KINETIC THEORY OF IMPACT IONIZATION IN SEMICONDUCTORS

L. V. KELDYSH

P. N. Lebedev Physics Institute, Academy of Sciences, U.S.S.R.

Submitted to JETP editor March 23, 1959


The effect of impact ionization processes on the distribution function for electrons and holes in a strong electric field is studied. It is shown that the energy dependence of the impact ionization probability near the threshold is essentially different for crystals with small and high dielectric constants; the solution of the kinetic equation is considered in both these cases. Expressions are obtained for the equilibrium number of carriers in a strong field, the impact-ionization coefficient, the critical field, etc. The dependence of the breakdown field on temperature, on specimen thickness, and on the electron-lattice interaction law is found. The connection of the expressions obtained with the known breakdown criteria of Fröhlich and Hippe is established.

Increasing the electric field causes a decrease in the recombination speed, as a result of which the equilibrium number of carriers starts growing as the field increases long before the appearance of impact ionization.

THE electric breakdown of semiconductors apparently takes place as a result of the unlimited growth of carrier concentration with increasing field strength.\(^1\) In a stationary state the number of carriers, \(n\), is determined by the relationship

\[
n \langle w_i(E, T) - w_r(n, E, T) \rangle + n_0(E, T) = 0,
\]

where \(w_i(E, T)\) and \(w_r(n, E, T)\) are the impact ionization and recombination probabilities averaged over the distribution function, \(n_0(E, T)\) is the number of carriers of a given type created in unit volume of the semiconductor in unit time by thermal ionization and direct field extraction of valence electrons into the conduction band, \(E\) is the field strength and \(T\) the temperature.

With increasing field, as will be shown below, \(w_i\) decreases but \(w_r\) grows rapidly and, consequently, \(n\) increases. In the field \(E_C\) for which \(w_i = w_r\) the carrier concentration tends to infinity, which is the breakdown criterion for the case given. Thus, quantitative consideration of the behavior of a semiconductor in the pre-breakdown region, as well as a study of the mechanism of breakdown itself, requires the solution of the kinetic equation taking into account the processes of impact ionization and recombination. This is the aim of the present work.

Following the usual method,\(^2\) it is not difficult to obtain the following system of equations for determining the symmetric \(f_0(\epsilon)\) and antisymmetric \(f_1(\epsilon)\) parts of the distribution function, \(f(\mathbf{P})\)

\[
f(\mathbf{P}) = f_0(\mathbf{P}) + \frac{e\mathbf{E} \cdot \mathbf{P}}{m} f_1(\mathbf{P}) + ...
\]

\[
f_0(\mathbf{P}) = f_0(\mathbf{P}) - \frac{e\mathbf{E} \cdot \mathbf{P}}{\hbar \tau} \int \frac{1}{1 + w_i(\mathbf{P}) \frac{1}{\tau}(\mathbf{P})} \frac{dS(\mathbf{P})}{d\xi} d\mathbf{P},
\]

\[
\left[ \eta(x) + \left( \frac{E}{E_f} \right)^{\frac{1}{2}} \frac{1}{\tau(x)} \frac{1}{1 + w_i(x) \frac{1}{\tau}(x)} \right] \frac{dS(x)}{dx} = \frac{1}{N_1} \int \frac{1}{\xi_1 n(\mathbf{P})} \left[ w_i(x) + w_r(x) \right] f_0(x) + n_0(x, E, T)
\]

\[
\left[ w_i(x) + w_r(x) \right] f_0(x) + n_0(x, E, T)
\]

where \(\mathbf{P} = \epsilon / \epsilon_1\); \(\epsilon = P^2/2m\) is the energy; \(P\) is the momentum; \(\epsilon_1\) is the threshold ionization energy; \(\tau^{-1}\) is the frequency of collision with phonons; \(w_i(x)\) and \(w_r(x)\) are the total probabilities of impact ionization and recombination; \(w_i(x, x')\) is the probability of the creation by ionization of a carrier with energy \(x\) by a carrier with initial energy \(x'\); \(\lambda(x) = I(x)/I(1); I(x) = P \tau (x)/m\) is the mean free path; \(S(\mathbf{P})\) is the carrier current through the surface \(\epsilon(\mathbf{P}) = x \epsilon_1\), caused by the field and phonon interactions; \(\frac{1}{2} N_1\) is the total number of states with energy \(\epsilon < \epsilon_1\);

\[
\delta(x) = 4m \int B(q) \lambda_0 dq 
\]

\[
\sim \frac{\lambda \mathbf{P}_p}{\epsilon N_p + 1}.
\]
Here, $q$, $\omega_0$ and $N_q$ are the momentum, frequency, and number of phonons; $B_q$ is the square of the matrix element of the interaction of an electron with a phonon; $V$ is the normalization volume; $E_i$ is the field in which the mean energy of the carriers becomes of the order $\epsilon_i$; $\delta = \delta (1)$. The small value of $\delta$, the average fraction of the energy lost by an electron in one collision with the lattice, is a condition for the applicability of the approximation considered. For the parameter values of interest to us, $\delta \sim 10^{-2}$. The small quantity $\eta (x)$ we will neglect henceforth. For those valence crystals in which the electrons interact mainly with acoustic phonons $B_{ac}(q) \sim q$ and $\hbar \omega_0 q = \omega_0 q = \text{const.}$, where $\omega_0$ is the speed of sound. When the interaction is with optical phonons $B_{op}(q) = \text{const.}$ and $\hbar \omega_0 q = \text{const.}$.

From (5) and (6) it follows that

$$\delta_{ac}(x) = 4mc^2/kT$$
$$\beta_{ac}(x) = 1,$$
$$\beta_{op}(x) = \hbar \omega_0 q / (N_q + 1/2),$$
$$\beta_{op}(x) = 1,$$
$$\beta_{op}(x) = 4mc^2/kT \frac{3}{2} \frac{1}{2} \left( \frac{\hbar}{\omega_0 q} \right)^3.$$

Before proceeding to the solution of Eqs. (2) - (4), we make some remarks on the choice of the probabilities $w_f(x), w_i(x)$ and $w_i(x, x')$. The effect of recombination on the form of the distribution function is insignificant in view of the inequality $w_f \tau \ll 1$, which is well fulfilled. Therefore, the corresponding term in (4) can be considered as a small contribution and the fact that it is in general nonlinear has no effect. In the majority of cases, however, an important part is played by the so-called “radiationless” recombination associated with carrier capture into local states. Its probability for a sufficiently large number of carriers can be considered as independent of concentration. Such a capture can only take place for very slow carriers ($\epsilon \ll \hbar \omega_m$, where $\omega_m$ is the maximum lattice vibration frequency), which in fact are not included in the conditions considered, since for them $\delta (x)$ is not small. Therefore, it is most natural to include radiationless recombination, not in (4), but in the boundary condition for $S(x)$ at $x = 0$. The last two terms in (4) can be taken into account in a similar way. In fact, the probability of creation of a carrier with energy $\epsilon$ by thermal ionization and by the field decreases exponentially with increase of $\epsilon$. Likewise, the number of carriers with energies essentially exceeding the ionization threshold is exponentially small. Therefore, both $n_0(E, T)$ and $w_i(x, x')$ cause the creation of only very slow carriers $x \ll 1$, and we include them only in the boundary condition

$$\frac{S(0)}{N_i} - w_f(0) = \frac{n_0(E, T)}{N_i} + \sum_{x', x} \int_{x'}^{x_{th}} w_i(x', x') f_0(x') x^{\nu_i} dx' = 0,$$

where the summation takes into account the presence of two carrier types and the indices 1 and 2 refer to the creation probability of carriers of the same or opposite charge.

The quantity $w_i(x)$, as shown in Appendix 1, can increase near the threshold either linearly or quadratically, depending upon whether the value of the dielectric constant of the crystal is small or large, i.e.,

$$w_i(x) = p(x - 1) k_i(x),$$

where $k_i(x) = 1 + \sum_{n=1}^{\infty} k_{in}(x - 1)^n$.

The dimensionless quantity $p$ thus defined is rather large ($p \sim 10^2$).

Equations (3) and (4) take an essentially different form in the regions $x < 1$ and $x > 1$. It is natural, therefore, to solve them in each of these regions separately and then match the solutions obtained at $x = 1$. Below the ionization potential $w_i(x) = 0$, $w_f(x)$ is a small correction, and all the remaining terms in the right half of (4) do not depend on the value of $f_q(x)$ in this region. Equations (3) and (4) are consequently integrated in the general form. However, on the basis of the remarks made above, we will use a simplified form of the solution in which the value $\eta (x)$ is neglected, and $w_f(x), w_i(x, x')$, and $n_0(x, E, T)$ are taken into account only in the boundary condition (8).

$$S(x) = \text{const} = - \frac{N_i}{\tau (1)} \sigma (E) f_0 (1),$$

$$f_0(x) = f_0 (1) \exp \left\{ \left( \frac{E_i}{E} \right)^2 \int_{x}^{x_{th}} \frac{x^{\nu_i} (x')}{\nu_i (x') \frac{\sigma (E)}{x^{\nu_i} (x')}} dx' \right\}$$

$$\times \left\{ 1 + \left( \frac{E_i}{E} \right)^2 \frac{\sigma (E)}{\nu_i (1)} \int_{x}^{x_{th}} \frac{d x'}{\nu_i (1) x^{\nu_i} (x') \frac{\sigma (E)}{x^{\nu_i} (x')}} \right\}.$$

(10)
The constants of integration $f_0(1)$ and $\sigma(E)$ must be determined from the boundary conditions. The value of $\sigma(E) = -S(1)\tau(1)/N_0\delta f_0(1)$, giving the mean probability of impact ionization, is simply determined, as will be shown below, by solving the equation in the region $x > 1$. Knowing this value is sufficient to determine all the characteristics of the semiconductor in the strong electric field. In fact, we show below that in the region $x > 1$ the distribution function falls practically to zero for $x - 1 \sim (\delta/\rho)^{1/2} \sqrt{E/E_1} = \alpha$.

Consequently, the contribution of this region to all the observed quantities (number of carriers, conductivity, mean energy) is of the order $\alpha \ll 1$. In other words, all these quantities can be evaluated using the function (10) by averaging over the region $x < 1$.

The single quantity completely determined by the distribution of carriers over the ionization potential is the mean probability of impact ionization. But it is equal at the same time to $\int_0^1 \int_0^1 \frac{dE_0}{E_0} \frac{dE_1}{E_1} \sigma(E)$, proportional to the ratio $\frac{\alpha(E)}{E/E_1}$.

The total number of carriers $n$ and the mean impact ionization and recombination probabilities are determined from the following relationships

$$n = N_1 \int_0^\infty f_0(x) x n^2 dx = f_0(1) N_1 \exp \left[ \frac{(E_1/E_1)^2}{E_1} \right] \frac{\Phi_0}{\Phi_{00}} \left( \frac{E}{E_1} \right),$$

$$\omega_r(1) = \frac{S(1)}{n} = \frac{\delta}{\tau(1)} \frac{n}{n} \sigma(E) \frac{\exp \left[ -(u E_1/E_1)^2 \right]}{\Phi_{00}(E/E_1) + (E_1/E_1)^2 \sigma(E) \exp \left[ -(u E_1/E_1)^2 \right]} \Phi_{00}^2 \left( \frac{E}{E_1} \right),$$

$$\omega_r(0) = \frac{N_1}{n} \omega_r f_0(0) = \omega_r^0 \frac{1 + (E_1/E_1)^2 \sigma(E) \exp \left[ -(u E_1/E_1)^2 \right] \xi(E/E_1)}{\Phi_{00}(E/E_1) + (E_1/E_1)^2 \sigma(E) \exp \left[ -(u E_1/E_1)^2 \right] \Phi_{00}^2(E/E_1)},$$

$$\omega^a = \frac{1}{\sqrt[3]{6}} \frac{x^b(x)}{x^b(x) b} dx,$$

$$\xi(z) = \frac{1}{6} \int_0^z \frac{x^b(x')}{x^b(x') b} dx'$$

The integral $\xi(z)$ in general diverges at the lower limit. This is associated with the fact that we have considered in (8) the number of recombinations as proportional to $f_0(0)$. In fact, as follows from the reasons given, this number is determined by the mean value of $f_0(E)$ in the region of very small energies ($E \ll \hbar \omega_m$). Consideration of this fact should lead to the exclusion of the integral for $x \sim \hbar \omega_m/\epsilon_1$. The exact value of this limit for calculating $\xi(z)$ has no significance, since the divergence of the integral is logarithmic.

Using the expressions (11) - (13) and the evident relationships

$$-S(1) = N_1 \int_0^\infty \omega_r(1) f_0(x) x n^2 dx,$$

$$\omega_r(1) = 2N_1 \int_0^\infty \omega_r(0) x n^2 dx = 2\omega_r(0),$$

the boundary condition (8) can be rewritten in a form completely analogous to relation (1) (index e refers to electrons and h to holes),

$$[\omega_{re}(E) - \omega_{re}(E)] n_e(E) - \omega_{rh}(E) n_h(E) - n_{re}(E, T) = 0,$$

$$[\omega_{rh}(E) - \omega_{rh}(E)] n_h(E) - \omega_{re}(E) n_e(E) - n_{rh}(E, T) = 0.$$ (17)

If $\omega_r$ can be taken as independent of $n$, then (see reference 4)

$$n_r(E) = \frac{n_{re}(E, T)}{\omega_{re}(E)} = \frac{1 + (n_{re}(E, T)/n_{re}(E, T) - 1) r(E)}{1 - r(E) - r(E)},$$

where

$$r(E) = \frac{\omega_r(E)}{\omega_r[1],}$$ (18)

We study the behavior of this expression on increasing the field. In the region $E \ll E_1$, we have $r(E) \ll 1$ and (18) takes the usual form $n(E) = n_0/\omega_r(E)$. Thus, the growth of $n$ is determined mainly by the function $\Phi_{00}^2(E/E_1)$. We notice that $\sigma(E)$ becomes of the order of unity only when $E$ is comparable with $E_1$. In particular, for the cases mentioned above, involving acoustic and optical phonons in valence crystals, $n$ is proportional to $E^{3/2}$ and $E^3$ respectively. This behavior is occasioned by the fact that when the field becomes
sufficiently large \( E \approx 10^3 \text{ v/cm} \), the mean electron energy starts to increase, the relative number of slower electrons decreases, and there is an associated decrease in the recombination velocity. When the field approximates to \( E_i \approx 10^3 \text{ v/cm} \), \( w_i(E) \) begins to increase rapidly and at a field \( E_{c_1} \), determined by the condition

\[
r_{c_1}(E_c) + r_h(E_c) = 1, \tag{19}\]

breakdown occurs. As a rule, the values of \( E_i \) are different for electrons and holes, therefore \( E_{c_1} \) is smaller is taken here. A direct comparison of \( \varepsilon \) and \( \tau \) as of the order of a microsecond, and \( \tau \approx 10^{-12} \text{ sec} \), we obtain the following estimate of the critical field \( E_{c_1} \approx E_{1}u/5 \) (see reference 5).

If the region in which the field acts is sufficiently small \( (\varepsilon < 1 \text{ cm}) \), then the lifetime of carriers is determined not by recombination but by their departure from this region. The carrier concentration depends in this case on the distance \( t \) from the boundary of the specimen and the number of carriers \( n_s \) flowing through this boundary in unit time \( t \) (see reference 5):

\[
n_s(E, t) = \frac{n_{so}}{\tau_d} \exp [\nu_s(E) t] \times \frac{\nu_s(E) \exp [-\nu_s(E)L] - \nu_s(E) \exp [-\nu_s(E)L] - \nu_s(E) \exp [-\nu_s(E)L]}{\nu_s(E) \exp [-\nu_s(E)L] - \nu_s(E) \exp [-\nu_s(E)L]}.
\]

Here \( \nu_d \) is the drift velocity of the carriers in the field \( E, L \) is the dimension of the region, and \( \kappa(E) \) is the so-called impact-ionization coefficient.

\[
\kappa(E) \equiv \frac{\nu_d(E)}{v_d} = \frac{V - E_i}{i(t)} \varepsilon (E) \times \exp \left[ \frac{-u E_i}{E_i} \right] \times \frac{\phi^{(1)}(E/E_i) + (E/E_i) \phi^{(2)}(E/E_i)}{\phi^{(1)}(E/E_i) + (E/E_i) \phi^{(2)}(E/E_i)}.
\]

The temperature dependence of \( E_{c_2} \) is determined mainly by the quantity \( E_i \approx \sqrt{\varepsilon} / i(1) \). For acoustical phonons \( E_{c_2} \sim \sqrt{T} \), for optical \( E_{c_2} \sim \varepsilon \).

In conclusion, we make a series of remarks on the connection of the parameter we have introduced, \( u \), with the known breakdown criteria of Frohlich and Hiepel. Since \( \lambda (1) = 1 \), the integrand of \( u^2 \) in (14) can be either of the order of unity if \( \lambda(x) \) does not increase with increasing energy, or increases sufficiently slowly, or much greater than unity, if the mean free path increases rapidly with growth of \( x \). In the first case, which apparently, obtains always in valence crystals, breakdown occurs in fields of the order \( E_i \), i.e., when the mean carrier energy becomes of the order \( E_i \). In form this condition agrees with Frohlich's criterion, although the primary idea of this criterion was somewhat different. In the second case, which has been well studied and of which ionic crystals are an example, the integrand of \( u^2 \) in (14) attains a maximum for small energies and, therefore,
E_c \sim uE_1 \gg E_1. \text{ In ionic crystals, for energies } 
\epsilon \approx \hbar \omega_q, \text{ the mean free path } \lambda(x) \text{ is proportional } 
\text{to } x, \text{ and consequently } u \sim (\epsilon_1/\hbar \omega_q)^{1/2}. \text{ The } 
\text{breakdown field is determined by the relationship} 
E_c \sim uE_1 \ln^{-1/2} \left( \frac{\beta}{\hbar \omega_q} \ln^{-1/2} \frac{\hbar \omega_q}{\tilde{\omega}} \right). \quad (27)

In other words, breakdown occurs in fields for which 
E_c \sim uE_1 \sim \hbar \omega_q \exp \left( - \frac{\hbar \omega_q}{\tilde{\omega}} \right). 
\text{which agrees qualitatively with Rippel's criterion.}^9

Starting from (5) and (6), it is not difficult to 
verify that Fröhlich's criterion is applicable to 
crystals in which B(q) \sim q \sim 0 \text{ increases less rapidly than } 
[q(1+2N_q)^{-1}]^{-1}, \text{ and Rippel's criterion applies in the opposite case.}

We proceed now to the solution of our basic 
problem — finding the distribution function in the 
region \( x > 1 \) \text{ where the process of impact ionization} \text{ is important. It will be convenient here to} 
introduce new units of energy and current 
\[ y = \frac{x-1}{u}, \quad s(y) = \frac{\alpha}{\beta} \gamma_j S(x)/N_i, \quad (28) \]
\text{where } \alpha_j = (\beta E^2/\hbar^2)^{1/2}, \quad \gamma_j = (E_i/E)^{u}, \quad \beta_j = \exp \gamma_j. \quad (29)

The system (3) and (4), expressed in these vari­
ables, takes the following form:
\[ \frac{dx(y)}{dy} - \frac{x_k}{\lambda(x)} y f_j(y) = 0, \quad (30a) \]
\[ 1 + \beta_j \lambda(x) y f_j(y) - \frac{x_k}{\lambda(x)} y f_j(y) = 0. \quad (30b) \]

An essential fact, on which all further discussion is 
based, is the smallness of \( \alpha \), evident directly 
from (29). In the most interesting region of the field \( \alpha \ll 0.1. \) The functions \( I_q(x) \) and \( S(x) \) are 
essentially different from zero only in the narrow 
region \( x - 1 \ll \alpha \), outside which they fall off 
exponentially. In this region the coefficients 
of (30), in the arguments of which the substitution of \( y \) for \( x \) has not been made, are very slowly 
changing functions of \( y \) and with great accuracy 
can be taken as the first terms of corresponding 
series of degree \( x - 1 = \alpha y \). In the zero order 
approximation in \( \alpha \), which we mainly use, all 
these are unity. Further, in this approximation, 
(30) can be solved exactly only in particular — 
although perhaps the most interesting — cases. 
We therefore now describe a general method al­
lowing us to determine with adequate accuracy the 
quantity \( \sigma(E) \) of direct interest to us.

Equations (30) for any values of the parameters have 
two linearly independent solutions; one ex­
ponentially decreasing at infinity, the other in­
creasing. Apparently, only the first of these is 
physically permissible. It is determined to within 
an arbitrary multiplier, but the ratio of \( s(y) \) to 
\( f_q(y) \) at any point, including at \( y = 0 \), is 
strictly defined. Therefore, the requirement of a solution 
decreasing at infinity is equivalent to the problem of 
determining the value of \( \sigma(E) \). We eliminate 
from (30) the function \( f_q(y) \). Then for the current 
\( s(y) \) we obtain a second-order equation
\[ \frac{d}{dy} \left[ \lambda(x) y f_j(y) \right] - \left( 1 + \beta_j \lambda(x) y f_j(y) \right) s(y) = 0, \quad (31) \]
\[ F(y) = \int_0^{\infty} \frac{x_k^2(y) \gamma_j}{\lambda(x)} \left[ 1 + \beta_j \lambda(x) y f_j(y) \right] \right| dy'. \quad (32) \]

The quantity \( \sigma(E) \) which is sought is derived 
from its solution in the following manner:
\[ \sigma(E) = \int \frac{1}{\int_0^{\infty} \frac{x_k^2(y) \gamma_j}{\lambda(x)} \left[ 1 + \beta_j \lambda(x) y f_j(y) \right] \right| dy}. \quad (33) \]

and is a function of the parameters \( \beta_j \) and \( \gamma_j \). 
Instead, to find this function, we invert the prob­
lem, take fixed values of \( \sigma \), and seek the inverse 
function \( \beta_j = \beta_j(\sigma, \gamma_j) \). In these circumstances 
(31) appears as a typical eigenvalue problem; the 
given boundary conditions at \( y = 0 \) \text{ [Eq. (32)]} and 
at infinity \( s(y) \rightarrow 0 \) \text{ require the finding of the} 
value of the parameter \( \beta_j \), for which the equation 
has a nontrivial solution. Solving then the expres­
sion obtained for \( \sigma \), we find the relationship of 
importance to us \( \sigma = \sigma(\beta_j, \gamma_j) = \sigma(E) \). The eigen­
value of interest to us must, apparently, be the 
smallest, since, from its physical meaning, the 
function \( f_q(y) \) cannot tend to zero anywhere ex­
etcept at infinity. But, from (30a) and the conditions 
x_k(x)/\lambda(x) > 0 \text{ and } s(\infty) = 0 \text{ it follows that} 
\( s(y) \) also has no zeros in the interval \( (0, \infty) \). 
These properties, by virtue of the oscillation 
theorem, are possessed only by the eigenfunction 
corresponding to the lowest eigenvalue.

One of the most accurate and at the same time 
simplest ways of finding the lowest eigenvalue is 
the variational method. Equation (31) with 
the boundary condition (32) is equivalent to the fol­
lowing variational problem:
\[ -\beta_j = \min_{u, v} \int_0^{\infty} \left[ \lambda(x) y f_j(y) \right]^2 + \frac{x_k^2(y)}{\lambda(x)} \gamma_j F(y) \int_0^{\infty} \frac{x_k^2(y)}{\lambda(x)} \gamma_j F(y) \right| dy = \sigma \]
considered, however, one can also hope to obtain a reasonable approximation for the function (although this is not a necessity), since its quantitative behavior is very simple. Directly from (31) it is apparent that \( s(y) \) is a smooth function monotonically decreasing with increase of \( y \).

Evaluation of the integrals entering into (33) and the subsequent solution requires as a rule rather cumbersome expressions. Therefore we will utilize this method only in cases when an accurate solution cannot be found.

We proceed now to the actual solution of (30) in different cases.

I. \( j = 1 \). As already remarked above, this case corresponds to semiconductors with not very large dielectric constants \( (\mu \sim 1) \). In the null approximation with respect to \( \alpha \), the system (30) takes the following form

\[
\frac{ds(y)}{dy} - y f_0(y) = 0, \\
\frac{1}{1 + \beta_1 y} \frac{d \tilde{f}_0(y)}{dy} + \tau_1 f_0(y) - s(y) = 0. 
\]

(34)

The integration of these equations is carried out in Appendix II and leads to the following results:

\[
f_0(y) = \frac{\text{const}}{\sqrt{z}} \left\{ W_+(z^2) + \rho W_-(z^2) \right\} \exp \left\{ -\frac{1}{4} \beta_1 \tau_1 \left( y + \frac{1}{2 \beta_1 \tau_1} \right)^2 \right\},
\]

(35)

\[
s(y) = -\frac{1}{\sqrt{\beta_1} \sqrt{z_0}} \left\{ \frac{W_+(z^2) - \rho W_-(z^2)}{W_+(z_0^2) + \rho W_-(z_0^2)} - \frac{\beta_1 \tau_1}{2} \right\} f_0(y),
\]

(36)

where

\[
z = \beta_1 \tau_1 \left( y + \frac{1}{2 \beta_1} \left( 1 + \frac{\beta_1 \tau_1^2}{4} \right) \right), \quad \rho = \frac{1}{4} \beta_1 \tau_1 \left( 1 - \beta_1 \tau_1^2 \right).
\]

(37)

The various limiting cases of these formulae are also treated in Appendix II. Here we only remark that the quantity \( \beta_1 \gamma_1^2 = \delta (E_i / E)^2 \) for the fields of interest to us is small, which allows formulae (35) - (37) to be greatly simplified.

II. \( j = 2 \). The case of large values of the dielectric permittivity. In Appendix I it is shown that \( \mu \) can be considered as large if the condition

\[
\mu \gg \varepsilon_0^3 \hbar \sqrt{2 \varepsilon_0 \varepsilon_i / m},
\]

(38)

is satisfied, where \( \varepsilon_0 \) is the electronic charge. By inspection this criterion can be interpreted as follows; the quantity \( \mu \) is considered large in crystals in which the binding energy of Coulombic levels \( \alpha \varepsilon_0 / \mu^2 \hbar^2 \) is smaller than the width of the region \( \varepsilon_1 \) where impact ionization takes place. In the semiconductors of most interest — germanium and silicon — the binding energy of the Coulomb levels is about \( 10^{-3} \) ev, and the quantity \( \alpha \varepsilon_0 / \mu^2 \hbar^2 \) \( \sim 0.1 \) ev. It is natural, therefore, to suppose that they belong to the case \( j = 2 \). The intermediate case when

\[
w_i(x) = \frac{\rho}{\alpha (x - 1)} + c_1 (x - 1)^2 \]

and \( \alpha c_1 \sim 1 \),

also leads to the case \( j = 2 \) by the transformation \( y' = y + 1 / 2 \alpha c_1 \) and, therefore, will not detain us.

Putting \( j = 2 \) in the original system (30) gives

\[
\frac{ds(y)}{dy} - y^2 f_0(y) = 0, \quad \frac{1}{1 + \beta_2 y^2} \frac{d \tilde{f}_0(y)}{dy} + \gamma f_0(y) - s(y) = 0.
\]

(39)

It is not possible to find the general solution of these equations for arbitrary values of the parameters. We proceed, therefore, in the following manner: we divide the integral of possible values of the field into two partially overlapping regions, in one of which \( \beta_2 \ll 1 \), and in the other \( \gamma_2 \ll 1 \). In the first region, which is apparently the one of greater interest, we find an analytical solution of (39), and in the second utilize the general method described above for determining the function \( \sigma(E) \).

We will start with a proof that these regions do, in fact, overlap, and thus the combination of the solutions we obtain contains the solution of the problem for any values of the parameters \( E \) and \( p \). To do this, we remark that the product \( \beta_2^2 / \gamma_2 \sim \sqrt{\rho \delta^2} \ll 1 \), and therefore for all values of \( E \) one of the quantities \( \beta_2 \) and \( \gamma_2 \) must be small. The regions overlap when \( \beta_2 \sim \gamma_2 \sim (p\delta^2 / \varepsilon_0^2)^{1/2} < 1 \), although the margin in the latter inequality is not very large. The considerations given have an obvious physical significance. The energy relaxation time due to collisions with phonons is much greater than the momentum relaxation time. In ionization collisions these times are of the same order. Therefore, if the ionization processes play an important role in establishing momentum equilibrium \( (\beta_2 \gg 1) \), then the energy relaxation is determined only by them \( (\gamma_2 \ll 1) \). On the other hand, if phonons make a significant contribution to establishing energy equilibrium \( (\gamma_2 \gg 1) \), then the momentum loss in ionization for time \( \tau \) is insignificantly small \( (\beta_2 \ll 1) \). The region \( \beta_2 \ll 1 \) is of the greatest interest, since it corresponds to a field \( E < \varepsilon_1 \), and in general this condition, as was shown above, is satisfied even by breakdown fields. We commence the discussion with this case.

a) \( \beta_2 \ll 1 \). Region of comparatively weak fields.
In the zeroth approximation with repect to $\alpha_2$ and $\beta_2$ (the method used, taking into account the corresponding corrections, is given in Appendix III) a second-order equation for $f_0(y)$ can be obtained from (39):

$$d^2f_0(y)/dy^2 + \gamma df_0(y)/dy - y^2f_0(y) = 0. \tag{40}$$

By a series of elementary substitutions this can be reduced to the Whittaker equation and its solution takes the following form:

$$f_0(y) = \text{const} \cdot y^{-\gamma/2} \exp \left( -\frac{1}{2} \frac{y^2}{\gamma} \right) W_{\gamma/2}(y^2). \tag{41}$$

Using the known behavior of the Whittaker function, it is not difficult to discover the behavior of $f_0(y)$ at zero and at infinity:

$$y > 1: f_0(y) \approx \text{const} \cdot y^{-\gamma/2} \exp \left( -\frac{1}{2} \frac{y^2}{\gamma} \right), \tag{42}$$

$$y < 1: f_0(y) \approx \text{const} \cdot \frac{\Gamma(\frac{\gamma}{2})}{\Gamma(\frac{\gamma}{2} + \frac{1}{2})} \left( 1 - \frac{1}{2} \gamma y \right) + \frac{\Gamma(-\frac{\gamma}{2})}{\Gamma(\frac{\gamma}{2} + \frac{1}{2})} y + O(y^2). \tag{43}$$

With the help of the last formulae and the second of Eqs. (39) the quantity $\sigma(E)$ of interest to us can be determined:

$$\sigma(E) = -\frac{1}{\gamma_1} \left( \frac{d}{dy} \ln f_0(y) + \gamma_1 \right) \bigg|_{y=0} = \frac{2}{\gamma_1} \frac{\Gamma(\frac{\gamma}{2} + \frac{1}{2})}{\Gamma(\frac{\gamma}{2} + 1)} - \frac{1}{2}. \tag{44}$$

In the limiting cases of small $(\gamma_2 \gg 1)$ and large $(\gamma_2 \ll 1)$ fields we have

$$\sigma(E) = \begin{cases} \frac{2\Gamma(\frac{\gamma}{2})}{\gamma_1 \Gamma(\frac{\gamma}{2} + 1)} \left( \frac{\gamma_1}{\gamma} \right)^{\frac{\gamma}{2}} \frac{e^{\frac{\gamma_1}{\gamma} y}}{y^\gamma} \quad & \gamma_2 \gg 1 \end{cases},$$

$$\sigma(E) = \begin{cases} \frac{2\Gamma(\frac{\gamma}{2})}{\gamma_1 \Gamma(\frac{\gamma}{2} + 1)} \left( \frac{\gamma_1}{\gamma} \right)^{\frac{\gamma}{2}} \frac{e^{\frac{\gamma_1}{\gamma} y}}{y^\gamma} \quad & \gamma_2 \ll 1. \end{cases} \tag{45}$$

The distribution function (41) in the latter case $(\gamma_2 \ll 1)$ is close to that obtained by Heller.\textsuperscript{11}

b) $\gamma_2 \ll 1$. Region of very strong fields. In the zero approximation with respect to $\alpha_2$ and $\gamma_2$ we must solve the following variational problem:

$$-\beta_2 = \min \int_0^\infty \left( \begin{array}{c} \text{const} \cdot y^{-\gamma/2} \exp \left( -\frac{1}{2} \frac{y^2}{\gamma} \right) W_{\gamma/2}(y^2) \\ -\frac{1}{\gamma_2} \left( \frac{d}{dy} \ln f_0(y) + \gamma_2 \right) \\ \lim_{y \to 0} \frac{1}{\gamma_2} \frac{d}{dy} s(y) \end{array} \right) dy.$$ \tag{46}

The detailed method of solution is given in Appendix III. Here we confine ourselves to a summary of the results. The connection between the quantity $g = (\gamma_2 y)^{-\gamma/2}$ and $\beta_2$ is given in parametric form by the two relationships

$$\beta_2 = g \left( \gamma_2 \phi_1(v) - \phi_2(v) \right), \quad g^2 = -\frac{dv}{\phi_1(v)} \left/ \frac{dv}{\phi_2(v)} \right., \tag{47}$$

where

$$\phi_1(v) = \left( \frac{\gamma_2 y^2}{\gamma} \right)^{\frac{\gamma}{2}} \left( \frac{\gamma_1}{\gamma} \right)^{\frac{\gamma}{2}} \frac{e^{\frac{\gamma_1}{\gamma} y}}{y^\gamma},$$

$$\phi_2(v) = \left[ \frac{\gamma_1}{\gamma} \right]^{\frac{\gamma}{2}} \left( \frac{\gamma_1}{\gamma} \right)^{\frac{\gamma}{2}} \frac{e^{\frac{\gamma_1}{\gamma} y}}{y^\gamma}.$$ \tag{48}

The corresponding function $s(y)$ takes the form

$$s(y) = \sigma K_v(x), \quad x = \frac{2\gamma_0 \sin \nu \varepsilon}{3\varepsilon}, \tag{49}$$

where $K_v(x)$ is the Macdonald function. $\nu$ varies from $\pi$ to $\frac{\pi}{2}$. In the limiting case when $\beta_2 \to 0$, $\nu$ tends to $\frac{\pi}{2}$ and the solution (49) agrees with that which is obtained from (41) in the limit as $\gamma_2 \to 0$. Thus, the solutions we have obtained in fact join up in the region where the conditions $\beta_2 < 1$ and $\gamma_2 \ll 1$ are simultaneously satisfied. In the other limiting case when $\nu \to \frac{\pi}{2}$, $g$ tends to infinity. Thus, $\beta_2 \sim g^2 \sim (\gamma_2 x)^{-2}$. In fact, for very large fields

$$\sigma(E) = \sqrt{\frac{\gamma_1}{\gamma}} \frac{\gamma}{\gamma_1} \frac{E}{E_0}. \tag{50}$$

APPENDIX I

The probability of the creation of electrons with momenta $p_1$ and $p_2$ and a hole with momentum $p_3$, due to an ionizing collision of an electron with an original momentum $p_0$, can always be written in the form

$$\omega_i = \frac{2\pi}{h} \sqrt{\frac{3}{2}} \times \frac{1}{\varepsilon} M(p_0; p_1, p_2)^2 \delta \left( \varepsilon_v(p_0) - \varepsilon_v(p_1) - \varepsilon_v(p_2) - \varepsilon_v(p_3) \right) \times \frac{1}{\varepsilon} \left( \varepsilon_v(p_0) - \varepsilon_v(p_1) - \varepsilon_v(p_2) - \varepsilon_v(p_3) \right). \tag{A.1}$$

All the conservation laws can be satisfied only for sufficiently large values of $p_0$. The ionization threshold is determined by the condition

$$\varepsilon_v(p_0) = \varepsilon_v(p_1) = \varepsilon_v(p_2) \equiv \min \left( \varepsilon_v(p_0) + \varepsilon_v(p_1) + \varepsilon_v(p_2) - \varepsilon_v(p_3) \right).$$

The condition of a minimum in the right-hand side of the equality means that at the threshold $\nabla \varepsilon_v(p_1) = \nabla \varepsilon_v(p_2) = \nabla \varepsilon_v(p_3) = \varepsilon_v$, i.e., the speeds of all the final particles are equal. Close to the threshold the argument of the energy $\delta$--function can be developed in a power series of the departure of the momenta from their values $p_m(p_0)$ determined from the minimum condition written down above. The coefficients of the corresponding quadratic form, after transforming to principal axes, we will label $m_p^{-1}(p_0)$. After introducing new variables of integration according to the formulae

$$p_k - p_{km}(p_0) = \sqrt{2m_k(p_0)} \left( \varepsilon_v(p_0) - \varepsilon_v(p_3) \right) \pi_k$$
etc., (A.1) acquires the following form:

$$\omega_i(p_0) = \frac{2\pi}{\hbar} \left( \sum_{k=0}^{\infty} n_k^2 \right)^{-1/2} \left| M(p_0; p_1, p_2) \right|^2 \delta \left( 1 - \sum_{k=0}^{\infty} n_k^2 \right) \frac{d\pi}{d\Omega} =$$

$$= \frac{1}{\hbar} \left( \sum_{k=0}^{\infty} n_k^2 \right)^{-1/2} \delta \left( 1 - \sum_{k=0}^{\infty} n_k^2 \right) \frac{d\pi}{d\Omega} \left| M(p_0; p_1, p_2) \right|^2 \left( \epsilon_0 - \epsilon_{in}(p_0) \right)^2.$$

(A.2)

In the Born approximation $M(p_0; p_1, p_2)$ is simply the matrix element of the corresponding interaction energy. Since the momenta which are exchanged by the particles participating in the reaction are of the order $\sqrt{m} \epsilon_1$, the collision parameters making the principal contribution to ionization at the threshold are of the order $\frac{n}{\sqrt{m} \epsilon_1}$. In this region the interaction potential must be of Coulomb order $e^2/r$, since the polarization of the medium only occurs at large distances. Therefore

$$e^2 - \frac{1}{13} \frac{1}{12} V^{\frac{1}{13}} \left( \frac{m}{e^2} \right)^{\frac{1}{12}}$$

(A.3)

This expression is almost the same as the result obtained by Tevordt from an exact analysis of a somewhat simplified model. In deriving (A.3) we have neglected the difference between the slowly varying quantities $m^*_{k}(p_0)$ and their values at the ionization threshold, and have replaced all the $m^*_{k}$ by some mean value $m$. Also we took the speed of the final particles $v$ as small compared with the speed of the primary $V \epsilon_1(p_0)$. When the Born approximation is inapplicable, the ionization cross-section for slow electrons differs from the Born multiplier $\frac{1}{V} \left( \epsilon_0 - \epsilon_{in}(p_0) \right)^2$ as determined by the long-range part of the Coulomb interaction. In the region of small energies

$$\omega_i(p_0) \sim \frac{e^m}{\hbar^2} V \left( \frac{\epsilon_0 - \epsilon_{in}}{e_i} \right)^2 = \frac{e_i}{\hbar} \left( \frac{e^m}{\mu^2 \epsilon_i} \right)^2 V \left( \frac{\epsilon_0 - \epsilon_i}{e_i} \right),$$

(A.4)

The energies of interest to us are of the order $\epsilon_i + \alpha \epsilon_1$. Consequently, if $\mu$ is large enough so that $\mu^2 \alpha \epsilon_1 h^2/e^2m \gg 1$ the Born approximation is applicable for them and Formula (A.3) can be used. In cases where $\mu \sim 1$, the situation is completely analogous to that which exists in gases and the ionization cross section close to the threshold decreases linearly.

APPENDIX II

We transform the system (34) by introducing new variables according to the formulae

$$z = \beta \frac{1}{2 \epsilon_1} \left[ y + \frac{1}{2 \beta} \left( 1 + \frac{\beta \gamma_1}{\epsilon_0} \right) \right], \quad \rho = \frac{1}{4} \beta \gamma_1 \left( 1 - \frac{\beta \gamma_1}{\epsilon_0} \right),$$

$$\chi_1(z) = \left( \frac{1}{2} - \frac{\beta \gamma_1}{2} \right) f_0(y) + V\gamma s(y) g z \frac{f_0(y) + V\gamma s(y)}{\epsilon_0 - \epsilon_i},$$

$$\chi_2(z) = \left( \frac{1}{2} - \frac{\beta \gamma_1}{2} \right) f_0(y) - V\gamma s(y) g z \frac{f_0(y) - V\gamma s(y)}{\epsilon_0 - \epsilon_i}.$$
\[ s(y) = -(1 - \sqrt{\frac{2}{\pi} \gamma_1}) f_0(y) / \sqrt{\pi}, \quad \sigma(E) = 1 / \sqrt{\frac{2}{\pi} \gamma_1} - 1/2. \]  
\[ (A.8) \]

b) \( \rho \gg 1 \). Region of relatively small fields. For this case it is most convenient to start directly from (A.6). We transform afresh to the independent variable \( y = (4p)^{3/2} (z - 2p) \) and neglect the small quantity \( \rho^{3/2} y^3 / 2 \). The functions \( \chi_{1,2}(y) \) then satisfy the equation

\[ d^2 \chi_{1,2}(y) / dy^2 - \left( y \pm \sqrt{\frac{3}{\pi} \gamma_1} \right) \chi_{1,2}(y) = 0. \]  
\[ (A.9) \]

Consequently,

\[ \chi_{1,2}(y) = \left( y \pm \sqrt{\frac{3}{\pi} \gamma_1} \right)^{\gamma_1} K_{\gamma_1} \left( y \pm \sqrt{\frac{3}{\pi} \gamma_1} \right)^{\gamma_1}, \]  
\[ (A.10) \]

\[ f_0(y) \approx \frac{(4p)^{3/2}}{\gamma_1} \exp \left[ -1/2 \left( \gamma_1 (y + 1/\gamma_2) \right) \right] + O(\gamma_2), \]  
\[ s(y) = -(1 - \sqrt{\frac{2}{\pi} \gamma_1}) f_0(y). \]  
\[ (A.11) \]

\[ \sigma(E) = -3\gamma_1 \left( \gamma_1 / \gamma_2 \right)^{1/2} K_{\gamma_1} \left( \gamma_1 / \gamma_2 \right) ^{-1/2}. \]  
\[ (A.12) \]

If \( \gamma_1 \) is not small, then

\[ \sigma(E) = -1/2 \left( K_{\gamma_1} \gamma_1^2 / 12 \right) \left( \gamma_1^2 / 12 - 1 \right). \]
\[ (A.13) \]

This solution is easily obtained also from (34) under the condition \( \beta_1 \ll 1 \).

APPENDIX III

1. The evaluation of the corrections to the solution of (30) for \( j = 2 \) which are proportional to \( \beta_2 \), \( \alpha_2 \), and \( \alpha_2^3 \), can be performed in the following way. We introduce the new independent variable

\[ z = \left[ 2 \left\{ y \sqrt{\frac{2}{\pi} \gamma_1} \left[ 1 + \frac{3}{\gamma_2} \left( y \right)^{\gamma_1} \right] \right\} dy \right]^{1/3} \]  
and then by substitution

\[ f_0(y) = \frac{dz}{dy}^{-1} \exp \left[ - \frac{1}{2} \int \frac{\lambda(x)}{x \tau(x)} dy \right] \gamma_2 \gamma_1 \]  
\[ + \frac{d}{dy} \left( \frac{\lambda(x)}{x \tau(x)} \right) \gamma_1 \frac{\lambda(x)}{x \tau(x)} \frac{dz}{dy} \right] \gamma_2 \gamma_1 (x) \]  
\[ (A.14) \]

arrive at the equation in the normal form

\[ d^2 \varphi(z) / dz^2 - \left[ \left( 1 + \frac{1}{4} \gamma_2^2 \right) + z^2 \right] \varphi(z) = 0. \]
\[ (A.15) \]

We develop \( \Psi(z) \) in a power series of \( z \): \( \Psi(z) = \Psi(0) + \Psi_1(z) + \Psi_2(z^2) + \ldots \). In essence this series is an expansion of \( \Psi(z) \) in powers of \( \alpha_2 \) and \( \beta_2 \). : \( \Psi \sim \alpha_1 \), \( \Psi_2 \) contains terms proportional to \( \alpha_2^2 \) and \( \beta_2 \), etc. By the transformation

\[ z' = \left( 1 + \frac{1}{4} \gamma_2^2 \right)^{1/4} \left( z + \frac{1}{2} \gamma_2 \right)^{1/4} \]  
\[ (A.16) \]

Equation (A.15) leads to the previous form. We limit ourselves here to this preliminary treatment and shall not proceed to explicit expressions for \( \Psi(z), \Psi_0, \Psi_1 \) and \( \Psi_2 \) in view of their cumbersome nature.

2. As the variational function for the problem formulated in Eq. (46), we will choose the function

\[ s(y) = z^2 \gamma_1 \left( \gamma_1 / \gamma_2 \right) \]  
for \( z = (\gamma_1 / \gamma_2)^{1/2} \). This function has the correct behavior at zero, \( s(y) \sim y^3 \), and decreases monotonically with increase of \( y \), i.e., it satisfies the basic qualitative requirements for \( s(y) \). Also in the limiting cases of small \( \nu < 1 \) and large \( \nu > 1 \) values of \( \beta_2 \), it gives an accurate solution of (39) for \( \gamma_2 = 0 \). Of the two parameters \( \xi \) and \( \nu \), only one is disposable by virtue of the additional condition (46). We will take \( \nu \) as the independent variable. The parameter \( \xi \) is expressed in terms of \( \nu \) and \( \sigma \) in the following way:

\[ \nu_1 = \left( 2 \nu / 3 \pi \right)^{1/2} \sigma \nu. \]
\[ (A.17) \]

The evaluation of all the integrals entering into (46) is most conveniently carried out using the known integral forms of the Macdonald functions

\[ K_\nu(z) = \frac{\gamma^2}{2} \int_0^\infty \gamma^{-1} \exp \left\{ - \frac{t}{2} - \frac{\gamma^2}{2t} \right\} dt, \]
\[ \int_0^\infty \gamma^{-1} \exp \left\{ - \frac{t}{2} - \frac{\gamma^2}{2t} \right\} dt = \int \frac{e^{\nu} K_\nu(z) K_\nu(z) dz}{z^{1/2}}, \]
\[ (A.18) \]

As a result, (46) is reduced to

\[ -\beta_2 = g \min \left( -\beta_2^2 + \gamma_2^2 \right) \]  
\[ (A.19) \]

where the functions \( \Psi_1(\nu) \) and \( \Psi_2(\nu) \) are determined by (48). The condition for the minimum of this expression leads to the connection (47) between \( \beta_2 \) and \( g \).


7 B. I. Davydov and I. M. Shmushkevich, JETP 10, 1043 (1940).
10 R. Courant and D. Hilbert, Methods of Mathematical Physics, Interscience, 1943. Chapter IV.
13 L. D. Landau and E. M. Lifshitz, Квантовая механика (Quantum Mechanics), Pergamon, 1958 Sec. 118.
14 S. Geltman, Phys. Rev. 102, 171 (1956).

Translated by K. F. Hulme