Part II

Heavily Doped Semiconductors

A semiconductor is called heavily doped (HDS) if its impurity concentration $N$ satisfies the inequality $Na^3 >> 1$, where $a$ is the Bohr radius of the impurity state. This inequality makes a number of theoretical approximations possible. In semiconductors with a small effective mass ($n$-InSb, $n$-InAs, etc.) the heavy doping condition is fulfilled in a rather wide range of concentrations. For example, in $n$-InSb the value $Na^3 = 1$ is reached at $N \approx 5 \cdot 10^{15} \text{cm}^{-3}$. On the other hand, in $n$-Ge this happens only at $N \approx 3 \cdot 10^{19} \text{cm}^{-3}$. This means that the range of heavy doping in germanium is rather narrow.

As discussed in Chap. 3, localization of electrons at the Fermi level provides a physical criterion for distinguishing between the ranges of light and heavy doping. In a lightly doped semiconductor, electronic states at the Fermi level are localized and low-temperature conduction is activated, whereas in a heavily doped semiconductor, states are delocalized and conduction is of a metallic nature. Experimentally, the transition between metallic and activated conduction (the Mott transition) usually occurs at concentrations satisfying the condition $Na^3 \approx 0.02$. This means that there is a wide intermediate range of concentrations, $0.02 < Na^3 < 1$, which cannot be classified as a light doping region because electrons are delocalized, but which also does not satisfy the condition of heavy doping. As far as most physical properties are concerned, however, this range is quite similar to the heavy doping region, and therefore the description of electronic states presented in the subsequent chapters also has some relation to the intermediate region.

A weak point of the HDS theory is the necessity of making certain assumptions about correlations in the impurity distribution. The structure of electronic states near the bottom of the conduction band, which will be discussed in Chap. 11, is not very sensitive to such assumptions. However, they are of critical importance in the description of deep tails of the density of states (Chap. 12) and the metal-insulator transition induced by compensation (Chap. 13). As in Part 1 of this book, we shall start with a Poisson distribution of impurities and then consider some models which account for correlations. Impurities will be assumed to be immobile, which is

---

1. The first monograph covering extensive experimental material on heavily doped semiconductors was written by Fis	ul’ [11.1].
undoubtedly true at low temperatures. At high temperatures there may be mobile excitations (fluctuons) involving both electrons and impurities. As we shall not discuss this question, we refer the reader to an excellent review by Krivoglaz [11.2].

Our approach to the HDS theory will be presented as follows. We shall consider various relations between parameters, and assume the fulfillment of certain strong inequalities. Having solved problems in various limiting cases, we shall then interpolate between these solutions and obtain a complete description of the structure of electronic states and the nature of the processes they enter. However, this approach gives only a qualitative picture in the intermediate regions, which, of course, are precisely the regions where most experimental observations are made. Therefore we shall not aim here to achieve numerical agreement between theory and experiment, but leave this instead to future theories, which will very likely be based on good computing rather than on an analytical approach. Our goal will only be to provide a qualitative explanation of experimental facts.
11. Electronic States in Heavily Doped Semiconductors

In this chapter we use the quasiclassical method to study electronic states in a heavily doped semiconductor. The advantage of this method lies in its simplicity and physical transparency. With its help we can obtain a description of the density of states "tails" in the forbidden gap.

11.1 Linear Screening Theory

Experimental data indicate that electrons in HDS have a rather high mobility, i.e., that they are relatively free. At first glance this may appear strange. Indeed, the interaction energy between electrons and impurities is determined by the mean interimpurity distance, and therefore it is of order $e^2 N^{1/3}/\kappa$. This quantity grows with the concentration $N$. In a nondegenerate plasma the condition that electrons represent an ideal gas is $kT \gg e^2 N^{1/3}$, i.e., the kinetic energy $kT$ must be greater than the potential energy. This means that the gas becomes nonideal at high concentrations.

For a degenerate electron gas, however, the situation is quite different. The crux of the matter is that the energy of most electrons is of the order of the Fermi energy. If the semiconductor is not too highly compensated, then the concentration of electrons is close to that of impurities, and the Fermi energy $\mu$ is of order $\hbar^2 N^{2/3}/m$. This quantity also grows with $N$, and moreover, it grows faster than the interaction energy. Therefore the ideality condition, $\mu \gg e^2 N^{1/3}/\kappa$, is satisfied provided $Na^3 \gg 1$, where $a = \hbar^2 k/m e^2$. Thus, as the concentration increases, the electron gas becomes more ideal. (The energy of electron-electron interaction is also of order $e^2 N^{1/3}/\kappa$, and it is small compared to the Fermi energy.)

This fact determines the approach to the theory of electronic states in HDS. It is clear that one should start from the free-electron picture, i.e., assume that the wave functions are plane waves and that the quasimomentum $p$ is a good quantum number.

At zero temperature all states with $\epsilon < \mu$ must be filled. Quantum mechanics tells us that the number of states per unit volume, with energy in the interval $(\epsilon, \epsilon + d\epsilon)$, is proportional to the volume of the region these states occupy in the momentum space. If the energy spectrum is isotropic and quadratic, $\epsilon = p^2/2m$, then this region represents a spherical shell, and its phase volume equals $4\pi p^2 dp$. The number of states is obtained by dividing
this volume by \((2\pi\hbar)^3\), which is the phase volume corresponding to one state. It should also be taken into account that each state can contain two electrons with opposite spins. As a result, we obtain the following expression for the density of states:

\[
g_0(\epsilon) \, d\epsilon = \frac{8\pi p^2 dp}{(2\pi\hbar)^3} = \frac{8\pi p^2}{(2\pi\hbar)^3} \frac{dp}{d\epsilon} \, d\epsilon,
\]

or

\[
g_0(\epsilon) = \frac{(2m)^{3/2}\sqrt{\epsilon}}{2\pi^2\hbar^3}.
\]

The Fermi energy in this approximation is determined by the condition

\[
\int_0^{\mu_0} g_0(\epsilon) \, d\epsilon = n
\]

and equals

\[
\mu_0 = (3\pi^2)^{2/3} \, \hbar^2 n^{2/3}/2m.
\]

So far it has been assumed that the electron distribution is spatially homogeneous. Now we shall show how one can describe inhomogeneous distributions. An inhomogeneous electron distribution in a pure semiconductor arises near the boundary or near the contact with another material. In a doped semiconductor, spatially inhomogeneous electron distribution is an unavoidable consequence of fluctuations in the concentration of charged impurities. This phenomenon will be our object of study in this and the subsequent chapters.

Suppose a nonuniform electrostatic potential \(\phi(\mathbf{r})\) is present in the bulk of an HDS. In this case the quasimomentum is no longer a good quantum number. In order to find the wave functions and energy levels we must solve the Schrödinger equation. However, in the most important applications of the HDS theory one can circumvent this procedure. Suppose the potential \(\phi(\mathbf{r})\) is smooth in the sense that it changes little over the electron wavelength. (Since most electrons have approximately Fermi momentum, their wavelength is given by \(\hbar/\sqrt{2m\mu}\).) In this case the electron can be regarded as classical, i.e., it can be assumed to "feel" only the potential of the point where it is located. Ignoring the uncertainty relation, we can write a classical expression for the electron energy:

\[
\epsilon = \frac{p^2}{2m} - e\phi(\mathbf{r}).
\]

The approximation is convenient to plot graphically (Fig. 11.1 a,b). The bottom energy of the conduction band now coincides with the electron potential energy \(-e\phi\) and the band is curved.

In thermodynamic equilibrium the maximum energy \(\mu\) of occupied states cannot vary in space. What does vary is the maximum kinetic energy and the
concentration of electrons. In the regions where the bottom of the band is lower, the number of electrons is higher, and vice versa. Those regions where the bottom of the band is above the Fermi level (Fig. 11.1.b) contain no electrons at all, since these regions are classically inaccessible and we neglect tunneling, which has an exponentially low probability.

Using the rules of quasiclassical quantization, it is easy to express the electron concentration $n(r)$ in terms of $\phi(r)$. The maximum electron momentum at a point $r$ is given by $p_F = \sqrt{2m(\mu + e\phi)}$ and the concentration is $n = \frac{8\pi p_F^3}{3(2\pi\hbar)^3}$, whence we have

$$n(r) = \begin{cases} 
\frac{(2m)^{3/2}}{3\pi^2\hbar^3} [\mu + e\phi(r)]^{3/2} & \text{if } \mu + e\phi > 0, \\
0 & \text{if } \mu + e\phi < 0.
\end{cases} \tag{11.1.5}$$

The fluctuations in electron density produce their own potential, which tends to straighten out the conduction band, i.e., to oppose the external potential. This is what electron screening is all about. With the above approximations it is easy to construct a quantitative theory of the screening. To do this it is sufficient to write down the Poisson equation in which the electron density is related to $\phi$ by (11.1.5), viz.

$$\Delta \phi = -\frac{4\pi e}{\kappa} [\xi(r) - n(r) + n]. \tag{11.1.6}$$

Here $n$ is the average electron concentration and $e\xi(r)$ is the density of fixed charge, which is independent of the electron density. Equation (11.1.6) is a quasiclassical self-consistent equation of the Fermi-Thomas type. Among other things it describes bending of the band on a boundary between an HDS and another material. This equation is nonlinear and cannot be solved in general. However, in some cases a linear approximation is sufficient. Suppose the characteristic energies $|e\phi|$ are small compared to $\mu$. Then we can

---

2. Essentially the same type of equation was used in Chap. 3 when we studied long-range potential fluctuations. However, there we were dealing with a lightly doped semiconductor whose states were localized. Because of this our expression (3.4.10) for $n(r)$ was quite different from (11.1.5).
linearize (11.1.6) by writing \( n(r) \) in the form

\[
n(r) = n + \frac{dn}{d\mu} \phi_e,
\]

where \( dn/d\mu \) is evaluated in the approximation (11.1.3). This results in the following linear equation:

\[
\Delta \phi = \frac{\phi}{r_0^2} - \frac{4\pi e \xi(r)}{k},
\]

(11.1.7)

where

\[
r_0 = \left( \frac{4\pi e^2}{k} \frac{dn}{d\mu} \right)^{-\frac{1}{2}} = \frac{a}{2} \left( \frac{\pi}{3} \right)^{1/6} \left( \frac{1}{na^3} \right)^{1/6}.
\]

(11.1.8)

The length \( r_0 \) is usually called the Debye screening radius for a degenerate electron gas. Equation (11.1.7) has the following solution:

\[
\phi = \int K (r-r') \xi(r') \, dr', \quad K(r) = \frac{e}{kr} \, e^{-r/r_0}.
\]

(11.1.9)

For a single positive charge placed at the origin, we find

\[
\xi(r) = \delta(r), \quad \phi = \frac{e}{kr} \, e^{-r/r_0},
\]

(11.1.10)

which means that the Coulomb potential is screened out at the distance \( r_0 \). It should be noted that if one takes quantum effects into account, one obtains additional oscillating terms [11.3] which are small compared to (11.1.10) at short distances, but which at large distances decay much more slowly. This fact, however, is not essential in the theory of HDS, because the impurity scattering produces a substantial uncertainty in the momentum of electrons close to the Fermi level. Let us now discuss the conditions for the validity of the quasiclassical linear theory of one-charge screening. It is readily seen that the condition \( na^3 >> 1 \) guarantees fulfillment of the condition \( r_0 >> \hbar/\sqrt{m}. \mu \). The latter means that the electron wavelength is much shorter than the characteristic dimensions of the potential well. The same equality ensures linearity of screening. It is necessary that the potential energy be small compared to the Fermi energy at all points of the potential well. This is ensured by the condition \( e^2/r_0k << \mu \), which also reduces to \( na^3 >> 1 \). Finally, we note that the condition \( na^3 >> 1 \) is equivalent to the condition \( nr^3 >> 1 \). In order for the above theory to make any sense, the latter condition must clearly be fulfilled. Indeed, one must have a large number of electrons in the sphere of radius \( r_0 \) in order for a small variation in their density to produce screening.

Let us emphasize the main conclusions of the linear screening theory:

1) Screening radius \( r_0 \) does not depend on the magnitude of the potential to be screened. It is determined only by the electron concentration and the effective mass.
2) Each charge is screened independently. For example, the total potential produced by all charged impurities is given by the sum of the individual contributions (11.1.10), i.e., each impurity is screened on its own. This important conclusion follows from the linearity of (11.1.7).

Typical examples of nonlinear screening can be found in the theory of contact phenomena. In particular, the sharp dependence of the width of a space charge layer on applied voltage results from nonlinear screening. In Chap. 13, we shall encounter nonlinear screening in the description of a highly compensated semiconductor. A peculiar example of nonlinear screening produced by impurity-band electrons was considered in Sect. 3.4.

11.2 Density of States Near the Bottom of the Conduction Band

At an international conference on semiconductor physics in Exeter in 1962, Bonch-Bruevich, Kane, and Keldysh suggested independently that the density of states in heavily doped semiconductors can be described using the semiclassical picture. The corresponding calculations were reported in [11.4-8].

Let us first consider the consequences of this idea, deferring a rigorous justification of it until the next section. We shall consider electrons to be classical and calculate their density of states by (11.1.1), using (11.1.4) for the relation between the energy and momentum. This results in a local density of states $g(\epsilon, r)$ which depends on the coordinate $r$:

$$g(\epsilon, r) = \frac{(2m)^{3/2}}{2\pi^2\hbar^3} \sqrt{\epsilon + e\phi(r)} . \quad (11.2.1)$$

Here $\phi(r)$ is the total potential produced by all impurities, with electron screening taken into account.

In regions where $\epsilon < -e\phi$ there are no classically allowed states of energy $\epsilon$, and the density of states vanishes. In order to find the total density of states $g(\epsilon)$, one must integrate (11.2.1) over all classically allowed regions and divide the results by the total volume of the system $V_0$ (Fig. 11.2):

$$g(\epsilon) = \frac{(2m)^{3/2}}{2\pi^2\hbar^3 V_0} \int_{\epsilon > -e\phi} d\mathbf{r} \sqrt{\epsilon + e\phi(r)} . \quad (11.2.2)$$

![Fig. 11.2. Calculation of the quasiclassical density of states. The bottom of the conduction band is shown by the curved line. Classically allowed states, which fall in the interval $\Delta\epsilon$, are shaded](image-url)
In what follows it will be more convenient to deal with the electron potential energy \( V = -e\phi \). Coordinate integration in (11.2.2) can be replaced by averaging over all values of the potential energy, as if we were studying the potential at one point in space but varying the impurity configuration in all possible ways. This gives

\[
  g(\epsilon) = \frac{(2m)^{3/2}}{2\pi^2\hbar^3} \int_{-\infty}^{\infty} \sqrt{\epsilon - V} \ F(V) \ dV ,
\]

where \( F(V) \) is the distribution function for the potential energy, which is to be determined. As already discussed, the electron gas in HDS is weakly nonideal. Therefore the typical values of the electron potential energy must be small compared to the Fermi energy, and we can use the linear screening theory. The role of external charge in (11.1.7) and (11.1.9) is played here by inhomogeneities in the distribution of charged donors. (To be specific we shall discuss an \( n \)-type HDS, assuming low compensation.) Thus, we have

\[
  \xi(r) = N(r) - N ,
\]

where \( N \) is the mean donor concentration.

We shall assume that there is no correlation in the distribution of impurities, and the statistical law for their distribution corresponds to "white noise":

\[
  \langle \xi(r) \xi(r') \rangle = N\delta(r - r') ,
\]

where \( \langle \cdots \rangle \) denotes averaging over configurations. We assume that the distribution of impurities is Gaussian, i.e., all higher-order correlators decouple into products of pairwise correlators (this assumption will be justified below). According to (11.1.9) and (11.2.5), we have

\[
  \langle V(r) V(r') \rangle = e^2N \int K(r - r'') K(r' - r'') \ dr'' ,
\]

whence the mean-square potential energy (multiplied by a factor of 2) is given by

\[
  \gamma^2 = 2Ne^2 \int K^2(r) \ dr = 4\pi \frac{e^4}{k^2r_0^2}N\delta^3 .
\]

For linear screening \( V \) and \( \xi \) are related linearly (11.1.9). Therefore the higher correlators of \( V(r) \) also decouple into pairwise products. This implies that \( V(r) \), like \( \xi(r) \), is a Gaussian random variable, whence

\[
  F(V) = (\gamma\sqrt\pi)^{-1} e^{-V^2/\gamma^2} .
\]

A simple interpretation of (11.2.8) consists in the following. As shown in Sect. 3.3 for the potential produced by Gaussian fluctuations of impurities, the larger the range of a fluctuation, the larger is the potential. In our case
fluctuations longer than \( r_0 \) are completely screened by electrons. Therefore the main contribution to the potential must come from fluctuations on a scale of order \( r_0 \). This can also be seen from (11.2.7), where the main contribution to the integral comes from lengths of order \( r_0 \). Estimating the potential produced by fluctuations of range \( r_0 \) within the uniformly charged sphere approximation, we find

\[
V = e^2 \frac{\delta N}{\kappa r_0},
\]

where \( \delta N \) is the excess number of charged donors. These donors are assumed to be uniformly distributed within a sphere of radius \( r_0 \). The average number of donors in such a sphere is \( \bar{N} = N r_0^3 \). As is known from statistics, the mean-square value of \( \delta N \) is \( (\bar{N})^{1/2} \), i.e., \( (N r_0^3)^{1/2} \). Substituting this value into (11.2.9), we find an estimate for \( \gamma \) which agrees with (11.2.7) to within a numerical coefficient. Now it is easy to estimate the range of validity of the Gaussian statistics. As is well known, fluctuations are Gaussian if deviations in the number of impurities are small compared to their average number, viz. \( \delta N << \bar{N} \). From (11.2.9) we find a restriction on the electron potential energy:

\[
V << \bar{V}, \quad \text{where} \quad \bar{V} = \frac{e^2}{\kappa r_0} N r_0^3.
\]

Note that the necessary condition \( \gamma << \bar{V} \) follows from the basic inequality \( N a^3 >> 1 \). In the absence of compensation one has \( n = N \) and \( \bar{V} \approx \mu_0 \). In this case (11.2.10) is equivalent to the linear screening condition.

Substituting (11.2.8) into (11.2.3) we get

\[
g(\epsilon) = \frac{\sqrt{2} m^{3/2} \gamma^{3/2}}{\pi^2 \hbar^3} G_0 \left( \frac{\epsilon}{\gamma} \right),
\]

where

\[
G_0(X) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{X} e^{-y^2} (X - y)^{3/2} dy.
\]

The function \( G_0(X) \) is plotted in Fig. 11.3 [11.9].
For $\epsilon > 0$ and $|\epsilon| >> \gamma$ one has
\[
g(\epsilon) = g_0 - \frac{(2m)^{3/2} \gamma^2}{32\pi^2 \hbar^3 \epsilon^{3/2}} ,
\]  
whence it is seen that in this case $g(\epsilon)$ is close to the free-electron density of states $g_0$. For $\epsilon < 0$ and $|\epsilon| >> \gamma$ we have
\[
g(\epsilon) = \frac{m^{3/2} \gamma^{3/2}}{4\pi^2 \hbar^3} \left( \frac{\gamma}{\epsilon} \right)^{3/2} e^{-\gamma/\sqrt{\epsilon}} ,
\]
i.e., the density of states decreases exponentially into the forbidden gap.

It is easy to see that for $Na^2 >> 1$ the Fermi energy $\mu_0$ is much larger than $\gamma$, i.e., the Fermi level lies in the range of energies where the density of states is close to $g_0$. It is this fact which makes it possible to regard the screening of fluctuations as Debye screening.

Measurements of the Moss-Burstein threshold and other experimental evidence suggest that the Fermi energy in HDS is somewhat lower than $\mu_0$. This is due to deviations of the electron gas from ideality. Let $\mu = \mu_0 - \mu_{ee} - \mu_{ei}$, where $\mu_{ee}$ and $\mu_{ei}$ are corrections owing to interaction of electrons with each other and with impurities. The dominant correction results from the exchange interaction, for which (see, e.g., [11.10])
\[
\mu_{ee} = \frac{e^2}{\pi \hbar} \frac{\sqrt{2m\mu_0}}{\hbar} .
\]
The correction $\mu_{ei}$ can be obtained from (11.2.11), viz.:
\[
\mu_{ei} = \frac{1}{8} \frac{\gamma^2}{\mu_0} .
\]

It describes lowering of the Fermi level due to the tail of the density of states.

It is relatively straightforward to generalize the above results to the case of a compensated semiconductor [11.9] if one can still assume that the single-electron approximation is valid and that impurities are screened linearly. In this case, besides the positive donors, one has negatively charged acceptors of concentration $N_A$ and the electron concentration is $n = N_D - N_A$.

An obvious generalization is that in the expression for $\gamma$ we must replace $N$ by the total concentration of charged centers $N_D + N_A$. Inasmuch as we have $N_D < N_D + N_A < 2N_D$, we can ignore this correction in our order-of-magnitude estimates. A much more important correction results from the fact that in the expressions for the screening radius $r_0$ (11.1.8) and the Fermi energy $\mu_0$ we must replace $N_D$ by the electron concentration $n$, which can be substantially lower than $N_D$. With decreasing $n$ both the screening length $r_0$ and the characteristic value of the potential energy $\gamma$ increase, while the Fermi energy $\mu_0$ goes down. It is clear that the Debye approximation is valid only provided
\[ \alpha \equiv \frac{\gamma}{\mu_0} = \frac{4}{(3\pi)^{3/4}} \left( \frac{N}{n} \right)^{3/4} (Na^3)^{-1/4} \ll 1, \]  

(11.2.17)

where \( N = N_D \approx N_A \). This condition restricts the applicability of the above theory to highly compensated semiconductors. The case of \( \alpha > 1 \) is considered in Chap. 13, where it is shown that then the structure of electronic states is entirely different. Equation (11.2.11) can be easily generalized to the case of a quantizing magnetic field [11.9,11].

Let us now consider the conditions for the validity of the quasiclassical method. In the next section we shall give a rigorous derivation of (11.2.11) from which the nature of the approximations we have made will become evident. Below we confine ourselves to qualitative arguments.

As seen from (11.2.7) the main contribution to the mean-square potential \( \gamma \) results from fluctuations on the scale \( r_0 \). The typical potential wells are of size \( r_0 \) and depth \( \gamma \). The ground-state energy in a well is above the well's bottom by an amount of order \( E_s = \hbar^2/mr_0^2 \). Most of the states in such a potential can be described quasiclassically if \( \gamma \gg E_s \). This is the main condition for the validity of the above theory. As will be shown in the next section, the small parameter in the expansion leading to (11.2.11) is given by the ratio

\[ \left( \frac{E_s}{\gamma} \right)^{1/2} = \frac{1}{(Na^3)^{1/24}} \left( \frac{n}{N} \right)^{5/24}. \]  

(11.2.18)

This is a very poor parameter, of course: according to (11.2.17), the ratio \( n/N \) cannot be very small. Nevertheless, this expansion appears to be the only possible analytical approach to the density-of-states problem near the bottom of the band. At any rate, it allows a qualitative understanding of the physical nature of electronic states, which is the main purpose of this approach.

Short-range fluctuations have been shown [11.12] to restrict severely the applicability of (11.2.11) in the region \( \epsilon < 0, |\epsilon| \gg \gamma \), where the density of states is exponentially small. It turns out that for \( \epsilon < 0 \) the range of validity is limited by the following condition:

\[ |\epsilon| < \min \{ \epsilon_c, \mu_0 \}, \quad \text{where} \quad \epsilon_c = \frac{\gamma^2 ma^2}{\hbar^2 (Na^3)^{2/5}}. \]  

(11.2.19)

Since

\[ \frac{\epsilon_c}{\gamma} = (Na^3)^{1/60} \left( \frac{N}{n} \right)^{1/12}, \]

the above condition means, in practice, that the quasiclassical method is not applicable in the region where the density of states is exponentially small. New ideas and methods are clearly required to describe this region. One such
method is the optimal fluctuation method described in the next chapter. It will be shown there that the exponent in (11.2.14) retains its form even beyond the range (11.2.19).

11.3 Derivation of the Quasiclassical Formula for the Density of States

A rigorous derivation of (11.2.11) and corrections to it contains certain nontrivial points in the case of HDS. Many authors [11.4,13-15] have derived the quasiclassical expression (11.2.3) by assuming a smooth impurity potential. Conceptually, these derivations are no different from the one presented in the preceding section, although they are certainly more logical and rigorous. For example, if one assumes a Gaussian potential, then (11.2.11) can be derived in a simple and elegant way by using Feynman path integrals [11.13,15]. However, one can also obtain (11.2.3) without assuming that the potential is Gaussian [11.4,5]. It is important that in such derivations the corrections to the quasiclassical expression contain a mean-square value of the potential gradient, \[ \int (\nabla V)^2 d \tau \] [11.13], or of higher derivatives [11.14]. Because of this, these derivations are, strictly speaking, invalid in the case of a potential which behaves as \( r^{-1} \) at short distances from the center, since for such a potential the correction terms diverge.

This divergence can be avoided without assuming a Gaussian potential, if one performs a systematic expansion of the density of states in powers of the Planck constant [11.16] which is similar to the well-known expansion of the partition function [3.4]. In this expansion the zero-order term in \( \hbar \) indeed leads to (11.2.11). The correction turns out to be of order \( \hbar^2 \), and its ratio to the leading term of order \( a/r_0 \) [with \( r_0 \) given by (11.1.8)].

This result is easy to understand. The potential under consideration represents a superposition of screened Coulomb wells. The condition \( a/r_0 \ll 1 \) implies that each well contains many bound states — which is precisely the condition for validity of the expansion in \( \hbar \). However, for heavy doping and a degree of compensation that is not too high, when \( \mu_0 \gg \gamma \) and the screening is linear (and this is the only case we are considering in the present chapter), the opposite condition is always true, viz. \( r_0 \ll a \). In this case a single well possesses no bound states at all. [The inequality \( r_0 \ll a \) readily follows from (11.2.17) and \( Na^3 \gg 1 \).] Thus, the usual quasiclassical expansion fails in the case we are interested in.

Keldysh [11.17] proposed a derivation of (11.2.11) using a diagram technique. A similar derivation was later proposed independently by Efros [11.12]. The formalism in [11.12] allows one to obtain corrections to (11.2.11) in any order. The main idea of this technique is to include the long-range fluctuations of the impurity potential to all orders quasiclassically, while the short-range fluctuations are included quantum-mechanically, although by perturbation theory.
In the zeroth approximation of this modified perturbation theory we obtain (11.2.11), while subsequent approximations give corrections. A practical conclusion of this theory is that the expansion parameter is given by (11.2.18) and the validity range by (11.2.19). It follows, in particular, that (11.2.3) is valid in the case of HDS only when the distribution function is Gaussian. The derivation of (11.2.11) by this method is given below.

Consider a single-electron Green function

\[ G = \left\{ \frac{1}{\epsilon - H + i\delta} \right\}, \quad \delta \rightarrow +0, \]  

(11.3.1)

where

\[ H = \frac{p^2}{2m} + \sum_j e \phi(r - r_j), \]  

(11.3.2)

and \( \phi(r) \) is given by (11.1.10). The density of states \( g(\epsilon) \) is given by

\[ g(\epsilon) = -\frac{2}{\pi} \text{Im} \text{Sp} G. \]  

(11.3.3)

The simplest diagrams for \( G \) are shown in Fig. 11.4. Each wavy line corresponds to a factor

\[ \frac{4\pi e^2}{\kappa(q_0^2 + r_0^{-2})}, \]

a bundle of wavy lines corresponds to a factor \( N \), the law of momentum conservation holds in each bundle, and all independent momenta are integrated over. The free Green function is given by

\[ I_p = (\epsilon - \epsilon_p + i\delta)^{-1}, \quad \epsilon_p = p^2/2m. \]

![Fig. 11.4. Simplest diagrams for Green's function](image)

The diagram in Fig. 11.4a corresponds to an energy shift which is exactly compensated by the potential of electrons, and hence will not be considered. The self-energy part, corresponding to the diagram in Fig. 11.4b, is of the form

\[ \Sigma^{(1)} = \int \frac{d^3q}{(2\pi)^3} \frac{D(q)}{\epsilon - \epsilon_p - \hbar q + i\delta}, \]  

where

\[ D(q) = N \left[ \frac{4\pi e^2}{\kappa(q^2 + r_0^{-2})} \right]^2. \]  

(11.3.4)
Expression (11.3.4) is sufficient for developing the usual perturbation theory which is valid at high energies. It is interesting to note that perturbation theory is valid for $\epsilon/\gamma \gg 1$ and that the first correction to the density of states coincides with that in (11.2.13).

The main idea of the derivation is to regard the transferred momentum $\hbar q$ as small compared to $p$. If in the integral (11.3.4) we let $\epsilon_{p-\hbar q} \approx \epsilon_p$, the integral will converge and the typical value of $q$ will be of order $r_0^{-1}$, i.e., the expansion parameter is $\hbar/r_0p$. At small $\epsilon$ the largest contribution to the density-of-states integral results from $p \approx \sqrt{m\gamma}$, so that our expansion parameter is given by

$$\left(\frac{r_0\sqrt{m\gamma}}{\hbar}\right)^{-1} = \left(\frac{E_s}{\gamma}\right)^{1/2}.$$ 

This procedure, however, can be carried out only for diagrams containing two wavy lines in a bundle. For diagrams with more than two lines per bundle (Fig. 11.4c), neglect of the transfer momentum leads to a divergence. Indeed, each new line in a bundle corresponds to an integral of the form

$$\frac{e^2}{\kappa} \int \frac{L_{p-\hbar q}d^3q}{(q^2 + r_0^{-2})},$$

which diverges linearly if $L_{p-\hbar q}$ is replaced by $L_p$. The addition of each new line contributes, therefore a factor $Ge^2p/k\hbar$. At small $\epsilon$ the values $G \approx 1/\gamma$ and $p \approx \sqrt{m\gamma}$ dominate, so that

$$G \frac{e^2p}{\kappa \hbar} \approx \frac{\hbar}{a\sqrt{m\gamma}} \ll \frac{\hbar}{r_0\sqrt{m\gamma}} \ll 1.$$ 

Therefore in the first few orders of the expansion in $\hbar/pr_0$ it is sufficient to include only the diagrams containing two lines per bundle. The origin of this can be traced to the fact that for $r_0/a \ll 1$ the Born approximation is applicable to the potential $\phi(r)$ at all energies.

If we include only the diagrams having two lines in a bundle, then the diagram technique becomes identical to that for an elastic electron-phonon interaction, with the propagator $D(q)$ given by (11.3.5). The Green function satisfies Dyson's equation,

$$\epsilon - \frac{p^2}{2m} - G_p^{-1} = \int \Gamma(p, p - \hbar q, q) G_{p-\hbar q} D(q) \frac{d^3q}{(2\pi)^3}, \quad (11.3.6)$$

where $\Gamma(p, p - \hbar q, q)$ is the vertex part. As discussed above, we neglect the transfer momentum $\hbar q$ in the arguments of $\Gamma$ and $G$ in the right-hand side of (11.3.6). This gives

$$\epsilon - \frac{p^2}{2m} - G_p^{-1} = G_p\Gamma(p, p, 0) \frac{\gamma^2}{2}, \quad (11.3.7)$$
where
\[ \gamma^2 = 2 \int D(q) \frac{d^3q}{(2\pi)^3}. \] (11.3.8)

Definitions (11.3.8) and (11.2.7) coincide, of course. Next, we use the Ward identity [11.18],
\[ \Gamma(p, q, 0) = \frac{dG_p^{-1}}{d\epsilon}, \] (11.3.9)
to obtain the following closed-form equation for \( G_p \):
\[ \frac{dG}{dx} \frac{\gamma^2}{2} + Gx - 1 = 0, \] (11.3.10)
where \( x = \epsilon - p^2/2m + i\delta \). The solution of this equation, subject to the boundary condition \( G(x) = 1/x \) for \( x \to \infty \), is of the form
\[ G = \frac{1}{\gamma \sqrt{\pi}} \int_{-\infty}^{\infty} \frac{e^{-\epsilon^2/4}}{\epsilon - p^2/2m - V + i\delta} d\epsilon. \] (11.3.11)

Equation (11.3.3) can be represented in the form
\[ g(\epsilon) = -\frac{2}{\pi} \text{Im} \int \frac{d^3p}{(2\pi\hbar)^3} G_p. \] (11.3.12)

Substituting (11.3.11) into (11.3.12) we get
\[ g(\epsilon) = \frac{2}{\gamma \sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\epsilon^2/4} d\epsilon \int \frac{d^3p}{(2\pi\hbar)^3} \delta \left( \epsilon - \frac{p^2}{2m} - V \right). \] (11.3.13)

Performing the momentum integration in (11.3.13), we obtain (11.2.11).
The principle of calculating the higher-order approximations in \( \hbar/r_0^2 m \gamma \) consists in the following. The propagator \( D(q) \) is written in the form
\[ D(q) = D^0(q) + \tilde{D}(q), \] (11.3.14)
where
\[ D^0(q) = \frac{1}{2} \gamma^2 (2\pi)^3 \delta(q). \] (11.3.15)
The second term in (11.3.14) is then assumed to be small compared to the first. The technique of evaluating the corrections is developed in [11.12].
12. The Density-of-States Tail and Interband Light Absorption

In this chapter we describe the "optimal fluctuation" method, which permits one to calculate the density-of-states exponent in the forbidden gap, where the density of states is very low. We also discuss a simple modification of this method which allows to estimate the exponent using only the Poisson distribution and elementary quantum mechanics. A classification is proposed for the types of density of states which occur in doped semiconductors. The same method is then used to describe the interband light absorption.

12.1 The Optimum Fluctuation Method

An elegant method for determining the exponential decay law of the density of states in its tail region was proposed by Halperin and Lax [12.1,2], Zittartz and Langer [12.3], and I. M. Lifshitz [12.4]. The essence of this method, which we shall refer to as the optimum fluctuation method, (cf. the monograph by Lifshitz et al. [12.5]), is described in the following.

Suppose the characteristic size of a wave function much exceeds the average separation between impurities in a fluctuation. Quantum-mechanical averaging then smears out the discrete charge of individual impurities, and the fluctuation can be described by a smooth function of coordinates $\xi(r)$ which represents the deviation of the impurity concentration from its average value. To determine the density of states $g(\epsilon)$, one must sum over the contributions of all impurity configurations, i.e., over all potential wells containing an electronic level of given energy $\epsilon$. In the energy range we are interested in, these configurations have exponentially small probabilities, and the sum is dominated by the most probable configurations. Let us write the probability of a fluctuation $\xi(r)$ in the form $\exp(-\Omega(\xi))$. If $\xi_0$ is the most probable fluctuation among those which contribute to level $\epsilon$, then we can expect that as $\epsilon$ increases, the density of states falls as $\exp(-\Omega(\xi_0))$. In seeking $\xi_0$, it is sufficient to consider only those fluctuations in which $\epsilon$ is the lowest energy level, since the probability of a fluctuation in which this level corresponds to an excited state is certainly lower.

The optimum fluctuation method, as outlined, does not make it possible to evaluate the pre-exponential factor in the density of states, but it suffices to determine the leading term in the exponent, viz. $\ln[g(\epsilon)/g(0)]$. In practice, it is this quantity which is measured experimentally.
Calculating the preexponential factor turns out, of course, to be much more complicated. As in the usual method of steepest descent, one must take into account the states which are close to the extremum. Furthermore, one has to include the translational invariance, i.e., the fact that a localized state can exist at any point of the coordinate space. An approximate method for evaluating the prefactor was proposed by Halperin and Lax \[12.1\]. These authors assumed that the wave functions of all states corresponding to a given energy and arising on fluctuations which are close to the extremal fluctuation are, in fact, identical to the extremal wave function. Their predictions for the case of one-dimensional "white noise" were compared against the exact solution which is available in this case. The comparison showed a discrepancy by a factor of \(1/\sqrt{5}\) in the preexponential factors. This discrepancy can be substantially reduced by making corrections according to a method suggested in \[12.2\].

Simultaneously with the first paper by Halperin and Lax, work appeared by Zittartz and Langer \[12.3\] which was also devoted to the form of the density of states in the region where it is exponentially small. The latter authors considered potentials of the white-noise type (in one, two, and three dimensions) and obtained an exact result which differed from that of Halperin and Lax by a numerical factor in the preexponent. This factor can only be evaluated for the one-dimensional case, and here the result of Zittartz and Langer agrees with the exact-solution asymptotics. A rather complicated procedure was proposed for evaluating this number in the three-dimensional case.

In the work of Edwards and co-workers \[11.13, 2.41, 12.6\] the deep tail of the density of states was studied with the help of Feynman path integrals. This method, in principle, also permits one to evaluate the preexponential factor.

A convenient method of evaluating the preexponential factor for a white-noise potential was developed in \[12.7, 8\].

In all of the work cited above, except \[12.4\], the case of Gaussian statistics was investigated. Lifshitz's work \[12.4\] was not restricted in this way. Instead, he assumed that potentials of different impurity centers overlapped only weakly (their range being of the order of the shortest separation between the centers). This allows an important mathematical simplification, namely to regard the potential as locally related to the concentration \(\xi(r)\). As will be shown below, this approximation works in the HDS case only for a relatively narrow range of parameters, since the screening length here is larger than the mean separation between impurities.

We investigated the semiconductor case specifically \[12.9, 10\], and studied in detail the region where Gaussian statistics was not applicable. In \[12.9\] we also made a classification of possible situations which can arise in semiconductors for various relations between the mass of carriers and the impurity concentration. To this end, we used a simple approximate method proposed earlier by Kane \[11.6\] for determining the optimum fluctuation and
the density of states. It is assumed in this method that fluctuations have the shape of uniformly charged spheres. A sphere is characterized by two parameters: its radius \( R \) and its net charge \( eZ \). One of these parameters is chosen so as to make the electron ground-state energy in the fluctuation potential equal the given value \( \epsilon \), while the other parameter is varied to find the optimum fluctuation. It is easy to understand that in this method we lose only numerical coefficients which depend on the true shape of the optimum fluctuation. The simplicity of the method allows us to readily analyze and classify all possible cases. This method will be described in the next section and used to investigate the majority-carrier spectrum in HDS. Section 12.3 deals with the exact impurity distribution in the optimum fluctuation, and gives the values of the numerical coefficients omitted in Sect. 12.2. The minority-carrier spectrum is considered in Sect. 12.4.

It should be borne in mind that the entire theory developed in the present chapter is based on the effective-mass approximation. It follows therefore that the energy range we can consider in this way should not extend too far into the forbidden gap. In particular, this range must be much narrower than the width of the forbidden band.

### 12.2 The Uniformly Charged Sphere Approximation. The Spectrum of the Majority Carriers

Fluctuation-induced levels in the forbidden gap correspond either to a cluster of attractive impurities or to a void of repulsive ones. Let us first confine our attention to attractive impurities. Suppose the fluctuations represent homogeneous spherical formations of radius \( R \) with an enhanced impurity concentration. Assume also that the distribution of impurities is random. Then the probability that \( k \) impurities will be found in a volume where there are \( \bar{k} \) impurities on average, is given by Poisson's formula:

\[
W = \frac{(\bar{k})^k e^{-\bar{k}}}{k!}.
\]  

(12.2.1)

Assuming that \( \bar{k} \gg 1, k \gg 1 \), and keeping only the leading term in the logarithm of the probability, we get

\[
\ln W = k \ln \frac{\bar{k}}{k} + k - \bar{k}.
\]  

(12.2.2)

On average, there are \( \bar{k} = NR^3 \) impurity atoms\(^1\) in the sphere under consideration. The number \( k \) can be written in the form

\(^{1}\) In this discussion we shall avoid writing the numerical coefficients, as it would give the appearance of a higher accuracy than the described method actually provides. The expression for \( k \), as well as the other equations below, should be regarded as an order-of-magnitude estimate.
\[ k = Z + NR^3, \quad (12.2.3) \]

where \( Z \) is the excess number of impurities in the fluctuation. Substituting \( k \) and \( k \) into (12.2.2), we obtain the fluctuation probability as a function of \( R \) and \( Z \):

\[ \ln W(R, Z) = (Z + NR^3) \ln \frac{NR^3}{Z + NR^3} + Z. \quad (12.2.4) \]

Our task now is to find from the Schrödinger equation the ground-state energy \( \epsilon(R, Z) \) in the potential produced by a sphere with parameters \( R \) and \( Z \). The charge can be assumed to be smeared out over the sphere, since the characteristic dimensions of the ground-state wave function prove to be much larger than the mean distance between impurities. Having found \( \epsilon(R, Z) \), we can express \( Z \) in terms of \( R \) and \( \epsilon \). Substituting this \( Z \) into (12.2.4) gives \( W(R, \epsilon) \), which is the probability that a homogeneous aggregate of radius \( R \) will produce an electronic state of energy \( \epsilon \).

Next, we find the value \( R = r_{\text{max}} \) at which \( \ln W(R, \epsilon) \) has its absolute maximum at fixed energy. Deep in the forbidden gap, where the density of states \( g(\epsilon) \) is small while the number of impurities in the aggregate is large, this maximum is quite sharp, whence we can write

\[ \ln \left[ \frac{g(\epsilon)}{g(0)} \right] = \ln W(r_{\text{max}}, \epsilon). \quad (12.2.5) \]

This is the outline of our calculations.

We now introduce an auxiliary function \( W(R, V) \), defined as the probability that the impurity aggregate of radius \( R \) will produce a potential well whose maximum depth is \( V \). The energy \( V \), as well as the electron energy \( \epsilon \), will be measured downward from the bottom of the unperturbed conduction band. If \( R \ll r_0 \), where \( r_0 \) is the screening radius (11.1.8), then we have

\[ V = \left( \frac{e^2}{\kappa R} \right) Z \quad (R \ll r_0). \quad (12.2.6) \]

(We are including only the fluctuation potential produced by excess impurities. The average donor potential is neutralized by electrons.) If \( R \gg r_0 \), then not all the impurities in the aggregate contribute to the potential, but only those located at a distance less than the screening radius \( r_0 \). The concentration of excess impurities is given by \( Z/R^3 \) and their number in the range of interaction is \( (Z/R^3)r_0^3 \). Therefore instead of (12.2.6), we have in this case

\[ V = \frac{e^2}{\kappa r_0} \frac{r_0^3}{R^3} Z \quad (R \gg r_0). \quad (12.2.7) \]

Expressing \( Z \) through \( R \) and \( V \) from (12.2.6) and (12.2.7), and substituting into (12.2.4), we find \( \ln W(R, V) \).
The auxiliary function $W(R, V)$ bears a certain relation to the function we seek, namely $W(R, \varepsilon)$. Indeed, the electronic level $\varepsilon$ in the potential well produced by a uniformly charged sphere is close to the potential energy $V$ at the center of the sphere, provided $V \gg \hbar^2/mR^2$. Therefore for $R \gg r_q \equiv \hbar/\sqrt{mV}$, we can obtain $\ln W(R, \varepsilon)$ from $\ln W(R, V)$ by simply substituting $\varepsilon$ for $V$. However, for $R \leq r_q$, the function $Z(R, \varepsilon)$ in (12.2.4) must be calculated quantum-mechanically.

In this and the next sections we shall study the spectrum of majority carriers in a heavily doped and low-compensated semiconductor ($n$-type, for concreteness). As was shown in Chap. 11, for $\varepsilon < \varepsilon_c$ [cf. (11.2.19)], the density of states is given by the quasiclassical formula (11.2.11). Since, strictly speaking, we have $\varepsilon_c > \gamma$, the validity range of the optimum fluctuation method, viz. $|\ln [g(\varepsilon)/g(0)]| = \varepsilon^2/\gamma^2 \gg 1$, overlaps that of (11.2.11). Therefore for energies that are not too high, we should obtain the same formula again.

Let us introduce a characteristic length $r_2 = (V_k/N\varepsilon^2)^{1/2}$. The ratio $r_2/r_0$ equals $(V/V)^{1/2}$, where $V = (e^2/k\varepsilon)V_0^3$. According to (11.2.10), $V$ is the average potential energy produced by donors. Since $V > \gamma$, the ratio $r_2/r_0$ can be arbitrary in the range where the density of states is exponentially small. Plots of the function $\ln W(R, V)$ are shown schematically in Fig. 12.1 for two cases: $V < \bar{V}$ and $V > \bar{V}$.

Consider first the case $V < \bar{V}$. It can be readily seen that the range $R > r_2$ in Fig. 12.1a corresponds to Gaussian statistics. Indeed, Gaussian statistics is valid if $Z \ll NR^3$. If $R \ll r_0$, then we find from (12.2.6) that

$$\frac{Z}{NR^3} = \frac{V}{(e^2/\kappa R)NR^3} = \left(\frac{r_2}{R}\right)^2.$$  \hspace{1cm} (12.2.8)

Thus, for $R \gg r_2$ we have

$$\ln W = -\frac{Z^2}{NR^3}.$$  \hspace{1cm} (12.2.9)

Substituting (12.2.6) and (12.2.7) into (12.2.9) we get
\[ \ln W = - \frac{V^2}{\left( (e^2/kR) (NR^3)^{1/2} \right)^2} \quad (r_2 \ll R \ll r_0) \]  

(12.2.10)

and

\[ \ln W = - \frac{V^2 R^3 \kappa^2}{e^4 N r_0^4} \quad (R \gg r_0). \]  

(12.2.11)

Thus, for \( R \gg r_2 \) the function \( \ln W(R, V) \) increases with \( R \) until \( R = r_0 \), and then decreases again (Fig 12.1a). Fluctuations with \( R > r_0 \) turn out to be less probable because of screening. We have thus shown that the function \( \ln W(R, V) \) has a relative maximum at \( R = r_0 \). From the inequality \( r_q/r_0 = [(\gamma/V) (E_s/\gamma)]^{1/2} \ll 1 \) it follows that this maximum is located in the region \( R \gg r_q \), i.e., where the ground-state energy coincides with the bottom of the well \( V \). In this case, according to (12.2.5), we can substitute \( R = r_0 \) into (12.2.10) or (12.2.11) and find the corresponding contribution to the density of states:

\[ \ln \frac{g(\epsilon)}{g(0)} = - \frac{\epsilon^2}{\gamma^2}. \]  

(12.2.12)

This formula coincides with the quasiclassical result (11.2.11). However, as we shall see below, it has a much wider range of validity.

As seen from Fig. 12.1, the maximum at \( R = r_0 \) exists and (12.2.12) is valid only for \( \epsilon \ll V \). Substituting \( R = r_0 \) into (12.2.8), we find that for \( V = \bar{V} \) the fluctuations in a well of radius \( r_0 \) are no longer Gaussian. It is easy to show that values of \( V \) greater than \( \bar{V} \) cannot result from Gaussian fluctuations (the curve in Fig. 12.1b has no Gaussian region). However, even for \( \epsilon < \bar{V} \), fluctuations on the left of the minimum in Fig. 12.1a can turn out to be more probable than those with \( R = r_0 \). For this to be true, it is necessary (though not sufficient) that the inequality \( r_q < r_2 \) be fulfilled for \( V = \epsilon \). In this case \( \ln W(R, \epsilon) \) increases with decreasing \( R \) all the way down to \( R = \hbar/\sqrt{m \epsilon} \).

For \( R \ll r_2 \), using (12.2.8) and (12.2.4), we find

\[ \ln W = - Z \ln \frac{Z}{NR^3}. \]  

(12.2.13)

Substituting (12.2.6) into (12.2.13), we get

\[ \ln W(R, W) = - \frac{VR_k^2}{e^2} \ln \frac{V_k}{e^2 NR^2}. \]  

(12.2.14)

The function (12.2.14) increases with decreasing \( R \), as shown in Fig. 12.1a. In this case \( Z \) decreases, attaining unity at \( R = r_1 = e^2/V_k \). However, if \( V \gg E_0 \) (where \( E_0 = me^4/2 \hbar^2 \kappa^2 \)), then the well will prove to be shallow (will have no bound states) at a much larger value of \( R \) than \( R = r_1 \). Indeed, in this case \( r_1/r_q \approx (E_0/V)^{1/2} < 1 \).
It is easy to guess how the function \( \epsilon(R, Z) \) should behave for \( R \ll \hbar/\sqrt{m\epsilon} \). In this case the fluctuation represents an almost pointlike atomic nucleus whose lowest term realizes the energy \( \epsilon \). Including the correction due to a finite size of the nucleus, we have

\[
\epsilon = Z^2 E_0 \left[ 1 - \frac{R^2}{(a/Z)^2} \right]. \tag{12.2.15}
\]

Thus, for \( R \ll \hbar/\sqrt{m\epsilon} \) the ground-state energy \( \epsilon \) is quite different from the potential depth \( V \) (12.2.6). The latter tends to infinity as \( R \to 0 \), while \( \epsilon \) tends to a finite limit \( Z^2 E_0 \).

Inverting (12.2.15), we find

\[
Z = \sqrt{\frac{\epsilon}{E_0}} \left[ 1 + \frac{R^2}{a^2} \frac{\epsilon}{E_0} \right], \quad R \ll \frac{\hbar}{\sqrt{m\epsilon}} \approx a \sqrt{\frac{E_0}{\epsilon}}. \tag{12.2.16}
\]

Substituting (12.2.16) into (12.2.13), we see that \( \ln W(R, \epsilon) \to -\infty \) as \( R \to 0 \). This means that \( \ln W(R, \epsilon) \) must have another maximum (Fig. 12.2), since for \( R \gg \hbar/\sqrt{m\epsilon} \) this function is described by (12.2.14), where we can let \( V = \epsilon \). The second maximum is determined by competition between the combinatorics logarithm in (12.2.13), which favors less compact nuclei, and the finite-size correction in (12.2.16), which indicates that pointlike nuclei can provide the same energy at a lower \( Z \). Inasmuch as the finite-size correction enters (12.2.13) with a power-law dependence, it wins over the logarithm and the optimum nucleus turns out to be pointlike \( (r_{\text{max}} \ll \hbar/\sqrt{m\epsilon}) \). To find \( r_{\text{max}} \) therefore it is sufficient to substitute (12.2.16) into (12.2.13) and maximize the latter with respect to \( R \). As a result, for the optimum fluctuations we have

\[
r_{\text{max}} \approx \frac{\hbar}{\sqrt{m\epsilon}} \left\{ \ln \left[ \left( \frac{\epsilon}{E_0} \right)^2 \frac{1}{N a^3} \right] \right\}^{1/2}. \tag{12.2.17}
\]

It was assumed in this derivation that the argument of the logarithm in (12.2.17) is much larger than unity. This condition is satisfied for energies at which the density of states is governed by the maximum under consideration \( (\epsilon > E_1; \text{see below}) \).

![Fig. 12.2. Logarithm of the probability that an aggregate of radius \( R \) will form an electron state of energy \( \epsilon \) as a function of \( R \) \((E_1 < \epsilon < V)\). The corresponding classical curve is indicated by the dashed line](image_url)
Thus the size of the optimum aggregate is small compared to its Bohr radius, i.e., the aggregate represents a nearly pointlike nucleus.

Substituting \( Z = \sqrt{\epsilon/E_0} \) and \( r_{\text{max}} \) from (12.2.17) into (12.2.13), we obtain \([12.9,10]::\)

\[
\ln \frac{g(\epsilon)}{g(0)} = -\sqrt{\frac{\epsilon}{E_0}} \ln \left[ \left( \frac{\epsilon}{E_0} \right)^2 \frac{D(\epsilon)}{N a^3} \right], \tag{12.2.18}
\]

where

\[
D(\epsilon) \approx \left\{ \ln \left[ \left( \frac{\epsilon}{E_0} \right)^2 \frac{1}{N a^3} \right] \right\}^{3/2}. \tag{12.2.19}
\]

Comparing (12.2.18) with (12.2.12), we see that (12.2.18) takes place for \( \epsilon > E_1 \), while (12.2.12) holds for \( \epsilon < E_1 \), where

\[
E_1 = \frac{\gamma^{4/3}}{E_0^{1/3}} \left\{ \ln \left[ \left( \frac{\gamma}{E_0} \right)^{8/3} \frac{1}{N a^3} \right] \right\}^{2/3}. \tag{12.2.20}
\]

It is easy to see that we have \( E_1 < \bar{V} \). If \( \epsilon > \bar{V} \), then, as discussed, the maximum at \( R = r_0 \) disappears, while our arguments related to the maximum at \( R = r_{\text{max}} \) (12.2.17) remain unaffected. This means that (12.2.18) is valid for all energies \( \epsilon > E_1 \) for which the effective-mass method is applicable. Let us now discuss the assumptions made in the derivation of this formula.

In deriving (12.2.18) we regarded \( Z \) as a continuous variable. This is reasonable if the width of the maximum of \( \ln W(Z,\epsilon) \) as a function of \( Z \) is large compared to unity. It is easy to see that this is true, provided

\[
\left( \frac{\epsilon}{E_0} \right)^{1/4} \gg \ln \left[ \left( \frac{\epsilon}{E_0} \right)^2 \frac{1}{N a^3} \right]. \tag{12.2.21}
\]

We obtained (12.2.18) in the uniformly charged sphere approximation. In determining the size of the optimum fluctuation, we ignored numerical factors which in reality depend on a particular density distribution within the fluctuation. However, our assertion that \( r_{\text{max}} < a/Z \) does not depend on this approximation. Therefore the prelogarithmic factor in (12.2.18) contains no additional numbers. This is also confirmed by the theory developed in Sect. 12.3, in which the true distribution of charge in a fluctuation is taken into account and the function \( D(\epsilon) \) is found exactly. It should be added that within the effective-mass approximation the factor in front of the logarithm does not depend on the form of the spectrum. Indeed, the form of this factor follows from the relation \( \epsilon \propto Z^2 \), which is always true when the Hamiltonian is a quadratic function of the momentum components.

---

2. In this case \( E_0 \) should be interpreted as the binding energy of a single-charge shallow impurity.
Throughout this section we have assumed the single-electron approximation and the linear screening of impurities. However, the latter assumption was really necessary only in the derivation of (12.2.12). Since in a low-compensated semiconductor one has \( E_1 < \mu_0 \), where \( \mu_0 \) is the Fermi energy, the linear screening assumption places no restriction on the validity range of (12.2.12). On the other hand, the inclusion of electron-electron interaction does not affect the form of (12.2.18) even for \( \epsilon > \mu_0 \). Indeed, our aggregate represents a multielectron atom, whose lowest term has the required energy \( \epsilon \). It is well known [1.8] that for a large nuclear charge, outer-shell electrons have little effect on the energy of this term.

Thus, the asymptotic expression for the density of states at large energies is given correctly by (12.2.18). It is important, however, that the impurities were assumed to be randomly distributed. This assumption may be valid when centers are produced by irradiation at low temperatures. On the other hand, if the specimen was subjected to a high-temperature treatment, during which impurities were free to move, then their interaction could rise to a correlation. The diffusion coefficient for impurities is a very sharp function of the temperature. Since the cooling time of samples is usually quite short, one can assume that the impurity distribution corresponds to a snapshot of the plasma at some temperature \( T_0 \) at which the diffusion is quenched. This temperature is usually close to the melting point.

It proves important to take these correlations into account in calculating the asymptotics of \( g(\epsilon) \) for large \( \epsilon \). Indeed, we assumed that the probability of the formation of a nucleus containing \( Z \) impurities is given by \( \exp[-Z \ln(Z/NR^3)] \). However, at temperatures above \( T_0 \) the impurities repelled each other. A work \( Z^2e^2/\kappa R \) was required to gather them together. Thus, the probability of such a fluctuation should, in fact, be given by

\[
\exp \left( -Z \ln \frac{Z}{NR^3} \right) \cdot \exp \left( -\frac{Z^2e^2}{\kappa R kT_0} \right).
\]

(12.2.22)

The second factor will always dominate for a sufficiently large \( Z \). Let us find the size of an optimal fluctuation in this case. If \( R >> \hbar/\sqrt{m\epsilon} \), then the well is classical and \( \epsilon \approx Ze^2/\kappa R \). (We are considering only the values of \( R \) which are small compared to the screening length.) Hence we have \( Z \approx \kappa R \epsilon/e^2 \), and the probability of such a well, \( \exp(-\kappa R e^2/e^2 kT_0) \), decreases with increasing \( R \). If, on the other hand, \( R << \hbar/\sqrt{m\epsilon} \), then the well represents a pointlike nucleus, \( Z = \sqrt{\epsilon/E_0} \), and its probability, given by \( \exp(-\epsilon e^2/E_0 kRkT_0) \), decreases with decreasing \( R \). Thus, we find that the optimum fluctuation has the size \( R \approx \hbar/\sqrt{m\epsilon} \), which in order of magnitude equals the radius of the first orbit. In this case \( Z \approx \sqrt{\epsilon/E_0} \). Substituting these values of \( R \) and \( Z \) into the second factor in (12.2.22), we find [12.10-12]:

\[
\ln \frac{g(\epsilon)}{g(0)} = -C \frac{\epsilon^{3/2}}{E_0^{1/2}kT_0},
\]

(12.2.23)
where $C$ is a numerical coefficient. This gives the density-of-states asymptotics with the impurity correlation taken into account. A dimensionless equation obtained in [12.10] determines both the shape of the optimum fluctuation and the value of $C$. This equation has been solved numerically in connection with the polaron theory [12.23].

To summarize, we have considered the density of states in a heavily doped and low-compensated semiconductor, and showed that it is described by (12.2.12) and (12.2.18) in the absence of correlations, and by (12.2.12) and (12.2.23) when correlations are included.

### 12.3 Exact Distribution of Impurities in Optimum Fluctuations

#### 12.3.1 Derivation of Basic Equations

In the preceding section we carried out a qualitative analysis in which it was assumed that fluctuations represent uniformly charged spheres which differ from each other only in their radius and excess charge. We shall now describe a rigorous theory which allows us to determine the true shape of optimum fluctuations. We shall consider only uncorrelated impurity distributions. A generalization of this method to the case of a correlated distribution is given in [12.10].

Let us introduce a function $\xi(r) = N(r) - N$, where $N(r)$ represents the impurity concentration at $r$, averaged over a volume larger than $N^{-1}$. We shall find the probability of a fluctuation $\xi(r)$, assuming the absence of correlations in the impurity distribution. Consider a volume element $\Delta V$ which is small enough that the impurity concentration within it can be regarded as constant, but which still contains a large number of impurities $\tilde{N} = N(r) \Delta V \gg 1$. To count the number of states, we assume that impurities can be located only at lattice sites, whose concentration $p$ is large, $p \gg N(r)$. The number of sites in the volume, $\bar{p} \Delta V$, can accommodate $\tilde{N}$ impurities in $\Gamma_{\Delta V}$ different ways, where

$$
\Gamma_{\Delta V} = \frac{P!}{\tilde{N}! (P - \tilde{N})!}.
$$

The entropy in volume $\Delta V$ is therefore given by

$$
S_{\Delta V} = \ln \Gamma_{\Delta V} = \tilde{N} \ln \frac{pe}{\tilde{N}},
$$

and the entropy density by

$$
\sigma(r) \equiv \frac{S_{\Delta V}}{\Delta V} = N(r) \ln \frac{pe}{N(r)}.
$$

The total entropy of the crystal volume $V_0$ is given by

$$
S = \int_{V_0} d^3r \sigma(r).
$$
Let us find the entropy of such an impurity distribution when a fluctuation \( \xi(r) \) — in general comparable to \( N(r) \) — takes place in a volume \( V \) which constitutes a minor fraction of \( V_0 \), while in the rest of the volume one has \( \xi(r) \ll N \). From conservation of the particle number we have

\[
\int_{V_0-V} \xi(r) d^3r = - \int_V \xi(r) d^3r.
\]  

(12.3.5)

The entropy of such a fluctuation is given by

\[
S = \int_V (N + \xi) \ln \frac{p_e}{N + \xi} d^3r + (V_0 - V) N \ln \frac{p_e}{N} + \int_{V_0-V} \left[ \frac{d\sigma}{dN(r)} \right]_{N(r)=N} \xi d^3r.
\]  

(12.3.6)

Transforming the last term with the help of (12.3.5), we find that the additional entropy, compared to that of a uniform impurity distribution, \( S_0 = NV_0 \ln (p_e/N) \), is given by

\[
S - S_0 = \int_V \left[ (N + \xi) \ln \frac{N}{N + \xi} + \xi \right] d^3r.
\]  

(12.3.7)

If \( \xi(r) \) falls fast enough so that the integral (12.3.7) converges, the integration can be extended to infinity. Then the probability that some part of the volume will contain the fluctuation \( \xi(r) \) is given by \( \exp (-\Omega[\xi]) \), where

\[
\Omega[\xi] = - \int \left[ (N + \xi) \ln \frac{N}{N + \xi} + \xi \right] d^3r.
\]  

(12.3.8)

Let us now introduce two functionals \( V_r[\xi] \) and \( \lambda_0[\xi] \), the former being the potential energy of a test electron at \( r \), and the latter the electron ground-state energy, determined by the Schrödinger equation

\[- \frac{\hbar^2}{2m} \Delta \Psi + V_r[\xi] \Psi = \lambda_0 \Psi.\]  

(12.3.9)

In the linear screening approximation (11.1.9) one has

\[-V_r[\xi] = e \int \xi(r') K(r-r') d^3r', \quad \text{where}

(12.3.10)

\[K(r) = \frac{e}{\kappa r^2} e^{-r/r_s}.\]

The optimum fluctuation \( \xi(r) \) must minimize \( \Omega[\xi] \) subject to the condition

\[\lambda_0(\xi) = -\epsilon.\]  

(12.3.11)

The corresponding equation is of the form

\[\delta(\Omega[\xi] + \beta \lambda_0[\xi]) = 0,\]  

(12.3.12)

where \( \beta \) is a Lagrange multiplier to be found from (12.3.11).
Varying (12.3.8) and taking into account that
\[ \delta \lambda_0(\xi) = -\int |\Psi(r')|^2 d^3r' e \int K(r' - r'') \delta \xi(r'') d^3r'', \]
we obtain, finally, the following equation [12.4]:
\[ \ln \frac{N + \xi}{N} - \beta \int |\Psi(r')|^2 e K(r - r') d^3r' = 0. \]  
(12.3.14)

Together with (12.3.9) this equation determines the optimum fluctuation \( \xi \).
The leading term in the logarithm of the density of states is given by
\[ \ln \frac{g(\xi)}{g(0)} = -\Omega(\xi). \]
(12.3.15)

Next we shall give a rigorous derivation of (12.2.12) and (12.2.18) and evaluate the numerical factors which depend on the shape of the fluctuation.

12.3.2 Derivation of (12.2.12)
The optimal fluctuation corresponding to (12.2.12) is described by Gaussian statistics. In this case \( \xi \ll N \), (12.3.8) is of the form
\[ \Omega(\xi) = \frac{1}{2N} \int \xi^2 d^3r, \]
(12.3.16)
and in place of (12.3.14) we have
\[ \xi = N\beta e \int |\Psi(r')|^2 e K(r - r') d^3r'. \]
(12.3.17)

Suppose the characteristic size of the wave function is \( \lambda \ll r_0 \). Then we have
\[ \xi = N\beta e K(r) \quad (r \gg \lambda), \]
(12.3.18)

and the potential energy of the optimum well (12.3.10) is of the form
\[ V_r(\xi) = -N\beta e^2 \int K(r') K(r - r') d^3r', \] or\n\[ V_r(\xi) = -\frac{e^4 N\beta r_0}{\kappa^2} 2\pi e^{-r/r_0}. \]
(12.3.19)
(12.3.20)

The potential energy falls off at distances of order \( r_0 \), and for \( \epsilon \gg E_s = \hbar^2/2mr_0^2 \) the ground-state energy is close to the bottom of the well. Therefore we can determine \( \beta \) from the following condition:
\[ V_r(\xi)|_{r=0} = -\epsilon. \]
(12.3.21)

This gives
\[ \beta = \frac{\epsilon\kappa^2}{2\pi e^4 Nr_0}. \]
(12.3.22)
Substituting (12.3.22) into (12.3.18), and (12.3.16) into (12.3.15), we obtain (12.2.12). The numerical coefficient in $\gamma$ turns out to be identical to that in (11.2.7).

Let us now find the electron wavelength in the ground state. Note that for $r \ll \lambda$ we have, according to (12.3.17),

$$\xi = \frac{N\beta e^2}{\kappa \lambda} \quad (r \ll \lambda),$$

(12.3.23)

and $\xi$ does not depend on $r$. This means that the fluctuation core $(r < \lambda)$ represents a uniformly charged sphere. Self-consistency requires that the potential of this sphere give rise to an electronic state with a wavelength of the order of the sphere radius. A sphere of radius $\lambda$ containing an excess density (12.3.23) produces a potential well of depth $N\beta(e^2/\kappa)^2\lambda$. Equating this quantity to $\hbar^2/2m\lambda^2$, we find the radius $\lambda$, viz.

$$\lambda \approx r_0 (E_s/\epsilon)^{1/3}.$$  

(12.3.24)

The inequality $\lambda \ll r_0$ assumed in the derivation of (12.3.18) is satisfied for $\epsilon > E_s$.

In the work of Halperin and Lax [12.1], equations (12.3.9) and (12.3.17) were solved numerically. It is interesting to compare the results of their calculation with the above analytical solution. The optimum potential and the wave function calculated for $\nu = \epsilon/E_s >> 1$ are shown in Fig. 12.3a,b. The figure displays the dimensionless potential $U(r) = V_r r^2/2\pi e^4 N\beta r_0$ and the dimensionless wave function $S(r) = \sqrt{4\pi} \Psi(r) r r_0^{1/2}$. In accordance with (12.3.20), in the region where $S'(r)$ is small, $U(r)$ coincides with $\exp(-r/r_0)$ (shown by dashed lines). Not having derived the exact form of the wave function, we can only make an approximate comparison of (12.3.24) with Fig. 12.3. To do this, we assume that the wavelength $\lambda$ is proportional to the coordinate of the maximum of $S(r)$. The ratio of wavelengths obtained in

![Fig. 12.3a, b. Dependencies of the dimensionless potential $U(r)$ and the wave function $S(r)$ on $r/r_0$, calculated numerically in [12.1]. Dashed line shows the potential calculated from (12.3.20), where it is different from $U(r)$, (a) $\nu = 10$, (b) $\nu = 10^3$](image-url)
this way for \( \nu = 10 \) and \( \nu = 10^3 \) equals \( \lambda_{10}/\lambda_{1000} = 4.4 \). According to (12.3.24) this ratio is given by \((100)^{1/3} = 4.6\). We see that the analytical and the numerical solutions agree well with each other.

### 12.3.3 Density-of-States Asymptotics in the Absence of Impurity Correlations

Let us rewrite (12.3.14) in the form

\[
\xi = N \left[ e^{x(r)} - 1 \right],
\]

(12.3.25)

where

\[
\chi(r) = \frac{\beta e^2}{\kappa} \int \frac{\left| \Psi(r') \right|^2}{|r-r'|} e^{-|r-r'|/\rho_0} d^3 r'.
\]

(12.3.26)

We shall now show that in the region of high energies, (12.3.9) and (12.3.25) possess a spherically symmetric solution. This solution has the property that the characteristic length of the exponential decay of \( \xi \) is small compared to the electron wavelength \( \lambda \) in the ground state. [In order of magnitude this length coincides with \( r_{\text{max}} \) (12.2.17).] Moreover, it turns out that \( \lambda \ll \rho_0 \). If we assumed all this from the outset, we would conclude that the fluctuation represents an almost pointlike nucleus of an ion with the charge

\[
Z = \int \xi(r) d^3 r,
\]

(12.3.27)

and the ground-state wave function is of the form

\[
\Psi(r) = \frac{1}{\sqrt{\pi \lambda^{3/2}}} e^{-r/\lambda}
\]

(12.3.28)

with \( \lambda = a/Z \). Since the ground-state energy must equal \( \epsilon \), we have

\[
Z = \sqrt{\frac{\epsilon}{E_0}}, \quad \lambda = a \sqrt{\frac{E_0}{\epsilon}}.
\]

(12.3.29)

Let us now prove that this solution indeed satisfies (12.3.9) and (12.3.25). To do this, we substitute (12.3.28) into (12.3.26) and find the explicit form of \( \chi(r) \) and \( \xi(r) \). To determine \( \beta \), we then substitute \( \xi \) into (12.3.27) and use (12.3.29).

Substituting (12.3.28) into (12.3.26) and expanding in \( r/\lambda \), we get

\[
\chi(r) = t - \frac{r^2}{x^2}, \quad \text{where}
\]

\[
t = \frac{\beta e^2}{\kappa \lambda} \quad \text{and} \quad x^2 = \frac{3}{2} \frac{\lambda^2}{t}.
\]

(12.3.30)

We see that the above assumption of a sharp decline of \( \xi(r) \) at short distances is equivalent to assuming \( t \gg 1 \). If \( t \gg 1 \), then the exponent in (12.3.25) is large for \( r < \lambda \) and we have
\[ \xi(r) = N e^t e^{-r^2/x^2}. \] (12.3.32)

Substituting (12.3.32) and (12.3.29) into (12.3.27), we obtain a transcendental equation for \( t \), viz.

\[ t = \ln \left( \frac{\epsilon}{E_0} \right)^2 \frac{t^{3/2}}{Na^3} \left( \frac{2}{3\pi} \right)^{3/2}, \] (12.3.33)

whence for sufficiently high energies we have

\[ t \approx \ln \left( \frac{\epsilon}{E_0} \right)^2 \frac{1}{Na^3} \gg 1, \] (12.3.34)

which justifies the above approximation. Thus, the described solution does, indeed, exist.

To evaluate \( \Omega(\xi) \), we note that \( \xi \gg N \) and (12.3.8) can be written in the form

\[ \Omega(\xi) = \int \xi \left( \ln \frac{\xi}{N} - 1 \right) d^3r. \] (12.3.35)

Substituting (12.3.32) into (12.3.35), we readily obtain

\[ \Omega(\xi) = Z(t-1) - \frac{3}{2} Z. \] (12.3.36)

This expression contains terms of order \( Zt \) and \( Z \). Therefore we cannot simply substitute \( Z = \sqrt{\epsilon/E_0} \), but must include corrections of order \( t^{-1} \). The ground-state energy, corrected for the finite nucleus size, is of the form [12.13]:

\[ \epsilon = Z^2 E_0 \left( 1 - \frac{4}{3} \frac{\langle r^2 \rangle}{\lambda^2} \right), \] (12.3.37)

whence we have

\[ Z = \sqrt{\frac{\epsilon}{E_0}} \left( 1 + \frac{2}{3} \frac{\langle r^2 \rangle}{\lambda^2} \right). \] (12.3.38)

where

\[ \langle r^2 \rangle = \frac{1}{Z} \int \xi(r) r^2 d^3r. \] (12.3.39)

Substituting (12.3.32) into (12.3.39) and (12.3.38), we obtain

\[ Z = \sqrt{\frac{\epsilon}{E_0}} \left( 1 + \frac{3}{2t} \right). \] (12.3.40)

Corrections to (12.3.33) obtained with the help of (12.3.38) are of order \( t^{-1} \) and can be neglected. Substituting (12.3.40) into (12.3.36), we find
\[ \Omega \{ \xi \} = \left( \frac{\epsilon}{E_0} \right)^{1/2} (t-1). \]  

(12.3.41)

Substituting (12.3.41) into (12.3.15) yields a result of the form (12.2.18) with

\[ D(c) = \left( \frac{2}{3\pi} \right)^{3/2} t^{3/2} e^{-1}, \]  

(12.3.42)

where \( t \) is determined by (12.3.33).

It should be noted, however, that the exponential decay of \( \xi(r) \) described by (12.3.32) takes place only for \( r < \lambda \). At larger distances the fall is more gradual. For \( r >> \lambda \) we find from (12.3.26)

\[ \chi(r) = \frac{\beta e^2}{kr} e^{-r/r_0}. \]  

(12.3.43)

Substituting \( \beta e^2/kr \) from (12.3.31), we see that \( \chi(r) \ll 1 \) for \( r >> \lambda \), and hence

\[ \xi(r) = N t \left( \frac{\lambda}{r} \right) e^{-r/r_0} \quad (r >> \lambda). \]  

(12.3.44)

The potential produced by this portion of the charge at the center of the fluctuation leads to a classical level shift by the amount

\[ \frac{e^2}{\kappa} \int \frac{\xi(r') d^3 r'}{r} \approx \frac{e^2}{\kappa} t N \lambda r_0. \]  

(12.3.45)

(The shift can be regarded as classical since the main contribution to the integral comes from the region where \( r \approx r_0 >> \lambda \).) The theory described is valid provided the right-hand side of (12.3.45) is small compared to \( \epsilon \). A necessary condition for this is \( \epsilon > E_1 \) where \( E_1 \) is given by (12.2.20). It is, of course, also necessary that \( \lambda < r_0 \), i.e., that \( \epsilon > E_F \). These two conditions ensure a large value of the expression under the logarithm (12.3.33). We thus conclude that the validity range of (12.2.18) was estimated correctly in the preceding section.

**12.4 The Spectrum of the Minority Carriers**

As will be shown in the next section, it is the minority-carrier spectrum which determines the frequency dependence of the interband light absorption in degenerate semiconductors at frequencies below the threshold. The problem of calculating the density of states has, in this case, two important peculiarities:

1) The dominant impurities represent centers of repulsion for minority carriers (e.g., a charged donor for holes).
2) The effective masses of electrons and holes are usually quite different. This adds a lot of variety to the problem [11.8]. Our derivation of (12.2.12) was essentially dependent on the condition \( \gamma > E_s \equiv \hbar^2/m_r \), which ensured that a typical potential well of size \( r_0 \) contained quantum levels. When dealing with minority carriers, we must use their mass in the expression for \( E_s \), but in the expression for \( r_0 \) we must substitute the mass of majority carriers, which effect the screening. Below we shall analyze the various situations which can arise in this case.

### 12.4.1 Classical Case (\( \gamma \gg E_s \))

Consider the spectrum of holes in an \( n \)-type semiconductor with the assumption that the hole mass \( m_h \) is heavier than the electron mass \( m_e \). First, assume that there is no compensation at all, i.e., that there are only donor impurities present, and furthermore assume that the heavy-doping condition, \( Na^3 \gg 1 \), is fulfilled for electrons (\( a = \kappa \hbar^2/m_e e^2 \) is the electron Bohr radius). In this case it is easier to satisfy the condition \( \gamma \gg E_s \), which takes the form \( \hbar^2/m_h r_0^2 \ll \gamma \), or \( (m_h/m_e)(Ne_0^3)^{1/6} \gg 1 \).

Since the donor potential is repulsive for holes, deep hole levels are produced by donor voids and not clusters. In Gaussian statistics both voids and clusters are equally probable. It is therefore easy to see that our derivation leading to (12.2.12) remains valid. According to (11.2.10), Gaussian statistics are applicable for \( \epsilon << \bar{V} \). For higher energies the void-cluster symmetry breaks down.

In the absence of acceptors the hole spectrum does not extend arbitrarily deep into the forbidden band, and has a cutoff point. This point lies above the top of the valence band of an impurity-free crystal, separated by an energy of order \( \mu \), where \( \mu \) is the electron Fermi level measured from the bottom of the conduction band. Hole levels close to the cutoff point arise in regions of space where there are virtually no donors. Such a fluctuation, in fact, represents a macroscopic impregnation of an intrinsic semiconductor. At zero temperature the Fermi level in the volume of this fluctuation is close to the bottom of the conduction band. Since in thermodynamic equilibrium the Fermi level is uniform over the entire system, both the conduction and the valence band must bend locally (Fig. 12.4), so that their position in the fluctuation region is shifted to higher energies by an amount close to \( \mu \). This

![Fig. 12.4. Band bending in a volume depleted of electrons. Occupied states of the conduction band are shaded](image)
band bending simply represents a contact potential at the junction between a pure and a doped semiconductor. Obviously, no fluctuation of the donor concentration can raise the bottom of the valence band by an amount greater than μ.

In order to find the form of the density of states near the cutoff point, we must estimate the minimum size of the donor void. In the case of a junction between a heavily doped and a lightly doped semiconductor the potential variation occurs smoothly over a region which is usually referred to as the space-charge region. For an abrupt junction the length of this region turns out to be of the order of the screening radius on the lightly doped side. In order for the contact potential difference to reach its macroscopic value, the size of the fluctuation must exceed the length of the space-charge region; the latter thus represents the minimum fluctuation size.

Suppose that a fluctuation gives rise to a region where the donor concentration N is substantially lower than the average concentration N. If the size of this region is greater than its screening radius rs, then the resultant band bending ε is related to N by μ - ε = \( \hbar^2 N^{2/3} / m_e \) (cf. Fig. 12.4). According to Sect. 11.1, \( r_s = r_0 (N/N) \). With the help of (12.2.1) (where we let \( k = Nr_s^3 \), \( k \ll k \)) we can estimate the probability of such a fluctuation to be \( \exp(-Nr_s^3) \). Expressing \( r_s \) in terms of \( (\mu - \epsilon) \), we find the hole density of states in the form

\[
\ln \frac{g(\epsilon)}{g(0)} \approx -Nr_0^3 \left( \frac{\mu}{\mu - \epsilon} \right)^{3/4} \tag{12.4.1}
\]

The energy \( \epsilon \) is measured upward from the edge of the valence band of a pure crystal. Note that in the derivation of (12.4.1) the electron gas in the fluctuation is assumed to be ideal. A necessary condition for this to be true is \( \tilde{N}a^3 \gg 1 \), or \( \mu - \epsilon \gg E_0 \) (all these formulae involve the electron effective mass). Since \( E_0 \ll \mu \), the decay of \( g(\epsilon) \) is described by (12.4.1) in a rather wide interval.

If we allow for the presence of compensating acceptors in the semiconductor, then the density of states will be nonvanishing everywhere in the forbidden gap. At a sufficiently high value of \( \epsilon \) we can use (12.2.18), where \( N, a, \) and \( E_0 \) should be replaced by \( N_A, a_h = \hbar^2 k / m_h e^2 \), and \( E_h = \hbar^2 / 2m_h a_h^2 \), respectively (details can be found in [12.9]).

### 12.4.2 Quantum Case (γ << Ec)

In typical semiconductors the mass of heavy holes is much larger than that of electrons. The quantum condition \( \gamma \ll Ec \) is, therefore often realized for the electron spectrum in a \( p \)-type semiconductor. We shall again be assuming the heavy doping condition \( Na^3 \gg 1 \). Since this condition is equivalent to \( N_A r_0^3 \gg 1 \), the inequality \( \gamma \ll Ec \) implies that \( r_0 \ll a \), with \( a \) being the electron Bohr radius. (Recall that the screening is governed by holes, and \( Ec = \hbar^2 / me^2 r_0^2 \)). Consequently, a single donor impurity has no bound states.
In the range of Gaussian fluctuations \(|Z| \ll N_A R^3\) we can use (12.2.6) and (12.2.7) with \(Z\) of opposite sign. Negative values of \(Z\) (i.e., regions of lower acceptor concentration) thus correspond to positive \(V\) and \(\epsilon\), which are now measured downward from the bottom of the conduction band.

As already discussed, for \(R >> r_q = \hbar / \sqrt{m_e \epsilon}\) the graph of \(\ln W(R, \epsilon)\) coincides with the curves in Fig. 12.1. Suppose, first, that \(\epsilon < E_s\), i.e., \(r_q > r_0\). Then, for \(R > r_q\), the function \(\ln W(R, \epsilon)\) decreases with increasing \(R\) (Fig. 12.5).

![Graph](image)

**Fig. 12.5.** (Solid line) \(\ln W(R, \epsilon)\), (dashed line) \(\ln W(R, V)\)

If the radius \(R\) of a fluctuation satisfies the condition \(r_0 << R << \hbar / \sqrt{m_e \epsilon}\), then the value of \(\epsilon\) can result only from a narrow well with a shallow level, viz. \(V \approx (\hbar^2 / m_e R^2) >> \epsilon\). Substituting (12.2.7) into this condition and expressing \(Z\) in terms of \(R\), we get

\[
|Z| = \frac{aR}{r_0^2}, \quad \text{and} \quad \ln W(R, \epsilon) = -\frac{Z^2}{N_A R^3} = -\frac{a^2}{N_A R r_0^4}. \tag{12.4.2}
\]

We see that for \(R << \hbar / \sqrt{m_e \epsilon}\), the function \(\ln W(R, \epsilon)\) decreases with decreasing \(R\) (Fig. 12.5). It is easy to show that it also continues to fall when \(R \leq r_0\). The optimum fluctuation radius is hence given by \(R = \hbar / \sqrt{m_e \epsilon}\). Substituting this expression into (12.4.3), we obtain

\[
\ln \frac{g(\epsilon)}{g(0)} = -\left(\frac{\epsilon}{E_0}\right)^{1/2} \frac{E_s}{V} C, \quad \bar{V} = \frac{\epsilon^2}{\kappa r_0} N_A r_0^3. \tag{12.4.4}
\]

Expression (12.4.4) is well known in the theory of disordered systems. It corresponds to a Gaussian potential with vanishing correlation radius (white noise). A wide variety of theoretical methods have been applied to the derivation of this expression [12.1-9,14].

The numerical coefficient \(C\) was found [12.1,14] to be \(C \approx 3/16\pi\). The white-noise problem is the only three-dimensional problem in which the preexponential factor in the density-of-states tail was found exactly by the optimum fluctuation method. The resultant \(g(\epsilon)\) is of the form:
\[ g(\varepsilon) = \frac{G}{E_0} \left( \frac{E_s}{V} \right)^2 \left( \frac{2m\varepsilon}{\hbar} \right)^{3/2} \exp \left[ -\left( \frac{\varepsilon}{E_0} \right)^{1/2} \left( \frac{CE_s}{V} \right) \right], \]  

(12.4.5)

where \( G \) is a numerical coefficient, first evaluated in [12.8]. The law (12.4.5) governs the minority-carrier density of states for HDS in the quantum case.

Thus, in contrast to the classical case, where the exponential decay of \( g(\varepsilon) \) begins at \( \varepsilon = \gamma \) and is described by (12.2.12), in the quantum case this decay begins at \( \varepsilon = E_0(\overline{V}/E_s)^2 \) and is given by (12.4.5).

The further behavior of \( g(\varepsilon) \) depends on the relation between \( E_s \) and \( \overline{V} \). If \( E_s \ll \overline{V} \), then the lengths \( r_q \) and \( r_3 \) are comparable when \( \varepsilon = E_s \), and (12.4.4) goes over into (12.2.12). From there on \( g(\varepsilon) \) behaves just as in the classical case. If, on the other hand, \( E_s \gg \overline{V} \), then the Gaussian statistics are no longer valid for \( r_q > r_0 \), and (12.2.12) does not hold at all. In the absence of attractive centers the electron spectrum has a cutoff at \( \varepsilon \approx \overline{V} \). In the presence of compensation the density of states does not vanish at high energies, being governed by fluctuations of donors. The asymptotic energy dependence is still described by (12.2.18). The transition from (12.4.4) to (12.2.18) occurs in a rather complicated fashion (studied in detail in [12.9]).

### 12.5 The Theory of Interband Light Absorption

#### 12.5.1 On the Relation Between the Density of States and the Interband Light Absorption Coefficient (ILAC)

Interband light absorption experiments near the fundamental absorption threshold represent an important tool for studying fluctuation levels. Figure 12.6 shows the typical experimental data obtained for \( n \)-GaAs.

We see that the ILAC falls by 4 orders of magnitude as the energy of light quanta is reduced even slightly below the forbidden gap width \( (E_g = 1.51 \text{ eV at } T = 77 \text{ K}) \). With increasing electron concentration the curves are shifted towards shorter wavelengths and become less steep. This shift is associated with filling of the conduction band (the Moss-Burstein effect). In \( n \)-type HDS at low temperatures all electron states below the Fermi level are occupied, and the transition occurs between the top of the valence band and the Fermi level (Fig. 12.7). The threshold energy thus turns out to be greater than \( E_g \) by an amount which equals either the Fermi energy \( \mu \) (if we include indirect transitions, i.e., those which require a change in the momentum) or \( \mu(1 + m_e/m_h) \) (if only direct transitions are included; cf. Fig. 12.7). The indirect transitions are made possible by the presence of impurities, and in the case of HDS their probability is not likely to be much lower than that of the direct transitions. In what follows, therefore, when speaking of the threshold energy we shall mean the minimum energy of an indirect transition.
At finite temperatures the ILAC can differ from zero even at frequencies below the threshold, owing to holes in the Fermi sea. Since the probability of a hole with energy \( \mu - \epsilon \) is proportional to \( \exp(-\epsilon/kT) \), the ILAC falls with the frequency \( \omega \) as \( \exp((\mu+E_g-H\omega)/kT) \).

This mechanism of absorption becomes less and less probable as the temperature is lowered.\(^4\) The only absorption mechanism which persists at zero temperature corresponds to transitions from the fluctuation levels above the top of the valence band. (For concreteness, we are still considering an \( n \)-type semiconductor.)

The absorption of a light quantum of energy \( \hbar \omega < E_g + \mu \) leads to the formation of a hole whose energy lies above the unperturbed valence band edge by an amount not less than \( \epsilon_h = E_g + \mu - \hbar \omega \). Figure 12.8

3. If only direct transitions are taken into account, then the ILAC is given by

\[
\exp\left(-\mu - \frac{m_h}{m_e+m_h} (E_g-H\omega) / kT \right).
\]

4. Similarly, the absorption mechanism corresponding to transitions with a simultaneous absorption of phonons is also frozen out at low temperatures.
schematically shows such transitions in the presence of a fluctuating potential which bends the edges of both bands. The number of fluctuation-induced levels for holes decreases exponentially with $\epsilon_h$, which results in an exponential decay of the ILAC with decreasing frequency.

However, the ILAC is proportional not only to the probability of the existence of a deep hole level, but also to that of the interband transition. The latter may also be exponentially low, since the hole level is produced by local clustering of negatively charged acceptors, and the electron wave function can be small in this location. As was shown in Sect. 12.2, electronic states near the Fermi level undergo only a weak distortion from Gaussian fluctuations of the potential (this fact allowed us to regard the screening as linear). In this case the probability of a transition to the Fermi level is not small, and hence the exponential factor in the ILAC can be obtained by merely replacing $\epsilon$ by $\epsilon_h = E_g + \mu - h\omega$ in the density-of-states exponent.

In the region of non-Gaussian fluctuations the dominant contribution to the hole density of states results from clusters of acceptors forming multicharge ionic nuclei for which $\epsilon$ is the ground-state energy. In this case, the electron wave function of energy $\mu$ may be significantly depleted for $\mu < \epsilon$. However, this depletion occurs in a relatively small region of space, and it can be shown [12.10] not to affect the ILAC frequency dependence.

We conclude that in a heavily doped and low-compensated semiconductor with degenerate current carriers, the frequency dependence of the ILAC reproduces the minority-carrier density of states (e.g., holes for the above example of an $n$-type semiconductor).

12.5.2 Light Absorption Induced by Gaussian Fluctuations in the Absence of Carrier Degeneracy

Consider a heavily doped semiconductor whose Fermi level lies deep in the forbidden gap. This may be due either to a high temperature or high compensation. A related situation happens when the absorption of light quanta induces electron transitions into an empty band whose extremum is at $\Gamma$, while the filled Fermi sea of electrons is located at a side extremum. In this case (which is realized, e.g., in germanium) the Fermi level lies well below the bottom of the empty band into which electrons make their transition.
Next, let the sample contain both charged donors and acceptors with a total concentration \( N_t = N_D + N_A \). The impurity potential will be regarded as coulombic at short distances, and then rapidly decaying beyond a certain distance \( r_0 \). Such a decay may result either from electron screening or from correlations in the impurity configuration. The only important assumptions in what follows are (i) that impurity correlations for Gaussian fluctuations of size \( R << r_0 \) can be neglected and (ii) that length \( r_0 \) is large enough so that the fluctuating potential can be regarded as classical (in the terminology of Sect. 12.4) both for electrons and holes, i.e.,

\[
\frac{\hbar^2}{m_e r_0^2} << \gamma, \quad \frac{\hbar^2}{m_h r_0^2} << \gamma, \quad \gamma = \frac{e^2}{\kappa r_0} (N_t r_0^3)^{1/2}.
\] (12.5.1)

In this case the density of states falls according to the law (12.2.12), and the optimum fluctuations represent wells of size \( r_0 \) containing many levels. It is easy to see that in this case the frequency dependence of the absorption coefficient does not reproduce the density of states. This is evident from Fig. 12.9.

![Fig. 12.9. Absorption of a quantum of deficit \( \Delta \) in a nondegenerate semiconductor. Straight solid horizontal lines indicate the edges of unperturbed bands, curved lines correspond to boundaries of the bands bent by the impurity potential.](image)

Suppose we are studying a transition which results in the formation of an electron of energy \( \epsilon_e \) and a hole of energy \( \epsilon_h \) (these energies are measured upward from the respective band edges). In this case the deficit of the absorbed quantum equals \( \Delta = \epsilon_h - \epsilon_e \). (The deficit of a quantum is defined as the amount by which its energy is short of the forbidden gap width, i.e., \( \Delta = E_g - \hbar \omega \).)

As seen from Fig. 12.9, the electron and the hole generated by a quantum with \( \Delta > 0 \) are spatially separated. Indeed, the region to the right of Point \( A \) is classically inaccessible to the electron, and that to the left of Point \( B \) is classically inaccessible to the hole. The absorption coefficient involves an overlap integral of the electron and hole wave functions. Let the electron and hole masses tend to infinity for a fixed \( r_0 \). The probability of tunneling in this case will vanish, and hence no quanta with \( \Delta > 0 \) will be absorbed. At the same time the density of states depends only on the probability of having the necessary potential well, and is expressed by the classical formula (12.2.12), into which the mass does not even enter. Thus the absorption problem in this case obviously requires special consideration. We are essentially dealing here
with the Franz-Keldysh effect in a random field of charged impurities (instead of the usual case of an external electric field).

We shall solve this problem by the optimum fluctuation method [12.16], i.e., we shall find the impurity configuration which makes the largest contribution to the absorption of light with a given deficit. At large deficits, the absorption coefficient \( K(\Delta) \) is exponentially small, and the maximum corresponding to the optimum fluctuation is very sharp. Therefore the leading term in \( \ln K(\Delta) \) is determined by the contribution of the optimum fluctuation. This method, of course, does not allow us to determine the preexponential factor in \( K(\Delta) \).

We shall describe here a simplified derivation, analogous to the uniformly charged sphere approximation of Sect 12.2, which gives the exponent in \( K(\Delta) \) only to within a numerical factor. This factor can be determined with the help of a theory developed in [12.16,17].

Suppose that a fluctuation of the impurity concentration in a volume of linear dimensions \( R \) gives rise to a uniform electric field \( E \). We assume that \( R \) and \( E \) are related by \( eER = \Delta \). This is necessary in order for a quantum of deficit \( \Delta \) to be absorbed in the field of this fluctuation. (Of course, such a quantum can also be absorbed when \( R > \Delta/eE \), but a fluctuation of this size is certainly less probable.) The excess number of impurities \( Z \) necessary to produce the required fluctuation is determined from the condition \( E = Ze/\kappa R^2 \), viz.

\[
Z = \frac{E \kappa R^2}{e} = \frac{\kappa R \Delta}{e^2}.
\]

Numerical factors, which depend on the shape of the fluctuation, have been omitted here. The contribution to the absorption coefficient from such a fluctuation is proportional to

\[
\exp \left( -\frac{Z^2}{N_t R^3} \right) \exp \left( -\frac{R \sqrt{m_e \Delta}}{\hbar} \right).
\]  

(12.5.2)

The first factor in (12.5.2) represents the probability of finding \( Z \) excess impurities in a volume \( R^3 \). The second factor corresponds to the probability of electron tunneling through the distance \( R \) (we assume for simplicity that holes are heavier than electrons, so that hole tunneling can be neglected).

Expressing \( Z \) in terms of \( R \) and \( \Delta \), we rewrite (12.5.2) in the form

\[
\exp \left( -\frac{\kappa^2 \Delta^2}{e^4 N_t R} - \frac{R \sqrt{m_e \Delta}}{\hbar} \right),
\]  

(12.5.3)

whence it is evident that for a given \( \Delta \) the tunneling probability grows with decreasing \( R \), whereas the probability of formation of the fluctuation decreases.
The size of the optimum fluctuation $\tilde{R}$ is found by maximizing the exponent in (12.5.3):

$$\tilde{R} = a \left( \frac{\Delta}{E_0} \right)^{3/4} \frac{1}{(N_t a^3)^{1/2}},$$

(12.5.4)

where $a$ is the electron Bohr radius. Substituting (12.5.4) into (12.5.3) we find the contribution to $K(\Delta)$ from the optimum fluctuation. In accordance with the above discussion we have, finally,

$$\ln \frac{K(\Delta)}{K(0)} = -\beta \left( \frac{\Delta}{E_0} \right)^{5/4} \frac{1}{(N_t a^3)^{1/2}}.$$  

(12.5.5)

Here $\beta$ is a numerical factor which cannot be found by such a simplified argument. In [12.16] it is shown that if $m_h \gg m_e$, then $\beta = 2/5\sqrt{\pi}$. Subsequently, Merkulov and Perel' [12.17] showed that when the electron and hole masses are comparable, one must use a reduced mass $m_h m_e / (m_h + m_e)$ in (12.5.5), while $\beta$ remains unchanged. From (12.5.5) it is evident that the behavior of the absorption coefficient is substantially different from that of the density of states (12.2.12).

Let us investigate the range of validity of (12.5.5). The Coulomb nature of the impurity potential (at least for distances of order $\tilde{R}$) was essential for the above derivation. The result is valid therefore only if $\tilde{R} < r_0$, or, taking into account (12.5.4), if

$$\Delta < \Delta_1, \text{ where } \Delta_1 = E_0 \left( \frac{r_0}{a} \right)^{4/3} (N_t a^3)^{2/3}.$$  

(12.5.6)

On the other hand, validity of (12.5.5) requires the condition $[\ln[K(\Delta)/K(0)]] \gg 1$, or

$$\Delta \gg \Delta_2, \text{ where } \Delta_2 = E_0 (N_t a^3)^{2/5}.$$  

(12.5.7)

The first of the inequalities (12.5.1) ensures that $\Delta_1 \gg \Delta_2$, i.e., that a range exists where (12.5.5) is valid, and the larger $r_0$ is, the wider the range is.

As can be seen from Fig. 12.9, the fluctuation we have considered consists of a potential well for an electron and a well located nearby for a hole. It was assumed in the derivation that the electron energy is lowered and the hole energy raised by the amount $\Delta$, i.e., that each well has a level whose energy is of the order of its depth. This is ensured by the inequality $(\hbar^2 / m_e \tilde{R}^2) \ll \Delta$, which follows from (12.5.7) and hence imposes no further restrictions.

Consider now what happens when $\Delta > \Delta_1$. If $\tilde{R} > r_0$ then for $R = r_0$ the second term in (12.5.3) is small compared to the first and tunneling is negligible. As discussed in Sect. 12.2, the probability of having a potential well of size $R$ and depth $\Delta$ decreases with increasing $R$ when $R > r_0$ (the
potential of remote impurities does not reach the center). Therefore the size of an optimum cluster equals \( r_0 \), like that in the theory of the density of states. Substituting \( R = r_0 \) into (12.5.3), we find

\[
\ln \frac{K(\Delta)}{K(0)} = -\frac{\Delta^2}{\gamma^2} \quad (\Delta > \Delta_1),
\]

(12.5.8)

which reproduces the density of states. For some parameters it is possible that already for \( \Delta < \Delta_1 \), non-Gaussian atomlike fluctuations will make a larger contribution to the ILAC than the above-considered Gaussian fluctuations. In this case there will be no validity range at all for (12.5.8).

In conclusion we would like to mention that the optimum fluctuation method described above was used by Merkulov and Perel' [12.17] to develop a theory for electroabsorption in non-degenerate HDS.

### 12.5.3 Discussion of Experimental Results

At frequencies below threshold, the experimentally observed ILAC in semiconductors is usually well described by the following expression:

\[
\ln \frac{K(\omega)}{K(\omega_{th})} = -\frac{\hbar(\omega_{th} - \omega)}{\Delta_0},
\]

(12.5.9)

where \( \hbar \omega_{th} \) is the threshold energy and \( \Delta_0 \) is a characteristic energy independent of the frequency. The first difficulty we encounter is that none of the above-considered cases of absorption induced by fluctuation levels involved a linear dependence \( \ln K(\omega) \propto (\omega - \omega_{th}) \). Moreover, we do not know of any reasonable way to obtain such a dependence at low temperatures. To put it more precisely, a linear dependence can be obtained as an interpolation between two regions where the ILAC obeys different laws, but as far as we know, no consistent combination of parameters gives rise to a law of the form (12.5.9) which would be valid by itself in a wide frequency range. It is possible, however, that such a law is not required to explain the experimental data, since all measurements have been carried out in a rather narrow frequency interval, in which dependences like \( \Delta ^{5/4} \) and \( \sqrt{\Delta} \ln \Delta \) are hard to distinguish from linear ones.

Unfortunately, as yet there is no good quantitative theory which would explain the ILAC tails. In our view, in degenerate \( n \)-GaAs samples, which have been studied most extensively [12.18], these tails are associated with fluctuations of the acceptor concentration and with hole levels above the top of the valence band [12.9,19].

This conclusion should not be regarded as final, however. The hypothesis that the ILAC tails are dominated by minority impurities is difficult to verify quantitatively, because of the experimental uncertainty in the degree of compensation when it is low.
Radiative recombination experiments in epitaxial GaAs carried out by Alferov et al. [12.20], also suggest the importance of the role played by minority impurities in the formation of density-of-states tails.

A different approach to describing experiment was proposed by Dow and Redfield [12.21]. They suggested that the ILAC tail should by understood as a consequence of the Franz-Keldysh effect in a random electric field. Their results, however, differ from those obtained in Sect. 12.5.2 for the following reasons: (1) Dow and Redfield included the interaction between the electron and the hole generated during an interband transition. According to their numerical calculations, the dependence of ILAC on the quantum deficit $\Delta$ in a uniform electric field $E$ is of the form:

$$K(\Delta) \propto e^{-\text{const} \Delta / E},$$  \hspace{1cm} (12.5.10)

which differs substantially from the Franz-Keldysh asymptotics,

$$K_E(\Delta) \propto \exp \left( - \frac{4\sqrt{2m} \Delta^{3/2}}{3\hbar eE} \right).$$  \hspace{1cm} (12.5.11)

The latter does not take into account the electron-hole interaction. (2) The authors assumed a Gaussian distribution function for the random field. The merit of their work is that (12.5.10) allows an explanation of the linear dependence (12.5.9), but this very fact also gives grounds for the main objection to their theory.

Merkulov and Perel' [12.22] obtained an analytic solution to the ILAC problem for the case of a uniform electric field, including the electron-hole interaction. They showed that the linear dependence (12.5.10) takes place only in a relatively narrow range of $\Delta$ in the vicinity of $\Delta \approx 2E_{\text{ex}}$, with $E_{\text{ex}}$ being the exciton binding energy. At larger $\Delta$ one still has (12.5.11), which also follows from simple physical arguments. Because of this we feel that the explanation proposed by Dow and Redfield cannot be universal.
13. The Theory of Heavily Doped and Highly Compensated Semiconductors (HDCS)

At sufficiently high degrees of compensation all heavily doped semiconductors undergo a transition from metallic to activated conduction. The purpose of this chapter is to calculate critical values of the compensation at which this transition occurs, as well as the activation energy $\epsilon_1$ in the nonmetallic phase. The results strongly depend on whether the impurity distribution in the semiconductor is correlated or it is of a purely Poisson form.

13.1 Uncorrelated Impurity Distribution

13.1.1 Qualitative Discussion

To be specific, we shall be discussing $n$-type semiconductors. In this case each acceptor gains one electron from a donor and is charged negatively. The remaining conduction-band electrons of concentration $n = N_D - N_A$ are located in a random potential of charged donors and acceptors.

As discussed in Sect. 11.2, the quasiclassical theory, which is based on linear screening, is applicable to HDCS only provided the inequality (11.2.17) holds, i.e., the characteristic magnitude of the random potential is small compared to the Fermi energy. In the two-dimensional case, the potential relief can be pictured as a hilly terrain at the bottom of a sea of electrons; condition (11.2.17) implies that the mountain peaks very seldom rise above the sea level (Fig. 12.8).

Suppose that we let the acceptor concentration $N_A$ increase, approaching $N_D$. The Fermi energy $\mu_0$ then decreases as $n^{2/3}$, while the potential relief rises because of weaker electron screening [increased $r_0$ in (11.2.7)]. In the above two-dimensional picture, the case of high compensation corresponds to a dried-out water basin in which only isolated puddles remain at the deepest spots. As we shall show below, in the three-dimensional case electrons gather together, forming metallic droplets whose density remains finite as $n \to 0$. Like the puddles in a dry basin, these droplets are isolated from one another. Therefore the static conductivity of a macroscopic sample becomes activated in a wide range of temperatures. Figure 13.1 displays experimental data [13.1] which show that the compensation indeed destroys metallic conduction.

In this chapter we describe a theory [13.2, 3] which makes it possible to determine the critical degree of compensation at which metallic conduction disappears, and its dependence on the concentration of dominant impurities;
for the nonmetallic phase the theory gives the dependence of the activation energy on the degree of compensation and on the concentration of the dominant impurities.

We shall see below that the transition between metallic and activated conduction represents a quasiclassical version of the Anderson transition discussed in Chap. 2. The critical electron concentration corresponding to the transition is given by

\[ n_c = \frac{\beta N^{2/3}}{a}, \tag{13.1.1} \]

where \( \beta \) is a numerical coefficient. An order-of-magnitude estimate for \( n_c \) can be obtained by turning the inequality (11.2.17) into an equation. Experimental data indicate that \( \beta \approx 0.5 \) (see Sect. 13.3).

The form of the potential relief and the structure of the electronic states present a much more complicated problem in the case of above-critical compensation, when \( \alpha \gg 1 \) [cf. (11.2.17)]. We encountered this problem in Sect. 3.4 when dealing with a highly compensated but lightly doped semiconductor. The main difficulty is associated with a very large potential produced by long-range fluctuations in the impurity concentration. A typical fluctuation in a volume with linear dimensions of order \( R \) lowers (or raises) the electron energy by the following amount:
\[ \gamma(R) = \frac{e^2(N_t R^3)^{1/2}}{\kappa R} \quad (N_t = N_A + N_D). \] (13.1.2)

As \( R \to \infty \), this quantity diverges, which means that electronic screening must be taken into account even when the electron concentration is very low. This argument can, of course, be directly extended to the HDCS case. The electron potential energy becomes comparable to the Fermi energy already for \( n \approx n_c \). Clearly, for \( n \ll n_c \), the humps of the potential relief will become so large that the electron density will be highly inhomogeneous (Fig. 13.2). The linear screening theory described in Sect. 11.1 therefore becomes totally inapplicable. On the other hand, the main ideas of the nonlinear screening theory developed in Sect. 3.4 remain fruitful in the HDCS case. As shown in Sect. 3.4, fluctuations whose size is greater than \( r_s \) are screened by electrons, and hence the typical value of the electron potential energy fluctuation equals \( \gamma(r_s) \), where

\[ r_s = \frac{N_t^{1/3}}{n^{2/3}}, \quad \gamma(r_s) = \frac{e^2 N_t^{2/3}}{\kappa n^{1/3}} \] (13.1.3)

[cf. (3.4.8) and (3.4.9)]. In the arguments leading to (13.1.3) the electron density of states was averaged over long distances. The result is therefore independent of the local form of the density of states on smaller scales. Thus, in order to justify (13.1.3), we must show that a typical volume \( r_s^3 \) contains many electronic droplets.

The short-range structure of electronic states turns out to be different than in the case of light doping. Inside a well of size \( r_s \), there are other humps and wells of shorter range. Electrons will, of course, fall into these wells. Therefore the electron density must contain harmonics of all wavelengths smaller than \( r_s \). Suppose the shortest range \( L \) of the impurity-potential fluctuations satisfies \( L \ll r_s \). The electrons will then form droplets of size \( L \) located in the deepest short-range potential wells, which in turn are located in the deepest wells of longer range (Fig. 13.2).

---

**Fig. 13.2.** Energy diagram of a compensated semiconductor. The meandering line shows band bending, the upper solid line indicates the position of the bottom of the conduction band in the absence of the impurity potential, the lower solid line shows the Fermi level, and the dash-dotted line the percolation level. Regions occupied by electrons (electronic droplets) are shaded.
The characteristic depth of a potential well of range \( L \) is determined by its impurity charge. The excess number of donors inside such a well is of order \( (N_i L^3)^{1/3} \), whence \( \gamma(L) = e^2 \kappa^{-1} L^{-1} (N_i L^3)^{1/2} \). The number of electrons in a droplet cannot be larger than the excess donor number, otherwise the well will turn into a hump. If the number of electrons in a droplet of size \( L \) equals \( (N_i L^3)^{1/2} \), their concentration in that droplet is given by \( \tilde{n}(L) = (N_i L^3)^{1/2}/L^3 \). With decreasing \( L \) the electron droplets break into smaller ones — of higher electron concentration.

It should be remembered that the number of electrons in a well is limited not only by its impurity charge, but also by the Pauli principle. The number of quantum states in a well of depth \( \gamma(L) \) is given by \( N_q(L) = L^3 \{ m \gamma(L) \}^{3/2}/h^3 \), so that for small \( L \) we have \( N_q(L) < (N_i L^3)^{1/2} \), and the quantum limitation becomes more restrictive. Let us introduce a length \( R_q \) (quantum) defined by \( N_q(R_q) = (N_i R_q^3)^{1/2} \), viz.

\[
R_q = \frac{a}{(N_i a^3)^{1/3}}.
\]

The charge limitation is of importance for wells which are larger than \( R_q \), whereas the quantum restriction dominates for smaller wells. Our discussion above makes sense only provided \( R_q \ll r_s \). If \( \alpha \gg 1 \) [see (11.2.17)], this condition is fulfilled.

The length \( R_q \) can be alternatively defined as the value of \( L \) at which the electron Fermi energy corresponding to the concentration \( \tilde{n}(L) \) becomes comparable to the well's depth \( \gamma(L) \). It is easy to see that fluctuations of the impurity potential on the scale \( R \ll R_q \) do not give rise to smaller droplets. Indeed, humps and wells produced by such fluctuations are of the amplitude \( \gamma(R) \), which is less than the Fermi energy of electrons in the droplet [the latter equals \( \gamma(R_q) \)]. Such humps do not form islands in the electron sea (Fig. 13.2). To put it differently, droplets of size \( R_q \) do not split further, since the smaller droplets together would not have enough quantum states to accommodate all the electrons of the parental droplet.

Thus, we have proven that electrons break down into droplets of size \( R_q \). It can be readily seen that this size is much larger than the mean separation between impurities; moreover, a well of size \( R_q \) and depth \( \gamma(R_q) \) contains many levels [\( \hbar^2/mR^2_q \ll \gamma(R_q) \)]. The electron concentration in a droplet,

\[
\tilde{n} = \frac{(N_i R_q^3)^{1/2}}{R_q^3} = \frac{N_i}{(N_i a^3)^{1/3}},
\]

is independent of the average electron concentration \( n \). With respect to their mutual interaction, the droplet electrons form a weakly nonideal Fermi gas, since \( \tilde{n}a^3 \gg 1 \). The electron kinetic energy is of the order of

\[
\gamma(R_q) = E_0 (N_i a^3)^{4/9},
\]

which is small compared to \( \gamma(r_s) \).
In the above arguments we assumed zero temperature. In practice, they remain valid provided \( kT \ll \gamma(R_q) \). This inequality represents the Fermi degeneracy condition for the electron gas. For \( kT \gg \gamma(r_s) \) we have an ideal Boltzmann gas, and the random potential structure described above is completely destroyed. In the temperature range \( \gamma(R_q) \ll kT \ll \gamma(r_s) \) the long-range structure of the potential persists, but the droplet size depends on temperature. The thermodynamic properties of the electron gas in this range were studied by Gulyaev and Plesskii [13.4].

### 13.1.2 Basic Equations and the Asymptotic Properties of the Potential

From the above qualitative analysis of the structure of the potential and the electronic states, it should be clear that a mathematical description of such a system should be based on a self-consistent equation of the Thomas-Fermi type (11.1.6), where the electron concentration is related to the potential \( \phi(r) \) by (11.1.5) and the excess impurity charge density \( e \xi(r) \) is given by

\[
e \xi(r) = e[N_D(r) - N_D - N_A(r) + N_A].
\]

We can regard \( \xi(r) \) as a Gaussian random function, since qualitatively, the most important contribution to the electron potential results from Gaussian fluctuations. Correlators of \( \xi(r) \) decouple into pairwise products of the form

\[
\langle \xi(r) \xi(r') \rangle = (N_D + N_A) \delta(r-r'),
\]

since we are not including here correlations associated with the interaction between impurities.

Let us introduce a distribution function \( F(V) \) for the potential energy \( V = -e\phi \). The Fermi level \( \mu \) is determined by the following equation:

\[
n = \frac{(2m)^{3/2}}{3\pi^2 \hbar^3} \int_{-\infty}^{\mu} (\mu - V)^{3/2} F(V) \, dV.
\]

The resultant system of equations was solved in Sect. 11.2 assuming \( \alpha \ll 1 \). This condition allowed us to linearize (11.1.6). As should be clear from the above qualitative analysis, such a linearization is not permissible in the case of high compensation. We can only rewrite the system of equations in dimensionless variables and thus formulate a mathematical problem containing no adjustable parameters. The dependence of the relevant quantities on parameters of the problem will thus be entirely contained in the units in which these quantities enter the dimensionless equation. In addition, one can determine the asymptotic properties of \( F(V) \) by using the optimum fluctuation method of Chap. 12.

We introduce the dimensionless variables:

\[
\chi = \frac{2^{2/3} V}{\gamma(r_s)}, \quad x = \frac{2^{1/3} r}{r_s},
\]
\[ \eta = \frac{-2^{2/3} \mu}{\gamma(r_s)}, \quad f(x) = \frac{\xi}{n}, \] (13.1.10)

where \( r_s \) and \( \gamma(r_s) \) are defined by (13.1.3). In these variables we have [cf. (11.1.5), (11.1.6)]

\[ \Delta_x \chi = 4\pi [f(x) + 1 - u(x)], \] (13.1.11)

where

\[ u(x) = \begin{cases} 
\frac{3^{3/2} \alpha^2}{\pi^{1/2} 4 \sqrt{2}} (-\eta - \chi)^{3/2}, & \chi < -\eta, \\
0, & \chi > -\eta.
\end{cases} \] (13.1.12)

The random function \( f(x) \) is Gaussian and has the following correlator:

\[ \langle f(x) f(x') \rangle = 2\delta(x-x'). \] (13.1.13)

All higher-order correlators decouple into products of pairwise correlators.

From (13.1.11) and (13.1.12) it follows that for \( \alpha \gg 1 \) the potential energy \( \chi(x) \) is close to \(-\eta\) in regions where \( u(x) > 0 \). Moreover, in these regions the following condition holds to within terms of order \( \alpha^{-1} \):

\[ u(x) = f(x) + 1. \] (13.1.14)

This means that in regions where electrons are present, they completely neutralize the impurity charge.

Consequently, in the zeroth-order approximation in \( \alpha^{-1} \) we have a peculiar mathematical problem: to find those regions in space where the impurity charge is completely neutralized by electrons. This problem can be formulated as follows.

Let the charge density \( \rho(x) \) be a continuous function given in some region of space (possibly in the entire space). The potential \( \chi_0 \) produced by \( \rho(x) \) is given by

\[ \chi_0(x) = -\int \frac{\rho(x')d^3x'}{|x-x'|}. \] (13.1.15)

It is necessary to find the regions \( O \) possessing the following properties:

1) \( \rho(x) > 0 \) everywhere in \( O \).
2) The potential \( \chi(x) \) produced by all charges outside \( O \) is constant everywhere in \( O \) and equal to \(-\eta\) (the electric field vanishes in \( O \)).
3) \( \chi(x) \geq -\eta \) everywhere outside \( O \).

We are interested in the case when \( \rho(x) \) represents the total impurity charge \( (\rho = f + 1) \), and the chemical potential \( \eta \) is determined by the
neutrality condition. The latter means that there is no net uncompensated charge (i.e., the total net charge outside $O$ vanishes).

We have just formulated a new potential-theory problem. A similar problem for gravitation could be posited thus: inside a body of arbitrary shape and arbitrary mass distribution, cut a cavity in such a way that all the masses remaining outside the cavity, combined, produce inside it a given constant potential (the force of gravity should be zero at all points inside the cavity).

It can be proven [11.16] that the solution to this problem is unique, if it exists. One can also write down a functional [13.2] which is bounded from below and has its minimum at the sought function $u(x)$. With the help of this functional one can prove certain theorems related to the existence of a solution [11.16].

The system of equations (13.1.11), (13.1.13), and (13.1.14), which determines the chemical potential $\eta$ to zeroth order in $\alpha^{-1}$, contains no parameters. Therefore the characteristic amplitude of the electrostatic potential is of order $\gamma(r_s)$, and so is the energy by which the Fermi level is lowered with respect to the average value of the potential, as explained in Sect. 13.1.1. According to (13.1.10), we have

$$\mu = -\eta \frac{e^2 N_B^{2/3}}{\kappa n^{1/3}}. \quad (13.1.16)$$

The optimum fluctuation method described in Chap. 12 permits us to find the asymptotic form of the distribution function $F(V)$ both at large positive and large negative values of $V$. To this end, we shall use (13.1.11-13). The probability that the Gaussian random function defined by (13.1.13) has a given value is proportional to $\exp(-\Omega(f))$, where

$$\Omega(f) = \frac{1}{4} \int f^2 dx. \quad (13.1.17)$$

This result also follows from (12.3.16), where we must replace $N$ by $N_t$ and substitute (13.1.10).

The essence of the optimum fluctuation method is to find a function $f$ which minimizes the functional $\Omega(f)$, subject to the additional constraint that the functional $\chi(f)$ [which represents a solution to Poisson's equation (13.1.11) at some point in space] equals $\chi$. Then, if $\Omega(f) \gg 1$, we have

$$\ln \left[ \frac{F(\chi)}{F(0)} \right] = -\Omega(f). \quad (13.1.18)$$

We place the coordinate origin at a point where $\chi(f) = \chi \gg 1$. Note that owing to spherical symmetry the extremal function should have no angular dependence. In regions where $\chi > -\eta$ there are no electrons and we can drop the last term in (13.1.11). Electrons begin to appear only at a distance $x_0 \gg 1$ from the fluctuation's center, where the potential is nearly vanishing. They rapidly (at a distance of order unity) neutralize the fluctuation, so that
the region filled with electrons makes only a small contribution to the potential \( \chi(f) \) of the center of fluctuation. In spherical coordinates we thus have

\[
\chi(f) = - \int_{x < x_0} \frac{(f + 1) d^3x}{x} = -4\pi \int_0^{x_0} (f + 1) x \, dx.
\]  

(13.1.19)

The potential must nearly vanish at \( x_0 \). In the spherically symmetric case this implies that the net charge in the entire region \( x < x_0 \) must be zero, i.e.,

\[
\int_0^{x_0} (f + 1) x^2 \, dx = 0.
\]  

(13.1.20)

The extremal function \( \tilde{f} \) is determined from the following equation:

\[
\delta (\Omega(f) + \tau \chi(f)) = 0,
\]  

(13.1.21)

where \( \tau \) is a Lagrange multiplier to be found from the condition \( \chi(\tilde{f}) = \chi \). Varying (13.1.19), we obtain

\[
\delta \chi(f) = -4\pi [1 + f(x_0)] x_0 \delta x_0 - 4\pi \int_0^{x_0} \delta f \, x \, dx.
\]  

(13.1.22)

The variation \( \delta x_0 \) is found by varying (13.1.20), viz:

\[
[f(x_0) + 1] x_0^2 \delta x_0 = - \int_0^{x_0} \delta f \, x^2 \, dx.
\]  

(13.1.23)

Finally, varying \( \Omega(f) \) and substituting the resultant expressions into (13.1.21), we find

\[
\tilde{f} = \frac{2\tau}{x} \left[ 1 - \frac{x}{x_0} \right].
\]  

(13.1.24)

From (13.1.20) we have \( \tau = -x_0 \), and from the condition \( \chi(\tilde{f}) = \chi \) we get \( x_0 = \sqrt{\chi/2\pi} \). Substituting (13.1.24) into (13.1.18), we have for \( V \gg \gamma(r_s) \):

\[
\ln \left( \frac{F(V)}{F(0)} \right) = -\frac{\sqrt{2}}{3\sqrt{\pi}} \chi^{3/2} = -\frac{\sqrt{2}}{3\sqrt{\pi}} \left( \frac{V \kappa n^{1/3}}{e^2 N_B^{2/3}} \right)^{3/2}.
\]  

(13.1.25)

Next, we consider the function \( F(V) \) at large positive values of the quantity \( \epsilon = \mu - \nu \), i.e., at values of the potential energy lying much below the Fermi energy. In this case we shall confine ourselves to the uniformly charged sphere approximation (Sect. 12.2), since the numerical coefficient in \( \Omega(f) \) is unknown anyway. According to (12.2.10), the probability that a Gaussian potential fluctuation within a sphere of radius \( R \) will lower the electron potential energy by the amount \( \epsilon \) is given by
\[
\exp \left\{-\frac{\epsilon^2}{\left(\frac{e^2}{kR}N_R R^3\right)^{1/2}}\right\},
\]

provided the impurity charge is not screened by electrons. The larger \( R \) is, the larger this probability is, and hence the optimum fluctuation is determined by the electron screening radius. If the bottom of the well is below the Fermi level by the amount \( \epsilon \), then the electron concentration in the well is given by \( \bar{n} \approx \left(\frac{\sqrt{m_e}}{h}\right)^3 \). According to (11.1.8), the screening radius is then of order \( \tilde{r}_s = a^{1/2}\bar{n}^{-1/6} \). Larger fluctuations \((R > \tilde{r}_s)\) in the impurity concentration are neutralized by electrons. Therefore we should substitute \( R = \tilde{r}_s \) into (13.1.6), which gives

\[
\ln \left[ \frac{F(\epsilon)}{F(0)} \right] = - \left[ \frac{\epsilon}{\gamma(R_q)} \right] \lambda. \tag{13.1.27}
\]

A method for determining the numerical coefficient \( \lambda \) was formulated in [13.3].

### 13.2 Correlated Impurity Distribution

There is no doubt that for a sufficiently high compensation, electronic states also become localized in the case of a non-Poisson impurity distribution. There must certainly be a metal-insulator transition in this case, but the critical concentration \( n_c \), as well as the length and energy scales characterizing the potential distribution in an HDCS, may have a strong dependence on correlations in the impurity distribution.

In the work of Keldysh and Proshko [11.7] it is shown that correlations in the configuration of impurities give rise to a peculiar nonelectronic screening. The key idea of this work is that the impurity distribution in a compensated sample obtained by pulling from the melt can be viewed as snapshot of the distribution which existed at the freeze-out temperature \( T_0 \) for the impurity diffusion. The impurity diffusion coefficient decays exponentially with the temperature, with a large activation energy. The value of \( T_0 \) is therefore quite well defined. In the high-temperature plasma, charged donors and acceptors screen each other. Because of this correlation, the fluctuation of charge in a large volume proves to be much smaller than that corresponding to a Poisson distribution of charged centers. Therefore, the results of the preceding section should be revised if the length \( r_s \) introduced there happens to be larger than the impurity screening radius \( r_i \).

If one assumes that at the temperature \( T_0 \) impurities form a weakly nonideal plasma, one can easily evaluate the fluctuating potential distribution. In an ideal plasma the potential distribution produced by donors and acceptors is Gaussian, viz.

\[
F(V) = (\sqrt{\pi}\gamma_i)^{-1} e^{-\gamma^2/\gamma_i^2}, \quad \gamma_i^2 = 16\pi \frac{e^4 N_D r_i}{\kappa^2}, \tag{13.2.1}
\]
where \( r_i = (\kappa k T_0 / 8\pi N_D e^2)^{1/2} \) is the Debye screening radius for plasma. [An estimate for \( \gamma_i \) can be obtained by substituting \( R = r_i \) into (13.1.2).] According to Keldysh and Proshko, this is the potential in which electrons are found at low temperatures. From the theory developed in the preceding section it is easy to see that the potential produced by electrons themselves can be neglected, provided \( r_i < R_q, r_s \). In this case the occupation of wells of size \( r_i \) is governed by the Pauli principle, and the total charge of electrons which a well can accommodate is small compared to the impurity charge of the well. If \( r_i \) satisfies the condition \( \hbar^2/\pi r_i \ll \gamma \), then the typical wells contain many levels and the Fermi level is determined by (13.1.9), where one must substitute (13.2.1). If \( \hbar^2 n^{2/3} / m \ll \gamma \), then one has isolated electron droplets. In this case, from (13.1.9) one can easily derive an equation for \( \mu \):

\[
\eta^2 = \ln \left[ \frac{(m \gamma_i)^{3/2}}{8\pi^2 \hbar^3 n} \right],
\]

where \( \eta = -\mu/\gamma \). For very low \( n \) one has

\[
\eta \approx \left\{ \ln \left[ \frac{(m \gamma_i)^{3/2}}{8\pi^2 \hbar^3 n} \right] \right\}^{1/2}.
\]

It is easy to evaluate the electron concentration \( \bar{n} \) in a droplet. To do this one must find out which values of \( \mu - V \) are important in the integral (13.1.9). The corresponding values of the concentration are given by \( n \approx [m(\mu - V)]^{3/2} / \hbar^3 \). As a result, one finds (to within logarithmic factors):

\[
\bar{n} = N_D (N_D a^3)^{-1/4} \left( \frac{r_i}{a} \right)^{3/4}.
\]

As for a random impurity distribution, \( \bar{n} \) does not depend on \( n \). Formulae (13.2.4) and (13.1.5) coincide for \( r_i = R_q \).

If \( R_q < r_i < r_s \), then the longest range of potential fluctuations is still equal to \( r_i \) and the amplitude of typical fluctuations is of order \( \gamma_i \). However, in this case one must take into account effects of the electron potential on the structure of droplets [13.2]. The droplet size and the electron concentration in a droplet turn out to be the same as in the absence of impurity correlations. Lowering of the Fermi level is described by an expression of the form (13.2.3) but with a slightly different term under the logarithm sign. Finally, if \( r_i > r_s \), then the impurity correlations are not important and one can apply the theory of Sect. 13.1.

From the above it follows that any correlation in the impurity distribution will necessarily become important at a sufficiently high degree of compensation, when \( r_s \) becomes equal to \( r_i \). At still higher compensations the range of typical fluctuations remains equal to \( r_i \). Further lowering of the Fermi level is, therefore, proportional to \( |\ln n|^{1/2} \) rather than \( n^{-1/3} \). The
situation is different, however, for narrow-gap semiconductors. At
temperature $T_0$ these materials may have a large number of intrinsic
electrons and holes, which also take part in screening. These carriers
recombine when the temperature is lowered (after the freezing-out of the
impurity diffusion), and the impurity charge becomes partially unscreened.
This phenomenon was considered by Gal'pern and Efros [13.5]. Their main
result consists in the following. Let $p$ be the intrinsic carrier concentration at
$T_0$ and suppose $p << N_D$. The Fourier component of the density correlation
function $\langle \xi(r)\xi(r') \rangle$ is then of the form:

$$\langle \xi_q^2 \rangle = \frac{4N_D(q^2 + Q_p^2)}{V_0(q^2 + r_i^{-2})}, \quad (13.2.5)$$

where $Q_p^2 = \frac{8\pi pe^2}{\kappa k T_0}$ and $V_0$ is the normalization volume. The mean-
square potential energy is given by

$$\gamma_i^2 = 2\langle V^2 \rangle = V_0 \int \frac{dq}{(2\pi)^3} \langle \xi_q^2 \rangle \left[ \frac{4\pi e^2}{\kappa q^2} \right]^2. \quad (13.2.6)$$

If $p = 0$, then $\langle \xi_q^2 \rangle \to 0$ as $q \to 0$, which corresponds to the impurity
screening. Substituting (13.2.5) into (13.2.6) for $p = 0$, we get (13.2.1).
However, for $p \neq 0$ the integral (13.2.6) diverges at small $q$. The reason for
this divergence is that $\langle \xi_q^2 \rangle = 4p/V_0$ for $q \to 0$. At large distances the
correlator is of the form (13.2.8), with $2p$ substituted for $N_D + N_A$. The
situation is as if we had completely unrelated impurities of concentration
$2p$. From Sect. 13.1 we know that for $p \gg n$ these impurities give rise to
potential fluctuations whose range is of order $r_s$ and amplitude of order $\gamma(r_s)$, where

$$r_s = \frac{p^{1/3}}{n^{2/3}} \quad \text{and} \quad \gamma(r_s) = \frac{e^2 p^{2/3}}{\kappa n^{1/3}}. \quad (13.2.7)$$

For $p \gg n^{1/2}(N_D r_i)^{3/4}$ we have $\gamma(r_s) >> \gamma_i$ and the situation becomes
equivalent to that described in Sect. 13.1. Thus, when intrinsic carriers are
taken into account, we find that the correlation does not eliminate the long-
range fluctuations but only reduces the fraction of impurities which
participate in them. Even this reduction, of course, does not take place if
$p > N_D$. In this case the correlation can simply be neglected.

It is not so easy to estimate the value of $p$, since the forbidden gap width
is not very well known at high temperatures. Still, it is probably safe to say
that, e.g., in indium antimonide at $N_D < 10^{17} \text{cm}^{-3}$ the impurity correlation is
completely eliminated by intrinsic carriers.

### 13.3 Kinetic Properties of HDCS

Let us first discuss the static conductivity in HDCS. As was shown in
preceding sections, at a high compensation electrons form droplets which are
isolated from each other. Within each droplet the electron concentration is high and conduction is of a metallic nature. However, these droplets occupy only a small fraction of the total crystal volume and hence do not allow percolation. The droplets are separated from each other by potential barriers of height $\gamma(r_s)$ and thickness $r_s$. The tunneling transparency coefficient for this barrier contains a factor of the form

$$\exp \left[ -\sqrt{m \gamma(r_s)} \frac{r_s}{\hbar} \right],$$

(13.3.1)

which means that for $\alpha \gg 1$ (cf. Sect. 11.2) transmission is very low. Therefore if the temperature is not too low, conduction occurs via thermal excitation of electrons onto the percolation level $V_c$. That level, of course, lies above the Fermi level by an energy of order $\gamma(r_s)$; see Fig. 13.2. For the activation energy $\epsilon_1$ we thus have

$$\epsilon_1 = |V_c - \mu| = \theta_1 \frac{e^2 N^{2/3}}{\kappa n^{1/3}}.$$

(13.3.2)

The numerical coefficient $\theta_1$ is unknown, although in principle it can be evaluated with the help of the dimensionless equations in Sect. 13.1. One can also evaluate the coefficient $\beta$ in the expression (13.1.1) for the critical electron concentration at which the metal-insulator transition takes place. Note that for $\alpha \approx 1$ the potential relief is still quasiclassical. Therefore the transition under consideration represents a quasiclassical version of the Anderson transition. It occurs when the percolation level crosses the Fermi level. The possibility of such a transition was first pointed out by Ziman [5.57].

The above expressions correspond to the case when there is no correlation in the impurity distribution. If $\sqrt{m \gamma_i} r_i/\hbar \gg 1$, then the picture remains qualitatively the same even with correlation. Expressions (13.1.1) and (13.3.2), however, take a different form. If $r_i < R_q, r_s$, then the problem reduces to a single-electron one, as discussed in Sect. 13.2. In this case one can even determine the numerical coefficients. From Sect. 5.2 we know that for a Gaussian random function the critical volume fraction corresponding to the onset of percolation equals 0.17. From this it is easy to find [cf. (5.2.2)] the percolation level, $V_c = -0.67 \gamma_i$. To determine $\mu$, we substitute (13.2.1) into (13.1.9). According to (13.2.3), we find for a high compensation

$$\epsilon_1 \approx \gamma_i \left\{ \ln \left[ \frac{(m \gamma_i)^{3/2}}{8 \pi^2 \hbar^3 n} \right] \right\}^{1/2}.$$

(13.3.3)

Also, from (13.1.9) we find that $\mu = V_c$ for

$$n_c = \frac{\beta (2 m \gamma_i)^{3/2}}{3 \pi^{5/2} \hbar^3},$$

(13.3.4)
where

\[ \beta = \int_{0.67}^{\infty} (x - 0.67)^{3/2} e^{-x^2} \, dx = 0.11. \]

At sufficiently low temperatures the probability of activation to the percolation level becomes less than the probability of tunneling in the vicinity of the Fermi level. This happens when \( T < T_s \), where \( T_s \) is determined by the following condition:

\[ r_s \frac{\sqrt{m \gamma(r_s)}}{\hbar} = \frac{\gamma(r_s)}{kT_s} \quad (13.3.5) \]

or by a similar condition with \( \gamma_l \) and \( r_l \). In this range the conduction is dominated by hopping over droplets. The temperature dependence for this mechanism of conduction was studied in [13.6] without taking the Coulomb gap into account. At a sufficiently low temperature the presence of a Coulomb gap leads to a law of the form \( (10.1.18) \), with the localization radius \( a \) in \( (10.1.19) \) replaced by a length \( \hbar/\sqrt{m |\mu|} \) appropriate for tunneling under a barrier of height \( \mu \).

The metal-insulator transition induced by compensation was observed in n-Ge (Sasaki and Yamanouchi [13.7], Gadzhiev et al. [13.1], Zabrodskii et al. [9.17], and Zabrodskii [9.18,19; 13.8]), in GaAs (Redfield and Crandall [13.9]), in CdTe (Agrinskaya and Krymova [13.10] and Didkovskii and Khivrich [13.11]), and in n-InSb (Gershenzon et al. [13.12], Yaremenko [13.13]).

Although all these experiments were carried out in the range of concentrations where the parameter \( Na^3 \) was either close to or slightly exceeded unity, the experimental dependence \( n_e(N) \) did agree well with (13.1.1). Figure 13.3 [13.8] displays the values of \( n_e \) obtained in all experiments with compensated germanium that are known to us. The dash-dotted line corresponds to (13.1.1) with \( \beta = 0.5 \). We see that this dependence gives a reasonable agreement with the experimental data. As pointed out in [13.10,12], there is good agreement with (13.3.2). Nevertheless, the whole situation should not be considered satisfactory. The trouble is that the linear dependence of \( \ln \rho \) on \( T^{-1} \) is observed experimentally only in a very narrow range. This is hard to explain within the framework of the above theory, although such attempts have been made [9.18,19; 13.6]. The narrowness of the linear region may result from a correlation in the impurity configuration which suppresses all long-range quasiclassical fluctuations. Moreover, at very low temperatures one sometimes sees [13.12,13] a weakening in the temperature dependence of the resistivity, which is also hard to explain.

The metal-insulator transition in HDS can be induced not only by compensation, but also by a magnetic field. The latter promotes localization by squeezing the electronic wave functions. This effect is referred to as magnetic freezing-out of electrons (see also Chap. 7). In the instance of
HDCS the freeze-out effect gives rise to several interesting questions. In particular, one would like to know the dependence of the critical electron concentration \( n_c \), at which metallic conduction is terminated, on the magnetic field. On the nonmetallic side of the transition one would like to know the activation energy \( \epsilon_1(n, H) \). These problems were solved by Shklovskii and Efros [13.16] for the uncorrelated-impurities model, and by Efros and Yanchev [13.17] for a correlated distribution.

The case of uncorrelated impurities gives rise to an interesting theoretical problem of quantum nonlinear screening. So far in this book we have classified screening effects as linear or nonlinear. There is, however, another way of classifying them. The type of screening described in Sect. 11.1 on the basis of a quasiclassical equation can be termed classical. It is characterized by a large screening radius compared to the electron wavelength. In a quantizing magnetic field the wavelength perpendicular to the field is short, while that along the field is long. In this case the electron screening acquires certain interesting features. In particular, screening of a point charge becomes strongly anisotropic. This type of screening, which is naturally termed quantum screening, is described in the review by Horing [13.18] (see also [13.16]). Quantum screening, in turn, can be linear or nonlinear. In [13.16] the nonlinear screening theory described in Sect. 13.1 was generalized to the quantum case.
The strong inhomogeneity in the electron density, which is characteristic of the HDCS does not manifest itself merely in static conduction. Samples of $n$-InSb cooled to helium temperatures are widely used as detectors for submillimeter and millimeter waves. It turns out [13.19] that the detector sensitivity and the magnitude of photoresponse are both significantly enhanced by compensation. There have been many attempts [13.20-23] to explain these properties of $n$-InSb on the basis of the picture described in this section. However, one can only hope for a qualitative understanding here, since most of the samples studied had $N_D = 10^{15}$ cm$^{-3}$, i.e., $N_D a^3 \approx 1$.

Even before the development of the present theory, Gal’pern and Efros [13.20], in an attempt to explain absorption in the microwave (mw) range, proposed a model in which the HDCS represents metallic droplets embedded in a dielectric. This model is supported by the experimental results of Potapov et al. [13.21,24], who measured the Faraday effect at millimeter-wave frequencies and deduced the temperature dependence of the Hall mobility in the range between 4.5 and 50 K. Comparing it with the behavior of the dc Hall mobility, they found that the latter was an order of magnitude lower than the mw mobility. The mw mobility proved to be so high that the authors were able to measure the cyclotron resonance [13.24].

Anomalously low Hall mobility under dc current conditions is a common property of all HDCS at low temperatures. This is undoubtedly related to the inhomogeneity in the electron density. It is still unclear how to interpret this mobility [13.25]. Certainly it does not have the same simple meaning as in a homogeneous system. Since the electric field in millimeter waves is homogeneous over the system, electrons in isolated droplets take part in conduction, and they have a high mobility. Both mobilities coincide at higher temperatures, when the electron density becomes homogeneous.

The presence of high potential barriers in HDCS also allows to explain the effect of persistent (or frozen) photoconductivity. A qualitative explanation of this effect, given by Ryvkin and Shlimak [13.21], consists in the following. Electrons and holes, generated by light, rapidly move in opposite directions and then stay in different regions of space, separated by potential barriers. This strongly inhibits their recombination, giving rise to a long relaxation time for the photocurrent. A quantitative theory for this phenomenon was developed by Tkach [13.29] and Shik [13.30]. It should be also mentioned, finally, that Levanyuk and Osipov [13.31] used the picture of long-range fluctuations to develop a theory of the luminescence in HDCS.

13.4 Completely Compensated Semiconductor

The theory presented in Sect. 13.1, assumes that lowering of the Fermi level due to compensation is less than half the forbidden gap width $E_g$. An interesting question arises: what is the form of electronic states in the case when the compensation is so close that this condition is no longer fulfilled.
Below, we shall first discuss this question qualitatively, and then explain why it is of interest.

Consider a completely compensated semiconductor, containing $N$ donors and $N$ acceptors per unit volume. For the sake of simplicity, we shall assume that the heavy-doping condition is fulfilled both for electrons and holes, i.e., that the concentration $N$ satisfies both inequalities $Na_e^3 \gg 1$ and $Na_h^3 \gg 1$, with $a_e$ and $a_h$ being the Bohr radii for electrons and holes, respectively. In this case the electron and hole states are collectivized (impurity bands merge with the conduction bands), and the impurities are charged positively (donors) or negatively (acceptors). If donors and acceptors were distributed in space uniformly, then at zero temperature electrons and holes would all recombine with each other. The situation would be equivalent to an intrinsic semiconductor, in which the Fermi level lies in the middle of the forbidden gap.

Let us now include fluctuations in the impurity concentration, assuming no correlations. In doing so, we again encounter the problem of screening the long-range fluctuations. In a completely compensated semiconductor at low temperatures, there are no free carriers, either in the conduction or in the valence band. Therefore in this model nothing prevents the amplitude of random potential fluctuations from reaching the value $E_g$. It is easy to see, however, that the potential cannot swing by more than $E_g/2$. Indeed, if the bottom of the conduction band in some region is lowered (due to a positive net impurity charge) below the Fermi level, then electrons will appear in this region and prevent the band from falling much further (Fig. 13.4).

![Energy diagram of a completely compensated semiconductor.](image.png)

**Fig. 13.4.** Energy diagram of a completely compensated semiconductor. The upper and the lower straight lines indicate the unperturbed positions of the bottom of the conduction band and the top of the valence band; the middle line corresponds to the Fermi level. Meandering lines represent the band edges, which are modulated by the electrostatic potential of charged impurities. Percolation levels for electrons and holes are shown by dashed lines. Regions occupied by carriers (droplets) are shaded. $R_e$ and $R_h$ are, respectively, the sizes of electron and hole droplets.
Similarly, a negative net charge cannot raise the bottom of the conduction band by more than $E_g/2$, since in this case the top of the valence band rises above the Fermi level and holes emerge.

Let us determine the smallest size $R_g$ of a typical fluctuation which bends the bands by an amount of order $E_g$. From (13.1.2) we find

$$\gamma(R_g) = E_g, \quad R_g = \frac{E_g^2 k^2}{N e^4}.$$  \hspace{1cm} (13.4.1)

The number of electronic states in a well of size $R_g$ and depth $E_g$ is given by

$$N_q(R_g) = \frac{(m_e E_g)^{3/2} R_g^3}{\hbar^3}.$$  \hspace{1cm} (13.4.2)

The excess number of impurities in such a well equals $Z(R_g) = (NR_g^3)^{1/2}$. Provided the concentration of impurities is small compared to that of host atoms, the following inequality is well satisfied for typical semiconductors:

$$\frac{Z(R_g)}{N_q(R_g)} = \left(\frac{e^2 N^{1/3}}{\kappa E_g}\right)^{9/2} (Na_e^3)^{1/2} \ll 1,$$  \hspace{1cm} (13.4.3)

where $a_e = \hbar^2/2m_e e^2$. This means that all the electrons necessary to compensate the charge of a well of size $R_g$ can be accommodated in its lowest states. In other words, the bottom of such a well can sink below the Fermi level only by an amount that is small compared to the well’s depth (as shown in Fig. 13.4). It thus becomes clear that fluctuations of a size greater than $R_g$ are screened by electrons and holes, i.e., that $R_g$ plays the role of a screening radius for the impurity potential.

We see that long-range fluctuations turn the intrinsic semiconductor into a system similar to a semimetal. At $T = 0$ one still has a rather large number of electrons and holes. However, they are located in different regions of space, isolated from each other. Because of this, the static conduction in such a system is of a nonmetallic nature and occurs via thermal excitation of carriers into percolation levels.\(^2\)

Inside a well of size $R_g$ there still are humps and wells corresponding to short-range fluctuations. Of course, electrons and holes will fall into these wells (Fig. 13.4). Repeating the reasoning which led us to (13.1.4), we come to the conclusion that electrons and holes must form droplets with the characteristic dimensions $R_e$ and $R_h$, respectively, where

$$R_e = \frac{a_e}{(Na_e^3)^{1/9}}, \quad R_h = \frac{a_h}{(Na_h^3)^{1/9}}.$$  \hspace{1cm} (13.4.4)

1. Here, as in all qualitative arguments in this chapter, we omit numerical factors.
2. In narrow-gap or gapless semiconductors, inequality (13.4.3) is not satisfied. Then fluctuations give rise to metallic conduction at zero temperature, since either electrons or holes occupy a fraction of the volume large enough to provide percolation.
From (13.4.3) and the similar inequality containing \( a_h \), it follows that \( R_e, R_h \ll R_g \) and \( \gamma(R_e), \gamma(R_h) \ll E_g \).

The excess number of impurities in the droplets, \( (NR_e^3)^{1/2}, (NR_h^3)^{1/2} \), is small compared to that in a region of size \( R_g \), which is given by \( (NR_g^3)^{1/2} \). Therefore many droplets take part in the screening of each long-range fluctuation. Because of the different mass of electrons and holes, the electron and hole droplets differ in their size and charge. Also different is the number of droplets necessary to screen a positive or negative charge of the same magnitude. The charge symmetry of the screening is completely restored upon averaging over a distance larger than the mean separation between droplets. The statistical properties of the resultant long-range potential (including the potential due to electrons and holes) are symmetric with respect to the sign of the potential, since for a random impurity distribution one has the same probability for both positively and negatively charged fluctuations. From this it follows, in particular, that at \( T = 0 \) the Fermi level is located in the middle of the gap to within terms of order \( \gamma(R_e), \gamma(R_h) \).

In the preceding arguments it was assumed that the donor concentration \( N_D \) was exactly identical to the acceptor concentration \( N_A \). It is easy to see that the arguments remain valid, provided the extrinsic electron concentration \( n = N_D - N_A \) is small compared to the average concentration \( n_p \) of the electron-hole pairs induced by the band bending. The concentration \( n_p \) is obviously close to the mean net impurity charge concentration in the volume \( R_g^3 \), viz.

\[
    n_p = \frac{(NR_g^3)^{1/2}}{R_g^3} = \frac{N^2 e^6}{E_g^3 k^3}.
\]

If \( n >> n_p \), then \( r_s \ll R_g \) and \( \gamma(r_s) \ll E_g \) [as follows from (13.1.3)], which implies that the Fermi level is close to the conduction band. We see that as the degree of compensation is raised, both the screening radius and the amplitude of long-range fluctuations increase, and the Fermi level sinks deeper into the forbidden gap. As the Fermi level approaches midgap, intrinsic carriers begin to exercise screening, and a further lowering of the Fermi level does not occur.

According to (13.4.5), the situation when \( n < n_p \) can be realized more easily in narrow-gap semiconductors. As discussed in Sect. 13.2, in this case one more readily has an uncorrelated impurity distribution, which is necessary for the above theory to be applicable.

The picture of band bending shown in Fig. 13.4 was first proposed by Fritzschke [13.26] in connection with amorphous semiconductors. According to Fritzschke, this picture makes it possible to explain the apparent contradiction between the high density of states at the Fermi level, that is observed in microwave-absorption and field-effect experiments, and the low density of states in the forbidden gap that is observed by interband light absorption. Following his ideas, Shklovskii and Efros [13.3] developed the theory described above and suggested that a completely compensated
semiconductor could serve as a model for amorphous semiconductors. As seen from Fig. 13.4, the density of electronic states in the forbidden gap is very large in this model; however, because the potential is smooth, the "local" width of the forbidden gap is everywhere close to $E_g$. The interband light absorption coefficient must, therefore, fall very sharply with the quantum deficit [as described by (12.5.5)]. The assumption of smooth band bending also implies [5.65] that the optical width of the forbidden gap must be larger than twice the activation energy for static conductivity (the electrical width of the forbidden gap).

Recently, Overhof and Beyer [13.27] developed a model based on the idea of long-range fluctuations of the electrostatic potential to describe kinetic properties of hydrogenized silicon.

Ryvkin and Shlimak [13.28] also pointed out the analogy between properties of HDCS and amorphous semiconductors, and attributed its origins to the presence of a long-range potential in both systems. They suggested modelling amorphous semiconductors by using various means of producing compensating impurities in a crystalline semiconductor.
Part III

Computer Modelling
14. Modelling the Impurity Band of a Lightly Doped Semiconductor and Calculating the Electrical Conductivity

In Chap. 3 we discussed the analytic methods of describing the impurity band of a lightly doped semiconductor. Unfortunately, these methods hold only in the limiting cases of high and low compensations, where the small parameters of the problem are the quantities \((1-K)\) and \(K\). In the case of intermediate compensation such methods fail, and in order to describe the density of states over the whole energy interval one has to use computer simulations. Our group at the A. F. Ioffe Physical-Technical Institute in Leningrad began such studies in 1977.

Initially, the aim of these studies was to verify our theoretical understanding of the Coulomb gap (Chap. 10). For this reason the first studies [10.6, 7; 14.1] made use of a lattice model proposed by one of us [10.11], which is briefly described on page 233. We considered that model convenient for simulation, while the object of our research — the density of states near the Fermi level — was, according to theory, universal and therefore independent of the model and dependent only on the form of the bare interaction potential and the dimensionality.

The simulation method proposed in [14.1] was based on the direct minimization of the total energy. It is important to realize that the problem cannot be solved by examining all the possible sites, for that would require the computation and comparison of approximately \(2^N\) values of total energy, where \(N\) is the number of sites in a given array (each site can be empty or occupied). Any value of \(N\) that could conceivably describe a macroscopic system would foil the best computer. In [14.1] the energy was minimized only with respect to one-electron transitions. This minimization produces pseudoground states which have the property that their total energy increases if any one (but only one!) electron is transferred to any empty site. The method succeeded because pseudoground states proved few in number, and by analyzing them it was possible to single out the true ground states of fairly large arrays with weak size effects. This, in turn, made it possible to calculate the impurity band of a real semiconductor in a wide range of degrees of compensation [14.2], and also a number of electrical and optical phenomena — the subject matter of this part of the book.

As in the first part of the book, we are discussing such light doping that in the study of impurity band structure we can neglect the overlap of electronic states on adjacent donors. The localization radius \(a\) is the smallest characteristic length, and therefore the electron level shift can be considered
classical, that is, equal to \(-e\phi(r_i)\), where \(\phi(r_i)\) is the potential on the impurity center. Also, we neglect changes in the potential over the length \(a\) and regard the quantum-mechanical structure of the state, that is, the structure of its wave function, as equivalent to that of an isolated impurity. This model of the impurity band will henceforth be termed the classical model. To compute this model we must minimize the total energy of the electrostatic interaction of charged impurities (10.1.1) with respect to the occupation numbers \(n_i\) at a given electron concentration. Minimization produces the set of occupation numbers \(\{n_i\}\) and one-electron energies \(\{\epsilon_i\}\) [see (3.1.1)] which correspond to the ground state, and also the Fermi level \(\mu\).

In the case of hopping conduction, the approximation of a classical impurity band should certainly hold at donor concentration \(N_D < N_m\), where \(N_m\) is the concentration corresponding to the maximum activation energy \(\epsilon_3\) (Sect. 8.1). At \(K \leq 0.4\) the quantity \(N_m\) is determined by the condition\(^1\) \(N_m a^3 \approx 2 \times 10^{-3}\). (In the case of high compensation, \(N_m\) increases and the range of the classical impurity band broadens greatly.) Recall that the Mott transition takes place at a concentration \(N_M\) such that \(N_M a^3 = 2 \times 10^{-2}\). Thus there is a wide range of concentrations \(N_m < N_D < N_M\), which is very interesting for experimentation and where quantum effects play an important role. Nonetheless, we believe that the main qualitative properties of the classical impurity band [the two-peak structure of the density of states, the Coulomb gap, the general form of the dependence \(\mu(K)\)] are present in that range as well. We hope that computations of the classical impurity band will serve as a starting point for studying the role of quantum effects.

This chapter is organized as follows. In Sect. 14.1 we discuss the minimization of the total energy (10.1.1) \(^{14.2,4}\), on which all subsequent calculations are based. In Sect. 14.2 the calculated density of states and the Fermi level are compared with the results of the theory developed in Chap. 3 for the limits of high and low compensation; the form of the density of states at intermediate compensations is also discussed. A small modification of the computation program permits the calculation of the electric field distribution over neutral donors, which is necessary in order to find the Stark broadening of impurity transitions \(^{14.5}\) -- this problem is discussed in Sect. 14.3.

As we observed in Chap. 8, a knowledge of the impurity band structure is essential in calculating the hopping conduction's temperature dependence. The set \(\{\epsilon_i\}\) corresponding to the ground state must be used in calculating the percolation threshold with the bonding criterion \((8.3.1)\). This problem was also solved on a computer to determine the activation energy for hopping conduction in a wide range of degrees of compensation \(^{14.6}\). It was found that at low temperatures the hopping conduction is governed by the law \((10.1.18)\); the coefficient \(\beta\) of \((10.1.19)\) was also calculated \(^{14.7}\). These

---

\(^1\) In low-compensated Ge with Sb impurities one has \(N_m = 8 \times 10^{15} \text{cm}^{-3}\) (see Fig. 8.1); hence \(N_m a^3 = 1.7 \times 10^{-3}\). For \(p\)-Ge with Ga impurities and \(K = 0.4\) one has \(N_m = 2.6 \times 10^{15} \text{cm}^{-3}\) and \(N_m a^3 = 1.9 \times 10^{-3}\) \(^{14.3}\).
questions are discussed in Sects. 14.4 and 14.5. Section 14.6 is devoted to calculating the percolation level in the conduction band, which we need in order to evaluate corrections to the activation energy \( \epsilon_i \) for band conduction that are due to finite donor concentrations. Finally, in Sect. 14.7 we discuss certain other applications of the energy minimization procedure.

### 14.1 Minimization of the Total Energy and Calculation of the Density of States

The following procedure was used to minimize the total energy and calculate the density of states. First, a random number generator set the coordinates of \( N \) donors and \( KN \) acceptors in a cube of edge \( L \). Experiments were conducted at \( N = 100, 200, 400, 800, 1600 \) and \( K = 0.01, 0.03, 0.1, 0.3, 0.5, 0.7, 0.8, 0.9, 0.93, 0.95, 0.98 \), so that \( KN \) was always an integer. The length \( L \) was fixed at the value \((N/N_D)^{1/3}\) such that at different \( N \) the donor concentration was identical and equal to \( N_D \). Next, \((1-K)N\) electrons were randomly distributed among \( N \) donors, thus ensuring the electric neutrality of the system.

Further, the energies \( \epsilon_i \) of all donors were calculated using (3.1.1). Among the occupied donors the one with the maximum energy \( \epsilon_p \) and among the empty donors the one with the minimum energy \( \epsilon_q \) were chosen (Fig. 14.1). Control was then transferred to the so-called \( \mu \)-subroutine (\( \mu \)-S), whose function was fulfill the conditions (10.1.4) for all donors, i.e., to form the Fermi level \( \mu \). If it turned out that \( \epsilon_p > \epsilon_q \) (as was usually the case with a random initial distribution of electrons) the occupation numbers of donors \( p \) and \( q \) were changed so that \( n_p \) became zero and \( n_q \) became 1. This

---

**Fig. 14.1a–e.** Density-of-states functions \( g_p(\epsilon) \) and \( g_q(\epsilon) \). (a) Infinite-array density of states, (b, c) energy levels for two different realizations of donor and acceptor coordinates, (d) \( g_i(\epsilon) \), (e) \( g_\mu(\epsilon) \)
corresponds to the transfer of an electron from donor \( p \) to donor \( q \). After that the energies \( \epsilon_i \) were calculated from scratch, new maximum and minimum values \( \epsilon_p \) and \( \epsilon_q \) were determined, and if \( \epsilon_p > \epsilon_q \), an electron was again transferred from \( p \) to \( q \). The subroutine \( \mu \)-S terminated its operation only when it found \( \epsilon_p < \epsilon_q \), i.e., when all occupied donors were lower in energy than all empty donors.

Next, the energy was minimized with respect to the transfer of an electron from one site to another. The change in the total energy (10.1.1) as one electron is transferred from an occupied site \( i \) to an empty site \( j \) can be found from (10.1.5). By definition of the ground state, this quantity has to be positive, i.e.,

\[
\Delta_i^j = \epsilon_j - \epsilon_i - e^2/kr_{ij} > 0.
\] (14.1.1)

Inequality (14.1.1) was tested for each pair consisting of an occupied and an empty donor. If the inequality did not hold, then an electron was transferred from donor \( i \) to donor \( j \), \( \mu \)-S was accessed again, and the testing of inequality (14.1.1) continued.

It is important to note that every electron transfer effected by the program lowered the total energy (10.1.1) of the system. In the end, minimization produced a set of occupation numbers \( \{n_i\} \) and energies \( \{\epsilon_i\} \) which fulfilled conditions (10.1.4) and (14.1.1) and consequently defined a pseudoground state with the property that the transfer of any electron from one site to another would decrease the total energy of the system. However, since the total energy could still be lowered through the simultaneous transfer of two or more electrons, a true ground state was still not in evidence. Moreover, if the initial occupation numbers at a given configuration of donor and acceptor coordinates are altered, the minimization program may lead to a different pseudoground state with different occupation numbers and energies. Thus, a different configuration of donors and acceptors can lead to many pseudoground states.

The study then proceeded to the properties of pseudoground states. A series of \( M \) minimizations was performed for fixed donor and acceptor coordinates and varying initial electron distributions produced by the random number generator. The results are presented in Table 14.1 for the case of \( K = 0.5 \). Here \( N \) is the number of donors and \( q \) the number of series of \( M \) minimizations. Different series have different donor and acceptor coordinates. The quantity \( \Delta H \) represents the energy dispersion for pseudoground states, averaged over the different series. (In each series we evaluated the difference between maximum and minimum energies of pseudoground states, and then this difference was averaged over all series.) We see that the dispersion is very small. The greatest difference in the total energies of pseudoground states does not exceed the average energy of a single particle. At \( N = 200 \) and 400 the same pseudoground state is frequently repeated within the same series, and it is this repeated pseudoground state that has the lowest energy in the series. The last column in Table 14.1 shows the number of repetitions of the
Table 14.1. Results of the investigation of pseudoground states at $K = 0.5$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$q$</th>
<th>M</th>
<th>$\Delta H \left[ \frac{e^2 N_D^{1/3}}{\kappa} \right]$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>6</td>
<td>50</td>
<td>0.31</td>
<td>25</td>
</tr>
<tr>
<td>400</td>
<td>4</td>
<td>30</td>
<td>0.47</td>
<td>7</td>
</tr>
<tr>
<td>800</td>
<td>3</td>
<td>10</td>
<td>0.57</td>
<td>1</td>
</tr>
<tr>
<td>800</td>
<td>1</td>
<td>50</td>
<td>0.69</td>
<td>2</td>
</tr>
<tr>
<td>1600</td>
<td>1</td>
<td>18</td>
<td>1.18</td>
<td>1</td>
</tr>
</tbody>
</table>

The lowest-energy state in a given series, averaged over all series. We believe this lowest-energy state is the true ground state. Our reasoning runs as follows: with a random initial distribution of electrons among the donors, the probability that minimization will lead to the true ground state should be no less than the probability of arriving at any pseudoground state. Granted, there are many pseudoground states and only one true ground state, which is very difficult to hit, given a sufficiently large $N$. However, if the lowest energy comes up repeatedly, then this should be the energy of the ground state. Indeed, if we suppose that there is an energy that is still lower, then we must also suppose that the probability of hitting this hypothetical state is for some reason much smaller than for the lowest-energy state we obtained. This is the principle of locating the ground states; it was first formulated in [14.1].

Based on that principle, we think we succeeded in finding the ground state at $N = 200$ and 400 and failed to do so at $N = 800$ and 1600, where the pseudoground state with the lowest energy occurs only once or twice in a series, and there is every reason to believe that increasing the number of minimizations in a series will yield a state with energy that is lower.

The important result is that at $N = 200$ and 400 the density of states found by averaging over only the ground states corresponding to different donor and acceptor configurations was the same as that found by averaging over the first-encountered pseudoground state for each impurity configuration. This result held for the range of single-particle energies $|\epsilon - \mu| > 0.15 e^2 N_D^{1/3} \kappa^{-1}$. In arrays of $N = 800$ and 1600, where we failed to identify the ground state, we found that the density of states $g(\epsilon)$ of the lowest-energy pseudoground state did not differ from that of any other pseudoground state within the accuracy of our calculations ($\approx 10\%$). This fact, as well as the absence of a size effect in the energy range $|\epsilon - \mu| > 0.15 e^2 N_D^{1/3} \kappa^{-1}$ as we increased $N$ from 400 to 1600, convinced us that in the stated energy range, the $g(\epsilon)$ averaged over pseudoground states differs little from the $g(\epsilon)$ corresponding to the true ground state of an infinite system. (Let us note that size effects were small only in the case of...
intermediate compensation. Peculiar size effects (discussed below) did appear at very low or very high compensations, but they were not related to the selection of a true ground state among the pseudoground ones.)

The results for the density of states, the Fermi level, and all other physical quantities presented below were found by averaging over one pseudoground state at each impurity configuration. A more detailed investigation [14.1] of the properties of pseudoground states in the lattice model led to the same results but was perhaps more convincing, since in a lattice model with no level dispersion the ground state is known a priori. This ground state is ordered (Sect. 10.1.1) and it can be verified that our method does yield it.

The computation of the density of states \( g(\epsilon) \) and the Fermi level \( \mu \) was carried out in the following manner. Each realization of impurity coordinates was subjected to one minimization procedure which yielded the set of energies \( \{\epsilon_i\} \), as well as the quantity

\[
\mu' = \frac{1}{2} (\epsilon_p + \epsilon_q),
\]

(14.1.2)

which we termed the Fermi level for a given realization. The value of the Fermi level \( \mu_N \) for an array of \( N \) donors was obtained by averaging the quantity \( \mu' \) over different realizations of impurity coordinates. The number of realizations depended on \( N \) and was approximately 150, 100, 50, 30, and 15 for \( N = 100, 200, 400, 800, \) and 1600, respectively. The density of states and other quantities were averaged over this same number of realizations.

Two different procedures were used to average the density of states. One of them (referred to as simple averaging) consisted of finding the fraction of donors having energies \( \epsilon_i \) in a given range and then averaging this fraction over many realizations of an array of \( N \) donors. At high compensations this procedure proved inconvenient: as we shall see, at high compensations the Fermi level fluctuates strongly from one realization to another. Consequently, the energy intervals near the Fermi level contained empty donors in some realizations and filled donors in others. This resulted in a smeared Coulomb gap which did not reflect the macroscopic sample’s density of states.

For this reason another averaging procedure — \( \mu \)-averaging — was developed. Instead of sorting over intervals and averaging over realizations the energies \( \epsilon_i \), this was done to the differences \( \epsilon_i - \mu' \), where \( \mu' \) is the Fermi level for the realization at hand. The graph of the density of states \( g_\mu(\epsilon) \) obtained in this way has the intervals of positive and negative values \( \epsilon_i - \mu' \) separated by the point \( \epsilon = \mu_N \). In other words, this procedure suppresses fluctuations in the Fermi level by shifting the energies of all realizations in such a way that the Fermi level of each realization matches the average Fermi level, and only then averaging the density of states over realizations (see Fig. 14.1).

The nature of the Coulomb gap is such that the density of states with energy \( \epsilon \) is formed by interactions at distances of order \( \epsilon^2/\kappa \epsilon \). Consequently, in an array of \( N \) donors the model of the macroscopic density of states makes
sense only if \( \epsilon \geq \epsilon N \), where \( \epsilon N \equiv N^{-1/3}e^{2N_{D}^{1/3}k^{-1}} \). At low energies the above procedures of averaging the density of states describe size effects differently. It is easily demonstrated [14.1] that the density of states \( g_{\mu}(\epsilon) \) vanishes exactly in the energy interval \( |\epsilon - \mu| < (1/2\sqrt{3})\epsilon N \) and has only a slight dependence on \( N \) at much higher energies (at \( N = 1600 \) the region where the size effects are of importance is so small that it is practically absent from the graph). The density of states \( g_{\mu}(\epsilon) \), obtained via simple averaging, differs from zero at \( \epsilon = \mu \) and is strongly dependent on \( N \) at intermediate compensations with \( |\epsilon - \mu| < \epsilon N \). In the range of energies were \( g_{\mu}(\epsilon) \) and \( g_{\mu}(\epsilon) \) coincide, they represent the macroscopic density of states.

### 14.2 Density of States and the Fermi Level

Figure 14.2 displays the Fermi level \( \mu N \) plotted as a function of \( N^{-1/3} \) and shows the procedure of extrapolating to an infinite array. It appears logical that the size effect (that is, the dependence of the results on \( N \)) is due to the fact that donors near the edges of the cube behave differently from those

![Fig. 14.2. Dependence of the average Fermi energy (in units of \( e^{2N_{D}^{1/3}k^{-1}} \)) on the number of donors \( N \) at different degrees of compensation \( K \).](image-url)
inside the volume. The fraction of donors near the surface, and consequently the correction to the Fermi level, should be proportional to $N^{-1/3}$ — which is the measure of the $x$ axis in Fig. 14.2. Figure 14.3 shows the extrapolated values of the Fermi level $\mu$ as a function of the degree of compensation. Figures 14.4 and 14.5 present the density of states averaged over realizations. Each density-of-states graph represents a smoothed histogram based on some 50 energy intervals. Individual histogram bars usually differ from the smoothed curve by much less than 10%; the other distribution functions cited below are equally accurate. Let us now discuss the results.
14.2.1 Intermediate Degrees of Compensation

The most pronounced characteristic of the density-of-states graphs in Fig. 14.4 is the two well-defined peaks, which were predicted by Pollak [10.1] and Knotek and Pollak [10.2]. The states gradually shift from the lower to the higher peak as compensation increases. At the same time, despite a significant change in the Fermi level, the distance between the peaks remains essentially constant in a wide range of compensations ($0.1 \leq K \leq 0.7$).

As can be seen from Fig. 14.4, the density of states $g_\mu(\varepsilon)$ tends to zero as the energy approaches the Fermi level, regardless of compensation. In Fig. 14.5 the Coulomb gap region at $K = 0.5$ is enlarged. On that graph we can see the density of states $g_\sigma(\varepsilon)$, which differs from zero at $\varepsilon = \mu$. At $|\varepsilon - \mu| \geq 0.15 e^2 N_D^{1/3} \kappa^{-1}$ the two calculated densities of states $g_\sigma(\varepsilon)$ and $g_\mu(\varepsilon)$...
match, which is also the case at all compensations in the interval $0.1 \leq K \leq 0.7$. As we noted earlier, this indicates that for $|e-\mu| > 0.15 e^2 N_D^{1/3} \kappa^{-1}$ both densities of states describe an infinite array quite well. We believe that even in the presence of high compensation, $\mu$-averaging describes the macroscopic density of states in the same energy interval, even though in the graphs of $g_\epsilon(e)$ not included here, the Coulomb gap is almost invisible because of strong fluctuations in the Fermi level.

In [14.1] the density of states for the lattice model was extrapolated to an infinite array. We found that at low energies the law (10.1.6) holds and that the constant $\alpha_3$ is close to the value $3/\pi$ obtained by solving the self-consistent equation (see [10.7, 11; 14.1.8]).

It is interesting to verify the universality of this law, i.e., the independence of $\alpha_3$ from a particular model of the disordered system. To that end we can compare the densities of states for the lattice model [14.1] and the impurity band [14.2]. Both are represented in Fig. 14.5 in different dimensionless units. In the lattice model the unit of length is the lattice constant and the unit of energy is the interaction energy of two electrons on adjacent sites. In the impurity band model these units are $N_D^{-1/3}$ and $e^2 N_D^{1/3} \kappa^{-1}$ respectively. Let us note that in both systems of units the Coulomb law appears as $e = 1/r$. The nature of the Coulomb gap is such that in any system of units where the Coulomb law is of this form, one has $g(e) = \alpha_3 (e-\mu)^2$ and the coefficient $\alpha_3$ is independent of the unit system. From Fig. 14.5 we see that at high energies the densities of states are quite dissimilar, but near the Fermi level they coincide, which supports the idea of universality (the arrays compared are of approximately equal size).
This universality is further supported by the fact that at low energies, the density of states does not depend on the degree of compensation, as is evident from Fig. 14.4.

Thus, the modelling performed by our research group indicates that the one-electron density of states vanishes at the Fermi level in accordance with a universal quadratic law (10.1.6). We were not able to detect the polaron effect, which should result in an exponential decrease in the density of states [see (10.1.8)].

To further demonstrate that the behavior of the density of states at low energies is related to long-range interaction, we conducted yet another computer experiment at \( K = 0.5 \). In the expression (10.1.1) for the total energy and in all subsequent formulae, the Coulomb potential \( 1/r_{ij} \) was replaced by a potential of the form \( r_{ij}^{-1} \exp (-r_{ij}/r_0) \), where \( r_0 \) was the radius of an artificially introduced screening which we varied from \( r_0 = 0.5 N_D^{-1/3} \) to infinity. The corresponding change in the density of states, shown in Fig. 14.6, confirms that the density of states near the Fermi level is determined by long-range interaction and that the Coulomb gap disappears as \( r_0 \) gets smaller.

![Fig. 14.6. Disappearance of the Coulomb gap as the "screening radius" \( r_0 \) is diminished at \( K = 0.5 \) and \( N = 130 \). Dashed line corresponds to \( r_0 = 2.5 N_D^{-1/3} \) and solid line to \( r_0 = 0.67 N_D^{-1/3} \).](image)

### 14.2.2 Strong Compensation \((1 - K << 1)\)

As we can see from Fig. 14.3, the Fermi level falls sharply when \( K \to 1 \), which is in qualitative agreement with the analytic theory presented in Sect. 3.3. In Fig. 14.7 the experimental dependence \( \mu(K) \) is plotted as a function of \((1 - K)^{-1/3} \). It is natural to assume that the last three points on
the graph \( K \geq 0.95 \) lie near the asymptotic straight line which corresponds to the law:

\[
\mu(K) = -C \frac{e^2N_D^{1/3}}{\kappa} (1-K)^{-1/3}. \tag{14.2.1}
\]

Then for the coefficient \( C \) we obtain \( C = 2.8 \). Figure 14.7 also shows the theoretical dependence (3.4.3), which accounts only for proximate donors. It appears reasonable to assume that at \( K \geq 0.9 \) the differences between this dependence and the calculated results are due to the large-scale potential (Sect. 3.4). This conclusion is borne out by the strong size effect visible in Fig. 14.2 at \( K \geq 0.9 \). Indeed, the slope of the dependence of \( \mu_N \) versus \( N^{-1/3} \) should be proportional to the thickness of the surface layer, that is, to the depth at which surface effects are significant. In the presence of strong compensation this depth should be of the order of the magnitude of the radius of nonlinear screening and should increase rapidly with compensation, which does in fact occur (see Fig. 14.2).

![Graph showing dependence of \( \mu \) on \( (1-K)^{-1/3} \).]

In order to ascertain directly the role of the large-scale potential, the Coulomb potentials were multiplied by \( \exp(-r/r_0) \). At \( r_0 < N_D^{-1/3} \) the introduction of this factor should artificially suppress the large-scale potential and permit us to describe the ground state using the "pair" approximation. In this approximation, instead of (3.4.3) for \( \mu \) we get [14.2]:

\[
\mu = -\frac{e^2}{\kappa R_p} \exp \left(-\frac{R_p}{r_0}\right), \quad \text{where}
\]

\[
R_p = \left[ \frac{3}{2\pi} (1-K) \right]^{1/3} N_D^{-1/3} \tag{14.2.3}
\]

is the maximum distance between donors in a donor pair with an electron. In Fig. 14.8, (14.2.2) is compared to the results of a computer experiment [14.5].
for $K = 0.95$ and $N = 800$. As $r_0$ decreases, the computed points lie closer and closer to the theoretical curve, which indicates the applicability of the pair approximation at low values of $r_0$ and the correctness of the procedure in general. Using Fig. 14.8 it is possible to roughly evaluate the spatial scale $r_s$ of the potential that causes $\mu$ to differ from (3.4.3) at $r_0^{-1} = 0$. Let us take $r_s$ to be that value of $r_0$ at which the difference between the computed result and (14.2.2) is half the difference at $r_0^{-1} = 0$. Then we find that $r_s$ is of order $2N_D^{-1/3}$. On the one hand this confirms that in the real Coulomb case the deviations from the pair model are caused by a potential on a scale greater than $N_D^{-1/3}$. On the other hand, a comparison of this result with (3.4.8) for $r_s$ at $K = 0.95$ shows that on the right-hand side of this formula there should be a small numerical coefficient of order 0.3. Apparently both (3.4.8) and (14.2.1) hold only at very high compensations, $K \geq 0.95$. Chenskii [14.9] arrived earlier at similar conclusions on the basis of an approximate analytic theory of the highly compensated semiconductor.

14.2.3 Low Compensation

The lowest degrees of compensation in our numerical experiments were $K = 0.01$ and $K = 0.03$. At $N = 3000$ we found $\mu = 0.92$ (in $e^2N_D^{1/3}k^{-1}$ units) for $K = 0.01$ and $\mu = 0.89$ for $K = 0.03$. The difference between our maximum value 0.92 and the theoretical maximum 0.99 [see (3.2.11)] may be caused by either the finite degree of compensation or the size effect. Recently, Hearn et al. [14.10] modelled the impurity band at $K = 0.01$ using the method described in this chapter. Their array of donors was much larger,

2. Only in these two instances did we use an array of $N = 3000$. 
\[ N = 64000, \] but they studied only one realization and obtained \( \mu = 0.93 \). If we assume that at such values of \( N \), variation in \( \mu \) from one realization to another can be neglected, then, comparing the result obtained in [14.10] with that cited above (\( \mu = 0.92 \) at \( N = 3000 \)), we conclude that the size effect cannot explain the discrepancy. The other explanation is that the degree of compensation is not low enough. Despite this discrepancy, we think our results are in reasonable agreement with theory.

Another important characteristic of the impurity band which describes the correlation of charged centers is the distribution of internal electric fields. We will discuss this topic in the next section.

### 14.3 Distribution of Electric Fields on Neutral Donors

Static electric fields created by ionized donors and acceptors cause Stark shift and splitting of the levels of those impurity centers that have remained neutral and thus contribute to intra-impurity optical absorption. Since electric fields vary from one impurity center to another, this effect leads to the broadening of spectral lines for impurity optical absorption and photoconduction. In many cases this is the dominant mechanism of broadening.

The usual approach to calculating Stark broadening of spectral lines in lightly doped, compensated semiconductors consists in the following (see the bibliography in [14.5]). It is assumed that the distribution of charged donors and acceptors is completely chaotic (uncorrelated) with respect to the neutral donors which absorb radiation. In this case, the field distribution that determines the shape of spectral lines is well known: it is described by Holzmark's formula [14.11], in which the concentration of charged particles is set equal to \( 2N_A \).

This approach is valid only when the energy \( kT \) of the thermal motion is much greater than \( e^2N_D^{1/3}\kappa^{-1} \). However, the opposite inequality, \( kT \ll e^2N_D^{1/3}\kappa^{-1} \), usually holds when one measures the optic and photoelectric spectra of shallow impurities. In this case the state of the remaining, uncompensated electrons of concentration \( N_D - N_A \) is determined by minimum of the Coulomb energy, and the configurations of neutral and charged impurities are strongly correlated. Kal'fa and Kogan [3.12] and Golka et al. [14.12] noted the major role of this correlation in the problem of Stark line broadening. A marked broadening of lines was observed [14.12] in the photoconduction spectra of \( n \)-GaAs and \( n \)-CdTe with shallow donors as the temperature was raised from 4K to 10K. The authors attributed this phenomenon to the transition from a correlated positioning of electrons to an uncorrelated one.

At a low degree of compensation, the distribution function of the magnitude of the electric field on a neutral donor \( F(E) \) can be calculated analytically [14.5]. We need only remember that for \( K \ll 1 \) there is one
positively charged donor near almost every acceptor. Only a small fraction of acceptors (2.6%) have either two charged donors or none at all (a 2-complex or a 0-complex). If we ignore these relatively rare charge configurations, we find that all charges are grouped into dipoles of a size (arm length) of the order of \( N_D^{-1/3} \), while most of the neutral donors are separated from the dipoles by distances of the order of \( N_A^{-1/3} >> N_D^{-1/3} \), i.e., they are found in the dipolar field.

The interaction energy between two adjacent dipoles is of the order of \( K e^2 N_D^{1/3} / k \). This energy is normally insufficient to rotate the dipole or increase its arm length, because such a process would result in the ionization of another donor that is not the closest to the acceptor, but, say, a next nearest neighbor. This would require an energy of order \( e^2 N_D^{1/3} k^{-1} \), which is much larger than the dipole-dipole interaction energy. Thus, despite the dipole-dipole interaction, almost every dipole is directed from an acceptor to the nearest donor. This means that the dipoles are chaotically oriented and the distribution function for their arm lengths is of the form

\[
 f(r) = 4\pi r^2 N_D \exp \left[ -\frac{4\pi}{3} N_D r^3 \right]. \tag{14.3.1}
\]

Since acceptor impurities are chaotically distributed within the crystal, so are the dipole centers. Consequently, at low degrees of compensation, \( F(E) \) represents the distribution function of the field created within the crystal by chaotically positioned and randomly oriented dipoles of concentration \( N_A \), whose arm lengths are distributed according to (14.3.1). The Markoff method was used in [14.5] to find \( F(E) \):

\[
 F(E) = \frac{4}{\pi E_{md}} \frac{(E/E_{md})^2}{[1 + (E/E_{md})^2]^2}, \tag{14.3.2}
\]

where the field \( E_{md} \) corresponds to the maximum of \( F(E) \) (the most probable field) and equals

\[
 E_{md} = 2.51 K e N_D^{2/3} k^{-1}. \tag{14.3.3}
\]

At the same time, at high temperatures, when ionized donors are "torn" from acceptors and are positioned in an uncorrelated manner relative to them, the most probable field is defined by the Holzmark distribution [14.11] and equals

\[
 E_{mH} = 4.16 e (2N_A)^{2/3} k^{-1}. \tag{14.3.4}
\]

The ratio \( E_{md}/E_{mH} = 0.38K^{1/3} << 1 \). Thus the correlation between donors and acceptors at \( K << 1 \) — manifested by the formation of dipoles — should result in a marked reduction in the magnitude of the most likely electrical field.

Kall'fa and Kogan [3.12] pointed out that for \( 1 - K << 1 \) the correlation may have the opposite result. If we neglect the potential of scale \( r \geq N_D^{-1/3} \), we find that the electrons are found on donor pairs with the distance between
donors being \( R \leq R_p \), where \( R_p \) is defined by (14.2.3). This implies that near every neutral donor there should be a charged donor, and the distance between the two is much less than \( N_D^{-1/3} \). Accordingly, in the pair approximation, the field on a neutral donor is created for the most part by the charged donor of the pair, and

\[
F(E) = \begin{cases} 
0, & E < E_p \\
\frac{e^2}{\kappa R_p^2} = \frac{eN_D^{2/3}}{\kappa} \left[ \frac{2\pi}{3(1-K)} \right]^{2/3}, & \text{for} \ 0 < E < E_p \\
\frac{3}{2} \frac{E_p^{3/2}}{E^{5/2}}, & E > E_p.
\end{cases}
\]

(14.3.5)

In this case the most probable field coincides with the minimal field \( E_p \). If \( 1 - K \ll 1 \), then \( E_p(K)/E_{nth}(K) = 0.25(1-K)^{-2/3} \gg 1 \). Although the accuracy of the pair approximation is unknown, it appears reasonable that for \( 1 - K \ll 1 \) the typical fields should be greater than those produced by the same number of donors and acceptors in a random configuration.

In order to verify the above theoretical ideas, we calculated \( F(E) \) directly by appending a simple subroutine to the computer program described in Sect. 14.1. Having found the pseudoground state at a given realization of donors and acceptors, the program directly summed the field components created on each neutral donor by each charged donor and acceptor. The computed fields were sorted — thus defining \( F(E) \) — and then averaged over many realizations. Figure 14.9 displays the functions \( F(E) \) for \( N = 1600 \)

![Graphs showing distribution of electric fields](image)

**Fig. 14.9a, b.** Distribution of electric fields on neutral donors at various degrees of compensation \( K \). Field \( E \) is measured in units of \( eN_D^{2/3} \kappa^{-1} \) and \( F(E) \) in units of \( (eN_D^{2/3} \kappa^{-1})^{-1} \). Values of \( K \): (1) 0.1, (2) 0.3, (3) 0.5, (4) 0.7, (5) 0.9, (6) 0.95
and several values of \( K \). The most probable field \( E_{mN} \) was taken as the simplest characteristic of \( F(E) \) for an array of \( N \) donors. The field \( E_m \) found by extrapolating \( E_{mN} \) to \( N = \infty \) is plotted in Fig. 14.10 as a function of compensation. The figure also shows the dependences of \( E_{md} \) and \( E_{mH} \) on \( K \), as expressed by (14.3.3) and (14.3.4). For all \( K \leq 0.9 \) we see that \( E_m \) is much less than the field \( E_{mH} \) which would result from an uncorrelated distribution of charged impurities with respect to neutral ones. The effect of correlation is very large; for instance, at \( K = 0.5 \) the ratio \( E_{mH}/E_m \approx 4 \).

For \( K \leq 0.5 \) the computer experiment yields results that match the theoretical formula (14.3.3) derived for \( K \ll 1 \). This is true not only for the value of the most probable field, but also for the entire shape of \( F(E) \) [14.5]. The fact that (14.3.3) is applicable even for \( K = 0.5 \) indicates that the charge correlation in the impurity band consists mostly in the formation of donor-acceptor dipoles. This is probably due to the fact that the mean distance between an acceptor and the nearest donor \( 0.55N_D^{-1/3} \) (the average dipole length) is somewhat less than the mean distance between dipoles even at \( K = 0.5 \).

For \( K > 0.9 \) the field \( E \) rapidly increases as \( K \) approaches unity, which is in a qualitative agreement with Kal'fà and Kogan's prediction based on the concept of donor pairs. But even at \( K = 0.95 \) the field \( E_m \) is still less than the field \( E_{mH} \) that corresponds to chaotic charge distribution. This result is rather unexpected and is probably due to the fact that a significant fraction of uncompensated electrons are located on large aggregates of donors (and not pairs) in which the fields of charged donors compensate each other [14.5].

Let us now discuss Stark broadening of the most intense line \((n = 1 \rightarrow n = 2)\) in the spectrum of hydrogenlike impurities. An electric
field \( E \) shifts the two lines by \( \pm 3 \) (\( eEa \)). The shift of the third line contains no linear term in \( E \), but involves two contributions, one of which is proportional to \( E^2 \) and the other to second spatial derivatives of the electrostatic potential on the donor. Usually it is accepted that the first two lines form wide and low wings and that only the third central line is observable. Kogan and Nguyen [14.13] used the same minimization program and found the distribution of \( E^2 \) and of the second derivatives of the potential. This enabled them to investigate the shape of the central line for various dopings and compensations. Their main conclusions were that the linewidth is significantly narrower and grows faster with increasing \( K \) than in the case of randomly distributed electrons. These conclusions are borne out qualitatively by experimental data.

### 14.4 Activation Energy of Hopping Conduction

Scattering of the donor levels determines the temperature dependence of hopping conduction in a lightly doped semiconductor. As discussed in Chaps. 8 and 9, to find the activation energy we must compute the percolation threshold \( \xi = \xi_c \) for the problem with the bonding criterion

\[
\frac{2r_{ij}}{a} + \frac{\epsilon_{ij}}{kT} \leq \xi, \tag{14.4.1}
\]

where \( \epsilon_{ij} \) is given by the expression (10.1.17) that excludes self-action. For \( T \to \infty \) the percolation threshold equals \( \xi_c^0 = 2r_c/a \). At a finite temperature the quantity \( \Delta \xi_c = \xi_c - \xi_c^0 \) contains the whole exponential temperature dependence of the resistivity, viz.

\[
\rho(T) = \rho_0 \, e^{\xi_c} = \rho_0 \exp \left( \frac{1.73}{N_D^{1/3}a} \right) e^{\Delta \xi_c(T)}. \tag{14.4.2}
\]

where \( \rho_0 \) is a power function of \( N_D \) and \( T \).

In [14.6,7] the threshold value \( \xi_c \) was found by direct testing for percolation between the opposite faces of a cube with the bonding criterion (14.4.1). This testing used the donor coordinates as well as the sets \( \{ \epsilon_i \} \) and \( \{ n_i \} \) generated in the process of minimizing the total energy as outlined in Sect. 14.1. The random site percolation problem with the bonding criterion (14.4.1) contains physical parameters \( a, T, N_D, \) and \( \kappa \). It is convenient to rewrite (14.1.1) introducing dimensionless variables \( r' = N_D^{1/3}r \) and \( \epsilon_{ij}' = \epsilon_{ij} (e^2 N_D^{1/3}/\kappa)^{-1} \) and also the dimensionless temperature

\[
t = \frac{kT\kappa}{e^2 N_D^{1/3}} \frac{2}{N_D^{1/3}a}. \tag{14.4.3}
\]

Then, from (14.4.1) we get
\[ r_{ij}^* + t^{-1} \epsilon_{ij}^* \leq \eta, \quad \eta = \xi \frac{N_d^{1/3} a}{2}. \] (14.4.4)

The distribution of \( \epsilon_{ij} \) depends solely on \( K \). Consequently, from (14.4.4) it is apparent that for fixed \( K \) the quantity \( \eta_c = \xi_c N_d^{1/3} a / 2 \) is a universal function of \( t \). This function must be determined numerically by solving the percolation problem with bonding criterion (14.4.4) in a cube of dimensionless edge length \( L' = LN_d^{1/3} = N^{1/3} \) at various values of \( t \).

In Chaps. 8 and 9 we found that at \( t \gg 1 \) the temperature dependence of hopping conduction should obey the law (8.1.1). At \( t \ll 1 \) we had variable-range hopping conduction. In this section we shall restrict ourselves to relatively high temperatures, \( t^{-1} \ll 0.5 \), and thus investigate the activation energy \( \epsilon_3 \). The case of low temperatures, \( t \ll 1 \), will be discussed in the next section.

The procedure used to calculate \( \eta_c \) was similar to those used in earlier studies [5,13,26,56]. The coordinate origin was placed at a vertex of the cube, with the axes \( x', y', z' \) directed along the three adjacent edges. At the face \( x' = 0 \) a layer \( 0 < x' < qr_c \) was earmarked, where \( r_c = 0.865 \) and \( q \) is a numerical coefficient (calculations were carried out with \( q = 0.5 \) and \( q = 1 \)). All donors in that layer were termed "first generation" donors and their numbers were apportioned to array \( A \). In addition, all these donors were assigned a label. Then from the unlabeled donors we selected those that satisfied the bonding criterion (14.4.4) with respect to at least one donor from array \( A \). These were termed "second generation" donors and were labeled in the same manner. Then array \( A \) was emptied of first generation donors and filled with second generation ones. Once again we selected all unlabeled donors bonded to at least one donor in array \( A \), thus forming the third generation. This procedure was repeated step by step in such a fashion that array \( A \) always contained donors of the latest generation and all used donors were marked. At every step array \( A \) was checked to see whether it contained any donor from the thin layer \( L' > x' > L' - qr_c \) at the opposite face of the cube. Such an occurrence indicated percolation. If array \( A \) turned out to be empty at some step, then we registered the absence of percolation. By varying \( \eta \) it was possible to find \( \eta_c \) through successive approximations.

After finding the pseudoground electron state for each realization of the donor and acceptor array, the quantity \( \eta_c \) was calculated at \( t^{-1} = 0, 0.02, 0.05, 0.1, 0.2, 0.3, \) and 0.5. Then for each nonzero value of \( t^{-1} \) we calculated the quantity responsible for the temperature dependence of hopping conduction, viz.

\[ \Delta \eta_c = \eta_c(t^{-1}) - \eta_c(0) = \frac{N_d^{1/3} a}{2} \Delta \xi_c. \] (14.4.5)

This quantity was averaged over random realizations of the donor and acceptor array. Calculations were carried out for \( N = 100, 200, 400, 800 \) at \( K = 0.1 \) and 0.3, and for \( N = 100, 200, 400, 800, \) and 1600 at \( K = 0.5, 0.7 \) and 0.9.
Fig. 14.11. Dependences of \( \Delta \eta_c = \frac{1}{2} N_D^{1/3} a \Delta t \) (right scale, - - - ) and the activation energy \( \varepsilon \) in units \( e^2 N_D^{1/3} \kappa^{-1} \) (left scale, - - - - ) on the dimensionless inverse temperature. Curves 1 and 2 were obtained without taking adiabatic processes into account for \( K = 0.5 \), \( N = 1600 \) and \( K = 0.3 \), \( N = 800 \), respectively. Curves 3 and 4 include adiabatic processes and correspond to \( N = 100 \) for \( K = 0.5 \) and 0.3, respectively.

Fig. 14.12. Same as Fig. 14.11 without the inclusion of adiabatic processes. (a) \( K = 0.9 \), \( N = 1600 \), (b) \( K = 0.7 \), \( N = 800 \), (c) \( K = 0.1 \), \( N = 800 \). Dotted line represents extrapolation to \( N = \infty \) for \( K = 0.9 \).

The results of calculating \( \Delta \eta_c \) with the highest values of \( N \) are shown in Figs. 14.11 and 14.12 (in Fig. 14.11, Curves 3 and 4 should be ignored for the time being). According to (14.4.2) and (14.4.5), one has \( \Delta \eta_c \propto \ln [\rho(T)/\rho_0] \), whence, given a proper choice of scale, the graphs of \( \Delta \eta_c(t^{-1}) \) will yield the dependence \( \ln [\rho(T)/\rho_0] \) versus \( T^{-1} \). These graphs are also convenient for finding the local activation energy \( \varepsilon \equiv d(\Delta t)/d(kT)^{-1} \). According to (14.4.4) and (14.4.5), we have

\[
\varepsilon(t^{-1}) = \frac{e^2 N_D^{1/3}}{\kappa} \frac{d(\Delta \eta_c)}{d(t^{-1})}.
\]

The functions \( \varepsilon(t^{-1}) \) in units of \( e^2 N_D^{1/3} \kappa^{-1} \) are also plotted in Figs. 14.11 and 14.12.

The dependence of the results on the number of donors showed no systematic size effect in the range \( 0.1 \leq K \leq 0.7 \). We therefore assumed that in this range the relation \( \eta_c(t^{-1}) \) corresponding to \( N = \infty \) is the same as that corresponding to the largest arrays investigated (Figs. 14.11, 12). On the other hand, for \( K = 0.9 \) there was a noticeable size effect. However, since this effect changed sign at \( t^{-1} = 0.15 \), the extrapolation to \( N = \infty \) was very crude. The dotted curve in Fig. 14.12 shows the result of this extrapolation.
In all graphs in Figs. 14.11,12 we can distinguish two regions. In the high-temperature region \( (0 \leq t^{-1} \leq 0.05) \) increasing \( t^{-1} \) results in a rapid decrease in the local activation energy \( \epsilon \) according to the following law:

\[
\epsilon(t^{-1}) = \epsilon(0) - ct^{-1}e^{2N_D^{1/3}k^{-1}},
\]

where \( c \) is a numerical coefficient. In the usual low-temperature region \( (0.15 \leq t^{-1} \leq 0.5) \) the energy \( \epsilon \) is practically constant and at \( K \gtrsim 0.3 \) it is smaller than \( \epsilon(0) \) by a factor of 1.5 to 2.8. We define \( \epsilon_u \) as the median energy \( \epsilon(t^{-1}) \) in this region. There is an area of smooth transition between the two regions \( (0.05 \leq t^{-1} \leq 0.15) \).

Due to the shunting effect of electrons activated in the conduction band, measurement of the hopping conduction in samples with \( N_D < N_m \) is usually feasible only for \( t^{-1} \geq 0.15 \). On the other hand, the very high resistance of the samples at low temperatures usually restricts the region in which hopping conduction can be measured to \( t^{-1} \leq 0.5 \). Thus the temperature range \( 0.15 \leq t^{-1} \leq 0.5 \) will be our operative region; \( \epsilon_u \) corresponds to precisely that region.

Thus, the presence on the theoretical curve of a high-temperature region, where \( \epsilon \) is a relatively fast decreasing quantity, does not contradict the fact that one usually observes a unique value of \( \epsilon \). There is hope, however, that in those cases where the shunting action of free electrons is reduced (for example, when dealing with hopping conduction over deep impurities) the transition area \( (0.05 \leq t^{-1} \leq 0.15) \) may be observable.

Let us discuss the theoretical interpretation of the shape of the curves \( \Delta \eta_c(t^{-1}) \). In Sect. 8.3, in order to calculate \( \xi_c \) in the high-temperature region, we developed a perturbation method with respect to the energy term in (14.4.1). As \( t^{-1} \to 0 \), the first-order perturbation theory yields (8.3.11):

\[
\epsilon(0) = \epsilon_p, \quad \text{where}
\]

\[
\epsilon_p = <\epsilon_{ij}>|_{r_n=r_e},
\]

and the symbol \( <...>|_{r_n=r_e} \) denotes averaging over all pairs of length close to \( r_e \). In [14.6] the energy \( \epsilon_p \) was calculated on a computer by directly selecting such pairs. The values thus obtained proved to be in good agreement with the values of \( \epsilon(0) \), which shows the applicability of the perturbation theory as \( t^{-1} \to 0 \). In that same study [14.6] it was demonstrated that the perturbation theory also yields the coefficient \( c \) in (14.4.7). The resulting value of \( c \) agrees with numerical experiment.

Thus, the perturbation theory successfully describes the high-temperature region. Unfortunately, we cannot yet interpret with sufficient rigor the slight variation of \( \epsilon(t^{-1}) \) in the region \( 0.15 \leq t^{-1} \leq 0.5 \) which leads to a relatively

---

3. For example, in \( n \)-germanium with antimony concentration \( N_D = 3 \times 10^{15} \) cm\(^{-3} \) the operative region corresponds to the temperature range 4.3 - 1.3 K.
well-defined energy $\varepsilon_u$. As shown in Chap. 8, the existence of a second activation energy at $K \to 1$ is due to the large-scale potential. However, we think that in its pure form this explanation should hold only for $K$ very close to unity. Possibly, for intermediate degrees of compensation, a major contribution to the existence of $\varepsilon_u$ results from extended tails of the density of states (Fig. 14.1). Indeed, as we saw in Sect. 8.3, in the case of a "rectangular" density of states (8.3.12) the perturbation-theory region passes into the Mott-law region relatively quickly, making it impossible to single out an extended range with an approximately constant activation energy $\varepsilon_u$ different from $\varepsilon_p$ [5.14]. In the presence of tails it appears logical that the origin of $\varepsilon_u$ is connected with a nonuniform narrowing of the energy band of the states around the Fermi level that determine conduction. At first, the edges of the band are located in the tails, and the band contracts rapidly as the temperature is lowered. Then the band’s edges reach the energies at which the density of states $g(\varepsilon)$ is large, and the width of the band becomes approximately pinned in some range of temperatures. This may be the cause of the relatively constant energy $\varepsilon_u$.

Energies $\varepsilon(0) = \varepsilon_p$ and $\varepsilon_u$ are plotted as functions of $K$ in Fig. 14.13, which also shows the experimental data (discussed in greater detail below). The first calculation of $\varepsilon_3$ at $K = 0.5$ (in which the numerical information about the impurity band structure was used) was performed in [14.2] via the perturbation-theory method (14.4.9). It turned out that the calculated value exceeded the experimental one by a factor of 4 (see Fig. 14.13). When it was discovered in [14.6] that at low values of $t^{-1}$ there is a transition from $\varepsilon(0) = \varepsilon_p$ to a much lower energy $\varepsilon_u$, it became clear that $\varepsilon_p$ does not tie in with the observed values of $\varepsilon_3$ and that the latter should instead be compared with $\varepsilon_u$.

Fig. 14.13. $K$ dependences of calculated and experimental values of the activation energy (in units $e^2N_d^{1/3}k^{-1}$). (1) $\varepsilon_p$, (2) $\varepsilon_u$, (3) $\varepsilon_A$. Experimental data: (a) $p$-Ge [14.3], (b) $n$-Ge [4.9]. Vertical arrows represent extrapolation to the classical case, slanted arrows correspond to changing $K_0$ according to [14.14].
In Chap. 8 we outlined a general framework in which the existence of a temperature-independent activation energy $\epsilon$ was related to the smallness of the parameter $t^{-1}$. The latter permits us to neglect the second term in (14.4.7), whence it follows that $\epsilon_3 = \epsilon(0)$. In the operative region the parameter $t^{-1}$ is indeed small compared to unity, and prior to the computer modelling we could assume that $\epsilon_3 = \epsilon(0)$. However, modelling of the impurity band revealed the coefficient $c$ to be so large that in the operative region the second term in (14.4.7) is comparable to the first. Consequently, the activation energy $\epsilon_3$ is very different from $\epsilon(0)$, and the explanation of its temperature independence is no longer universal. It now depends on the specific structure of the classical impurity band.

For $0.1 \leq K \leq 0.9$ the calculated $\epsilon_u$ still exceeds the observed values of $\epsilon_3$ by a factor of 1.4-1.7. An attempt was made in [14.6] to eliminate this substantial discrepancy within the framework of an idea advanced by Knotek and Pollak [10.1,2]. They demonstrated that the model of an equivalent resistor network which uses (10.1.17) for $\epsilon_{ij}$ may overstate the values of $\epsilon_{ij}$ and, consequently, $\epsilon_3$. That model assumes that while an electron hops from donor $i$ to donor $j$ all other electrons are frozen, i.e., the distribution of electrons among donors corresponding to $T = 0$ is conserved. As discussed in Sect. 10.2, Knotek and Pollak investigated two types of processes that account for electron-electron Coulomb correlations. They termed one type adiabatic and the other simultaneous. We will devote most of our attention to adiabatic processes. Consider four donors $i$, $j$, $c$, and $d$ — which at $T = 0$ have two electrons: electron 1 on donor $i$ and electron 2 on donor $c$ (in Fig. 10.4 donors $i$, $j$ were labeled $a$ and $b$). All other electrons are assumed to be frozen in the positions they occupied at $T = 0$; in other words, we will ignore the correlation between their motion and the motion of the two electrons we have singled out. Knotek and Pollak's idea was to consider the transitions $i \rightarrow j$ in those segments of time in which electron 2 is on donor $d$.

Let us denote the activation energy of the probability of an $i \rightarrow j$ transition in those time segments by $\epsilon_{ij}(c \rightarrow d)$. According to Knotek and Pollak, it is given by

$$
\epsilon_{ij}(c \rightarrow d) = \max \left\{ 0, \left[ \epsilon_j - \epsilon_i - \frac{e^2}{k} \left( r_{ij}^{-1} - r_{jd}^{-1} + r_{jc}^{-1} + r_{id}^{-1} - r_{ic}^{-1} \right) \right] \right\}.
$$

(14.4.10)

The fraction of time electron 2 spends on donor $d$ is proportional to $\exp \left( -\left( \epsilon_d - \epsilon_c - \frac{e^2}{k r_{cd}} \right) / k T \right)$ and typically it is very small. However, for some configurations of the "main" pair $(i, j)$ and the "auxiliary" pair $(c, d)$ the Coulomb interaction of electrons may make the energy $\epsilon_{ij}(c \rightarrow d)$ so much smaller than $\epsilon_{ij}$ [see (10.1.17)] that in the expression for the time-averaged frequency of $i \rightarrow j$ transitions,

$$
\Gamma_{ij} = \Gamma_{ij}^0 \exp \left( -2r_{ij} / a \right) \left[ \exp \left( \frac{-\epsilon_{ij}}{k T} \right) + \exp \left( \frac{-\epsilon_{ij,cd}}{k T} \right) \right],
$$

(14.4.11)
where

\[ \epsilon_{ij,cd} \equiv \epsilon_d - \epsilon_c - \frac{e^2}{kr_{cd}} + \epsilon_{ij}(c \rightarrow d), \quad (14.4.12) \]

the second term may dominate, i.e., there may hold the inequality \( \epsilon_{ij,cd} < \epsilon_{ij} \). As a result, the activation energy for \( R_{ij} \) may become lower than \( \epsilon_{ij} \). The new value of \( \epsilon_{ij} \) which accounts for adiabatic processes will be denoted \( \epsilon_{ij}^A \). The auxiliary pair with the lowest \( \epsilon_{ij,cd} \) turns out to be the most important. If this energy is lower than \( \epsilon_{ij} \), then it determines \( \epsilon_{ij}^A \), that is,

\[ \epsilon_{ij}^A = \min \left\{ \epsilon_j - \epsilon_i - \frac{e^2}{kr_{ij}}, \min \epsilon_{ij,cd} \right\}. \quad (14.4.13) \]

The second type of process discussed by Knott and Pollak consists of simultaneous two-electron transitions \( i \rightarrow j, c \rightarrow d \) in a single quantum act. The activation energy of such a process

\[ \epsilon_j - \epsilon_i + \epsilon_d - \epsilon_c - \frac{e^2}{k} \left( r_{cd}^{-1} + r_{ij}^{-1} - r_{jd}^{-1} + r_{jc}^{-1} - r_{ic}^{-1} + r_{id}^{-1} \right) \]

is less than or equal to \( \epsilon_{ij,cd} \) for the adiabatic process. However, the transition frequency includes an additional small factor \( \exp(-2r_{cd}/a) \), so that for low impurity concentrations and not-too-low temperatures, simultaneous two-electron transitions play a less important role than adiabatic hops and can be neglected.

Expression (14.4.13) was obtained for the case when one of the donors \( i, j \) is occupied and the other is empty. In [14.6] it was supplemented by an expression for the energy \( \epsilon_{ij}^A \) for pairs of two occupied or two empty donors. Then these two expressions for \( \epsilon_{ij}^A \) were used instead of (10.1.17) in calculating the dependence of \( \Delta n_i^A \) on \( t^{-1} \). The results are partially presented in Fig. 14.10. New values of the activation energy \( \epsilon_{ij}^A \) in the operative region are plotted in units of \( e^2N_D^{1/2}k^{-1} \) against degrees of compensation in Fig. 14.13 (Curve 3). It appears that the absolute error is less than 0.1.

In Fig. 14.13 we compare the final theoretical curve (3) with experimental data for \( p \) and \( n \)-Ge. The figure shows the experimental values of \( \epsilon_3 \) in units of \( e^2N_D^{1/2}k^{-1} \) obtained by Fritzche and Cuevas [14.3] for a series of \( p \)-Ge samples with \( N_A = 2.66 \times 10^{15} \text{ cm}^{-3} \) (see also Fig. 4.3). The values of \( \epsilon_3 \) were determined in the temperature range 5.1.3 K, which corresponds to \( t^{-1} \) ranging from 0.18 to 0.68, that is, practically the operative region. Compensation was produced by irradiating \( n \)-type samples with equal doses of slow neutrons. Nuclear reactions caused by the neutrons produced acceptors (Ga) and donors in the ratio \( N_D = K_0N_A \), where according to Fritzche and Cuevas, \( K_0 = 0.4 \). If the donor concentration \( N_D^0 \) in the original, low-compensated \( n \)-type sample is known, then one can measure the hole concentration \( p \) at a relatively high temperature after the irradiation and easily find the values of \( N_A \) and \( K \) after irradiation:
\[
N_A = \frac{N_B^0 + p}{1 - K_0},
\]
\[
K = K_0 + \frac{N_B^0}{N_A}.
\]

(14.4.14)

Given equal dosage, the concentration \(N_A\) is the same for all irradiated samples. The relatively large Bohr radius of the acceptor \((a = 90\text{Å})\) means that samples with \(N_A = 2.66 \times 10^{-5}\text{cm}^{-3}\) are not quite classical. This was demonstrated by Gadzhiev and Shlinak [7.16] and Knotek [8.8], who studied the changes in \(\epsilon_3\) due to wave-function compression in a strong magnetic field for a series of \(p\)-type samples with \(K = 0.4\) and various values of \(N_A\). Extrapolation to extremely strong magnetic fields led to the value \(\epsilon_3(K = 0.4) = 0.6 \ e^2 N_B^{1/3} K^{-1}\), which can be considered classical. No extrapolation was carried out for samples with \(K > 0.4\). Clearly, however, the upward shift of data points resulting from extrapolation to the classical case should be lower in such samples, because the increasing dispersion of levels due to higher \(K\) diminishes quantum effects. The actual shift at \(K = 0.4\) and the hypothetical shifts at \(K > 0.4\) are represented by vertical arrows in Fig. 14.13.

For \(K \leq 0.5\) the discrepancy between experimental and theoretical values of \(\epsilon_3(K)\) never exceeds 0.15. This discrepancy becomes even smaller if we accept the point of view of Thomas and Convington [14.14], who used different cross sections of nuclear reactions and argued that \(K_0 = 0.26\) and not 0.4. The values of \(K\) and \(N_A\) should then be recalculated in accordance with (14.4.14). This leads to a different shift of experimental points, indicated by slanted arrows. The agreement between experiment and theory improves, but a significant discrepancy still remains at \(K \geq 0.7\). The reason for this discrepancy is unclear.

In Fig. 14.13 we also provide the results of a study by Davis and Compton [4.9] on an \(n\)-germanium sample with an antimony concentration \(N_D = 6.7 \times 10^{15}\text{cm}^{-3}\) (sample E3). Studies of the effect of uniaxial stress on a sample with very similar parameters (Chap. 6) suggest the classical nature of sample E3. Reasonable agreement with the theoretical curve (3) is in evidence, but we shall not discuss Davis and Compton's data further, because they created compensation by irradiating with fast neutrons. It is not clear in this case how strong correlations in the configuration of acceptors are; they are not included in the theory.

So far we have discussed only the activation energy \(\epsilon_3\). However, the dependence \(\Delta \eta_c(t^{-1})\) derived above has an interesting consequence for the resistivity \(\rho_3\) [see (8.1.1)]. From Figs. 14.11 and 14.12 we see that in the operative region the function \(\Delta \eta_c(t^{-1})\) can be written in the form \(\Delta \eta_c = \tilde{\eta}(K) + \epsilon_u t^{-1}\), where \(\tilde{\eta}(K)\) is a positive quantity that can be graphically determined as an intercept on the ordinate axis of the straight segment of the curve \(\Delta \eta_c(t^{-1})\) in the operative region, extrapolated to \(t^{-1} = 0\). By applying (14.4.2) and (14.4.5) to the value of \(\rho_3\) found in the operative region, we get
\[ \rho_3 = \rho_0 \exp \left( \frac{\xi_c^0 + 2\tilde{\eta}(K)}{N_D^{1/3}a} \right) = \rho_0 \exp \left( \frac{\alpha(K)}{N_D^{1/3}a} \right), \quad (14.4.15) \]

where

\[ \alpha(K) = 1.73 + 2\tilde{\eta}(K). \quad (14.4.16) \]

In [14.6] the following values of \( \alpha(K) \) were found: \( \alpha(0.1) = 1.76 \), \( \alpha(0.3) = 1.79 \), \( \alpha(0.5) = 1.81 \), \( \alpha(0.7) = 1.9 \), \( \alpha(0.9) = 2.1 \). We see that \( \alpha \) is an increasing function of \( K \). Qualitatively, this explains the increase of \( \rho_3 \) with \( K \) observed at \( K \geq 0.4 \) by Fritzsche and Cuevas [14.3], and Davis and Compton [4.9]. Quantitatively, the calculated increase in \( \ln \rho_3 \) at \( K = 0.7 \) is in adequate agreement with their data, but at \( K = 0.9 \) it is almost two times lower than the experimental value [14.3]. This discrepancy is probably of the same nature as that between the theoretical and experimental values of \( \epsilon_3 \) at \( K = 0.9 \).

We see from Table 6.1 that in many semiconductors at intermediate compensation the values of \( \alpha \) found from the experimental dependence of \( \ln \rho_3 \) on \( N_D^{1/3} \) exceed 1.73 by an amount of order 0.1. This can be explained by (14.4.16). Comparing theory and experiment is of especially great importance for a series of samples of pure Ge crystals irradiated with various doses of neutrons and having \( K = K_0 \). If \( N_A \) is obtained from the expression

\[ N_A = p/(1 - K_0) \]

and \( K_0 \) is assumed to be 0.4, then an analysis of the dependence \( \rho_3(N_A) \) for these samples yields \( \alpha_{\text{expt}}(0.4) = 1.9 \) (see Table 6.1). If we assume \( K_0 = 0.26 \), then all values of \( N_A \) are diminished and we find \( \alpha_{\text{expt}}(0.26) = 1.77 \). On the other hand, interpolating between the calculated values cited above, we find \( \alpha_{\text{theor}}(0.26) = 1.78 \) and \( \alpha_{\text{theor}}(0.4) = 1.80 \). Thus, for \( K_0 = 0.26 \) the agreement between theory and experiment is better, which is an additional argument in favor of \( K_0 \) being 0.26.

Recently, the value of \( K_0 \) was measured again by Zabrodskii [14.15]. He carefully studied the changes in carrier concentration in the decay of products of neutron-irradiated germanium, and found \( K_0 = 0.30 \).

### 14.5 Variable-Range Hopping Conduction

If we assume that in the energy range \( |\epsilon - \mu| > 0.15e^2N_D^{1/3}\kappa^{-1} \) the polaron effect is small, then it makes sense to simulate variable-range hopping on the assumption that individual electrons rather than polarons will do the hopping. Since VRH always involves a band of states around the Fermi level with the width \( \epsilon_{\text{max}} = kT \xi_c \gg kT \) [see (9.1.10)], the range of temperature in which the single-electron approach works is bounded from below not by the inequality \( kT > 0.15e^2N_D^{1/3}\kappa^{-1} \), but by the much weaker inequality

\[ kT > 0.15e^2N_D^{1/3}/k\xi_c, \]

which does indeed hold in many experiments.

Consequently, it makes sense to study variable-range hopping by using the Miller and Abrahams resistor network and general percolation methods. In
[14.7] this was done for \( K = 0.5 \) and the temperature range \( 1 \leq t^{-1} \leq 20 \). The percolation threshold \( \xi_c \) was determined by testing for percolation between opposite faces of a cube using the same bonding criterion (14.4.1) and the same procedure as that used in connection with \( \epsilon_3 \) conduction (Sect. 14.4). In studying the dependence \( \eta_c (t^{-1}) \) at low temperatures, that is, at large values of \( t^{-1} \) \((t^{-1} \geq 5)\), we found a strong size effect. This is not surprising, considering that at low temperatures the hop length becomes appreciably larger than \( N_D^{-1/3} \). As a result, the cube’s edge length is covered in several hops. This means that the arrays under investigation are not macroscopic enough to find the percolation threshold, although they are sufficiently large to permit a quite accurate description of the impurity band structure. The “percolation” size effect makes \( \eta_c \) dependent on \( N \), but also on \( q \) — the quantity that determines the thickness of the layers where percolation begins and ends. The trouble is that at low values of \( N \) and \( q \) the number of donors in such a layer is small. For instance, for \( N = 100 \) and \( q = 1/2 \) each layer contains on the average 9 donors. At high values of \( t^{-1} \) many of these donors do not take part in percolation because of the constraints the bonding criterion imposes on the donor’s energy. Consequently, the “coupling” between the infinite cluster and the cube’s faces may prove inadequate, which would make \( \xi_{cN} \) greater than \( \xi_{c \infty} \). In this case, as \( q \) is increased, the quantity \( \xi_{cN} \) should diminish.

Figure 14.14 shows the results of calculations at \( t^{-1} \geq 5 \), for which \( \eta_c \) is dependent on the number of donors \( N \). Two values of \( q \) were used in these calculations: \( q = 0.5 \) (light circles) and \( q = 1 \) (dark circles). One can see that the values of \( \eta_{c \infty} \) obtained by extrapolating to \( N = \infty \) using the law

![Fig. 14.14. Extrapolation of \( \eta_{cN} \) to \( N = \infty \) for variable-range hopping conduction. (○) \( q = 1/2 \), (●) \( q = 1 \)](image-url)

![Fig. 14.15. Dependence of \( \eta_c \) on \( t^{-1} \) (○, top scale) and on \( t^{-1/2} \) (x, bottom scale) for VRH conduction](image-url)
\( \eta_cN = \eta_c\infty + \text{const} \cdot N^{-1/3} \) do not depend on \( q \). At \( t^{-1} = 1 \) and \( t^{-1} = 2 \) the values of \( \eta_c \) were practically independent of \( N \) and \( q \).

In Fig. 14.15 we plot the values of \( \eta_c \) extrapolated to \( N = \infty \) against \( t^{-1} \) (circles) and \( t^{-1/2} \) (crosses). We see that in the whole region \( t^{-1} \geq 1 \) the crosses fall near a straight line, i.e., law (10.1.18) holds. The value of coefficient \( \beta_1 \) in (10.1.19), judging by the slope of this straight line, equals 2.8.

In order to verify the accuracy of the value of \( \beta_1 \) obtained in the framework of the single-electron approach, we ran the program described in Sect. 14.4 to calculate \( \eta_c \) at \( t^{-1} = 5 \) and \( N = 100 \), taking account of adiabatic two-electron processes. As it turned out, the value of \( \eta_c \) fell only by 20\%, which shows that adiabatic processes do not alter the law (10.1.19) and have relatively little effect on the value of \( \beta_1 \).

In Chap. 9 we derived, following Pollak [9.3] and Hamilton [9.10], the expression (10.1.18) for the case when \( g(\epsilon)c|\epsilon - \mu|^2 \). However, that derivation neglected the very important factor \(-e^2/kr_{ij}\) in (10.1.17), which corresponds to the exclusion of self-interaction. Direct application of this derivation to the case of the Coulomb gap is therefore inconsistent, and one needs a more rigorous consideration. Below we shall analytically derive law (10.1.18) using an approximate solution of the percolation problem; an approximate value of the coefficient \( \beta_1 \) will be found concurrently [14.7].

Let us introduce a function \( F(\omega, r) \) such that \( 4\pi F(\omega, r)r^2dr\,d\omega \) is the probability of finding in a unit of volume a pair of donors whose separation lies in the interval \( (r, r + dr) \) and whose transition energy (10.1.17) is in the interval \( (\omega, \omega + d\omega) \). Such a pair is considered bonded if it satisfies the inequality (14.4.1), where \( \xi = \xi_c \). Let us use the two bonded donors as poles for a sphere of radius \( r/2 \) and introduce the quantity \( \Theta \) to represent the ratio of the volume contained in all such spheres to the total volume of the system:

\[
\Theta = \int F(\omega, r) \theta \left( \xi_c - \frac{\omega}{kT} - \frac{2r}{a} \right) \frac{\pi r^3}{6} d^3r \, d\omega, \tag{14.5.1}
\]

where \( \theta \) is a step function \([\theta(x) = 1 \text{ if } x > 0]\). It is a simple matter to find \( \Theta \) in the limiting case \( t^{-1} = 0 \) where \( \xi = \xi_c^0 = 2r_c/a \) and \( r_c = 0.865 N_D^{-1/3} \) is the percolation radius (Sect. 6.1). Dropping the factor \( \omega/kT \) in the argument of the \( \theta \) function and using the normalization condition \( \int_0^\infty F(\omega, r)\,d\omega = N_D^2/2 \), we get

\[
\Theta = \frac{\pi^2}{3} N_D^2 \int_0^{r_c} r^5dr = 0.23. \tag{14.5.2}
\]

Further, we will assume that the fraction of the total volume, \( \Theta \), is an invariant of the percolation problem, i.e., that it is independent of \( t \). Then, given a known function \( F(\omega, r) \), we can solve (14.5.1) for \( \xi_c \) and find the
dependence \( \xi_c(t) \), that is, \( \ln \sigma(T) / \sigma_3 \). To the first order in \( t^{-1} \) we find

\[
\xi_c - \xi_c^0 = \frac{\int_0^\infty F(\omega, r_c) \omega \, d\omega}{kT \int_0^\infty F(\omega, r_c) \, d\omega},
\]

(14.5.3)

which coincides with the result of the first-order perturbation theory (8.3.10).

In the region \( t^{-1} \gg 1 \) the density of states in the immediate vicinity of the Fermi level is of importance. As shown in Sect. 14.2, it is described by (10.1.6) with \( \alpha_3 = \frac{3}{\pi} \), which follows from the self-consistent equation. Neglecting the interaction between different donor pairs, we can write

\[
F(\omega, r) = \frac{1}{2} \int_0^\infty g(\epsilon_1) \, g(\epsilon_2) \, \delta(\epsilon_{12}(r) - \omega) \, d\epsilon_1 \, d\epsilon_2,
\]

(14.5.4)

where \( \epsilon_{12}(r) \) is given by (10.1.17). By substituting (10.1.6) into (14.5.4) and solving for \( \xi_c \) from (14.5.1), (14.5.2) and (14.5.4), we obtain the law (10.1.18) with \( \beta_1 = 2.6 \), which is very close to the value \( \beta_1 = 2.8 \) found from Fig. 14.15.

A comparison between experimental data on variable-range hopping conduction and (10.1.18) for \( \beta_1 = 2.8 \) was given in Sect. 10.1.

### 14.6 Activation Energy \( \epsilon_1 \) of Band Conduction

In the temperature range where conduction is determined by free electrons, the resistivity of a lightly doped semiconductor is given by (4.1.4). As discussed in Sect. 4.1, the activation energy \( \epsilon_1 \) of compensated samples in the limiting case \( N_D \rightarrow 0 \) equals the ionization energy \( E_0 \) of an isolated impurity. Experiment shows that \( \epsilon_1 \) diminishes as \( N_D \) increases, except in the case of high compensation. Pearson and Bardeen [14.16] suggested that this phenomenon could be related to the effect of the potential of charged impurities on free electrons. Dimensional considerations indicate that the most general expression for this effect, given sufficiently light doping and sufficiently low temperature, can be written in the form

\[
\epsilon_1 = E_0 + f(K) \frac{e^2 N_D^{1/3}}{\kappa},
\]

(14.6.1)

where \( f(K) \) is a universal function. To find \( f(K) \) Pearson and Bardeen, apparently interested in the case of low compensation, considered the free-electron gas of concentration \( N_D \) found in a field of charged donors with that same concentration. They calculated the electrostatic energy of the gas per electron. To do this, they subdivided the volume into Wigner-Seitz cells of radius \( (4\pi N_D/3)^{-1/3} \) and, assuming a uniform electron gas, calculated the
interaction energy between the sphere and the central donor, as well as the electrostatic energy of the sphere itself. To these two energies they added the correlation energy per electron that corresponds to the formation of a face-centered Wigner lattice in a uniform electron gas. By dividing the sum by $e^2 N_D^{1/3} \kappa^{-1}$ they found $f = -2.65$.

Castellan and Seitz [14.17] and Debye and Conwell [14.18] noted a major flaw in Pearson and Bardeen’s approach [14.16]. They observed that at low values of $K$ and low temperatures, practically all electrons are on donors, so that the overwhelming majority of donors remain neutral. At a finite degree of compensation $K$ and low temperature, there remain $KN_D$ charged donors. Therefore Debye and Conwell [14.18] suggested that $N_D$ in Pearson and Bardeen’s calculations be replaced by $KN_D$; they obtained $f(K) = -2.65K^{1/3}$.

In our view such a substitution is not quite consistent, since at low temperatures the concentration of free electrons is not $KN_D$ but is much lower. Also, all calculations based on Pearson and Bardeen’s approach ignored the acceptor potential.

Neumark [14.19] proposed another method of calculating $f(K)$. He reasoned that the screening of a charged donor’s Coulomb potential according to the law

$$ e \phi(r) = \frac{e^2}{\kappa r} e^{-r/r_0} $$

leads to an upward shift in the level that corresponds to the ground electron state on a donor. For $r_0 \gg a$ this shift is given by

$$ \Delta E = \int \Psi^2(r) \frac{e^2}{\kappa r} \left[ 1 - e^{-r/r_0} \right] d^3r = \frac{e^2}{\kappa r_0} \left[ 1 + 0 \left( \frac{a}{r_0} \right) \right], $$

where $\Psi(r)$ is the ground-state wave function. At $kT \ll E_0$, when the free-electron concentration is very low, screening comes about because of the redistribution of localized electrons among donors. If $kT$ is greater than the characteristic energy-level dispersion $\gamma$ (at intermediate compensation $\gamma \approx e^2 N_D^{1/3} \kappa^{-1}$), then Debye’s linear screening theory is applicable and $r_0 > N_D^{-1/3}$. However, in this case the correction to the activation energy is of little interest, since its calculation yields a quantity that is less than $kT$. Conversely, if $kT < \gamma$, then screening by localized electrons is essentially nonlinear. Neumark [14.19] followed Falicov and Cuevas [14.20] in accepting that in such a case $r_0$ is of the order of the mean distance between electrons, that is

$$ r_0 \approx (N_D - N_A)^{-1/3}. $$

The substitution of (14.6.4) into (14.6.3) leads to (14.4.1) with $f(K) = -\beta(1 - K)^{-1/2}$, where $\beta$ is a positive numerical coefficient. Leloup et al. [14.21] compared this formula with their experimental data for $n$-type InP with intermediate compensation and found that $\beta = 1.85$. 
In our opinion, (14.6.4) does not allow for a reliable determination of \( f(K) \), since for \( 1 - K \ll 1 \) this expression is very different from that for the nonlinear screening radius \( r_s \) obtained in Sect. 3.4 [see (3.4.8)]. The discrepancy is due to the fact that Falicov and Cuevas did not consider that large fluctuations of the charged impurities’ potential make for a free-electron distribution that is far from uniform, i.e., electrons are "banished" from regions with a characteristic size much greater than \( (N_D - N_A)^{-1/3} \). Nor does Neumark’s approach allow us to find the numerical value of \( f(K) \) at intermediate compensation, when (14.6.4) and (3.4.8) are of the same order of magnitude: as we saw in Sect. 14.3, nonlinear screening in the first approximation stipulates that the empty donor is the one nearest to the acceptor. Obviously such a phenomenon cannot be described within the framework of the linear screening formula (14.7.2), which is isotropic and contains no random element at all.

Another approach to calculating the correction to \( \epsilon \) was discussed in Sect. 8.2 for the case of high compensation. Here we will apply it to an arbitrary degree of compensation. The approach consists in seeking \( \epsilon_1 \) as the difference between the mobility threshold in the conduction band and the Fermi level. If the concentration \( N_D \) is different from zero, the mobility threshold does not coincide with the unperturbed bottom of the conduction band, because the latter is warped by spatial fluctuations of the charged impurities’ potential.\(^4\) The Fermi level is also different from the level of an isolated donor because of the dispersion of levels (Sect. 14.2). A combination of these two effects determines the correction to \( \epsilon_1 \).

In Sect. 8.2 we found that it becomes much easier to locate the mobility edge if in the potential profile corresponding to the bottom of the conduction band the electron can be considered a classical particle. In this case the mobility edge coincides with an important characteristic of the potential, namely the percolation level \( V_c \). Then for energy \( \epsilon_1 \) we get [see, for example, Fig. 3.5 and (8.2.13)]:

\[
\epsilon_1 = E_0 + V_c - \mu .
\]

Whether we can consider as classical the behavior of a particle of mass \( m \) in a random potential diagram of amplitude \( \gamma \) and spatial scale \( r \) depends on the magnitude of the parameter \( Q^2 \), defined as the ratio of \( \gamma \) and the characteristic separation between levels:

\[
Q^2 = \frac{\gamma}{\hbar^2/mr^2}.
\]

---

4. It is important to note that in the calculation of the mobility edge, the potential relief of the bottom of the conduction band can be considered static, i.e., while an electron is moving in the band we can ignore the donor-to-donor transitions of other electrons. This is so because at temperatures that correspond to \( \epsilon_1 \) conduction, the time of jumping the barrier in the band is proportional to \( \exp(\gamma/kT) \) and is much less than the donor-to-donor transition time, which contains a factor of order \( \exp(2N_D^{-1/3}a) \).
If \( Q^2 \gg 1 \) the particle can be considered classical, since the probability \( W \) of tunneling through the potential barrier is exponentially small:\n\[ W \propto \exp(-Q) \ll 1. \] According to (8.2.12), at \( 1 - K \ll 1 \) one has
\[
Q = \text{const} \frac{1}{(N_D a^3)^{1/6} (1-K)^{5/6}} \gg 1.
\] (14.6.7)

On the other hand, we can use (3.3.10) and (3.3.13) to show easily that \( Q \gg 1 \) at low compensation as well.

Thus the classical condition is fulfilled at both high and low compensations. At an intermediate compensation \( Q \approx (N_D a^3)^{-1/6} \). At sufficiently low concentrations that correspond to \( N_D a^3 \leq 10^{-3} \) one has \( (N_D a^3)^{-1/6} \gg 3 \). Thus there is an a priori hope that at intermediate compensation the classical approach holds for these concentrations at least as a first approximation.

We conclude that (14.6.5) is applicable at all degrees of compensation. For sufficiently low values of \( N_D \) we can write
\[
\mu = \frac{e^2 N_D^{1/3}}{\kappa} \phi(K), \quad V_c = \frac{e^2 N_D^{1/3}}{\kappa} \psi(K),
\] (14.6.8)
where \( \phi \) and \( \psi \) are universal functions of \( K \). Then for \( f(K) \) we have
\[
f(K) = \psi(K) - \phi(K).
\] (14.6.9)

Function \( \phi(K) \), calculated according to the energy minimization program, is plotted in Fig. 14.3. In [14.22] a direct computation of the percolation level \( \psi(K) \) was attempted using the Monte Carlo method in a wide range of degrees of compensation. A subroutine was appended to the energy minimization program; it worked as follows. After a pseudoground state of \((1-K)N\) electrons on \( N \) donors in a cube of edge \( L = (N/N_D)^{1/3} \) was found, the subroutine constructed within the cube a simple \( n \times n \times n \) cubic lattice with a smaller period \( L/n \), where \( n \) is an integer (\( n = 18 \) at \( K < 0.8 \) and \( n = 16 \) at \( K \geq 0.8 \)). This auxiliary lattice was used to test for percolation. At each of its sites the potential energy of an electron \( V_i \) was calculated in the field of all charged donors and acceptors. Given a parameter \( V \) an array was built up:
\[
A_i = \begin{cases} 
1 & \text{if } V_i \leq V, \\
0 & \text{if } V_i > V,
\end{cases}
\] (14.6.10)
and, using an algorithm analogous to that described in Sect. 14.4, a value \( V = V_c \) was found at which percolation "over ones" between the opposite faces of the cube first came about. The value of \( V_c \) was averaged over many different realizations of donor and acceptor coordinates.
We will not dwell on the dependence of mean values of \( V_c \) on \( N \) and \( n \), or on the extrapolation of \( V_c(N,n) \) to an infinite array, as these questions are discussed in detail in [14.22]. We note, however, that the dependence of \( V_c(N,n) \) on two parameters, \( N \) and \( n \), makes extrapolation very difficult. Consequently, the absolute error in the determination of \( f(K) \) in [14.22] was of the order of 0.1 at \( K \leq 0.9 \) and 0.2 at \( K > 0.9 \).

The graph of \( f(K) \) obtained in [14.22] is displayed in Fig. 14.16, together with the theoretical dependences of Pearson and Bardeen [14.16], Debye and Conwell [14.18], Neumark [14.19], and Leloup et al. [14.21], as well as some experimental data. It is notable that \( f(K) \) as found in [14.22] begins to rise quickly with the growth of \( K \) above \( K = 0.9 \). Such behavior of \( f(K) \) is due to an increase in the amplitude of the random potential and a rapid lowering of the Fermi level into the forbidden band (see Fig. 14.3). Qualitatively it follows the theoretical prediction that \( f(K) \) should increase, as \( K \to 1 \), proportional to \((1-K)^{-1/3}\) [see (8.2.9)]. However, calculations show that the region in which this asymptotic law is applicable is restricted to \( K > 0.93 \).

A comparison of the calculated values of \( \varepsilon_1 \) with the values of \( \varepsilon_3 \) from Fig. 14.13 shows that the relation (8.2.10), which should hold in the limit \( K \to 1 \), is still not fulfilled in the investigated region of high compensations. Apparently even higher values of \( K(K > 0.95) \) are necessary for this relation to hold.

---

**Fig. 14.16.** The function \( f(K) = (\varepsilon_1 - E_0) (e^2 N_{D}^{1/3} / \kappa)^{-1} \). Solid line represents the results of [14.22]; dotted line corresponds to the theory of Pearson and Bardeen [14.16]; dash-dotted line to the theory of Debye and Conwell [14.18]; dashed line represents the expression of Neumark [14.19] and Leloup et al. [14.21]. Experimental data: (1) Shlimak and Emtez [8.11], (2) Davis and Compton [4.9]
From Fig. 14.15 we see that the rapid growth of $f(K)$ as $K \to 1$ comprises the main qualitative difference between [14.22] and the theoretical predictions of other authors. At intermediate values of $K$ there are marked quantitative discrepancies.

Before we compare experimental data, let us note that the correction to $\epsilon_1$ found in [14.22] makes sense only if $kT \ll \gamma$, i.e., when the system is close to the ground state. At an intermediate compensation one has $\gamma \approx e^2 N_D^{1/3}/k$ and the region in which the results of [14.22] are applicable is determined by the inequality $kT \ll e^2 N_D^{1/3}k^{-1}$. Experimentally, however, hopping conduction already takes over from band conduction at $kT = (0.3-0.5)e^2 N_D^{1/3}k^{-1}$. Accordingly, at intermediate compensation the result of [14.22] can be only qualitatively compared with the values of $\epsilon_1 - E_0$, where $\epsilon_1$ should be found from the data for the low-temperature portion of the temperature interval that corresponds to band conduction. For $K \leq 0.1$ comparison with (14.6.1) is even less propitious. In order for the Fermi level to coincide with its value (3.2.11), which corresponds to $T = 0$, we must have $kT < e^2 N_D^{1/3}k^{-1}[\ln K]$ (see Sect. 8.1). Thus at $K > 0.9$ we can expect quantitative agreement with experimental data a priori.

Figure 14.16 shows data from two experimental studies. Shlimak and Emteev (see [8.11] and Fig. 8.3) investigated highly compensated $n$-Ge samples doped with arsenic. For $K \leq 0.9$ there is an obvious discrepancy between these data and the calculated results of [14.22], while for $K > 0.9$ the experimental data points approach the theoretical curve. In Fig. 14.16 we also have the results of an analysis of Davis and Compton's data [4.9] for an $n$-Ge sample lightly doped with antimony ($N_D = 6.7 \times 10^{15}$ cm$^{-3}$). It is apparent that their results do not qualitatively agree with theory, although the discrepancy between the experimental and theoretical values of $\epsilon_1$ does not exceed $kT$. The values of $E_0$ used to analyze the experiments were found from the temperature dependence of conduction in extremely lightly doped samples. As in [14.23], $E_0$ was taken equal to 12.7 meV for As and 9.7 meV for Sb. It should be emphasized that the results of the experimental data analysis are very sensitive to the value of $E_0$.

The discrepancy between the experimental and theoretical values of $\epsilon_1$ at $K \leq 0.9$ can be due to two circumstances. First, (14.6.5) is exact only at $T = 0$. At finite temperatures, conduction is determined by electrons with energies of the order of $V_c + kT$, and the effective activation energy $\epsilon_1$ may exceed the value found from (14.6.5) by $kT$. No experiment contradicts this point of view, since the difference between the experimental and theoretical values (14.6.5) does not exceed the values of $kT$ that are characteristic of the region in which $\epsilon_1$ is determined.
Secondly, the samples that are compared with theory have \( N_D a^3 \approx 10^{-3} \), in which case parameter \( Q \) does not exceed 3 and therefore the potential is not altogether classical. There is a question about the difference between the mobility edge and the percolation level if parameter \( Q \) is large but finite. If we compare such a case with the purely classical one \( (Q = \infty) \), we encounter two new effects. On the one hand, an electron may tunnel under the potential barriers, which promotes a lowering of the mobility edge compared to the percolation level. On the other hand, when an electron traverses a saddle point of the potential, say, moving along the \( x \) axis, it is localized in a finite region in the \( y \) and \( z \) direction. From the uncertainty principle it must, therefore, have at the saddle point a non-zero kinetic energy of perpendicular motion, which raises the mobility edge above the percolation level. The second effect appears to be the more significant one. Indeed, let us consider the case of \( Q \ll 1 \), the opposite of the classical one. In this case the electron wave function effectively averages out potential fluctuations and the mobility edge falls near the unperturbed bottom of the conduction band, i.e., it lies much higher than the percolation threshold, which is independent of \( Q \) and is of the order of \( \gamma \). If we assume that given a fixed value of \( \gamma \) the mobility edge depends monotonically on parameter \( Q \), we find that at large values of \( Q \) the mobility edge lies above the percolation threshold. This may be one of the reasons why the observed values of \( \epsilon_1 \) exceed the theoretical ones.

14.7 Some Other Applications of the Minimization Program

The energy minimization program described in preceding sections has also proved useful for computing other low-temperature phenomena.

A large number of questions concern the distribution function \( F(\omega) \) of the excitation energy of compact pairs consisting of an empty and an occupied donor (Sect. 10.1). Here \( \omega \) is the work necessary to transport an electron from one donor of the pair to the other. As discussed in Sect. 10.1, \( F(\omega) \) should logarithmically tend to zero as \( \omega \to 0 \) because of the dipole-dipole interaction between pairs. As a consequence, the quantity \( C(T)/T \) should also logarithmically tend to zero, where \( C \) is the heat capacity of the impurity band electrons. In order to test the theoretical concepts presented in [14.24], an additional program was appended to the energy minimization program. This program sorts through all possible pairs and calculates the distribution function in pseudoground states. Using the lattice model described in Sect. 10.1, it was shown that at very low values of \( \omega \) the function \( F(\omega) \) decreases monotonically.

In [14.25] the heat capacity of impurity band electrons was calculated using the Monte Carlo method at finite temperatures \( T \geq 0.025 e^2 N_D^{1/3}/k k \), where \( k \) is the Boltzmann constant. It was shown that function \( C(T)/T \) has a maximum and decreases with temperature at \( T < 0.1 e^2 N_D^{1/3}/k k \), which was interpreted as an effect of the dipole-dipole interaction between pairs.
Compact donor pairs are responsible for the high-frequency conduction in semiconductors at low temperatures. Electromagnetic waves are absorbed as the electron transfers from one donor of the pair to the other. Two regimes are distinguished here (for more detail see [2.22]). At higher temperatures and lower frequencies one has the relaxation regime, in which the electron transition occurs at the expense of phonons, while the wave field modulates the equilibrium donor occupation numbers. In the other regime, the wave quantum is directly absorbed without the intercession of phonons. In both regimes the frequency-dependent conduction can be expressed using the distribution function of energies and lengths $F(\omega, r)$ defined in Sect. 14.5. It has been shown that taking account of electron-electron interaction changes the temperature dependence of the conductivity in the first regime [14.26] and the frequency dependence in the second [14.27]. For a detailed description, computer modelling is necessary. In [14.28-30] the minimization program was supplemented by a program that calculated $F(\omega, r)$. This allowed a quantitative comparison with experiment in a wide range of degrees of compensation. Such a comparison was performed both for the first [14.28,29] and the second [14.30] regimes, and good agreement was found [14.31].

The above impurity band model apparently applies to both shallow (hydrogenlike) and deep impurities. Its only restriction is that the donors and acceptors are assumed to be single-charged. In reality, deep impurities often have several charged states. Acceptors (in $n$-type semiconductors) are found in the state with the highest charge (for example, Cu impurities in Ge have a charge of 3), while the charged state of the donor is determined by the degree of compensation. In these cases, the impurity band structure and the conduction activation energy can be computed by the same programs with minor modifications in the expressions concerning both the total and one-electron energies.

In Table 14.2 we give an example of the results of calculations for the impurity band of an $n$-type semiconductor with single-charge donors and Z-charge acceptors. The values of the Fermi level are given in units of $e^2 N_b^{1/3} k^{-1}$ as a function of the compensation $K = Z N_d / N_d$ at $Z = 1,2,$ and 3 in an array of $N = 800$ donors. It is seen that for $K \ll 1$ the value of $\mu$

<table>
<thead>
<tr>
<th>$K$</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>0.88</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>0.1</td>
<td>0.84</td>
<td>0.80</td>
<td>0.73</td>
</tr>
<tr>
<td>0.5</td>
<td>0.31</td>
<td>0.04</td>
<td>-0.14</td>
</tr>
<tr>
<td>0.9</td>
<td>-3.18</td>
<td>-4.45</td>
<td>-5.17</td>
</tr>
</tbody>
</table>

Table 14.2. Position of the Fermi level in the case of compensation by multicharge acceptors. $K$ is the compensation ratio, and $Z$ is the charge of the acceptor.
remains practically unchanged for all values of \( Z \). On the other hand, as \( K \) enters the region \( K \geq 0.5 \), the higher the value of \( Z \), the faster the Fermi level sinks into the forbidden band. This is probably due to the fact that both the inhomogeneity in the impurity charge distribution and the amplitude of impurity potential fluctuations grow with increasing \( Z \).

In [14.28] a program similar to that described in Sect. 14.2 was used to compute the structure of a two-dimensional impurity band on the surface of the inversion layer in an MOS structure. A particularity of this problem is the fact that the combined charge of donors, acceptors, and electrons found on donors does not equal zero and is compensated by a charge on the metallic electrode. The computation leads to a density of states that can be observed through capacity measurements or from the dependence of the activation energy on the electron concentration in the inversion layer. Because of electron-electron interaction, this density of states is not identical to the one-electron density \( g(\epsilon) \) which we calculated in preceding sections. In particular, the observed density of states may prove very large or even change sign. This means that the traditional method of analyzing experimental data in this case yields not the one-electron density of states but some other quantity.

Let us emphasize once again that the computer modelling discussed in this part of the book applies only to the classical impurity band. The influence of quantum effects on the structure of the impurity band was completely ignored. However, as we have seen, electron-electron interaction renders even this simplified problem very difficult.

In the vicinity of the Mott transition, electron-electron interaction probably plays as major a role as it does with light doping. In this region, however, we also have delocalization of electron states because of quantum effects, which makes the problem immeasurably more difficult. Nevertheless we believe that a good understanding of the classical problem is of great importance for further progress in the study of doped semiconductors.
A.1 Evaluation of the Exponent in the Energy Overlap Integral $I_{ij}$ in a Magnetic Field

All terms in (4.2.5) have the same exponents. Therefore it is sufficient to evaluate the exponent corresponding to $\int F_i(r) F_j(r) \, dr$. We begin with the case of a strong field. Placing donor $i$ at the origin, we find

$$I_{ij} \sim \int_{-\infty}^{\infty} dx \, dy \, dz \, \exp \left\{ \frac{x^2 + y^2}{4\lambda^2} - \frac{|z|}{a_H} - \frac{(x - x_j)^2 + (y - y_j)^2}{4\lambda^2} - \frac{|z - z_j|}{a_H} + \frac{i}{2\lambda^2} (y_j x - x_j y) \right\} \times \exp \left\{ - \frac{|z_{ij}|}{a_H} - \frac{x_{ij}^2 + y_{ij}^2}{4\lambda^2} \right\} \quad \text{(A.1.1)}$$

In the case of a weak magnetic field it is convenient to choose a Cartesian coordinate system in which donor $i$ is at the origin and donor $j$ is on the $x$ axis, separated from $i$ by the distance $x_0 \equiv r_{ij}$. Let the vector $\mathbf{H}$ lie in the $x, z$ plane, making angles $\theta_1$, $\theta_2$, and $\theta_2$, respectively, with the $x$ axis and vectors $\mathbf{r}$ and $\mathbf{r} - \mathbf{r}_j$. Then we have

$$I_{ij} \sim \int d\mathbf{r} \, \exp \left\{ - \frac{r}{a} - \frac{r^3 a \sin^2 \theta_1}{24\lambda^4} - \frac{|r - r_j|}{a} \right. \\
- \frac{|r - r_j|^3 a \sin^2 \theta_2}{24\lambda^4} - \frac{i x_0 y}{2\lambda^2} \sin \theta \left. \right\} \quad \text{(A.1.2)}$$

The integrand in (A.1.2) has a sharp maximum in the vicinity of $(x_0/2, 0, 0)$. We can, therefore, expand the quantities $r$ and $|r - r_j|$ near this point and set the angles $\theta_1$ and $\pi - \theta_2$ equal $\theta$. As a result, we get

$$I_{ij} \sim \exp \left\{ - \frac{x_0}{a} - \frac{x_0^3 \sin^2 \theta}{96\lambda^4} \right\} \times \int_{-\infty}^{\infty} dx \, \exp \left\{ - \frac{x_0 a \sin^2 \theta}{8\lambda^4} \left( x - \frac{x_0}{2} \right)^2 \right\} \int_{-\infty}^{\infty} dz \, \exp \left\{ - \frac{2z^2}{x_0 a} \right\}$$
\[ \times \int_{-\infty}^{\infty} dy \exp \left\{ - \frac{2y^2}{x_0a} - \frac{ixy}{2\lambda^2} \sin \theta \right\} \propto \exp \left\{ - \frac{x_0}{a} - \frac{x_0^3a \sin^2 \theta}{24\lambda^4} \right\} . \]  

(A.1.3)

Returning to the original system in which the z axis is directed along \( \mathbf{H} \), we have, finally:

\[ I_{ij} \propto \exp \left\{ - \frac{r_{ij}}{a} - \frac{r_{ij} a (x_{ij}^2 + y_{ij}^2)}{24\lambda^4} \right\} . \]  

(A.1.4)

### A.2 Activation Energy \( \varepsilon_3 \) in the Case of Anisotropic Wave Functions

#### A.2.1 Shift of the Percolation Threshold for a Small Variation of the Surface \( Q_\xi \)

Let us consider first an auxiliary problem which was already encountered in Sect. 7.2. Suppose that we know the value of the percolation threshold \( \xi^0_\xi \) for a problem (Sect. 5.3) with the bonding criterion \( \xi^0_{ij} (r_{ij}) \leq \xi \). Let us introduce a small perturbation, \( \xi_{ij} (r_{ij}) + \Delta \xi (r_{ij}) \leq \xi \), to the bonding criterion. It is required to find the shift in the percolation threshold, \( \Delta \xi_c = \xi_c - \xi^0_c \), brought about by the perturbation. This problem has no exact solution for the case of an arbitrary function \( \xi^0_{ij} \). The approximate method, described below, is based on the fact that the critical number \( B_c \) of bonds per site depends only weakly on the shape of the surface \( Q_\xi \) defined by the equation \( \xi_{ij} (r_{ij}) = \xi \) (cf. Sect. 5.3). This implies an approximate invariance of the volume \( V_{\xi_c} \) bounded by \( Q_\xi \). Let us write the equations defining surfaces \( Q_{\xi} \) in the form

\[ r = r_0 (\theta, \phi, \xi) \]  

(A.2.1)

and

\[ r = r_0 (\theta, \phi, \xi) + \Delta r (\theta, \phi, \xi) \]  

(A.2.2)

for the unperturbed and perturbed problems, respectively. Here \( r, \theta, \phi \) represent spherical coordinates of points on the surface. For a fixed \( \xi \) the volume \( V_{\xi_c}^0 \) bounded by the surface (A.2.1) differs from the volume \( V_{\xi_c} \) bounded by (A.2.2). The quantity \( \Delta \xi_c \) is found from the following condition:

\[ V_{\xi_c}^0 = V_{\xi_c} . \]  

(A.2.3)

From (A.2.1) and (A.2.2) we have

\[ V_{\xi_c} - V_{\xi_c}^0 = \int r_0^2 (\theta, \phi, \xi^0_\xi) \Delta r (\theta, \phi, \xi_c) d \Omega . \]  

(A.2.4)

Taking \( V_{\xi_c} \) in the form \( V_{\xi_c}^0 + \frac{dV_{\xi_c}}{d\xi} |_{\xi=\xi^0_c} \Delta \xi_c \), using (A.2.4), and substituting into (A.2.3), we obtain
\[
\Delta \xi_c = \left( \frac{dV_\xi}{d\xi} \bigg|_{\xi = \xi_c^0} \right)^{-1} \int r_0^2 (\theta, \phi, \xi_c^0) \Delta r (\theta, \phi, \xi_c^0) \, d\Omega \tag{A.2.5}
\]

\[
= \left( \frac{dV_\xi}{d\xi} \bigg|_{\xi = \xi_c^0} \right)^{-1} \int r_0^2 (\theta, \phi, \xi_c^0) \frac{dr}{d\xi} \bigg|_{\xi_c^0} \Delta \xi \left[ r_0 (\theta, \phi, \xi_c^0), \theta, \phi \right] \, d\Omega .
\]

Using the ideas of the perturbation theory (Sect. 8.3), one can show that to first order in \(\Delta \xi\), (A.2.5) represents an exact result for an arbitrary perturbation in the special case when the unperturbed surface \(Q_{\xi}\) is a sphere or an ellipsoid. We stress again that in the general case the derivation of (A.2.5) was based on an approximate invariance of \(B_{\xi}\).

### A.2.2 Activation Energy \(\epsilon_3\)

In general the distribution function for \(\epsilon_{ij}\) depends on \(r_{ij}\). We denote this function \(f_{r_j}(\epsilon)\). As shown in Sect. 8.3, the activation energy \(\epsilon_3\) is governed by the \(\epsilon_{ij}\) of those resistances for which \(\xi_{ij}^0 = \xi_c^0\). In the isotropic case these resistances connect the sites with \(r_{ij} = r_c\) and hence \(\Delta \xi\) is a functional of \(f_{r_j}(\epsilon)\). In the anisotropic case the surface equation \(\xi_{ij}^0 = \xi_c^0\) can be written in the form \(r = r_0 (\theta, \phi, \xi_c^0)\). Consequently, the activation energy \(\epsilon_3\) is affected by the values of \(\epsilon_{ij}\) corresponding to resistances with different \(r_{ij}\). The most general expression for the correction \(\Delta \xi_c\) in the linear approximation is of the form

\[
\Delta \xi_c = \frac{1}{kT} \int d\Omega \, d\epsilon \, dr \, f_{r_j}(\epsilon) \, F(\theta, \phi) \, \delta \left[ r - r_0 (\theta, \phi, \xi_c^0) \right] \tag{A.2.6}
\]

\[
= \frac{1}{kT} \int d\Omega \, F(\theta, \phi) \, \epsilon \left[ r (\theta, \phi, \xi_c^0) \right] \, ,
\]

where \(F(\theta, \phi)\) is the unknown function and \(\epsilon(r)\) is the average value of \(\epsilon_{ij}\) for \(r_{ij} = r\), viz.

\[
\epsilon(r) = \int \epsilon f_{r_j}(\epsilon) \, d\epsilon . \tag{A.2.7}
\]

According to (A.2.6), the shift \(\Delta \xi_c\) is the same for all functions \(f_{r_j}(\epsilon)\) giving the same value of \(\epsilon(r)\). In particular, in evaluating \(\Delta \xi_c\) one can replace the original function \(f_{r_j}(\epsilon)\) by \(\delta[\epsilon - \epsilon(r)]\). This brings about the problem considered in Sect. A.2.1, namely that of a shift in the percolation threshold due to a small perturbation of the surface \(Q_{\xi}\). The solution of this problem determines \(\epsilon_3\).

Using the approximate recipe (A.2.5), we find \(F(\theta, \phi)\). Substituting \(\Delta \xi = \epsilon(r)/kT\) into (A.2.5) and comparing the result with (A.2.6), we have

\[
F(\theta, \phi) = r_0^2 (\theta, \phi, \xi_c^0) \left( \frac{dr_0}{d\xi} / \frac{dV_\xi}{d\xi} \right) \bigg|_{\xi = \xi_c^0} . \tag{A.2.8}
\]
Expression (A.2.8) is exact in the special case of a spherical surface \( Q_\xi \), when 
\[ F(\theta, \phi) = 1/4\pi. \]

### A.3 Estimate of the Coefficients in (9.2.2), (9.2.8) and (9.2.12)

Let us consider the problem of determining \( \xi_c \) for variable-range hopping conduction in the four-dimensional space of dimensionless variables \( s, \Delta \) [cf. (9.1.11)]. We construct a four-dimensional hypersurface \( Q_i \) around each site. In order to satisfy the bonding criterion, site \( j \) should also be located within \( Q_i \). We calculate the volume \( V_i \) bounded by \( Q_i \) and then determine \( \xi_c \) from the following condition:

\[
\frac{n(\xi_c)}{2} \bar{V}_i = B_c^{(4)}. \tag{A.3.1}
\]

Here \( \bar{V}_i \) is the average value of \( V_i \), \( n(\xi) \) is the concentration of sites in the three-dimensional \( s \) space [given by (9.1.15)], \( n(\xi)/2 \) is the number of sites per unit volume of the four-dimensional space bounded by the hyperplanes \( \Delta = \pm 1 \), and \( B_c^{(4)} \) is the critical number of bonds per site. If \( H = 0 \), then the equation for \( Q_i \) is of the form \( s_{ij} + \Delta_{ij} = 1 \) and we have (assuming a constant density of states near the Fermi level):

\[
V_i = \frac{4\pi}{3} \int_{\Delta_u \leq 1} (1 - \Delta_{ij})^3 \, d\Delta_j,
\]

\[
\bar{V}_i = \bar{V}_i^0 = \frac{4\pi}{3} \int \int \int (1 - \Delta_{ij})^3 \, d\Delta_j \frac{d\Delta_i}{2} = \frac{\pi}{5}. \tag{A.3.2}
\]

Let \( \xi_c^0 \) be the value of \( \xi_c \) at \( H = 0 \). Then, according to Sect. 9.1, we have \( n(\xi_c^0) = n_c = 5.3 \) and \( B_c^{(4)} = \frac{1}{2} n_c \bar{V}_i^0 = 1.66 \). In a weak magnetic field the equation for the surface \( Q_i \) is

\[
s_{ij} + \Delta_{ij} + \beta s_{ij}^3 \sin^2 \theta = 1, \quad \text{where} \quad \beta = \frac{a^4}{96\lambda^4} \left( \frac{T_0}{T} \right)^{\frac{1}{2}} \ll 1. \tag{A.3.3}
\]

The last term in the left-hand side stems from the correction (9.2.1) in which \( r_{ij} \) is replaced by \( s_{ij} a \xi_c^0/2 \) in accordance with (9.1.11) and (9.1.10). Evaluating the volume \( \bar{V}_i \) in the linear in \( \beta \) approximation, we find

\[
\bar{V}_i = \frac{1}{3} \int \int \frac{d\Delta_i}{2} \, d\Delta_j \int d\Omega \left[ 1 - \Delta_{ij} - \beta (1 - \Delta_{ij})^3 \sin^2 \theta \right]^3
\]

\[
= \bar{V}_i^0 - \frac{4\pi}{21} \beta \equiv \bar{V}_i^0 - \Delta \bar{V}_i. \tag{A.3.4}
\]

Let us assume that \( B_c^{(4)} \) is practically unchanged by the magnetic field. Then we write \( \xi_c \) in the form \( \xi_c = \xi_c^0 + \Delta \xi_c \), where \( \Delta \xi_c \ll \xi_c^0 \), and, using (9.1.15), linearize the product \( n(\xi_c) V_i/2 \) with respect to \( \Delta \xi_c \) and \( \Delta \bar{V}_i \) (cf. the
analogous procedure in Sect. 7.2. and Appendix A.2.1). Then (A.3.1) yields:

$$\Delta \xi_c = \frac{\xi_c^0}{4} \frac{\Delta \bar{V}_i}{\bar{V}_i^0} = \frac{5}{21} \beta \xi_c^0 = \frac{5}{2016} \frac{a^4}{\lambda^4} \left( \frac{T_0}{T} \right)^{3/4}. \quad (A.3.5)$$

Thus, we have $$t_1 = 5/2016$$.

In the case of a strong magnetic field it is more convenient to use the four-dimensional variable space (9.2.4). According to (9.2.6), the surface equation for $$Q_i$$ is of the form $$\tau_i^2 + \eta_i^2 + |\xi_i| + \Delta_i = 1$$, and it is easy to show that $$\bar{V}_i = \pi/4$$. Assuming that the critical number of bonds per site remains the same as for $$H = 0$$, we then find

$$\frac{n_{c1}}{2} \bar{V}_i = B_c^{(4)}, \quad n_{c1} = \frac{8 B_c^{(4)}}{\pi} \approx 4.2. \quad (A.3.6)$$

In contrast with the above estimates, in the Coulomb gap region one cannot ignore the term $$-e^2/\kappa r_{ij}$$ in the expression (10.1.17) for $$\epsilon_{ij}$$, which takes into account the elimination of self-interaction. When working with (10.1.17) it is convenient to use the invariant $$\Theta$$, proposed in Sect. 14.5, which allows an accurate determination of the coefficient $$\beta_1$$ in (10.1.19). To describe the effect of a weak magnetic field, one should add a correction $$r^3a / 18\lambda^4$$ [resulting from the expression (9.2.1) for $$\Delta \xi_{ij}$$ averaged over all directions of $$\bar{r}_{ij}$$] to the term $$2r/a$$ in the argument of the $$\theta$$ function in (14.5.1). Next, one should find the change in the invariant $$\Theta$$ brought about by this correction at a fixed $$\xi_c$$ and also the required correction $$\Delta \xi_{ij}$$ to $$\xi_c$$ which would preserve the value of $$\Theta$$ corresponding to $$H = 0$$. Such an estimate leads to an expression of the form (9.2.12) with $$t^{(2)} \approx 0.0015 [9.25]$. 
Activation energy $\varepsilon_1$ 76, 186, 345
-- $\varepsilon_2$ 79
-- $\varepsilon_3$ 76, 78, 180, 333
Adiabatic processes 250, 338, 343
Amorphous semiconductors 35, 216, 223
Anderson model 30
Anisotropy of hopping conductivity 150, 169, 175, 213
Bond problem 94
Bonding criterion 111, 180

Chemical shift 9
Composition fluctuations 199
Correlation in impurity distribution 55, 143, 303, 305
-- length (radius) in percolation theory 100, 113, 152, 219, 222
Coulomb gap 209, 216, 228
Critical exponents of percolation 95, 99, 100, 103, 105, 113, 119, 227
Critical number of bands 112, 114, 117, 164

$D^-$-bond 80
Dilute ferromagnet 98
Dipole-dipole interaction 234
Droplets in compensated semiconductors 298, 304, 306, 311

Effective medium method 92
Electron-hole density of states 234
Electronic polaron 235, 243, 341

Fermi level in the impurity band 54, 60, 66, 322
Fermi-Thomas approach 257, 259, 299
Film infinite cluster 217
Freezing-out of electrons 75, 80, 170

Gaussian fluctuations 62, 71, 260, 272, 279, 299
Gaussian potential 64, 201, 260, 273, 303
Green function of electron 265

-- of electron in magnetic field 162
Granular metal 244
Hall effect in highly compensated semiconductors 309
-- in hopping conduction 81
High frequency conduction 309, 351
Hopping conduction in strong electric fields 92, 226
Hubbard current correlations 246
-- model 29

Infinite cluster 97
-- backbone 126
-- density 97
Inversion layer 36, 352

Localization radius 6, 43, 243
Long-range potential 60, 62, 67, 182, 189, 196, 297, 327
Luttinger hamiltonian 12
Luttinger parameters 13

Magnetic length 157
Mean cluster size 105
Metal-non-metal transition in compensated semiconductors 296–298, 303, 307
Minimum metallic conductivity 37
Mobility edge 35
Mott-Anderson transition in semiconductors 36–38, 45, 52, 253
Mott's law 202

Nodes and links (SSDG) model 126, 136, 197
Non-linear screening 67, 69, 297, 308
Non-resonant scattering of tunneling electron 215

Percolation level 107, 188, 306, 310, 346
-- method 137, 144, 149, 204, 222
-- radius 112, 138
-- threshold 33, 96, 104, 195
-- distribution function 101
-- mean square fluctuation 102, 115
Pseudo ground state 316, 320
Saturation of band conduction 74
- of hopping conduction 185
Simultaneous transitions 250, 338
Site problem 96
Size effect 216, 335, 342
Solid solution 199
Specific heat 234, 350