smallness of the boundary concentration $N_A(0)$.

(4) *The approximation of close diffusion coefficients.* Putting $d_A = d_B = d_Q = d$ in the set of equations (8.2.14), we find that the total profiles for the concentrations N_A and N_B are described by conventional diffusion equations. Their solution yields

$$N_A^0 = N_A(0)\operatorname{erfc}\lambda/d$$
$$N_B^0 = N_B(\infty). \qquad (8.3.14)$$

The concentration of complexes is found from the third equation of (8.2.14):

$$Q^{(0)} = \frac{1}{2}\left\{\sqrt{\left(N_A^{(0)} - N_B^{(0)}\right)^2 + 2K\left(N_A^{(0)} + N_B^{(0)}\right)} - \left(K + N_A^{(0)} + N_B^{(0)}\right)\right\}. \qquad (8.3.15)$$

The plots for the solutions are presented in Figure 8.6 showing the distribution profiles of free components at various complexation constants K.

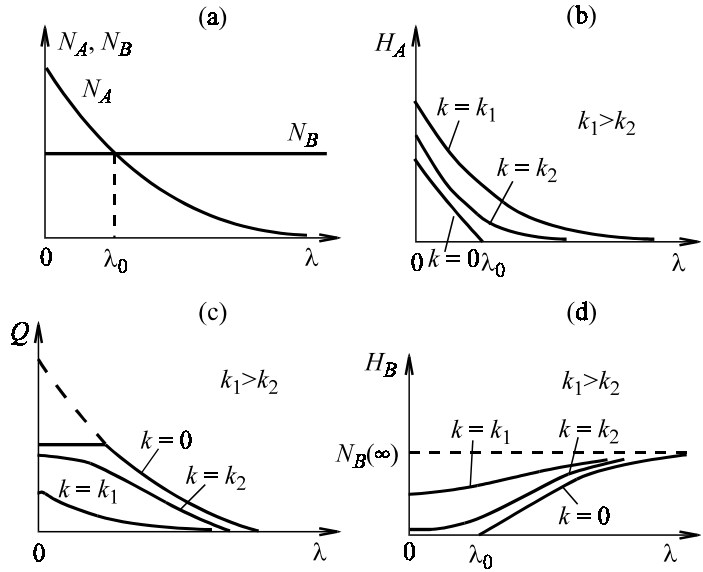

Figure 8.6. Distributions of components in sequential diffusion in the approximation of close diffusion coefficients.

8.3.2 Simultaneous diffusion

Let us consider the problem of diffusion of two components *A* and *B* migrating from fixed sources into a semi-infinite sample uniformly doped with component *B*. The boundary conditions are expressed by (8.2.16) at $H_A(\infty) = 0$. As was pointed out above, the boundary conditions for the set of equations (8.2.14) correspond to those for equations (8.2.1), when the boundary concentrations of, at least, two other components H_A, H_B, or Q are prescribed.

(1) *The strong complexation approximation* ($K = 0$). Suppose $N_A(0) < N_B(0)$, corresponding to the absence of the free component H_A at the sample boundary. The solution (Figure 8.7) is

$$N_A = Q = N_A(0)\operatorname{erfc}\lambda/d_Q = Q(0)\operatorname{erfc}\lambda/d_Q$$

$$H_B = \left[N_B(0) - N_B(\infty) - N_A(0)\right]\operatorname{erfc}\lambda/d_B + N_B(\infty) \qquad (8.3.16)$$

$$N_B = H_B + Q\,.$$

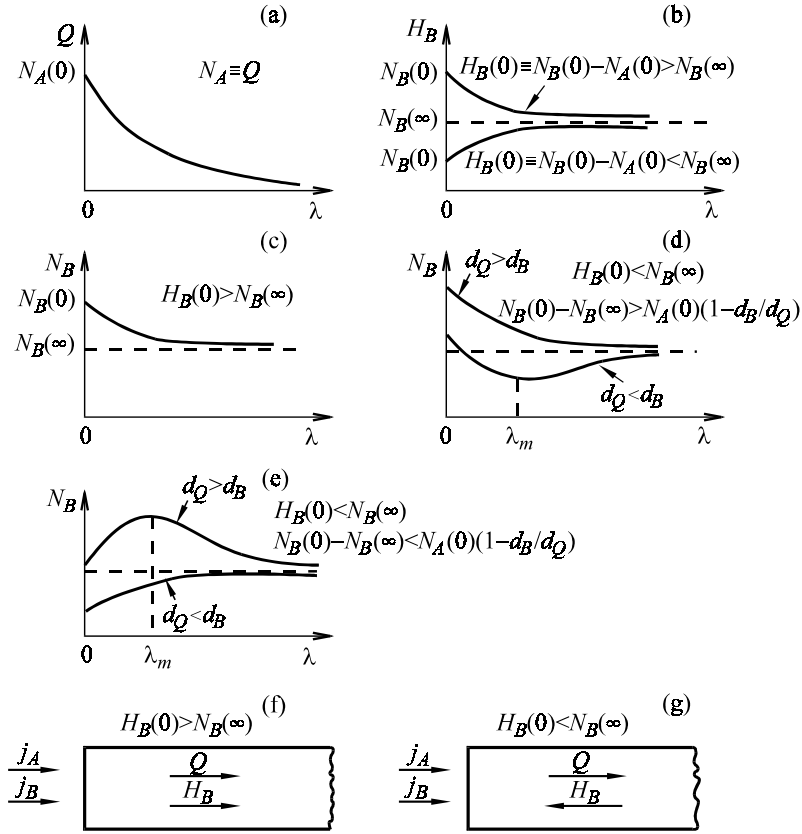

Figure 8.7. Distributions of components and simultaneous diffusion flows in the strong complexation approximation at $N_A(0) < N_B(0)$.

At $[N_B(0) - N_B(\infty) - N_A(0)] \geq 0$, the concentration of the free component at the sample boundary is higher than in the bulk, and the flow H_B migrates inward, together with the flow of complexes Q. For this reason, the N_B distribution falls monotonically with distance from the boundary. At $[N_B(0) - N_B(\infty) - N_A(0)] < 0$, the H_B flow migrates, in contrast, outward in the direction opposite to the Q flow, giving rise to an extremum in the H_B distribution. The extremum coordinate is

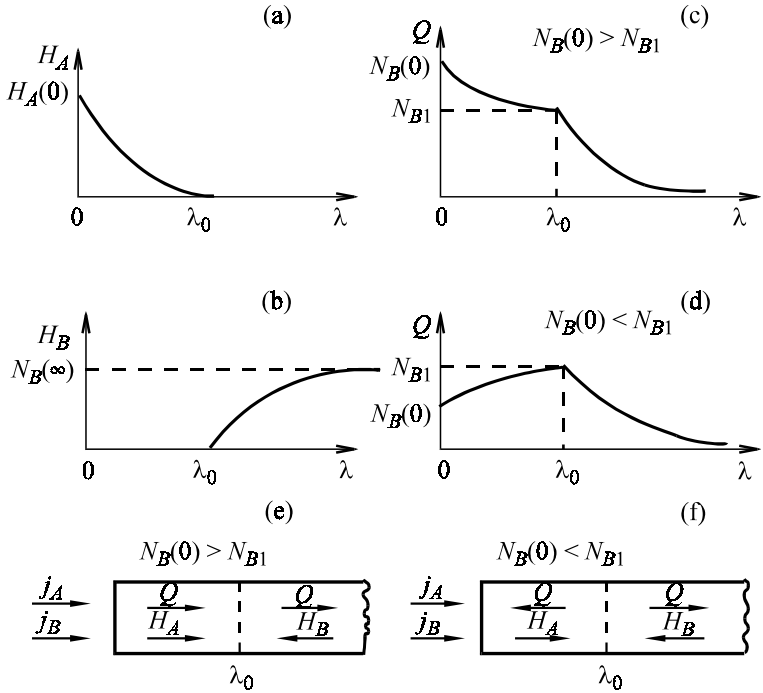

Figure 8.8. Distributions and flows of free components H_A, H_B, and Q in simultaneous diffusion at $N_A(0) > N_B(0)$.

$$\lambda_m^2 = \frac{d_B^2 d_Q^2}{d_Q^2 - d_B^2} \ln\left\{\frac{d_Q}{d_B}\left(1 - \frac{N_B(0) - N_B(\infty)}{N_A(0)}\right)\right\}.$$

Suppose $N_A(0) > N_B(0)$, i.e., there is the free component H_A on the sample surface. The solution to (8.2.14), illustrated in Figure 8.8, is

$$H_A \equiv H_A - Q = \begin{cases} \left[N_A(0) - N_B(0)\right]\left[1 - \dfrac{\operatorname{erfc}(\lambda/d_B)}{\operatorname{erfc}(\lambda/d_A)}\right], & \lambda \le \lambda_0 \\ 0, & \lambda > \lambda_0 \end{cases}$$

$$H_B \equiv H_B - Q = \begin{cases} 0, & \lambda \le \lambda_0 \\ N_B(\infty)\left[1 - \dfrac{\operatorname{erfc}(\lambda/d_B)}{\operatorname{erfc}(\lambda/d_A)}\right], & \lambda > \lambda_0 \end{cases} \tag{8.3.17}$$

$$Q = \begin{cases} N_B(0) + \left[N_{B1} - N_B(0)\right]\dfrac{\operatorname{erfc}(\lambda/d_B)}{\operatorname{erfc}(\lambda/d_A)}, & \lambda \le \lambda_0 \\ N_{B1}\dfrac{\operatorname{erfc}(\lambda/d_B)}{\operatorname{erfc}(\lambda/d_A)}, & \lambda > \lambda_0 \end{cases}$$

The parameters λ_0 and N_{B1} are found, as in Section 8.3.1, from the balance of total concentrations N_A and N_B or from equations (8.3.8)

$$\begin{aligned} &\left[N_A(0) - N_B(0)\right]S(\lambda_0/d_A) = N_B(\infty)T(\lambda_0/d_B) \\ &\left[N_{B1}(0) - N_B(0)\right]S(\lambda_0/d_Q) + N_{B1}T(\lambda/d_Q) = N_B(\infty)T(\lambda_0/d_B), \end{aligned} \tag{8.3.18}$$

where $S(x)$ and $T(x)$ are found from (8.3.7).

Figure 8.8 shows that the flows of free components H_A and H_B are directed toward the reaction front λ_0, as described in Section 8.3.1. In contrast to sequential diffusion, the Q flow can migrate, at $\lambda < \lambda_0$, either outward, with $N_{B1} > N_B(0)$, or inward, with $N_{B1} < N_{B0}$.

It is easy to obtain from (8.3.17) the distributions of total concentrations N_A and N_B (Figures 8.9 and 8.10). The analysis of the analytical expressions shows that the distribution N_B at $\lambda > \lambda_0$ depends on the sign of the expression $(N_{B1} - \tilde{N}_{B1})(d_B - d_Q)$, where

$$\tilde{N}_{B1} \equiv N_B(\infty)\frac{T(\lambda_0/d_B)}{T(\lambda_0/d_Q)}.$$

At $(N_{B1} - \tilde{N}_{B1})(d_B - d_Q) > 0$, an extremum appears in the N_B distribution at $\lambda > \lambda_0$ with the coordinate λ_{mB} (Figure 8.9*b*,*c*) defined as

$$\lambda_{mB}^2 = \lambda_0^2 + \frac{d_B^2 d_Q^2}{d_B^2 - d_Q^2}\ln\frac{N_{B1}}{\tilde{N}_{B1}}.$$

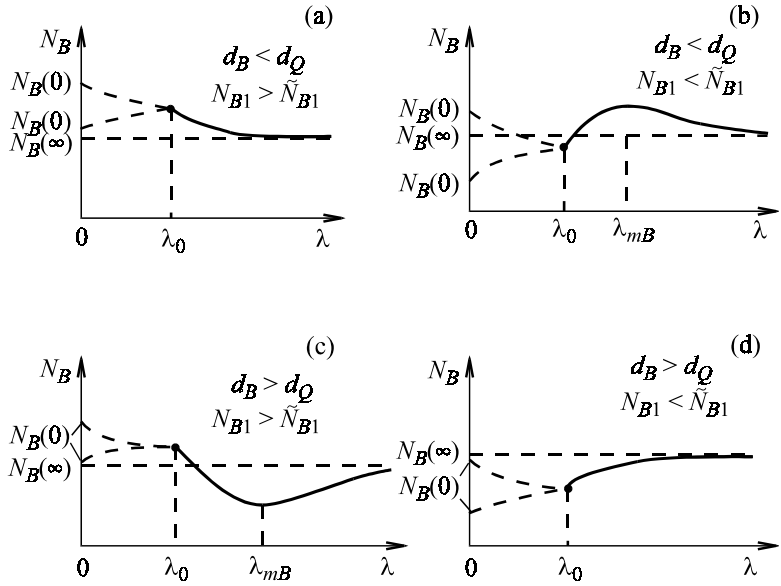

Figure 8.9. Possible distribution patterns for N_B in simultaneous diffusion in the strong complexation approximation at $N_A(0) > N_B(0)$.

With $(N_{B1} - \tilde{N}_{B1})(d_B - d_Q) < 0$, the function $N_B(\lambda)$ is monotonic at $\lambda > \lambda_0$ (Figure 8.9*a*,*d*).

The N_A distribution pattern at $\lambda < \lambda_0$ depends on the relations between the diffusion coefficients d_A^2 and d_Q^2 and between the quantities z, z_0, and z_{λ_0}:

$$z \equiv \frac{N_{B1} - N_B(0)}{N_A(0) - N_B(0)}, \qquad z_0 \equiv \frac{d_Q \operatorname{erf} \lambda_0/d_Q}{d_A \operatorname{erf} \lambda_0/d_A},$$

$$z_{\lambda_0} \equiv \frac{d_Q \operatorname{erf} \lambda_0/d_Q}{d_A \operatorname{erf} \lambda_0/d_A} \frac{\exp\left(\lambda_0^2/d_Q^2\right)}{\exp\left(\lambda_0^2/d_A^2\right)}.$$

When z falls within the range of z_0 and z_{λ_0}, the N_A distribution at $\lambda < \lambda_0$ has an extremum with the coordinate

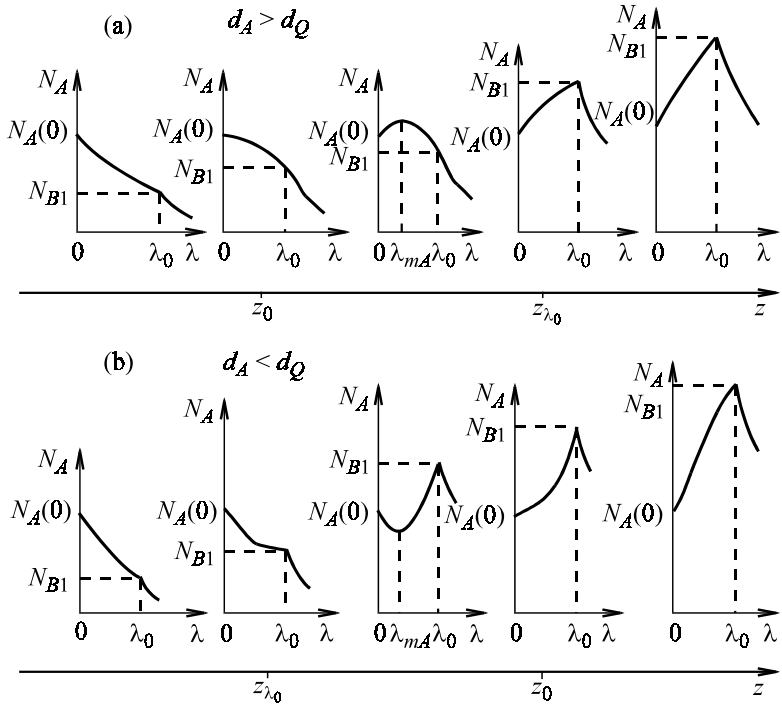

Figure 8.10. Possible distribution patterns for N_A in simultaneous diffusion in the strong complexation approximation at $N_A(0) > N_B(0)$.

$$\lambda_{mA}^2 = \lambda_0^2 + \frac{d_A^2 d_Q^2}{d_A^2 - d_Q^2} \ln \frac{z}{z_{\lambda_0}} = \frac{d_A^2 d_Q^2}{d_A^2 - d_Q^2} \ln \frac{z}{z_0},$$

which is a maximum at $d_A > d_Q$ and a minimum at $d_A < d_Q$.

The analysis of total concentration distributions of components in simultaneous diffusion shows that these distributions (Figures 8.9 and 8.10) reveal themselves even at relatively simple distributions of free components H_A and H_B and of complexes Q (Figure 8.8) and that they may be very diverse.

(2) *The weak complexation approximation* ($K > N_A$, N_B). As in Section 8.3.1, we have in the zero approximation:

$$N_A^{(0)} = N_A(0)\operatorname{erfc} \lambda / d_A$$

$$N_B^{(0)} = \left[N_B(0) - N_B(\infty)\right]\operatorname{erfc}\lambda/d_B + N_B(\infty) \qquad (8.3.19)$$

$$Q^{(0)} = 0.$$

To calculate the further approximations, additional restrictions should be imposed on the relations between the diffusion coefficients $d_A{}^2$ and $d_B{}^2$. Suppose $d_A << d_B$, then we have

$$Q^{(1)} = \frac{N_A(0)N_B(0)}{K}\operatorname{erfc}\lambda/d_A$$

$$N_A^{(0)} + N_A^{(1)} = N_A(0)\operatorname{erfc}\frac{\lambda}{\sqrt{d_A^2 + \left(d_Q^2 - d_A^2\right)\dfrac{N_B(0)}{K}}} \qquad (8.3.20)$$

$$N_B^{(0)} + N_B^{(1)} = \left[N_B(0) - N_B(\infty)\right]\operatorname{erfc}\lambda/d_B + \\ + N_B(\infty) + \frac{N_A(0)N_B(0)}{K}\frac{d_B^2 - d_Q^2}{d_B^2}\left(\operatorname{erfc}\lambda/d_B - \operatorname{erfc}\lambda/d_A\right).$$

At $d_A >> d_B$, we have

$$Q^{(1)} = \frac{N_A(0)\left\{N_B(0) - N_B(\infty)\right\}}{K}\operatorname{erfc}\lambda/d_B + \frac{N_A(0)N_B(\infty)}{K}\operatorname{erfc}\lambda/d_A$$

$$N_A^{(0)} + N_A^{(1)} = N_A(0)\operatorname{erfc}\frac{\lambda}{\sqrt{d_A^2 + \left(d_Q^2 - d_A^2\right)\dfrac{N_B(\infty)}{K}}} + \\ + \frac{N_A(0)\left\{N_B(0) - N_B(\infty)\right\}}{K}\left(1 - \frac{d_Q^2}{d_A^2}\right)\left(\operatorname{erfc}\lambda/d_B - \operatorname{erfc}\lambda/d_A\right) \qquad (8.3.21)$$

$$N_B^{(0)} + N_B^{(1)} = N_B(\infty) + \left[N_B(0) - N_B(\infty)\right]\operatorname{erfc}\frac{\lambda}{\sqrt{d_B^2 + \left(d_Q^2 - d_B^2\right)\dfrac{N_A(0)}{K}}} +$$

$$+ \frac{N_A(0)N_B(\infty)}{K}\frac{d_Q^2 - d_B^2}{d_A^2}\left(\operatorname{erfc}\lambda/d_A - \operatorname{erfc}\lambda/d_B\right)$$

At $d_A = d_B = d$, we obtain

$$Q^{(1)} = \frac{N_A(0)N_B(\infty)}{K}\operatorname{erfc}t + \frac{N_A(0)\left[N_B(0) - N_B(\infty)\right]}{K}\operatorname{erfc}^2 t$$

$$N_A^{(0)} + N_A^{(1)} = N_A(0)\operatorname{erfc}\frac{t}{\sqrt{1 + \left(\dfrac{d_Q^2}{d^2} - 1\right)\dfrac{N_B(\infty)}{K}}} +$$

$$+ \frac{N_A(0)\left\{N_B(0) - N_B(\infty)\right\}}{K}\left(1 - \frac{d_Q^2}{d^2}\right)\left(\operatorname{erfc}^2 t - \right. \qquad (8.3.22)$$

$$\left. - \frac{2}{\sqrt{\pi}}te^{-t^2}\operatorname{erfc}t + \frac{2}{\sqrt{\pi}}te^{-t^2} - \left(1 + \frac{2}{\pi}\right)\operatorname{erfc}t\right)$$

$$N_B^{(0)} + N_B^{(1)} = N_B(\infty) + \left[N_B(0) - N_B(\infty)\right]\operatorname{erfc}\frac{t}{\sqrt{1 + \left(1 - \dfrac{d_Q^2}{d^2}\right)\dfrac{N_A(0)N_B(\infty)}{K\left[N_B(\infty) - N_B(0)\right]}}} +$$

$$+ \frac{N_A(0)\left\{N_B(0) - N_B(\infty)\right\}}{K}\left(1 - \frac{d_Q^2}{d^2}\right)\left(\operatorname{erfc}^2 t - \right.$$

$$\left. - \frac{2}{\sqrt{\pi}}te^{-t^2}\operatorname{erfc}t + \frac{2}{\sqrt{\pi}}te^{-t^2} - \left(1 + \frac{2}{\pi}\right)\operatorname{erfc}t\right)$$

where $t \equiv \lambda/d$.

The above solutions are valid under the condition

$$\frac{K}{N_A}, \qquad \frac{K}{N_B} >> \left|1 - \frac{d_Q^2}{d^2}\right|.$$

(3) *The linear approximation* ($\Delta N_A << N_A$, $\Delta N_B << N_B$). According to (8.2.19), the diffusion coefficients D_{AA}, D_{AB}, D_{BA}, and D_{BB} are

$$D_{AA} = \frac{D_A K + D_Q N_B(\infty)}{K + N_B(\infty)} \equiv d_{AA}^2$$

$$D_{BB} = D_B \equiv d_{BB}^2$$

$$D_{AB} = 0$$

$$D_{BA} = (D_Q D_B)\frac{N_B(\infty)}{K + N_B(\infty)} \tag{8.3.23}$$

and the solutions for N_A and N_B have the form:

$$\begin{aligned} N_A &= N_A(0)\operatorname{erfc}\lambda/d_{AA} \\ N_B &= N_B(\infty) + \frac{1}{d_{BB}^2 - d_{AA}^2}\left[N_A(0)D_{BA}\operatorname{erfc}\lambda/d_{AA}\right] + \\ &\quad + \left\{\left(d_{AA}^2 - d_{BB}^2\right)\left[N_B(0) - N_B(\infty)\right] - N_A(0)D_{BA}\right\}\operatorname{erfc}\lambda/d_{BB} \end{aligned} \tag{8.3.24}$$

But at $({d_{AA}}^2 - {d_{BB}}^2)[N_B(0) - N_B(\infty)] - N_A(0)D_{BA} < 0$, the N_B distribution pattern shows an extremum with the coordinate

$$\lambda_m^2 = \frac{d_{AA}^2 d_{BB}^2}{d_{AA}^2 - d_{BB}^2} \ln\frac{d_{AA}}{d_{BB}}\left\{1 - \frac{\left(d_{AA}^2 - d_{BB}^2\right)\left[N_B(0) - N_B(\infty)\right]}{N_A(0)D_{BA}}\right\}.$$

A minimum is observed under the condition

$$\frac{N_A(0)D_{BA}}{d_{BB}^2 - d_{AA}^2}\left\{1 - \frac{d_{AA}}{d_{BB}}\left[1 - \frac{\left(d_{AA}^2 - d_{BB}^2\right)\left[N_B(0) - N_B(\infty)\right]}{N_A(0)D_{BA}}\right]\right\} > 0,$$

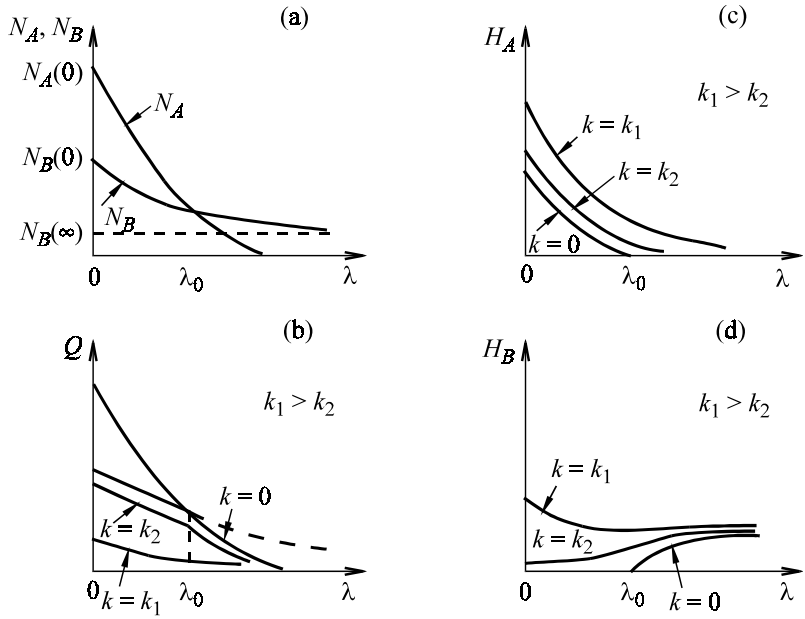

Figure 8.11. Distributions of components in simultaneous diffusion in the approximation of close diffusion coefficients at $N_A(0) > N_B(0)$.

otherwise, there is a maximum.

The applicability conditions for the linear approximation in this case are

$$\Delta N_B = \left|N_B(0) - N_B(\infty)\right| << N_B(\infty)$$
$$N_A(0) << N_B(\infty)\left|\frac{d_{AA}^2 - d_{BB}^2}{D_{BA}}\right|. \tag{8.3.25}$$

(4) *The approximation of equal diffusion coefficients*. Like in Section 8.3.1, we have

$$N_A^{(0)} = N_A(0)\operatorname{erfc}\lambda/d$$
$$N_B^{(0)} = \left[N_B(0) - N_B(\infty)\right]\operatorname{erfc}\frac{\lambda}{d} + N_B(\infty). \tag{8.3.26}$$

To calculate $Q^{(0)}$, one should use formula (8.3.15). The plots illustrating the solutions are given in Figure 8.11.

8.3.3 Interdiffusion

Consider the problem of outward interdiffusion from a semi-infinite sample uniformly doped with *A* and *B* impurities, respectively. The boundary conditions are obtained from (8.2.17) with $Q = 0$ at $t = 0$:

$$N_A\big|_{\lambda=-\infty} = N_A(-\infty), \qquad N_A\big|_{\lambda=+\infty} = 0$$
$$N_B\big|_{\lambda=-\infty} = 0, \qquad N_B\big|_{\lambda=+\infty} = N_B(\infty). \tag{8.3.27}$$

(1) *The strong complexation approximation* ($K = 0$). The solution to (8.2.14) was obtained in [7] and is illustrated in Figure 8.12. For the sake of definiteness, we will only consider the case with $N_A(-\infty)d_A > N_B(+\infty)d_B$. One can see from Figure 8.12 that the free components H_A and H_B migrate toward the reaction front, while complexes Q migrate away from the front into the sample bulk.

Similarly to simultaneous diffusion, the distribution patterns of total concentrations N_A and N_B (Figures 8.13 and 8.14) exhibit a considerable variation. The N_B distribution is defined by the relations between the diffusion coefficients d_B^2 and d_Q^2 and between the quantities Q_1 and $\tilde{Q}_{1B}$, where

$$\tilde{Q}_{1B} \equiv N_B(+\infty)\frac{d_Q \operatorname{erfc}\lambda/d_Q}{d_B \operatorname{erfc}\lambda/d_B}\exp\left[\lambda_0^2\left(\frac{1}{d_B^2}-\frac{1}{d_Q^2}\right)\right].$$

Under the condition $(Q_1 - \tilde{Q}_{1B})(d_B - d_Q) > 0$, the N_B distribution at $\lambda > \lambda_0$ shows an extremum with the coordinate λ_{mB}:

$$\lambda_{mB}^2 = \lambda_0^2 + \frac{d_B^2 d_Q^2}{d_B^2 - d_Q^2}\ln\frac{Q_1}{\tilde{Q}_{1B}}.$$

Under the condition $(Q_1 - \tilde{Q}_{1B})(d_B - d_Q) < 0$, the distribution $N_B(\lambda)$ at $\lambda > \lambda_0$ is monotonic.

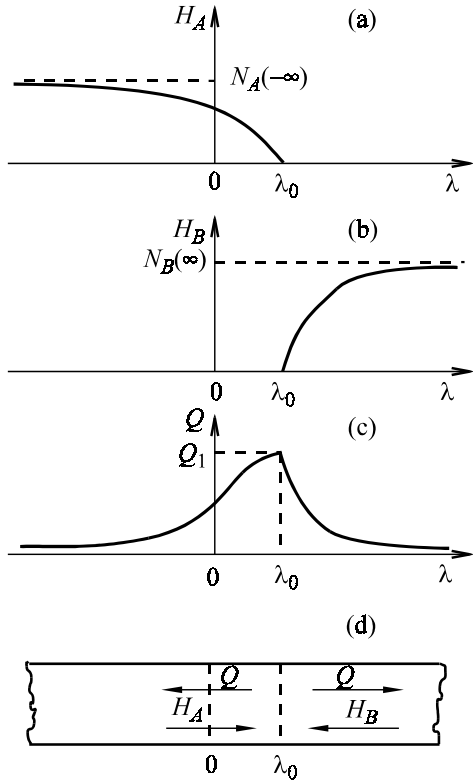

Figure 8.12. Distributions of free components H_A, H_B, and Q in interdiffusion in the strong complexation approximation at $d_A N_A(-\infty) > d_B N_B(+\infty)$.

The N_A distribution is defined by the relation between the diffusion coefficients d_A^2 and d_Q^2, as well as by the relation between Q_1 and $\tilde{Q}_{1A}$, where

$$\tilde{Q}_{1A} = N_A(-\infty)\frac{d_Q\left(1+\operatorname{erfc}\lambda/d_Q\right)}{d_A\left(1+\operatorname{erfc}\lambda/d_A\right)}\exp\left[\lambda_0\left(\frac{1}{d_Q^2}-\frac{1}{d_A^2}\right)\right]. \qquad (8.3.28)$$

If we have $(Q_1 - \tilde{Q}_{1A})(d_Q - d_A) > 0$, the N_A distribution at $\lambda < \lambda_0$ has two extrema with the coordinates $+\lambda_{mA}$ and $-\lambda_{mA}$. But at $(Q_1 - \tilde{Q}_{1A})(d_Q - d_A) < 0$,

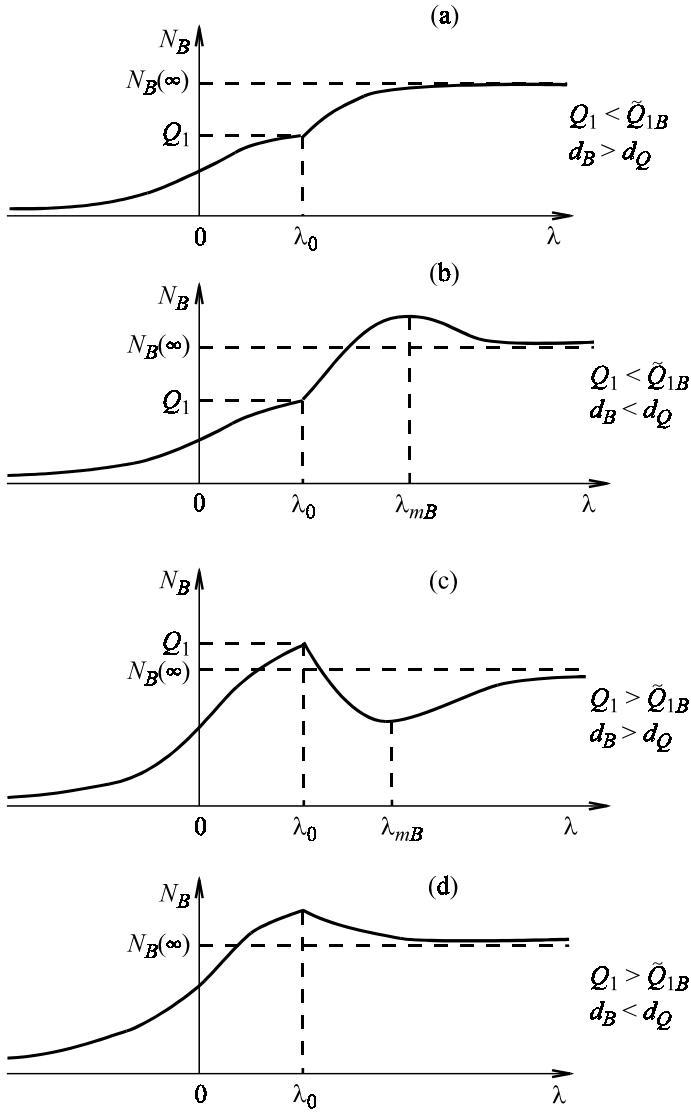

Figure 8.13. Possible distribution patterns of N_B in interdiffusion in the strong complexation approximation at $N_A(-\infty)d_A > N_B(+\infty)d_B$.

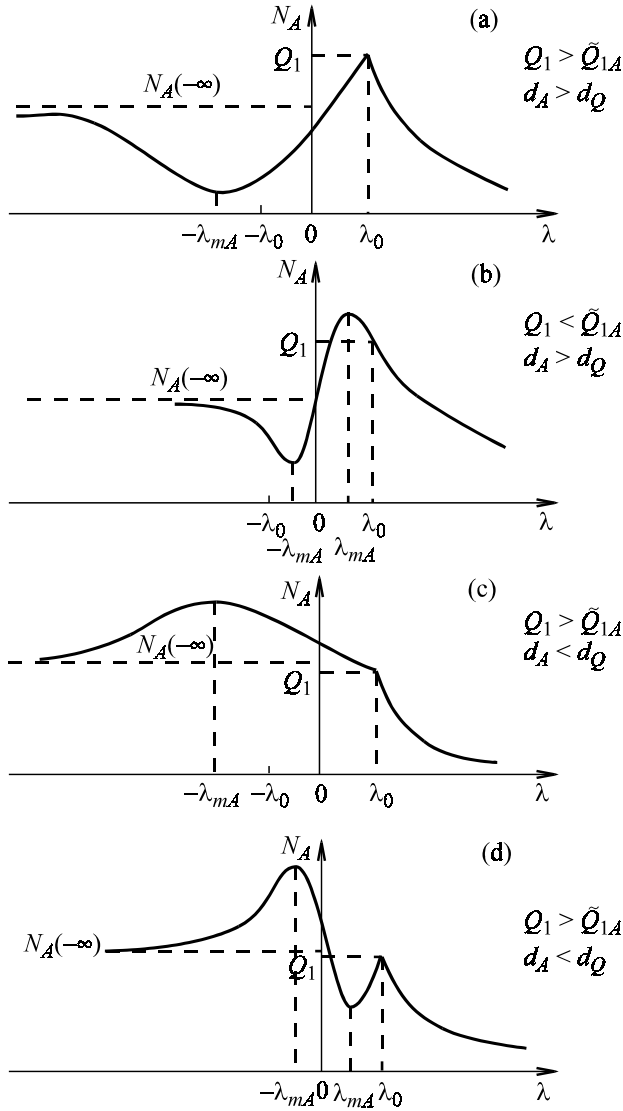

Figure 8.14. Possible distribution patterns of N_A in interdiffusion in the strong complexation approximation at $N_A(-\infty)d_A > N_B(+\infty)d_B$.

there is only one extremum with the coordinate $-\lambda_{mA}$, where λ_{mA} is

$$\lambda_{mA}^2 = \lambda_0^2 + \frac{d_A^2 d_Q^2}{d_A^2 - d_Q^2} \ln \frac{Q_1}{\widetilde{Q}_{1A}}.$$

(2) *The weak complexation approximation* ($K >> N_A, N_B$). At $K \to \infty$,

$$N_A^{(0)} = \frac{1}{2} N_A(-\infty)\left(1 - \operatorname{erfc} \lambda / d_A\right)$$
$$N_B^{(0)} = \frac{1}{2} N_B(+\infty)\left(1 + \operatorname{erfc} \lambda / d_B\right). \tag{8.3.29}$$

For the first order calculations in the perturbation theory, let us discuss the situations with $d_A << d_B$ and $d_A = d_B$ individually. At $d_A << d_B$, we have

$$Q^{(1)} = \frac{N_A(-\infty) N_B(+\infty)}{4K}\left(1 - \operatorname{erfc} \lambda / d_A\right)\left(1 + \operatorname{erfc} \lambda / d_B\right)$$

$$N_A^{(1)} = \begin{cases} \dfrac{N_A(-\infty) N_B(+\infty)}{2K}\left(\dfrac{d_Q^2 - d_A^2}{d_B^2}\right)\left(1 + \operatorname{erfc} \lambda / d_B\right) & \lambda << d_A \\ \dfrac{N_A(-\infty) N_B(+\infty)}{4K}\left(1 - \dfrac{d_Q^2}{d_B^2}\right)\left\{-\dfrac{\lambda}{d_A \sqrt{\pi}} \exp\left(-\dfrac{\lambda^2}{d_A^2}\right) + \right. & \lambda >> d_B \\ \left. + \dfrac{d_A^2}{d_B^2}\left(\operatorname{erfc} \lambda / d_A - 1\right)\right\} & \end{cases} \tag{8.3.30}$$

$$N_B^{(1)} = \begin{cases} \dfrac{N_A(-\infty) N_B(+\infty)}{2K}\left(1 - \dfrac{d_Q^2}{d_B^2}\right)\left\{\dfrac{\lambda}{d_B \sqrt{\pi}} \exp\left(-\dfrac{\lambda^2}{d_B^2}\right) + \right. & \\ \left. + \left(1 + \operatorname{erfc} \lambda / d_B\right)\right\} & \lambda << -d_A \\ \dfrac{N_A(-\infty) N_B(+\infty)}{4K}\left(1 - \dfrac{d_Q^2}{d_B^2}\right)\left(1 - \operatorname{erfc} \lambda / d_A\right) & \lambda >> -d_B \end{cases} \tag{8.3.31}$$

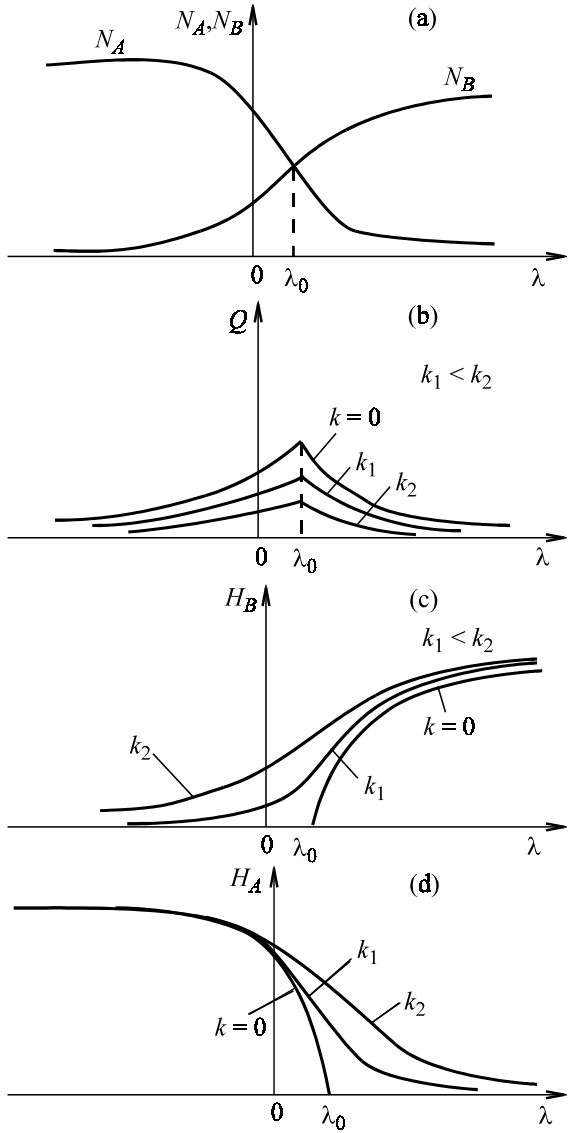

Figure 8.15. Distributions of components in interdiffusion in the approximation of equal diffusion coefficients.

At $d_A = d_B \equiv d$, we have

$$Q^{(1)} = \frac{N_A(-\infty)N_B(+\infty)}{4K}\left(1 - \mathrm{erfc}^2\,\lambda/d\right)$$

$$N_A^{(1)} = \frac{N_A(-\infty)N_B(+\infty)}{4K}\left(1 - \frac{d_Q^2}{d^2}\right)\left(1 - \mathrm{erf}^2\,\lambda/d\right) -$$

$$- \frac{2}{\sqrt{\pi}}\frac{\lambda}{d}\exp\left(-\frac{\lambda^2}{d^2}\right)\mathrm{erf}\,\lambda/d - \frac{2}{\sqrt{\pi}}\frac{\lambda}{d}\exp\left(-\frac{\lambda^2}{d^2}\right)$$

$$N_B^{(1)} = N_A^{(1)}.$$

(3) *The approximation of equal diffusion coefficients*. As in Section 8.3.1, we have

$$N_A^{(0)} = \frac{1}{2}N_A(-\infty)\left(1 - \mathrm{erfc}\,\lambda/d\right)$$

$$N_B^{(0)} = \frac{1}{2}N_B(+\infty)\left(1 - \mathrm{erfc}\,\lambda/d\right). \qquad (8.3.32)$$

To calculate Q^0, one should use formula (8.3.10). The plots illustrating the solutions are shown in Figure 8.15.

8.3.4 The allowance for the finite front thickness

In the above strong complexation approximation ($K = 0$) for diffusion involving the reaction $A + B \rightleftarrows Q$, the chemical reaction front had a zero thickness in the limit $K \to 0$. At small but finite K, the front thickness has a finite value. Consider an immobile complex Q. The initial equations have the same form as in (8.2.13).

For the sake of definiteness, let us discuss sequential diffusion with the initial and boundary conditions of (8.2.15). Considering the finite front thickness, let us subdivide the λ-axis into three regions:

(I) the near-surface region with $\lambda \le \lambda_0$;
(II) the complexation front region with $\lambda \approx \lambda_0$;
(III) the external region with $\lambda \ge \lambda_0$.

Taking into account $K << H_A(0)$ and $H_B(\infty)$, we obtain from (8.2.13) and (8.2.15) the following solutions for the subsurface region:

$$H_A^I = H_A(0) - M\operatorname{erf}\lambda/d_A\,, \qquad H_B^{II} = \frac{KQ^I}{H_A^I}\,, \qquad Q^I = \text{const} \tag{8.3.33}$$

$$H_A^{III} = 0\,, \qquad H_B^{III} = H_B(\infty) - N\operatorname{erfc}\lambda/d_B\,, \quad Q^{III} = 0\,, \tag{8.3.34}$$

where M and N are constants which will be defined later from the sewing conditions for the solutions in the various regions.

To calculate the component distributions at the reaction front, we will introduce, like in [13], the following denotations:

$$\begin{aligned} H_A^{II} &= K^{1/2}\overline{H}_A(\eta) + O\left(K^{1/2}\right) \\ H_B^{II} &= K^{1/2}\overline{H}_B(\eta) + O\left(K^{1/2}\right) \\ Q^{II} &= K^{1/2}\overline{H}_A(\eta)\overline{H}_B(\eta) + O\left(K^{1/2}\right), \end{aligned} \tag{8.3.35}$$

where $\eta = (\lambda - \lambda_0)/K^{1/2}$.

From (8.3.18), we obtain the equations for $\overline{H}_A(\eta)$ and $\overline{H}_B(\eta)$:

$$\begin{aligned} d_A^2\overline{H}_A'' + 2\lambda_0\left(\overline{H}_A\overline{H}_B\right)' &= 0 \\ d_B^2\overline{H}_B'' + 2\lambda_0\left(\overline{H}_A\overline{H}_B\right)' &= 0\,, \end{aligned} \tag{8.3.36}$$

where the differentiation was made with respect to η. By integrating (8.3.36), we obtain

$$\begin{cases} d_A^2\overline{H}_A' + 2\lambda_0\overline{H}_A\overline{H}_B = \alpha \\ d_B^2\overline{H}_B' + 2\lambda_0\overline{H}_A\overline{H}_B = \beta \end{cases}. \tag{8.3.37}$$

The sewing conditions for regions I and III yield

$$\overline{H}_B(\eta) \to 0\,, \qquad \overline{H}_A\overline{H}_B \to Q' \qquad \text{at} \quad \eta \to -\infty \tag{8.3.38}$$

$$\overline{H}_{\mathrm{A}}(\eta) \to 0\,, \qquad \overline{H}_{\mathrm{A}}\overline{H}_{\mathrm{B}} \to 0 \qquad \text{at} \quad \eta \to +\infty\,.$$

Hence, we obtain with (8.3.37)

$$\alpha = 0\,, \qquad Q' = \beta/2\lambda_0\,. \tag{8.3.39}$$

After the subtraction of the second equation in (8.3.37) from the first one, followed by integration, we obtain

$$-d_{\mathrm{A}}^2\overline{H}_{\mathrm{A}} + d_{\mathrm{B}}^2\overline{H}_{\mathrm{B}} = \beta\eta + \gamma K^{-1/2}\,, \tag{8.3.40}$$

where $\gamma K^{1/2}$ is an integration constant. Using (8.3.38), we can obtain from (8.3.40)

$$\begin{aligned} &\eta \to -\infty\,, \qquad \overline{H}_{\mathrm{A}} \to -\frac{\beta}{d_{\mathrm{A}}^2}\eta - \frac{\gamma}{d_{\mathrm{A}}^2}K^{-1/2} \\ &\eta \to +\infty\,, \qquad \overline{H}_{\mathrm{B}} \to -\frac{\beta}{d_{\mathrm{B}}^2}\eta + \frac{\gamma}{d_{\mathrm{B}}^2}K^{-1/2}\,. \end{aligned} \tag{8.3.41}$$

Using (8.3.41) and the solution expansions in regions I and III near λ_0, we perform the standard joining of the internal and external variables to obtain

$$\begin{aligned} &\beta = \frac{2}{\sqrt{\pi}}d_{\mathrm{A}}M\exp\left(-\lambda_0^2/d_{\mathrm{A}}^2\right) = \frac{2}{\sqrt{\pi}}d_{\mathrm{B}}N\exp\left(-\lambda_0^2/d_{\mathrm{B}}^2\right) \\ &\gamma = d_{\mathrm{B}}^2\left[H_{\mathrm{B}}(\infty) - N\mathrm{erfc}(\lambda_0/d_{\mathrm{B}})\right] = d_{\mathrm{A}}^2\left[-H_{\mathrm{A}}(0) + M\mathrm{erfc}\,\lambda_0/d_{\mathrm{A}}\right]. \end{aligned} \tag{8.3.42}$$

Taking into account that λ_0 is being evaluated in the zero approximation, we have

$$H_{\mathrm{A}}(0)S(\lambda_0/d_{\mathrm{A}}) = H_{\mathrm{B}}(\infty)T(\lambda_0/d_{\mathrm{B}})\,, \qquad \gamma = 0$$

$$M = \frac{H_{\mathrm{A}}(0)}{\mathrm{erfc}(\lambda_0/d_{\mathrm{A}})}\,, \qquad N = \frac{H_{\mathrm{B}}(\infty)}{\mathrm{erfc}(\lambda_0/d_{\mathrm{B}})}\,, \qquad \beta = 2\lambda_0 N_{\mathrm{B1}}\,.$$

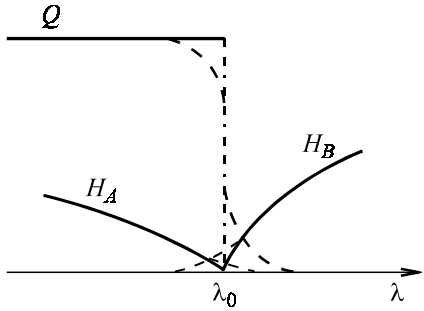

Figure 8.16. The distribution of components H_A, H_B, and Q at the reaction front for an immobile complex: dashed line – with the account of finite front thickness.

For convenience, let us transform the set of equations (8.3.37) as

$$\dot{X} + 2XY = 0\,, \qquad \dot{Y} + 2XY = 1\,, \tag{8.3.43}$$

where

$$X = \overline{H}_A/\omega_A\,, \quad Y = \overline{H}_B/\omega_B\,, \quad \eta = \omega_\eta t\,, \quad \omega_A = \frac{d_B}{d_A}\sqrt{2N_{B1}}$$

$$\omega_B = \frac{d_A}{d_B}\sqrt{2N_{B1}}\,, \qquad \omega_\eta = \frac{d_A d_B}{\lambda_0\sqrt{2N_{B1}}}$$

and carry out the differentiation in (8.3.43) with respect to t.

Let us find the solution to the set of equations (8.3.43) satisfying the boundary conditions

$$t \to -\infty, \quad X \to -t, \quad Y \to 0,$$

$$t \to +\infty, \quad X \to 0, \quad Y \to t.$$

The desired solutions are

$$X = \frac{\exp\left(-t^2\right)}{\sqrt{\pi}\left(1+\operatorname{erf} t\right)}\,, \qquad Y = \frac{\exp\left(-t^2\right)}{\sqrt{\pi}\left(1+\operatorname{erf} t\right)} + t\,. \tag{8.3.44}$$

Therefore, the solution for the reaction region has the form (Figure 8.16):

$$H_{\mathrm{A}}^{\mathrm{II}} = \frac{d_{\mathrm{B}}}{d_{\mathrm{A}}}\sqrt{2KN_{\mathrm{B1}}}\,X(t)$$
$$H_{\mathrm{B}}^{\mathrm{II}} = \frac{d_{\mathrm{A}}}{d_{\mathrm{B}}}\sqrt{2KN_{\mathrm{B1}}}\,Y(t) \qquad (8.3.45)$$
$$Q^{\mathrm{II}} = 2N_{\mathrm{B1}}X(t)Y(t),$$

where $t = [(\lambda - \lambda_0)\lambda_0/d_{\mathrm{A}}d_{\mathrm{B}}](2N_{\mathrm{B1}}/\mathrm{K})^{1/2}$.

The reaction front thickness is

$$\Delta\lambda_{\mathrm{F}} \approx \frac{d_{\mathrm{A}}d_{\mathrm{B}}}{\lambda_0^2}\sqrt{\frac{K}{N_{\mathrm{B1}}}}, \quad \text{or} \qquad \frac{K}{N_{\mathrm{B}}} << \frac{\lambda_0^4}{(d_{\mathrm{A}}d_{\mathrm{B}})^2}. \qquad (8.3.46)$$

The concentration of mobile complexes in region II is approximately constant, so the set of equations (8.3.36) is replaced by

$$\begin{cases} d_{\mathrm{A}}^2\overline{H}_{\mathrm{A}}'' - d_{\mathrm{B}}^2\overline{H}_{\mathrm{B}}'' = 0 \\ \overline{H}_{\mathrm{A}}\overline{H}_{\mathrm{B}} = Q(\lambda_0). \end{cases} \qquad (8.3.47)$$

By solving this set of equations in the same way as (8.3.36), we obtain

$$\overline{H}_{\mathrm{A}}^{\mathrm{II}} = \frac{K^{1/2}\beta_{\eta_0}}{d_{\mathrm{A}}^2}\left\{\sqrt{1+\left(\frac{\eta}{2\eta_0}\right)^2} + \frac{\eta}{2\eta_0}\right\}$$
$$\overline{H}_{\mathrm{B}}^{\mathrm{II}} = \frac{K^{1/2}\beta_{\eta_0}}{d_{\mathrm{B}}^2}\left\{\sqrt{1+\left(\frac{\eta}{2\eta_0}\right)^2} + \frac{\eta}{2\eta_0}\right\}$$
$$\frac{\beta}{2\lambda_0} = H_{\mathrm{A}}(0)S\left(\frac{\lambda_0}{d_{\mathrm{A}}}\right) = H_{\mathrm{B}}(\infty)T\left(\frac{\lambda_0}{d_{\mathrm{B}}}\right) = Q(\lambda_0)S\left(\frac{\lambda_0}{d_{\mathrm{Q}}}\right)$$
$$\eta_0^2 = \frac{Q(\lambda_0)d_{\mathrm{A}}^2d_{\mathrm{B}}^2}{\beta^2}, \qquad \eta = \frac{\lambda - \lambda_0}{K^{1/2}}.$$

The reaction region thickness is found from the condition $\eta \approx 2\eta_0$; hence,

$$\Delta\lambda_F = 2\sqrt{\frac{K}{Q(\lambda_0)}}\,\frac{d_A d_B}{\lambda_0 T\left(\lambda_0/d_Q\right)}. \qquad (8.3.48)$$

The condition for the existence of the reaction front is

$$\frac{\Delta\lambda_F}{\lambda_0} << 1. \qquad (8.3.49)$$

The critical value K_{cr} separating the regions of diffusion parameters, at which the reaction front can be formed, is described as

$$K_{cr} = Q(\lambda_0)\frac{\lambda_0^4}{4\left(d_A d_B\right)^2}T^2\left(\frac{\lambda_0}{d_Q}\right). \qquad (8.3.50)$$

To conclude, the above treatment can be easily extended to arbitrary boundary conditions (in simultaneous diffusion or interdiffusion). The principal result of these calculations for experimental profile analysis is focused in expressions (8.3.46) and (8.3.50) for immobile ($d_Q = 0$) and mobile ($d_Q \neq 0$) complexes, respectively. The condition necessary for the reaction front to be formed is $K << K_{cr}$.

The approach to the analysis of complex impurity profiles, developed in this chapter with reference to the reaction $A + B \rightleftarrows Q$, can be extended to an arbitrary number of diffusing components and quasichemical interactions involving them. An essential aspect is that this approach allows for the migration at times much longer than the characteristic times of establishing an equilibrium in quasichemical reactions. This permits diffusion interaction to be taken into account via the active mass law.

Similarly to the approach discussed above, we suggest the following treatment strategy for impurity profiles, making allowance for quasichemical interactions of diffusing components. The initial information must include possible types of reactions among particles. As in Section 8.2, one derives a set of equations to describe a consistent diffusion of particles at times much longer than the characteristic reaction times. Then, definite initial and boundary conditions are prescribed for the solution of the set of equations. Assuming that the reactions are sharply shifted to the right or to the left and using an exhaustive search, one obtains all possible distribution patterns for the prescribed initial and boundary conditions. Note that if there are several diffusing components and quasichemical reactions, the transition from total

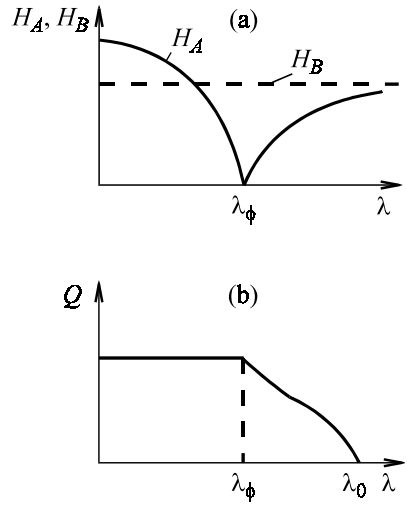

Figure 8.17. Qualitative impurity profiles of H_A and H_B (a) and of Q (b) for sequential diffusion, with account taken of the concentration dependence of the diffusion coefficients.

concentrations of particles to free particles and back is guaranteed by the theorem proved in [14].

The diffusion theory discussed in this chapter deals with constant diffusion coefficients independent of particle concentrations. The theory was extended to this case, too, in [15]. For the mathematical details, the reader is referred to this original work. Here, we will present only its qualitative results for the distributions H_A, H_B, and Q, illustrated in Figure 8.17. Qualitatively, they look like those for constant diffusion coefficients (see Figure 8.15), but there are some differences. Because of the finite values of $d(H_A^m)/d\lambda$ and $d(H_B^m)/d\lambda$ at the reaction front, the distributions H_A and H_B intercept the λ-axis at the right angle [$H_A \sim (\lambda_F - \lambda)^{1/m}$, $H_B \sim (\lambda - \lambda_F)^{1/n}$ at $\lambda \to \lambda_F$].

Here and below, the constants m, n, and q denote the concentration dependence of diffusion of the components A, B, and Q, respectively. The concentration dependence of the diffusion coefficient and the absence of Q at $\lambda \to \infty$ give rise to the front λ_0 in the distribution of Q behind the reaction front. One can see, therefore, that if the diffusion coefficient of one of the components depends on its concentration, there may appear a front in its distribution. The necessary condition for the front appearance is the absence of the respective component at $\lambda \to \infty$. A specific feature of such a front is its

existence in the distribution of only one component, in which case $d(Q^q)/d\lambda \to 0$ and $d(Q^{q-1})/d\lambda \to$ const at $\lambda \to \lambda_0$.

By plotting experimental impurity profiles for the front region in the coordinates $\ln|H - H_F|$, $\ln|\lambda - \lambda_F|$, and $[H_F \equiv H(\lambda_F)]$ along the slope attained by the straight line, one can calculate the values of m, n, and q for the front of a chemical reaction and that related to the concentration dependence of the diffusion coefficient. The treatment of impurity profiles for sequential diffusion to be presented in the next section can be easily extended to more complex diffusion mechanisms.

8.3.5 The physics of impurity diffusion with interactions

The solutions obtained in Section 8.3.4 demonstrate that the strong compensation approximation is the most informative and physically explainable approach. One can see from the solutions obtained in this approximation that the complicated profiles of total components N_A and N_B arise from the superposition of profiles of the free components H_A and H_B and of complexes Q.

The solutions of Section 8.3 show that the allowance for the mobility of complexes reveals a new important feature in the concentration profiles, namely, the possible appearance of a total concentration distribution maximum (Figures 8.3–8.5, 8.7–8.10, 8.13, and 8.14). The interpretation of extrema and bendings in the concentration profiles has been refined.

The figures just mentioned show that the reaction front is generally characterized by a bending in the total concentration distribution. The appearance of a maximum or a minimum indicates the presence of counterpropagating flows containing one kind of impurity (see, for example, Figure 8.3*f*). The directions of flows under various boundary conditions are illustrated in Figure 8.18. A common feature of all types of diffusion is the presence of H_A and H_B flows toward the λ_0 boundary and the formation of complexes there. Therefore, λ_0 can be interpreted (in the strong complexation approximation) as the complexation front. Note that the minimum in the distribution of complete component N_B in the case of an immobile complex coincides with the reaction front λ_0 (7.5.22), but this coincidence does not exist for a mobile complex.

The formation of the chemical reaction front is due to the possible subdivision of a sample into two regions, in which only one of the two free components, H_A or H_B, can migrate. For this reason, the reaction front arises only under boundary conditions permitting the existence of these two components in two different regions. If the boundary conditions do not meet this requirement, no reaction front is formed (see, for example, Figure 8.3).

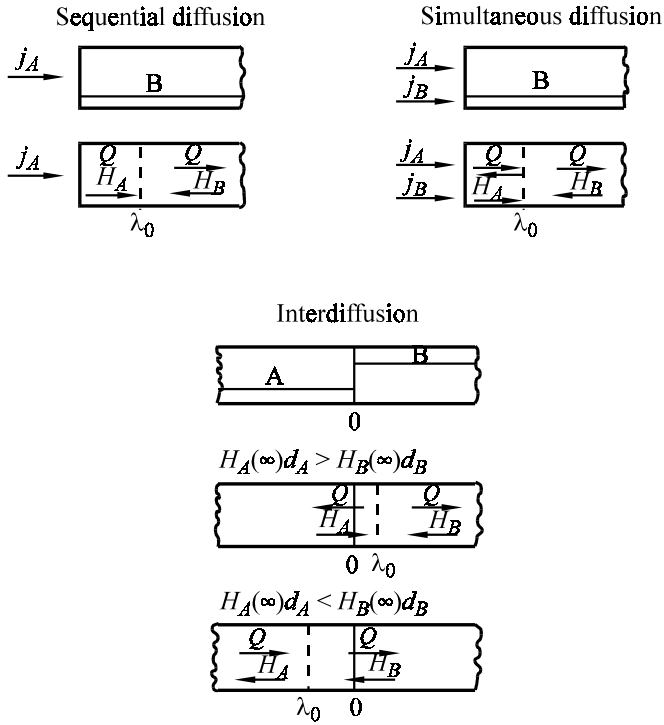

Figure 8.18. Schematic flows in various types of diffusion in the strong complexation approximation.

The comparison of solutions obtained in the strong and weak complexation approximations shows that in a weak complexation, the $Q^{(1)}$ flow and the related $H_A^{(1)}$ and $H_B^{(1)}$ flows are small, as compared with the $H_A^{(0)}$ and $H_B^{(0)}$ flows, so the features distinct in the strong complexation approximation are quite weak here. The $Q^{(1)}$, $H_A^{(1)}$, and $H_B^{(1)}$ flows indicate the tendency in the variation of concentration profiles and directions of flows, which are well-defined in a strong complexation.

There are only two flows, $H_A^{(0)}$ and $H_B^{(0)}$, in the zero approximation of a weak complexation. So, in the next approximation, which is practically a linear combination of these flows, only one extremum is possible. This is also characteristic of the linear approximation.

The solutions obtained in the approximation of close diffusion coefficients (Figures 8.6; 8.11; 8.15) reveal qualitative changes in the concentration profiles on transition from large complexation constants (weak complexation) to small ones (strong complexation) with arbitrary diffusion coef-

ficients D_A, D_B, and D_Q. Numerical estimations of the conditions for the appearance of a reaction front can be easily made using formulas (8.3.46) and (8.3.50).

8.4 DIFFUSION THEORY VERSUS EXPERIMENT

When the approach described above is applied to experimental data treatment, it is first necessary to make sure that the experimental profiles are nonstationary. In other words, the impurity profiles obtained at different times and plotted in the coordinates of concentration–Boltzmann variable $x/(2t^{1/2})$ must coincide with one another. To facilitate the interpretation of impurity distribution patterns, diffusing components should be measured individually and direct methods for concentration recalculation should be employed.

For a comparison with measured profiles, it is natural to use distributions of total concentrations N_A and N_B. To analyze the sources of specific features, it will be more convenient to turn to the free components H_A and H_B, as well as to complexes Q. Note that the well-known effect of solubility increase in complexation [16] is screened due to the prescribed boundary conditions for total components. The transition to the free components at the boundary makes this effect explicit.

8.4.1 Chemical diffusion of phosphorus into silicon

The distribution of phosphorus during its diffusion into silicon was studied experimentally in [17–20]. The following basic features of the impurity profile were revealed.

(1) The profile has an unsteady state diffusion-like pattern clearly seen in Figure 8.19 borrowed from [20].

(2) Phosphorus profiles are described by the standard function (7.5.2) at low surface concentrations. When these concentrations are above 10^{25} m^{-3}, there are considerable deviations from (7.5.2), and the deviation becomes greater with increasing phosphorus concentration, as is clear from Figure 8.20. At concentrations exceeding 3.5×10^{26} m^{-3}, dislocations and dislocation networks are formed at the surface.

(3) There are two specific features of electrically active phosphorus at

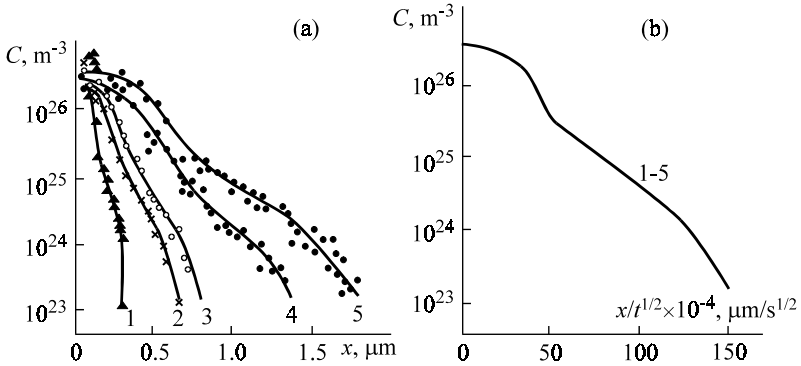

Figure 8.19. The distribution of phosphorus after diffusion at 1173 K for the times, min: 1 – 10; 2 – 30; 3 – 60; 4 – 120; 5 – 240; (a) in the conventional coordinates; (b) in the c = $x/t^{1/2}$ coordinates.

diffusion temperatures $T \le 1373$ K and surface concentrations 10^{25} m^{-3} ÷ 5×10^{26} m^{-3}: a shoulder in the subsurface region and a tail extending into the sample bulk (Figure 8.20).

(4) These specific features are smoothed out at diffusion temperatures $T \ge 1373$ K; however, the surface profile is not described by expression (7.5.2).

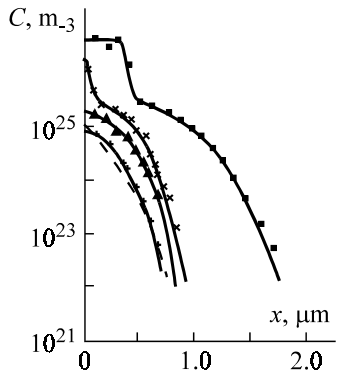

Figure 8.20. Phosphorus profiles in silicon at various surface concentrations after diffusion at 1373 K for 60 min; points – electrical measurements [18]; dashed line – function (7.5.2).

A qualitative model was suggested in [17] to explain the above features of phosphorus profiles. It was assumed that phosphorus migrates in silicon by two mechanisms—via interstitial migration and via mobile *E*-centers. Using this model and the approach discussed in Section 8.3.5, we will analyze possible distribution patterns of components during phosphorus migration.

The quasichemical interaction reactions involving three phosphorus states—at a site, at an interstice, and at an *E*-center—are

$$P_i + V \overset{\rightarrow}{\leftarrow} P_s, \tag{8.4.1}$$

$$P_s + V \overset{\rightarrow}{\leftarrow} E, \tag{8.4.2}$$

where P_i, P_s, V, and E are standard designations for atomic phosphorus at sites and interstices, for a vacancy, and for a vacancy-P atom complex, respectively.

The equations for consistent diffusion of components at times larger than the characteristic reaction times are derived exactly as in Section 8.2:

$$\begin{aligned} &d_V^2 V'' + d_E^2 E'' - d_i^2 P_i'' + 2\lambda\left(V + E - P_i\right)' = 0 \\ &d_E^2 E'' + d_i^2 P_i'' + 2\lambda\left(P_s + E + P_i\right)' = 0 \\ &P_i V = K_I P_s; \qquad P_s V = K_{II} E, \end{aligned} \tag{8.4.3}$$

where d_V^2 is the diffusion coefficient of vacancies, d_i^2 is the diffusion coefficient of interstitial P atoms, d_E^2 is the diffusion coefficient of an *E*-center, K_I is an equilibrium constant of reaction (8.4.1), and K_{II} is an equilibrium constant of reaction (8.4.2).

The initial and boundary conditions will be

$$V\big|_{\lambda=\infty} = V(\infty); \qquad P_i\big|_{\lambda=\infty} = P_s\big|_{\lambda=\infty} = E\big|_{\lambda=\infty} = 0 \tag{8.4.4}$$

$$P_i\big|_{\lambda=0} = P_i(0); \qquad P_s\big|_{\lambda=0} = P_s(0). \tag{8.4.5}$$

Following the procedure described in Section 8.3.5, let us calculate these initial and boundary conditions for the distribution of components when the reactions (8.4.1) and (8.4.2) shift to the right or to the left, using the set of equations (8.4.3).

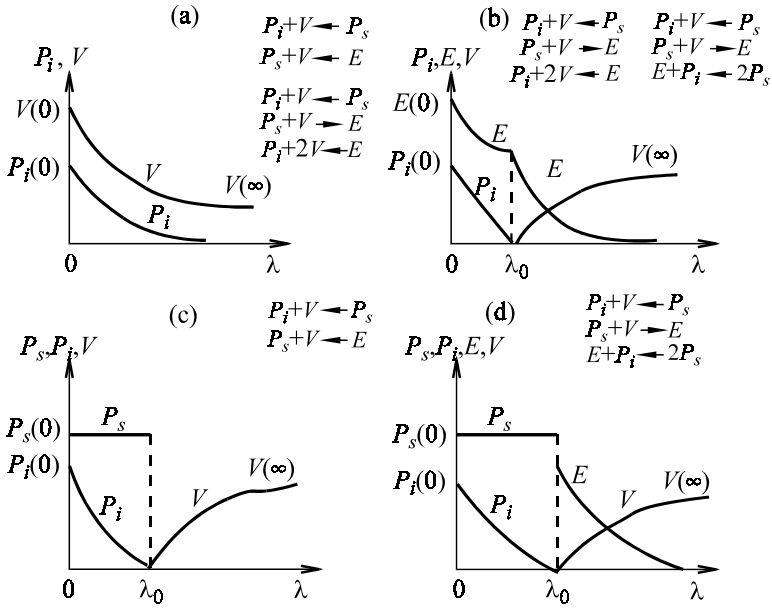

Figure 8.21. Possible distribution patterns of components after chemical diffusion of phosphorus into silicon.

(1) Reactions (8.4.1) and (8.4.2) are shifted to the right, toward P_s and E decomposition:

$$P_i + V \leftarrow P_s, \qquad P_s + V \rightarrow E;$$

or

$$K_I >> P_i, V \qquad K_{II} << P_s, V.$$

The set of equations (8.4.3) transforms to

$$\begin{aligned} &P_s = 0, \qquad E = 0, \\ &d_V^2 V'' + 2\lambda V' = 0, \\ &d_i^2 P_i'' + 2\lambda P_i' = 0. \end{aligned} \tag{8.4.6}$$

The solutions to this set of equations are illustrated qualitatively in Figure 8.21*a*. Vacancies and interstitial atoms migrate independently, without interacting with one another.

(2) Reaction (8.4.1) is shifted to the left and reaction (8.4.2) to the right, toward the formation of an *E*-center:

$$P_i + V \leftarrow P_s, \qquad P_s + V \rightarrow E,$$

or,

$$K_I >> P_i, V, \qquad K_{II} << P_s, V.$$

The set of equations (8.4.3) transforms to

$$\begin{aligned} &P_s = 0, \qquad P_s V = 0, \\ &d_V^2 V'' + d_E^2 E'' - d_i^2 P_i'' + 2\lambda\left(V + E - P_i\right)' = 0, \\ &d_E^2 E'' + d_i^2 P_i'' + 2\lambda\left(E + P_i\right)' = 0. \end{aligned} \tag{8.4.7}$$

This set of equations does not permit an unambiguous identification of concentration profiles at the initial and boundary conditions of (8.4.3) and (8.4.4). The missing information can be derived from the analysis of the total reaction

$$P_i + 2V \underset{\leftarrow}{\rightarrow} E, \tag{8.4.8}$$

for which the active mass law is $P_i V^2 = K_I K_{II} E$. One can see that the restrictions imposed on K_I and K_{II} are insufficient to identify the direction of shift of reaction (8.4.8). Some additional conditions are necessary.

If reaction (8.4.8) is shifted sharply to the left:

$$P_i + 2V \leftarrow E, \qquad \text{or} \qquad K_I K_{II} >> P_i V, V^2,$$

equations (8.4.7) transform to (8.4.6), and their solution corresponds to Figure 8.21*a*.

If reaction (8.4.8) is shifted sharply to the right:

$$P_i + 2V \rightarrow E, \qquad \text{or} \qquad K_I K_{II} << P_i V, V^2,$$

the set of equations (8.4.7) transforms to

$$\begin{aligned} &P_s = 0, \qquad P_s V = 0, \quad P_i V^2 = 0, \\ &d_V^2 V'' + d_E^2 E'' - d_i^2 P_i'' + 2\lambda\left(V + E - P_i\right)' = 0, \\ &d_E^2 E'' + d_i^2 P_i'' + 2\lambda\left(E + P_i\right)' = 0. \end{aligned} \tag{8.4.9}$$

The solution to (8.4.9) is illustrated in Figure 8.21*b* showing the formation of the reaction front. One can see that interstitial P atoms migrate to the reaction front from the surface and vacancies migrate to it from the crystal bulk. The mobile *E*-centers produced at the front migrate inward.

(3) Reaction (8.4.1) is shifted sharply to the right and reaction (8.4.8) to the left:

$$P_s + V \leftarrow E, \qquad \text{or} \qquad K_I << P_i, V$$

$$P_i + V \rightarrow P_s, \qquad \text{or} \qquad K_{II} >> P_s, V.$$

The set of equations (8.4.3) transforms to

$$\begin{aligned} &E = 0, \qquad P_i V = 0 \\ &d_V^2 V'' - d_i^2 P_i'' + 2\lambda\left(V - P_i\right)' = 0 \\ &d_i^2 P_i'' + 2\lambda\left(P_s + P_i\right)' = 0 . \end{aligned} \tag{8.4.10}$$

The solutions are illustrated qualitatively in Figure 8.21*c*. Vacancies and interstitial P atoms migrate toward the reaction front where reaction (8.4.1) representing the capture of P atoms by vacancies occurs.

(4) Reactions (8.4.1) and (8.4.2) are shifted sharply to the right:

$$P_i + V \rightarrow P_s, \qquad P_s + V \rightarrow E,$$

or

$$K_I << P_i, V, \qquad K_{II} << P_s, V.$$

Equations (8.4.3) transform to

$$\begin{aligned} &P_i V = 0, \qquad P_s V = 0, \\ &d_V^2 V'' + d_E^2 E'' - d_i^2 P_i'' + 2\lambda\left(V + E - P_i\right)' = 0 , \\ &d_E^2 E'' + d_i^2 P_i'' + 2\lambda\left(P_s + V + P_i\right)' = 0 . \end{aligned} \tag{8.4.11}$$

As in case (2), the information for an unambiguous identification of the distribution pattern is obtained from the analysis of the total reaction

$$E + P_i \underset{\leftarrow}{\rightarrow} 2P_s. \tag{8.4.12}$$

The active mass law for this is

$$EP_i = \frac{K_I}{K_{II}} P_s^2 .$$

Suppose reaction (8.4.12) is shifted sharply to the left:

$$E + P_i \leftarrow 2P_s, \qquad \text{or} \qquad \frac{K_I}{K_{II}} >> 1 .$$

Then, the set of equations (8.4.12) transforms to (8.4.11), and the distribution pattern changes to the one shown in Figure 8.21*b*.

If reaction (8.4.12) is shifted sharply to the right:

$$E + P_i \rightarrow 2P_s, \qquad \text{or} \qquad K_I << K_{II},$$

equations (8.4.11) take the form

$$\begin{aligned} &P_i V = 0, \qquad P_s V = 0, \qquad EP_i = 0 \\ &d_V^2 V'' + d_E^2 E'' - d_i^2 P_i'' + 2\lambda\left(V + E - P_i\right)' = 0 \\ &d_E^2 E'' + d_i^2 P_i'' + 2\lambda\left(P_s + V + P_i\right)' = 0 . \end{aligned} \qquad (8.4.13)$$

The solutions are illustrated qualitatively in Figure 8.21*d*.

A specific feature of the distributions thus obtained is that two reactions occur at the front. This is because of the assumption of immobile site phosphorus, $d_s = 0$. If we assume $d_s \neq 0$, this will lead to two reaction fronts (Figure 8.22). The front closest to the sample surface is for the formation of site phosphorus via the reaction $E + P_i \rightarrow 2P_s$, and the other front is for the formation of *E*-centers via the reaction $P_s + V \rightarrow E$. In the case of $d_s \rightarrow 0$, both fronts merge to yield the distribution patterns shown in Figure 8.21*d*.

The comparison of experimental (Figures 8.19 and 8.20) and theoretical (Figure 8.21) distributions shows that the characteristic features of the phosphorus profile in silicon are reproduced only in case (4), under the condition $E + P_i \rightarrow 2P_s$ (Figure 8.23). The formulas for component distributions in this case are

$$P_i = \begin{cases} P_i(0)\left(1 - \dfrac{\operatorname{erfc} \lambda/d_i}{\operatorname{erfc} \lambda_0/d_i}\right) & \lambda \leq \lambda_0 \\ 0 & \lambda > \lambda_0 \end{cases}$$

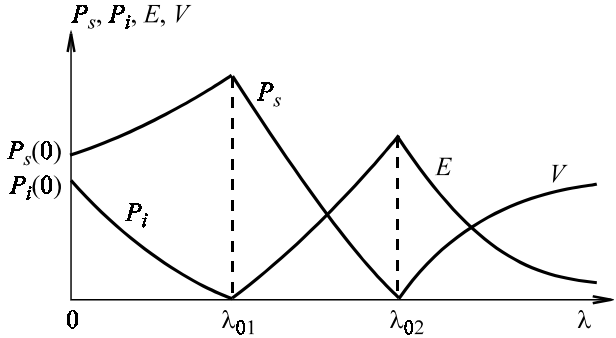

Figure 8.22. The distribution of components in P diffusion into silicon on the assumption of $d_s \neq 0$ under the following conditions: $P_i + V \rightarrow P_s$; $P_s + V \rightarrow E$; $P_i + E \rightarrow 2P_s$.

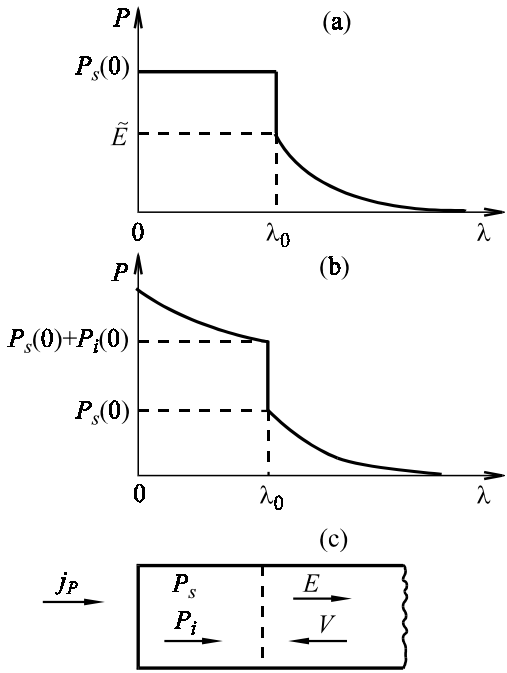

Figure 8.23. Distributions of electrically active phosphorus (a), its total concentration (b), and component flows (c) after chemical diffusion of phosphorus into silicon.

$$P_s = \begin{cases} P_s(0) & \lambda \le \lambda_0 \\ 0 & \lambda > \lambda_0 \end{cases}$$

$$E = \begin{cases} 0 & \lambda \le \lambda_0 \\ \tilde{E}\,\dfrac{\operatorname{erfc}\lambda/d_E}{\operatorname{erfc}\lambda_0/d_E} & \lambda > \lambda_0 \end{cases}$$

$$V = \begin{cases} 0 & \lambda < \lambda_0 \\ V(\infty)\left(1 - \dfrac{\operatorname{erfc}\lambda/d_V}{\operatorname{erfc}\lambda_0/d_V}\right) & \lambda > \lambda_0 \end{cases}. \qquad (8.4.14)$$

The constants λ_0 and $\tilde{E}$ are found from the set of equations derived from the balance conditions of the type of (8.3.8) at the reaction front:

$$\begin{aligned} \tilde{E} &= P_i(0)S(\lambda_0/d_i) + \tilde{E}T(\lambda_0/d_E) - V(\infty)T(\lambda_0/d_V) \\ P_s(0) &= 2P_i(0)S(\lambda_0/d_i) - V(\infty)T(\lambda_0/d_V). \end{aligned} \qquad (8.4.15)$$

Using the interpretation of the characteristic points of the P level, indicated in Figure 8.18, we can obtain from Figure 8.20

$$P_s(0) = 5\times10^{26}\ \mathrm{m}^{-3}, \qquad \lambda_0 = 4\times10^{-9}\ \mathrm{m/s}^{1/2},$$
$$\tilde{E} = 4\times10^{25}\ \mathrm{m}^{-3}, \qquad d_E^2 \equiv D_E = 5\times10^{-17}\ \mathrm{m}^2/\mathrm{s}.$$

In order to make numerical estimations, let us transform (8.4.15) by removing the term $P_i(0)S(\lambda_0/d_i)$. Multiplying the first equation in (8.4.9) by 2 and subtracting the result from the second equation, we find

$$V(\infty)T(\lambda_0/d_V) = P_s(0) + 2\tilde{E}\left[T(\lambda_0/d_E) - 1\right]. \qquad (8.4.16)$$

Taking into account $\lambda_0/d_E \sim 1$ and $\tilde{E} << P_s(0)$ from (8.4.16), we get the relation between the vacancy concentration in the sample bulk $V(\infty)$ and the diffusion coefficient of vacancies $D_V \equiv d_V^2$: $V(\infty)T(\lambda_0/d_V) = P_s(0)$.

By varying the diffusion coefficient D_V in the range 10^{-9}–10^{-14} m^2/s, we obtain the concentration values of vacancies, $V(\infty)$:

D_V, m^2/s	10^{-9}	10^{-10}	10^{-12}	10^{-14}
$V(\infty)$	1×10^{23}	3×10^{23}	3×10^{24}	3×10^{25}

These relatively large concentrations $V(\infty)$ indicate that the diffusion of phosphorus in silicon seems to involve vacancies bound into complexes, rather than free vacancies. Moreover, when vacancies are substituted by any other point defect containing no P atom in reactions (8.4.1) and (8.4.2), all the properties of impurity phosphorus described above remain unchanged.

8.4.2 Radiation-stimulated P diffusion into uniformly O_2-doped silicon

Phosphorus distribution in radiation-stimulated diffusion into silicon was studied in [21] at various oxygen contents in the initial samples. Diffusion was stimulated by bombardment with argon ions which were absorbed by a thin (~0.01 μm) surface layer. It was suggested in the analysis of post-radiation profiles that the diffusion involved vacancy–oxygen and vacancy–phosphorus complexes.

Let us analyze the impurity profiles obtained in [21] and shown in Figure 8.24 on the assumption that they are unstationary and diffusion-like profiles.

The interaction reactions between impurities and vacancies are

$$P_i + V \overset{\rightarrow}{\leftarrow} P_s, \qquad P_s + V \overset{\rightarrow}{\leftarrow} E, \qquad Q_s + V \overset{\rightarrow}{\leftarrow} Q, \qquad (8.4.17)$$

where Q is an oxygen–vacancy complex.

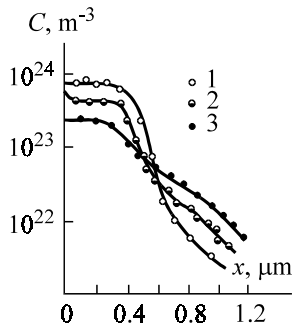

Figure 8.24. Phosporus distribution due to stimulated diffusion at 873 K as a function of oxygen content in silicon [21], N×10^{-22} m^{-3}: 1 – 0.12; 2 – 2.3; 3 – 6.8.

The equations for consistent diffusion are

$$
\begin{aligned}
&d_V^2 V'' + d_Q^2 Q'' + d_E^2 E'' - d_i^2 P_i'' + 2\lambda\left(V + Q + E - P_i\right)' = 0 \\
&d_Q^2 Q'' + 2\lambda\left(Q + Q_s\right)' = 0 \\
&d_E^2 E'' + d_i^2 P_i'' + 2\lambda\left(E + P_s\right)' = 0 \\
&P_i V = K_I P_s, \quad P_s V = K_{II} E, \quad Q_s V = K_{III} Q .
\end{aligned}
\tag{8.4.18}
$$

where Q_s is the diffusion coefficient of oxygen atoms and d_Q^2 is the diffusion coefficient of a complex.

The boundary conditions are written as

$$
\begin{gathered}
P_i\big|_{\lambda=0} = P_i(0), \qquad P_s\big|_{\lambda=0} = P_s(0), \qquad Q_s\big|_{\lambda=0} = Q_s(0) \\
P_i\big|_{\lambda=\infty} = P_s\big|_{\lambda=\infty} = 0, \qquad Q\big|_{\lambda=\infty} = Q(\infty).
\end{gathered}
\tag{8.4.19}
$$

The solution to the set of equations (8.4.18) under the boundary and initial conditions given in (8.4.19) and the comparison with experimental distributions presented in Figure 8.24 has shown [8] that an agreement between experimental and theoretical profiles is possible only if the three reactions (8.4.17) are shifted sharply to the right. The distribution of components for this case is illustrated in Figure 8.25 and described by the formulas

$$
P_i = \begin{cases} P_i(0)\left(1 - \dfrac{\operatorname{erfc}\lambda/d_i}{\operatorname{erfc}\lambda_0/d_i}\right) & \lambda \le \lambda_0 \\ 0 & \lambda > \lambda_0 \end{cases}
$$

$$
Q = \begin{cases} 0 & \lambda < \lambda_0 \\ Q(\infty)\left(1 - \dfrac{\operatorname{erfc}\lambda/d_Q}{\operatorname{erfc}\lambda_0/d_Q}\right) & \lambda > \lambda_0 \end{cases}
$$

$$
E = \begin{cases} 0 & \lambda < \lambda_0 \\ \tilde{E}\,\dfrac{\operatorname{erfc}\lambda/d_E}{\operatorname{erfc}\lambda_0/d_E} & \lambda > \lambda_0 \end{cases}
$$

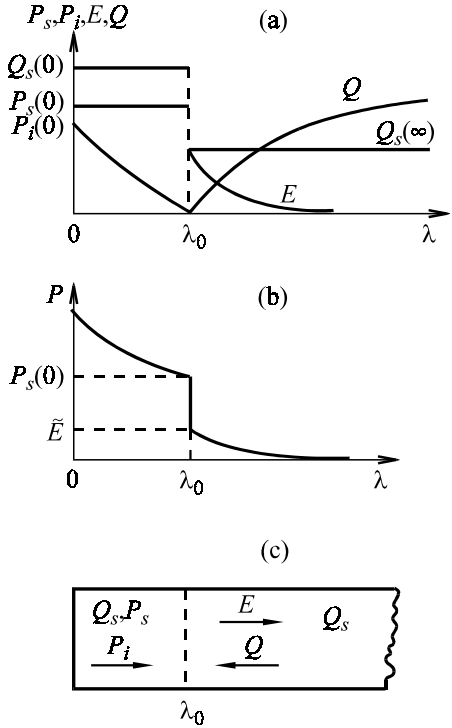

Figure 8.25. Theoretical distribution of components in stimulated diffusion of phosphorus into silicon.

$$Q_s = \begin{cases} Q_s(0) & \lambda < \lambda_0 \\ Q_s(\infty) & \lambda > \lambda_0 \end{cases}. \tag{8.4.20}$$

The constants λ_0, $\tilde{E}$, and $Q_s(\infty)$ are defined by the equations

$$\begin{aligned} Q_s(0) &= Q_s(\infty) + Q(\infty)T\left(\lambda_0/d_Q\right) \\ \tilde{E} &= Q(\infty)T\left(\lambda_0/d_Q\right) - \tilde{E}T\left(\lambda_0/d_E\right) - P_i(0)S\left(\lambda_0/d_i\right) \\ P_s(0) &= 2P_i(0)S\left(\lambda/d_i\right) - Q(\infty)T\left(\lambda_0/d_Q\right). \end{aligned} \tag{8.4.21}$$

Table 8.1. Model parameter values.

Profile number (Figure 8.24)	$Q(\infty) + Q_s(\infty)$, m^{-3}	$P_s(\infty)$, m^{-3}	E, m^{-3}	λ_0, $m/s^{1/2}$	d_E, $m/s^{1/2}$
1	0.12×10^{22}	8.3×10^{23}	1×10^{23}		
2	2.3×10^{22}	4.8×10^{23}	4×10^{23}	5×10^{-9}	5×10^{-9}
3	6.8×10^{22}	2.6×10^{23}	7×10^{23}		

The distributions described by (8.4.20) show that an interstitial P flow migrates from the sample surface to the reaction front λ_0 and a flow of Q complexes moves from the sample bulk outward. There are two reactions occurring at the front: $P_i + E \rightarrow 2P_s$ and $P_s + Q \rightarrow Q + E$. The E-centers produced at the front migrate inward. Vacancies are intermediate reaction products here. Comparing Figures 8.23 and 8.25, one can see that Q complexes act as vacancies in the latter case.

Let us make some numerical estimations using the theoretical results. The experimental profiles yield the following values for the model parameters (Table 8.1).

Note that the D_E value for radiation-stimulated diffusion at 873 K is approximately equal to that for chemical diffusion at 1273 K (Section 8.4.1).

Using (8.4.21), let us estimate the diffusion coefficient of a Q complex. By multiplying the second equation of (8.4.15) by 2 and summing it up with the third equation, we obtain

$$P_s(0) + 2\tilde{E}\left[1 + T\left(\lambda_0/d_E\right)\right] = Q(\infty)T\left(\lambda_0/d_E\right). \qquad (8.4.22)$$

Hence, with the assumption that $Q(\infty) \le Q(\infty) + Q_s$ and using the values of Table 8.1, we find $D_Q \sim 6\times10^{-13}$ m^2/s.

From the second equation in (8.4.21) and (8.4.22), the relation between $P_i(0)$ and d_i is

$$P_i(0)S\left(\lambda_0/d_i\right) = P_s(0) + \tilde{E}\left[1 + T\left(\lambda_0/d_E\right)\right]. \qquad (8.4.23)$$

When $P_i(0)$ varies within 10^{20}–10^{23} m^{-3}, d_i^2 varies within 10^{-13}–10^{-16} m^2/s. It also follows from equation (8.4.21) that the increase in $Q(\infty)$ due to a higher oxygen content corresponds to a higher value of $\tilde{E}$ at constant λ_0 and $P_i(0)$; this is what is, in fact, observed in experimental profiles.

8.4.3 Fe redistribution in B and P diffusion-doped silicon

Section 6.8 considered an indirect redistribution of impurity atoms over crystallochemical positions, mediated by the crystal electronic subsystem. But there is also a direct redistribution via a chemical interaction of migrating atoms. This effect was studied in [22] with reference to phosphorus and boron diffusion into silicon pre-doped with iron. The diffusion temperature for both phosphorus and boron was 1273 K and the diffusion times were 18 hours and 3–5 hours, respectively.

The impurity distribution patterns are shown in Figure 7.5. The characteristic features of the Fe profiles are the minima and maxima arising after the diffusion of boron and phosphorus. The non-monotonic character of Fe distribution was interpreted in [22] as being due to the formation of complexes involving P, B, and Fe. However, because of the theory limitations at that time, the maxima in Fe distribution remained unclear. Assuming the Fe distribution feature to be associated with quasichemical reactions among the diffusing components, the authors of [8] suggested the following interactions of Fe and P impurities in silicon:

$$\begin{aligned} &\mathrm{P_i} + V \overset{\rightarrow}{\leftarrow} \mathrm{P_s} \\ &\mathrm{P_s} + V \overset{\rightarrow}{\leftarrow} E \\ &\mathrm{Fe_i} + V \overset{\rightarrow}{\leftarrow} \mathrm{Fe_s} \\ &\mathrm{Fe_i} + \mathrm{P_i} \overset{\rightarrow}{\leftarrow} \left[\mathrm{Fe_i P_s}\right] \\ &\mathrm{Fe_i} + E \overset{\rightarrow}{\leftarrow} \left[\mathrm{Fe_i} E\right]. \end{aligned} \tag{8.4.24}$$

Bearing in mind that the diffusion mechanisms of B and P in silicon are similar, we believe that this model is also valid for the B–Fe–Si system.

Further, assuming the complexes $[\mathrm{Fe_iP_s}]$, $[\mathrm{Fe_i}E]$, and $\mathrm{Fe_s}$ to have a low mobility and taking into account the high migration rate of interstitial Fe, we have concluded that Fe decorates $\mathrm{P_i}$, E, and V involved in P diffusion. Indeed, a comparison of the Fe distribution (Figure 7.5*a*) and the distributions of components in P diffusion (Figure 8.21*f*) allows the following conclusion to be made. The maximum in the distribution of Fe is due to its decoration of E-centers. The descending region in front of the maximum indicates the decoration of P site atoms. The specific features of the Fe profile in B diffusion can be interpreted in a similar way.

It should be noted that the lower rate of P diffusion into a Fe-doped sample (Figure 7.5*b*) can be naturally accounted for, as in [22], by the binding of some P atoms to Fe atoms to produce complexes.

REFERENCES

8.1. G.P. Voevoda, I.G. Kirnis, P.G. Litovchenko *et al.*, *Vliyanie termo-obrabotki na glubokie urovni v kremnii i germanii p-tipa* (Thermal Treatment Effect on Deep Levels in *p*-Si and *p*-Ge). Kiev: AN USSR. Preprint No. 15, 12 p. (1982) (in Russian).

8.2. V.L. Vinetzky, S.A. Kholodar, *Statisticheskoe vzaimodeistvie elektronov i defektov v poluprovodnikakh* (Statistical Interaction of Electrons and Defects in Semiconductors). Kiev: Naukova dumka, 187 p. (1969) (in Russian).

8.3. N.P. Morosov, D.I. Telelbaum, *Phys. St. Sol., A* **51**: 624–640 (1979).

8.4. S.V. Bulyarsky, V.I. Fistul, *Termodinamika i kinetika vzaimodeistvuyushchikh defektov v poluprovodnikakh* (Thermodynamics and Kinetics of Interacting Defects in Semiconductors). Moscow: Nauka, 352 p. (1997) (in Russian).

8.5. V.S. Vavilov, V.F. Khisimov, B.N. Makushev, *Defekty v kremnii i na ego poverkhnosti* (Defects in and on Silicon). Moscow: Nauka, 211 p. (1990) (in Russian).

8.6. V.I. Fistul, M.I. Sinder, *FTP* **17**, No. 11: 1195–2002 (1983) (in Russian).

8.7. V.I. Fistul, M.I. Sinder, *FTP* **17**, No 11: 2003–2008 (1983) (in Russian).

8.8. V.I. Fistul, M.I. Sinder, *FTP* **18**, No. 5: 797–801 (1984) (in Russian).

8.9. V.V. Vaskin, V.A. Uskov, *FTP* **10**, No 4: 1239–1241 (1975) (in Russian).

8.10. I.M. Lifshitz, A.M. Kosevich, Ya.E. Geguzin, In: *Poverkhnostnaya diffuziya i rastekanie* (Surface Diffusion and Migration). Moscow: Nauka, 243–263 (1969) (in Russian).

8.11. V.V. Vaskin, V.A. Uskov, *FTT* **11**, No. 7: 1763–1769 (1969) (in Russian).

8.12. G.V. Shcherbedinsky, *Izuchenie vzaimodeistviya ugleroda s legiruyushimi elementami v troinykh sistemakh* (Carbon Interaction with Doping Elements in Ternary Systems). Avtoref. dokt. diss. Moscow: TZNIICHERMET, 33 p. (1973) (in Russian).

8.13. F.E. Fendel, *J. Fluid Mech.* **21**, No. 2: 281–304 (1965).

8.14. Ya.B. Zeldovich, *JFKh* **11**, No. 5: 685–687 (1938) (in Russian).

8.15. M.I. Sinder, *FTP* **20**, No. 2: 373–375 (1986) (in Russian).

8.16. M.I. Sinder, *FTP* **20**, No.2: 304–308 (1986) (in Russian).

8.17. D. Shaw, ed., *Atomic Diffusion in Semiconductors*. London–New York: Plenum Press, 684 p. (1973).

8.18. A.F. Willoughby, *Rep. Progr. Phys.* **41**: 1665–1705 (1978).

8.19. R.B. Fair, J.C. Tsai, *J. Electrochem. Soc.* **14**: 1107–1118 (1977).

8.20. M. Yoshida, *J. Appl. Phys.* **45**, No.4: 1498–1506 (1974).

8.21. V.E. Borisenko, L.D. Buiko, V.A. Labunov *et al.*, *FTP* **15**, No. 1: 3–7 (1981) (in Russian).

8.22. V.A. Uskov, E.V. Kurilchik, *Izvestiya AN SSSR, ser. Neorg. mat.* **6**, No. 10: 1887–1888 (1970) (in Russian).