High-temperature expansion of the thermodynamic functions of the electrons in the impurity band

A. A. Uzakov and A. L. Éfros

A. F. Ioffe Physicotechnical Institute, Academy of Sciences of the USSR, Leningrad

(Submitted 21 May 1981)
Zh. Eksp. Teor. Fiz. **81**, 1940–1946 (November 1981)

Expressions are derived for the energy and heat capacity of the electrons localized at impurity centers in the case in which the temperature is significantly higher than the impurity-band width. The case of a compensated semiconductor, in which the impurity-band width is determined by the random Coulomb fields of the charged donors and acceptors, is considered. The impurity concentration is assumed to be so low that the quantum overlap of the electronic states localized on neighboring impurities can be neglected. The analogy and difference between the energy of the impurity-band electrons and the correlation energy of a high-temperature plasma are emphasized.

PACS numbers: 71.55.Ht

1. The low-temperature properties of disordered systems with localized electronic states, such as spin and ordinary glasses, have recently been the subject of intensive studies in which the thermodynamic properties and, above all, the heat capacity are given much attention.[1] A unique example of systems with localized states is the impurity band of a slightly doped compensated semiconductor. Its distinguishing feature is the Coulomb interaction, which gives rise to distinctive low-temperature properties.[2]

The electronic heat capacity of semiconductors has been measured in the vicinity of the Mott transition[3] but, unfortunately, only in the case of weak compensation. At present, computer modeling of the thermodynamic properties of compensated semiconductors is being carried out by some members of our group. We hope also that the low-temperature heat capacity of compensated semiconductors will soon be experimentally studied.

In Ref. 2 the case in which the temperature is low compared to the impurity-band width is considered. (In a slightly doped semiconductor this width can be as high as 1–10 meV.)

It makes sense for the purpose of comparing the theory with the experimental data and the results of the computer modeling to consider the opposite limiting case, which corresponds to high temperatures. It is this problem that the present paper is devoted to. As shown below, the first terms of the expansions of the energy and the specific heat capacity in powers of T^{-1} can be found exactly with the aid of the method of functional integrals, and have a somewhat unusual form.

2. In the region of low concentrations the impurity-band width is determined by the random fields produced by the charged donors and acceptors. For definiteness, we shall consider a semiconductor of the n-type, i.e., we shall assume that the donor concentration N_D is higher than the acceptor concentration N_A. Then the system consists of negatively charged acceptors, positively charged donors, and neutral donors. The number of charged donors is equal to the number of acceptors, so that the system is neutral. The ratio N_A/N_D is called the degree K of compensation.

The quantity K varies in semiconductors within wide limits—from the case of weak compensation $K \ll 1$ to the case of strong compensation $1-K \ll 1$. We shall call the case in which neither the quantity K nor the quantity $1-K$ can be considered to be a small parameter the case of intermediate compensation. The quantities K and $1-K$ can in this case be replaced by unity in order of magnitude estimates.

In this case of intermediate compensation the impurity-band width is of the order of $e^2N_D^{1/3}/\varkappa$, where e is the electron charge and $\varkappa$ is the dielectric constant of the lattice. As in Ref. 2, here we consider the "classical" impurity band, i.e., impurity concentrations that are so low that the quantum overlap of the wave functions of the electrons occupying neighboring donors can be neglected. For small impurities such a situation is realized when the condition $N_Da^3 \ll 1$, where a is the Bohr radius of an electron on an impurity, is satisfied. The energy of the classical impurity band has the form

$$H=\frac{e^2}{\varkappa}\left[\frac{1}{2}\sum_{i\neq j}\frac{(1-n_i)(1-n_j)}{r_{ij}}-\sum_{i,\nu}\frac{1-n_i}{r_{i\nu}}+\frac{1}{2}\sum_{\nu\neq\gamma}r_{\nu\gamma}{}^{-1}\right], \qquad (1)$$

where $r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|$, $\mathbf{r}_i$ and $\mathbf{r}_j$ are the coordinates of the donors; $\mathbf{r}_\nu$ and $\mathbf{r}_\gamma$ those of the acceptors. The occupation number $n_i=0$ if the donor is ionized and $n_i=1$ if the donor is neutral.

The expression (1) describes the Coulomb interaction energy for all the charges of the system. To compute the thermodynamic energy, we must average (1) first over the occupation numbers with the aid of the Gibbs distribution and then over all possible donor and acceptor configurations.

In its physical content, the formulated problem is closest to the computation of the correlation energy of a slightly nonideal plasma. The electrons of the impurity band can migrate from donor to donor, effecting screening. Therefore, the sought thermodynamic energy should have the form

$$E=-\alpha(K)VN_D\frac{e^2}{\varkappa}q_s, \qquad (2)$$

where V is the columе of the system, $\alpha(K)$ is some dimensionless function of the degree K of compensation,

0038-5646/81/111028-04$02.40

$$q_s = \left(\frac{4\pi\beta(K) N_D e^2}{\varkappa T}\right)^{1/2} \tag{3}$$

is the reciprocal screening length, which contains another function $\beta(K)$ of the degree of compensation. Accordingly, the specific heat capacity is equal to

$$C = \frac{dE}{dT} = -\frac{1}{2}\frac{E}{T} \propto \frac{1}{T^{3/2}}.$$

The aim of the theory put forward below is to determine the functions $\alpha(K)$ and $\beta(K)$.

The formulated problem differs from the plasma problem for the following reasons:

a) The electrons can be on the donors only, and not anywhere. In itself the discreteness of the donor coordinates is not important, since the potential changes significantly over distances of the order of q_s^{-1}, and there are many charged donors in a sphere of radius q_s^{-1} at high temperatures. But the point is that only one electron can be on each donor, and this produces additional correlation between the charges. This can be stated differently. If the degree K of compensation is close to unity, then the electron (neutral-donor) concentration is low [it is equal to $(1-K)N_D$], and the positions that can be occupied by the electrons (i.e., the empty donors) are very many. Then we can neglect the effect of the additional correlation, and assume that the mobile carriers effecting the screening are electrons. Indeed, as will be shown below, as $K \to 1$, the screening length is determined by the electrons, and $\beta(K) = 1-K$. In the $K \ll 1$ case almost all the donors are occupied, and we can assert that the screening is effected by the holes, which in this case can move without impeding each other. Thus, for $K \ll 1$, we can assume that the mobile carriers are the holes, whose concentration is equal to KN_D. Indeed, for $K \ll 1$ we have $\beta(K) = K$. The general result, as will be shown below, has the form

$$\beta(K) = K(1-K), \tag{4}$$

and it is quite difficult to interpret it in terms of screening by mobile carriers.

b) The donor and acceptor are assumed to have Poisson distributions in space. We should, in averaging over the occupation numbers, assume the fluctuations in the donor and acceptor concentrations to be "frozen." This circumstance decreases the correlation energy, since the potential of the fluctuations having dimensions of the order of the screening length prevents the screening charge carriers from "accommodating" themselves to each other in the best way. This is the second reason why the problem under consideration does not reduce to the plasma problem at arbitrary values of K. In the limiting case $K \ll 1$, in which the carriers can be considered to be holes and the frozen charged-donor fluctuations can be neglected, the problem becomes equivalent to the problem of a plasma with stationary ions distributed according to the Poisson law. The role of these ions is played by the charged acceptors. The frozen acceptor-concentration fluctuations with dimensions of the order of the screening length give rise to a situation in which, for $K \ll 1$, the correlation energy is somewhat lower than the correlation energy of a plasma with stationary ions. As far as we know, the correlation energy of a plasma with stationary ions has not been computed before.

c) In the case of a plasma the correlation energy is a correction term in the expression for the total energy, and this term should be small compared to the dominant term in the entire region of applicability of the theory. In the case of localized electrons the kinetic energy is equal to zero, and the formula (2) gives the first nonvanishing approximation. As $T \to \infty$, the energy of the electrons in the impurity band and their heat capacity tend to zero.

3. Let us discuss the scheme of the computation of the correlation energy. As has already been indicated, the donor- and acceptor-concentration fluctuations with dimensions comparable to q_s^{-1}, i.e., significantly greater than the mean charged-impurity spacing, turn out to be important in the computation. From this we can draw the following conclusions:

a) The fluctuations can be considered to be Gaussian (the mean number of impurities in the fluctuation region is high compared to unity, while the excess number of impurities is small compared to the mean number). The conditions for the fluctuations to be Gaussian have the form

$$N_D q_s^{-3} \gg 1; \quad N_A q_s^{-3} \gg 1; \quad (N_D - N_A) q_s^{-3} \gg 1. \tag{5}$$

In the case of intermediate compensation the conditions (5) determine the region of applicability of the high-temperature expansion ($T \gg e^2 N_D^{1/3}/\varkappa$).

b) The concentrations, $N_D(\mathbf{r})$ and $N_A(\mathbf{r})$, of all the impurities can be considered to be smooth functions of the coordinates.

Let the functions $\eta(\mathbf{r})$ and $\psi(\mathbf{r})$ represent the deviations from the mean donor- and acceptor-concentration values:

$$N_D(\mathbf{r}) = N_D + \eta(\mathbf{r}), \quad N_A(\mathbf{r}) = N_A + \psi(\mathbf{r}). \tag{6}$$

Let us represent the charged-donor concentration $N_c(\mathbf{r}, t)$ in the form

$$N_c(\mathbf{r}, t) = N_A + \xi(\mathbf{r}, t). \tag{7}$$

(The mean number of charged donors is equal to the number of acceptors.) Let us note that the donor- and acceptor-concentration fluctuations are "frozen." They are frozen in during the sample preparation. The charged-donor concentration, on the other hand, depends on the time, since the electrons, executing thermal motion, migrate from donor to donor. We have by definition

$$\langle \eta(\mathbf{r}) \rangle = \langle \psi(\mathbf{r}) \rangle = \langle \xi(\mathbf{r}, t) \rangle = 0, \tag{8}$$

where the symbol $\langle \ldots \rangle$ denotes averaging over all possible impurity configurations. Such averaging is equivalent to averaging over all points in space, so that

$$\int \eta(\mathbf{r})\, d\mathbf{r} = \int \psi(\mathbf{r})\, d\mathbf{r} = \int \xi(\mathbf{r}, t)\, d\mathbf{r} = 0. \tag{9}$$

The random functions $\eta(\mathbf{r})$ and $\psi(\mathbf{r})$ are given by the correlation relations

$$\langle\eta(\mathbf{r})\eta(\mathbf{r}')\rangle=N_D\delta(\mathbf{r}-\mathbf{r}'),\quad \langle\psi(\mathbf{r})\psi(\mathbf{r}')\rangle=N_A\delta(\mathbf{r}-\mathbf{r}'),$$
$$\langle\eta(\mathbf{r})\psi(\mathbf{r}')\rangle=0. \tag{10}$$

The higher-order correlators break up into pair correlators.

Because of the neutrality of the system, the interaction energy of homogeneously distributed charges is equal to zero. The terms linear in the fluctuations are equal to zero owing to the property (9). Therefore,

$$H=\frac{e^2}{\varkappa}\left[\frac{1}{2}\iint\frac{\xi(\mathbf{r},t)\xi(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'-\iint\frac{\xi(\mathbf{r},t)\psi(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'+\frac{1}{2}\iint\frac{\psi(\mathbf{r})\psi(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'\right]. \tag{11}$$

The plan of the computations is as follows. We determine for given donor and acceptor concentrations the time-averaged values $\xi(\mathbf{r}, t)$ and $\xi(\mathbf{r}, t)\xi(\mathbf{r}, t)$ with the aid of the formula

$$\overline{A\{\xi\}}=\int D\xi A\{\xi\}\exp(-R_{min}\{\xi\}/T)\left[\int D\xi\exp(-R_{min}\{\xi\}/T)\right]^{-1}, \tag{12}$$

where $R_{min}\{\xi\}$ is the minimum work that must be done in order to produce the fluctuation $\xi(\mathbf{r})$ at fixed donor- and acceptor-concentration values. Then we find the configuration-averaged quantities:

$$L_1(|\mathbf{r}-\mathbf{r}'|)=\langle\overline{\xi(\mathbf{r},t)\xi(\mathbf{r}',t)}\rangle-N_A\delta(\mathbf{r}-\mathbf{r}'), \tag{13}$$

$$L_2(|\mathbf{r}-\mathbf{r}'|)=\langle\overline{\xi(\mathbf{r},t)}\psi(\mathbf{r}')\rangle. \tag{14}$$

The correlation energy is equal to

$$E=\frac{e^2}{2\varkappa}V\int\frac{L_1(r)}{r}d\mathbf{r}-\frac{e^2}{\varkappa}V\int\frac{L_2(r)}{r}d\mathbf{r}. \tag{15}$$

The subtraction of a term with the δ function should be done in the computation of the interaction between identical particles in order to eliminate the self-action (see Ref. 4, §116). According to (10), the contribution to the correlation energy from the last term in (11), a term which describes the interaction between the randomly distributed acceptors, is equal to zero.

4. The minimum work is given by the expression

$$R_{min}=H-T\Delta S, \tag{16}$$

where the last term in the expression (11) for H can be neglected, since it does not depend on ξ. The quantity ΔS is the entropy change due to the fluctuation $\xi(\mathbf{r})$. To compute ΔS, let us divide the whole system into regions each containing many impurities, but having dimensions that are small compared to q_s^{-1}, so that the functions $N_D(\mathbf{r})$ and $N_c(\mathbf{r})$ can be considered to be constants inside it. Let Ω be the volume of each region. Then the number $\mathcal{N}_D$ of donors inside a region is equal to $\mathcal{N}_D=N_D(\mathbf{r})\Omega$ while the number of charged donors is equal to $\mathcal{N}_c=N_c(\mathbf{r})\Omega$. Let us denote by Γ the number of ways $\mathcal{N}_D-\mathcal{N}_c$ electrons can be distributed over $\mathcal{N}_D$ donors. Then the entropy S_Ω of a region Ω is equal to

$$S_\Omega=\ln\Gamma,\quad \Gamma=\frac{\mathcal{N}_D!}{\mathcal{N}_c!(\mathcal{N}_D-\mathcal{N}_c)!}. \tag{17}$$

Using the Stirling formula, as well as (6) and (7), we find that $S_\Omega=\Omega\sigma(\mathbf{r})$, where the entropy density

$$\sigma(\mathbf{r})=(N_D+\eta)\ln(N_D+\eta)-(N_A+\xi)\ln(N_A+\xi)-(N_D-N_A+\eta-\xi)\ln(N_D-N_A+\eta-\xi). \tag{18}$$

Let us determine the total entropy

$$S=\int\sigma(\mathbf{r})d\mathbf{r} \tag{19}$$

and find ΔS by substituting (18) into (19) and expanding the resulting expression in powers of ξ up to second order. Then

$$\Delta S=-\frac{1}{2N_DK(1-K)}\int\xi^2d\mathbf{r}+\frac{1}{N_D(1-K)}\int\xi\eta d\mathbf{r}. \tag{20}$$

According to (9), ΔS is equal to zero in the first-order approximation.

The formula (20) differs from the normal Gaussian formula as a result of the presence of the second term, which indicates that the most probable distribution of the charged donors for a given configuration of the impurities is not the uniform (i.e., $\xi\equiv0$) distribution, but the distribution that preserves the mean fraction of charged donors $[\xi(\mathbf{r})=K\eta(\mathbf{r})]$.

5. To compute the functional integral (12), we must go over to the Fourier transforms with the aid of the formula

$$\xi(\mathbf{r})=\sum_{\mathbf{q}}\xi_{\mathbf{q}}e^{i\mathbf{q}\mathbf{r}},\quad \xi_{\mathbf{q}}=\xi_{-\mathbf{q}}^* \tag{21}$$

and separate the real and imaginary parts:

$$\xi_{\mathbf{q}}=\xi_{\mathbf{q}}'+i\xi_{\mathbf{q}}'',\quad \xi_{-\mathbf{q}}=\xi_{\mathbf{q}}'-i\xi_{\mathbf{q}}'',$$
$$\xi_{\mathbf{q}}'=\xi_{-\mathbf{q}}',\quad \xi_{\mathbf{q}}''=\xi_{-\mathbf{q}}''. \tag{22}$$

Similarly, we introduce the quantities ψ_q, ψ_q', ψ_q'', η_q, η_q', η_q''.

Using (11), (16), and (20), we obtain

$$\frac{R_{min}}{T}=V{\sum_{\mathbf{q}}}'(\xi_{\mathbf{q}}'^2+\xi_{\mathbf{q}}''^2)a_q-2V{\sum_{\mathbf{q}}}'(\xi_{\mathbf{q}}'X_{\mathbf{q}}'+\xi_{\mathbf{q}}''X_{\mathbf{q}}''), \tag{23}$$

where

$$a_q=\frac{q^2+q_s^2}{q^2N_DK(1-K)}, \tag{24}$$

the reciprocal screening length q_s is given by the formulas (3) and (4),

$$X_{\mathbf{q}}=\frac{4\pi e^2\psi_{\mathbf{q}}}{\varkappa Tq^2}+\frac{\eta_{\mathbf{q}}}{N_D(1-K)},\quad X_{\mathbf{q}}'=\operatorname{Re}X_{\mathbf{q}},\quad X_{\mathbf{q}}''=\operatorname{Im}X_{\mathbf{q}}. \tag{25}$$

The symbol Σ' indicates that the summation is performed only over the vectors lying in one half-space (e.g., in the $q_z>0$ half-space).

The functional integral (12) reduces, after the substitution of (23) into it, to Gaussian quadratures. As a result we obtain

$$\overline{\xi_{\mathbf{q}}'}=X_{\mathbf{q}}'/a_q,\quad \overline{\xi_{\mathbf{q}}''}=X_{\mathbf{q}}''/a_q, \tag{26}$$

$$\overline{\xi_{\mathbf{q}}'^2}=\frac{1}{2Va_q}+\frac{X_{\mathbf{q}}'^2}{a_q^2},\quad \overline{\xi_{\mathbf{q}}''^2}=\frac{1}{2Va_q}+\frac{X_{\mathbf{q}}''^2}{a_q^2}. \tag{27}$$

The averaging over the configurations is performed with the aid of the formulas (10), which are written in the form

$$\langle\psi_{\mathbf{q}}'^2\rangle=N_A/2V,\quad \langle\eta_{\mathbf{q}}'^2\rangle=N_D/2V,\quad \langle\psi_{\mathbf{q}}'\eta_{\mathbf{q}}'\rangle=0. \tag{28}$$

We find with the aid of (27) and (28) that

$$L_1(q)=\langle\overline{\xi_{\mathbf{q}}'^2}\rangle+\langle\overline{\xi_{\mathbf{q}}''^2}\rangle-N_A/V=-\frac{q^2q_s^2N_A(1+K)}{V(q^2+q_s^2)^2}. \tag{29}$$

Similarly,

$$L_2(q)=\langle\overline{\xi_q'}\psi_q'\rangle+\langle\overline{\xi_q''}\psi_q''\rangle=\frac{N_A}{V}\frac{q_s^2}{q^2+q_s^2}. \tag{30}$$

Using (15), (29), and (30), we find

$$E=-\sum_q{}'\frac{4\pi e^2q_s^2N_DK(1+K)}{\varkappa(q^2+q_s^2)^2}-\sum_q{}'\frac{8\pi e^2q_s^2N_DK}{\varkappa q^2(q^2+q_s^2)}. \tag{31}$$

Replacing the sum by an integral, we obtain the formula (2) with

$$\alpha(K)=\frac{(1+K)K}{4}+K. \tag{32}$$

6. The first term in (32) describes the interaction between the charged donors; the second term, the interaction between the charged donors and acceptors. Let us first consider the $K \ll 1$ case, in which almost all the donors are neutral. A unit volume contains N_A charged acceptors and as many stationary holes (ionized donors), which effect the screening. According to (2), (3), (4), and (32),

$$E=-VN_A\frac{e^2}{\varkappa}q_s\left[\frac{1}{4}+1\right],\quad q_s^{-1}=\left[\frac{\varkappa T}{4\pi N_Ae^2}\right]^{1/2}. \tag{33}$$

This is the solution to the problem for a plasma with stationary ions. The second term in (33) is the sum of the potentials produced by the screening atmosphere on each ion (acceptor). The first term describes the interaction between the mobile carriers (holes). If we replace the stationary ions by a background, then the second term in the square brackets will be absent, and the first will be equal to $\frac{1}{2}$ instead of $\frac{1}{4}$. Thus, the potential produced by the stationary ions reduces (in absolute value) the energy of interaction between the mobile carriers by a factor of two.

If the ions were mobile, then the square brackets in (33) should be replaced by unity, but then the screening length should be reduced by a factor of $\sqrt{2}$. Thus, the correlation energy for a plasma with stationary ions is in absolute value smaller that of a plasma with mobile ions. The term proportional to K^2 in the formula (32) describes the effect of the donor-concentration fluctuations on the energy of interaction between the charged donors.

Let us now consider the case $1-K \ll 1$. In this case the condition of applicability of the high-temperature expansion does not reduce to the condition $e^2N_D^{1/3}/\varkappa \ll T$. It is necessary that the correlation energy per electron be small compared to T, i.e., that

$$\frac{e^2}{\varkappa}\frac{q_s}{1-K}\ll T \quad \text{or} \quad T\gg\gamma\equiv\frac{e^2N_D^{1/3}}{\varkappa}(1-K)^{-1/3}. \tag{34}$$

The low-temperature impurity-band theory was developed by Shklovskiĭ and Éfros.[5] It was shown that the characteristic screening length in this case is of the order of $R_c=N_D^{-1/3}(1-K)^{-2/3}$, while the range of the impurity potential is of the order of γ. The condition (34) can be written in the form $q_s^{-1}\gg R_c$. When T is of the order of γ, the results of the high- and low-temperature theories coincide.

The authors take this opportunity to express their gratitude to S. D. Baranovskiĭ and B. I. Shklovskiĭ for fruitful discussions of the questions touched upon here.

[1] P. W. Anderson, B. I. Halperin, and C. M. Varma, Philos. Mag. **25**, 1 (1972).

[2] S. D. Baranovskiĭ, B. I. Shklovskiĭ, and A. L. Éfros, Zh. Eksp. Teor. Fiz. **78**, 395 (1980) [Sov. Phys. JETP **51**, 199 (1978).

[3] N. Kobayashi, S. Ikehata, S. Kobayashi, and W. Sasaki, Solid State Commun. **24**, 67 (1977); 32, 1147 (1979); W. Sasaki, J. Phys. Soc. Jpn. **49**, 31 (1980).

[4] L. D. Landau and E. M. Lifshitz, Statisticheskaya fizika (Statistical Physics), Vol. 1, Nauka, Moscow, 1976 (Eng. Transl., Pergamon Press, Oxford, 1980).

[5] B. I. Shklovskiĭ and A. L. Éfros, Zh. Eksp. Teor. Fiz. **60**, 867 (1971) [Sov. Phys. JETP **33**, 468 (1971)].

Translated by A. K. Agyei