PHYSICAL REVIEW VOLUME 85, NUMBER 2 JANUARY 15, 1952

Theory of Conductivity of Semiconductors*

GEORGE JAFFÉ
Louisiana State University, Baton Rouge, Louisiana

(Received February 26, 1951)

The author's older theory of conduction in polarizable media is generalized in such a way as to include rectification. An electronic component (positive holes or electrons) and an ionic component (negative or positive ions) are taken into account. The former are subjected to the boundary conditions of the Mott-Schottky theory, the latter to the boundary conditions of the author's theory of polarization. Significant deviations from the Mott-Schottky theory are caused by the ionic component.

In Sec. I the fundamental equations are established in a general way and the boundary conditions are discussed. In the remainder of the present paper only the limiting case is treated where all impurity centers are either completely associated or completely dissociated. In Sec. II(a) the general procedure for treating the time-dependent equations is sketched, and subsequently the special case of simple ac is treated in detail. An expression for the current is obtained without the omission of any terms [Eq. (21)]; however, its evaluation requires the introduction of a simplified "zero solution." Then the complete expression for the frequency dependence of the equivalent susceptance and conductance is established (Sec. II(b)). Finally, in Sec. III, the theory is compared with the results of ac measurements performed in this laboratory on locally manufactured selenium disks. The theory explains the frequency dependence of capacitance and conductance over the range from about 2000 to 200,000 cps, if the presence of two species of ions of different mobilities is assumed. From this agreement ionic data, i.e. the concentrations of the carriers, their mobilities, diffusivities, and the thickness of the depletion layer, are derived and discussed.

THE theory which has been applied with the most success to the understanding of the rectifying action of semiconductors is the so-called boundary layer theory. It was developed in a simpler form by Mott,[1] and in a more elaborate form by Schottky and his collaborators.[2] The underlying concept is that the unipolar resistance of the contact between metal and semiconductor is mainly due to a thin layer, the depletion layer, which is practically depleted of the mobile particles, "electrons" or "positive holes," as the case may be. The state inside the semiconductor is regulated by space charge and diffusion and is affected in a significant way by the relevant boundary conditions.

This model gives the essential features of the observed characteristics though it proves not to be flexible enough to account for the finer details. Thus it was recognized by Mott and Schottky that field effects, due to the intensity of the electric field near the boundary, and to the discrete nature of the perturbing charges, had to be introduced.

An entire group of phenomena, which are not explained by the theory as it stands, is related to the behavior of rectifiers with time, in particular the comparatively slow changes in conductivity which are observed. It has been suggested by several authors that these changes may be due to an ionic component of the current, as opposed to the electronic component so far studied.

Now, a theory of conduction in polarizable media based on exactly the same principles as the Mott-Schottky theory was developed by the author as early as 1932.[3] At that time the theory was applied to explain the well-known time effects connected with polarization, and the boundary conditions were chosen in such a way as to realize this aim. It was not the object of the investigation, as carried out then, to give account of rectification, and this effect was automatically excluded by the symmetrical nature of the boundary conditions. However, it is only necessary to introduce boundary conditions of a dissymetrical kind, such as used by Mott and Schottky, to make the theory include unipolar conduction, i.e., rectification.

It was pointed out in the quoted paper that the changes in conductivity observed in polarizable media can be explained if two kinds of carriers were assumed present. They were then called ions of the first and of the second kind, the former being characterized by the fact that they could not leave the dielectric and therefore could not carry a current across the boundary, whereas those of the second kind were subject to the usual boundary conditions and could carry a current across the boundary.

It is evident that this approach to the problem can be extended to semiconductors and that these lend themselves particularly well to description by such a model. If the normal carriers (electrons or positive holes) are identified with the "ions of the second kind," and the ions in the semiconductor with the "ions of the first kind," the semiconductor will show the phenomena of polarizability and relatively slow changes with time, as desired.

In the present investigation the older theory of the author will be generalized in such a way as to include

---

* Part of the work reported here was performed under contract with the Department of the Army, Signal Corps.

[1] N. F. Mott, Proc. Roy. Soc. (London) **A171**, 27, 281 (1939).

[2] (a) W. Schottky, Z. Physik **113**, 367 (1939); (b) W. Schottky and E. Spenke, Wiss Veröffentl Siemens-Werken **18**, 225 (1939); (c) W. Schottky, Z. Physik **118**, 539 (1942).

[3] G. Jaffé, Ann Physik **16**, 217, 249 (1933).

rectification. From what has been said above, this amounts to the same as saying: the Mott-Schottky theory will be generalized in such a way as to include mobile ions. However, our treatment, the older one as well as the present one, is more general than the Mott-Schottky approach in one other regard. Mott assumes the chemical equilibrium in the semiconductor to be established and known, and Schottky treats of it in a very elaborate way[4] under the specific assumption that it is "quasi-thermic," i.e., that it is not modified by the passage of the current. This assumption is evidently correct if the time which the dissociation equilibrium requires to establish itself is short compared to the time during which the concentration of the various particles may change as a result of their motion (in the diffusion field and in the electric field). Thus, the assumption of quasithermic equilibrium breaks down at least when we are dealing with ac and when its period becomes comparable with the chemical time of relaxation.[5] Now it is our object to include in our fundamental equations the possibility that the applied voltage varies with time in an arbitrary way. Then it will not be correct to calculate the dissociation equilibrium independently of the passage of the current. The way of avoiding this is to start from the equations for the rate of change of each species of particles, instead of starting from the equilibrium equations (as do Schottky and Spenke). This leads, in the case of the stationary state, to equations which establish the modification of the dissociation equilibrium by the diffusion field and the electric field.

The fundamental equations will be derived in this more general form because they are meant to serve as a basis for various problems in the theory of conductance of crystals, semiconductors, and solutions. However, in the present paper two limiting cases will be treated only, and further simplifying assumptions will be introduced. It will be assumed that there are present in the semiconductor either positive holes and negative ions only (case A : defect semiconductor), or electrons and positive ions only (case B: excess semiconductor). Furthermore, it will be assumed that the impurity centers are, in case A, completely associated with electrons and, in case B, completely dissociated from electrons. Under these restrictions the problem can be treated in a general way by a method of successive approximations for a wide class of time-dependent applied voltages. However, for the sake of brevity, only the case of simple ac will be treated explicitly here.

I. THE FUNDAMENTAL EQUATIONS

We are treating the linear case only and assume the two electrodes to be in the planes $x=0$ and $x=l$. Dealing with case A first, we designate by p the concentration (i.e., number per cc) of the positive holes, by n_A and n the concentrations of the "acceptors" in the neutral and charged states, respectively. Since we assume all particles, even the neutral acceptor molecules, to be mobile, the concentrations p, n_A, and n, as well as the field intensity E, are, in general, functions of x and t.

Each time a neutral acceptor molecule associates with ("catches") an electron, an A^- ion and one positive hole are formed. This "formation process" may be considered as being of the first order, since the electron concentration in the highest band is practically constant. Hence, the number of positive holes (and A^--ions) formed per sec and cc is given by $k_1 n_A$ where k_1 is a rate constant.

Each time an A^--ion gives up its electron to a positive hole, one A^--ion and one positive hole disappear (by "recombination" of the positive hole with the electron). This recombination process is of the second order, the number of positive holes (and A^--ions) disappearing per cc per sec being $k_2 pn$ where p_2 is a second rate constant.

Besides these "chemical" changes, the changes due to diffusion and to migration in the field must be taken into account. Let b be the mobility of the positive holes and D their coefficient of diffusion, then the complete equation of balance will be

$$\partial p/\partial t = k_1 n_A - k_2 np + D\partial^2 p/\partial x^2 - b\partial(pE)/\partial x. \quad \text{(I)}$$

If b' and D' have the corresponding significance for the negative ions, we obtain in a similar way

$$\partial n/\partial t = k_1 n_A - k_2 np + D'\partial^2 n/\partial x^2 + b'\partial(nE)/\partial x. \quad \text{(II)}$$

The neutral acceptor molecules do not move in the electric field but may diffuse, hence their equation of balance is

$$\partial n_A/\partial t = -k_1 n_A + k_2 np + D_A \partial^2 n_A/\partial x^2, \quad \text{(III)}$$

if D_A is the relevant coefficient of diffusion.

Finally, the electric field, E, is subject to the Poisson equation

$$\partial E/\partial x = 4\pi\epsilon/K(p-n) \quad \text{(IV)}$$

if ϵ is the elementary charge, and K the dielectric constant of the semiconductor.

Equations (I) to (IV) are the fundamental equations of the problem as treated here. They suffice to determine the four unknowns, namely p, n, n_A, and E, provided adequate boundary and initial conditions are given (see below).

Our theory is descriptive in the sense that the relevant "constants" b, D, etc., and the rate constants k_1 and k_2 are considered as known functions of the temperature. Of course, what is known of these constants from other arguments[6] can easily be incorporated into the present theory.

It will not be necessary to write down explicitly the fundamental equations for excess semiconductors (case

---

[4] See reference 2(b), Sec. A.

[5] It will be seen in what follows that the assumption becomes illegitimate even in the case of dc provided the inhomogeneities of the field and of the particle densities are sufficiently large.

[6] Compare N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, New York, 1940).

B). Let now n, p, and p_D stand for the concentrations of the free electrons, the positive ions, and the neutral "donator" molecules, respectively. The relevant formation process will now be the release of an electron by a neutral donator molecule whereby one free electron and one positive donator-ion is formed (rate: k_3p_D). The reverse recombination process consists in the formation of a neutral donator molecule by the capture of an electron by a positive ion (rate: k_4pn).

It is easy to show that, with the definitions given above, the fundamental equations for case B are obtained from (I) to (IV) by replacing in these equations n_A, k_1, k_2, D_A by p_D, k_3, k_4, D_D, respectively.

Our fundamental equations show in what way the quasithermic equilibrium is affected by the migration of the particles in the electric field and in the diffusion field. In the stationary case the left-hand sides of Eqs. (I) to (III) become zero. However, only if either the neutral molecules cannot diffuse ($D_A=0$ in case A), or if the ions have no mobility ($D'=b'=0$ in case A), do Eqs. (II) and (III) reduce to the mass action equation $np/n_A=k_1/k_2$.

In the general case this equation will be correct only if the differential coefficients with regard to time, as well as the local gradients representing the influence of diffusion and migration in the field, are sufficiently small.

It is easy to write down the fundamental equations in more general cases, e.g., if electrons, positive holes, acceptors, and donators are present at the same time. It can be deduced then that for each species of particles which are not mobile the corresponding mass action equation is obtained for the stationary case.

As for the boundary conditions, we have to distinguish between positive holes (electrons) on the one hand and ions on the other hand. For the electronic carriers we assume the asymmetric boundary conditions as introduced by Mott and Schottky. Thus we prescribe in case A

$$p=p_{\rm I} \text{ for } x=0; \quad p=p_{\rm II} \text{ for } x=l, \tag{1}$$

and in case B

$$n=n_{\rm I} \text{ for } x=0; \quad n=n_{\rm II} \text{ for } x=l. \tag{2}$$

For the ions we retain the boundary conditions of our older theory. These prescribe that no ionic current can cross the two boundaries, i.e., in case A

$$b'nE+D'\partial n/\partial x=0, \quad \text{for } x=0 \text{ and } x=l, \tag{3}$$

and in case B

$$bpE-D\partial p/\partial x=0, \quad \text{for } x=0 \text{ and } x=l. \tag{4}$$

It is reasonable to assume that the neutral impurity centers, though mobile by diffusion, cannot cross the boundaries. Hence, we have the boundary conditions

$$\partial n_A/\partial x=0 \text{ or } \partial p_D/\partial x=0, \quad \text{for } x=0 \text{ and } x=l. \tag{5}$$

Finally, the field E has to be determined subject to the condition that the voltage between the electrodes

$$V(t)=\int^{l} Edx \tag{6}$$

is a prescribed function of time, related to the applied voltage $U(t)$.[7]

The initial conditions, together with Eq. (6), determine the nature of the problem in question.

A few words should be said about the physical significance of the boundary conditions as stated. As far as the electronic component is concerned, there will be a "depletion layer" at the electrode, say at $x=0$, if the values of $p_{\rm I}$ or $n_{\rm I}$ (respectively, in cases A and B) are smaller than the undisturbed value which would hold at large distances from the electrode. However, the character of the depletion layer will differ both from that of Mott (no charges whatever), and that of Schottky (no electronic charges and constant ionic density). Since it is assumed that the ions are mobile but cannot carry their charges across the boundaries, the space charge distribution will vary under the influence of the applied voltage. In the dc case, and for the high resistance direction, the ions will run against the electrode until the ionic current is stopped. This will cause a space charge opposed in sign to the electronic component and, therefore, the flow of the latter will be enhanced. In the low resistance direction the electronic component will be decreased and the "infinite catastrophe" of the Schottky theory will be avoided.

The same situation will hold with ac as long as the frequency is so low that the average time which the ions take to travel over the distance l by diffusion is small compared to the period. If the frequency becomes so high that the reverse is true, the ionic flow is no longer impeded by the boundary layer, and the ions can contribute to the conductivity according to their density and mobility. This contribution can be appreciable because in the very nature of the depletion layer the density of the ions is high while that of the electronic carriers is low. It will be seen (see Sec. III) that in selenium this effect is so large that ionic effects completely dominate electronic effects at frequencies above a few thousand cps.

Finally, it should be pointed out that the ionic boundary conditions in the form (3) and (4) represent an ideal limiting case. Actually the discharge at the electrodes will be a time-dependent process. The theory here presented has been generalized in the direction that the discharge across the boundaries is not altogether prohibited, but is regulated by a rate constant and the concentration at the electrode. Then there will be the possibility of an ionic transport across the

---

[7] Since we are not going to treat the dc case here, the explicit relation between $V(t)$ and $U(t)$ will not be required. It is given in case (A) by $V=U-V_\epsilon$ and in case (B) by $V=U+V_\epsilon$, where V_ϵ represents Schottky's "diffusion potential" (see reference 6, pp. 70 and 168, and references 2(b) and 2(c)).

boundary, and its amount will depend on the joint actions of the diffusion and the discharge process. Thereby the theory becomes applicable to the case of electrolytic solutions.[8] However, this refinement of the theory is not required for semiconductors, at least not for the case of selenium here studied. The experimental results indicate that the limiting case represented by Eqs. (3) and (4) represents a sufficient approximation.

II. COMPLETE ASSOCIATION (DISSOCIATION)

(a) General Solution

Under this assumption the fundamental equations simplify considerably since all impurity centers are supposed to be ionized at all times. In order to make the formulations less cumbersome we shall treat only case A explicitly, and shall state the results for case B later.

In case A we have $n_A=0$ by assumption, hence it follows from Eq. (III) that $k_2=0$, i.e., the rate constant for recombination, must be negligibly small.

Under these circumstances the fundamental equations reduce to

$$\partial p/\partial t=D\partial^2 p/\partial x^2-b\partial(pE)/\partial x, \tag{Ia}$$

$$\partial n/\partial t=D'\partial^2 n/\partial x^2+b'\partial(nE)/\partial x, \tag{IIa}$$

$$\partial E/\partial x=(4\pi\epsilon/K)(p-n). \tag{IVa}$$

These equations are not linear and, therefore, can be treated only by successive approximations. If it is assumed that the voltage $V(t)$ is a periodic function which can be represented in the form[9]

$$V(t)=\sum_{m=0}^{\infty} V_m \exp(im\omega t), \tag{7}$$

Eqs. (Ia), (IIa), and (IVa) can be treated in a general way. If expressions similar to (7) are assumed for p, n, E, and the current density j, the differential equations are linearized. They can be solved in a formal way if the field is supposed to be known, and this leads to an alternate procedure for the actual determination of the unknowns similar to the one given below for a more special case.

We have carried through the calculations using infinite series as indicated,[10] however, for the sake of brevity we will restrict ourselves here to the simpler case where the series (7) reduces to its first two terms.

Even then the complete solutions for p, n, E, and j involve infinite series of the type (7), owing to the nonlinearity of the differential equations. This means physically that a semiconductor described by the present model will act as a nonlinear element, i.e., will respond to a simple harmonic emf not only with the same period but also with all its overtones. Again, since we are interested mainly in the simple ac response, we shall disregard all terms involving m higher than one.

Consequently, we assume not only that Eq. (7) reduces to two terms, but also that all unknowns are of the form

$$p=p_0+p_1\exp(i\omega t). \tag{8}$$

The case $V_1=0$ has to be treated separately. It represents the dc characteristic which has been developed to the second approximation.[11] However, since no new observations on the dc characteristic are available, we shall not treat the dc case explicitly here.

If the expressions for p, n, and E of the form (8) are substituted into Eqs. (Ia) and (IIa), these reduce to

$$i\omega p_1-D(d^2p_1/dx^2)=-b(dF_1/dx) \tag{9}$$

and

$$i\omega n_1+D'(d^2n_1/dx^2)=b'(dF_1'/dx), \tag{10}$$

where we have set

$$F_1=E_0p_1+E_1p_0, \tag{11}$$

$$F_1'=E_0n_1+E_1n_0. \tag{12}$$

If we treat F_1 and F_1' as known, we obtain the following solutions for p_1

$$p_1=\mathcal{E}_1\exp(\mu, x)+\mathfrak{F}_1\exp(-\mu_1x)+(b/2D)[G_1(x)+\bar{G}_1(x)] \tag{13}$$

and n_1

$$n_1=\mathcal{E}_1'\exp(\mu', x)+\mathfrak{F}_1'(-\mu^1, x)-(b'/2D')\times[G_1'(x)+\bar{G}_1'(x)]. \tag{14}$$

Here we have set

$$\mu_1=\rho_1(1+i),\quad \rho_1=(\omega/2D)^{\frac{1}{2}}, \tag{15a}$$

$$\mu_1'=\rho'(1+i),\quad \rho_1'=(\omega/2D')^{\frac{1}{2}}, \tag{15b}$$

and

$$G_1(x)=\exp(\mu_1x)\int_0^x \exp(-\mu_1y)F_1dy, \tag{16a}$$

$$\bar{G}_1(x)=\exp(-\mu_1x)\int_0^x \exp(\mu_1y)F_1dy. \tag{16b}$$

The definitions for G_1' and $\bar{G}_1'$ are obtained from (16a), (b) by substituting μ_1', $-F_1$, for μ_1, F_1. The constants of integration $\mathcal{E}$ and $\mathfrak{F}$ remain to be determined.

---

[8] This part of the theory and its application to electrolytic solutions will be published later.

[9] Since, for the sake of simplicity, the summation in Eq. (7) is not extended to $-\infty$, Eq. (7) does not represent the most general case. It holds if $V(t)$ is an arbitrary function of time which is either even or odd.

[10] Final report of the Signal Corps Project No. 152 B.

[11] See reference 10. In that report the theory is applied to the slow changes in time which are observed upon prolonged application of the direct voltage. The results are in qualitative agreement with observations, including the occasionally observed appearance of a current maximum, some time after the application of the field. However, the ionic mobilities necessary to explain such an event come out to be unreasonably large (of the order of $\frac{1}{10}$ of the electronic mobility), if a quantitative fit is attempted.

If now Eqs. (13) and (14) are substituted into the Poisson equation, an integral equation for E_1 is obtained which we shall indicate only in the abbreviated form

$$E_1 = V_1/l + E_1^+ + E_1^-. \tag{17}$$

The expressions for E_1^+ and E_1^- are somewhat lengthy[10] and need not be given here since they are not going to be used. They represent the space charge fields arising from the holes and negative ions, respectively, while V_1 is the applied ac voltage.

We are mainly interested in the current which will flow for a given value of V_1. The total current j_{tot} consists of the displacement current

$$j_{\text{dis}} = (K/4\pi)\partial E/\partial t$$

and the convection current

$$j_{\text{conv}} = \epsilon[(bpE - D\partial p/\partial x) + (b'nE + D'\partial n/\partial x)]. \tag{18}$$

The resultant value depends on the constants of integration which, in turn, depend on the boundary conditions. In case A the density p refers to positive holes and must be subjected to the conditions in Eq. (1). Since p_{I} and p_{II} are supposed to be independent of time, the boundary conditions must be realized in the dc component; consequently, we must demand $p_{\text{I}} = 0$ for $x = 0$ and $x = l$. This determines $\mathcal{E}_1$ and $\mathfrak{F}_1$. On the other hand, n_1 refers to negative ions. Hence the boundary conditions (3) will have to be used, which determines $\mathcal{E}_1'$ and $\mathfrak{F}_1'$ to be

$$\mathcal{E}_1' = \mathfrak{F}_1' = (b'/2D')[G_1'(l) - \bar{G}_1'(l)]/ [\exp(\mu_1'l) - \exp(-\mu_1'l)]. \tag{19}$$

The total current can be obtained in a variety of ways. The knowledge of E_1^+ and E_1^- is not required if the following procedure is adopted. Since the total current is the sum of the displacement current and of the convection current, and since $\partial E/\partial t$ reduces under our assumptions to $\partial E_1/\partial t$, we obtain by simple integration

$$(j_1)_{\text{tot}} = i\omega(K/4\pi)V_1/l + (1/l)\int_0^l j_{\text{conv}}dx, \tag{20}$$

where j_{conv} is given by Eq. (18). The integral is easily transformed if it is taken into account that p_{I} vanishes at $x = 0$ and $x = l$, and if the expressions (11), (12), (14), and (19) are used. The result is given by

$$\begin{aligned}(j_1)_{\text{tot}} = {} & i\omega(K/4\pi)(V_1/l) \\ & + (\epsilon b/l)\int_0^l F_1 dx + (\epsilon b'/l)\int_0^l F_1' dx \\ & - (\epsilon b'/l)[G_1'(l) + \exp(\mu_1'l)\bar{G}_1'(l)]/ [\exp(\mu_1'l) + 1].\end{aligned} \tag{21}$$

In case B an expression can be derived which is identical with (21) except that in the last term b, μ_1, G_1, $\bar{G}_1$ replace b', μ_1', G_1', $\bar{G}_1'$.

Our equations (13), (14), (17), and (21) for the densities, the field and the current density are rigorous consequences of the differential equations (9), (10), and the corresponding Poisson equation. However, they are formal solutions only, in the sense that they represent rather complicated integral equations for the unknowns. The value of this formal solution consists in the fact that it leads the way to an actual calculation of the unknowns by successive approximations. The procedure is as follows: we begin by assuming any reasonable zero solution for p_1, n_1, and E_1. In the simplest case, when $V_0 = 0$ and V_1 is small, a homogeneous field and no free space charge would represent such a zero solution. By the aid of this zero solution improved first-order solutions for p_1 and n_1 may be obtained from (13) and (14), and then for the field from (17), or from the Poisson equation.

In this way the procedure goes on by the alternate use of Eqs. (13) and (14) on the one hand, and of Eq. (17) on the other hand. Finally, the current density is obtained from Eq. (21). The value of the procedure will, of course, depend greatly on the choice for the zero solution. In the case of dc, only the field E_0 has to be assumed. The assumption of a constant field leads to Mott's formula, the assumption of a linear field to an improved Schottky formula which avoids the "infinite catastrophe." The zero approximation will be "adequate" if the next step in the alternate procedure does not modify it significantly.

b. The ac Characteristic

The second term of (21) represents the electronic contribution. It is not our object here to study the frequency dependence of the impedance as far as it is due to the electrons or positive holes, though our method is applicable. The ac behavior of the electronic component of semiconductors has been studied by Spenke,[12] who obtained results which are explicit for small frequencies. Furthermore, Schottky[13] has shown that the "internal time of relaxation" for semiconductors, as far as the electronic component is concerned, is very short, of the order of 10^{-10} sec. Therefore, up to frequencies almost as high as 10^{10} cps, the behavior of the electronic component can be described by equivalent values of the capacitance and resistance which are constants. On the other hand, Eq. (21) indicates a frequency dependence for much lower frequencies due to the ionic component.

For these reasons we shall assume that the frequency is limited to such values that the first two terms of Eq. (21) can be written in the form

$$(j_1)_{\text{el}} = (1/R_{\text{el}} + i\omega C_{\text{el}})V_1, \tag{22}$$

where C_{el} and $1/R_{\text{el}}$ are the constant effective values of the electronic contribution to the parallel capacitance and conductance, respectively.

---

[12] W. Spenke, Wiss. Veröffentl Siemens-Werke **20**, 40 (1941).

[13] W. Schottky, reference 2, p. 550 ff.

The last two terms of (21) represent the ionic contribution to the current. It will be seen that an exact evaluation would require the knowledge of E_0 and n_0, i.e., the solution for the dc case. If we should attempt the evaluation using the solutions obtained for this case,[10] the calculations would become prohibitively complicated. Therefore, we are going to introduce simpler approximations with the double object of simplifying the integrations and of making the knowledge of the dc solution unnecessary. This object is achieved by the simple assumption that in calculating the ac current, the inhomogeneity of the ionic distribution may be disregarded, though the inhomogeneity of the field should be retained.

Consequently we assume, as zero approximation for n,

$$n_0=c_0, \quad n_1=0, \tag{23}$$

where c_0 is the undisturbed value.

A few words should be said regarding the validity of these assumptions. If we suppose that the applied ac voltage V_1 is sufficiently small, our second assumption (23) becomes legitimate since a weak V_1 will not contribute noticeably to the space charge distribution. Nor is it inconsistent to retain the inhomogeneity of the field as, in the case of small voltages, it is much more marked than the inhomogeneity of the ionic distribution.[14]

The first assumption (23) is less justified and would be quite inadequate if the case of an applied bias were to be studied. Therefore, we have to assume that there is no applied dc voltage present. Even then there may be a space charge distribution of the negative ions near the electrodes. Its value will depend strongly on the boundary conditions which we impose on n, and will affect the current indirectly. However, it is easy to see that this influence is limited to low frequencies. The discussion given in Sec. I shows that the conditions near the electrodes become of less and less importance as the frequency increases. Therefore, the first assumption (23) becomes adequate as those values of the frequency are reached where the ionic current begins to rise strongly from its low value at low frequencies. In the applications we are going to study this region exclusively.

As for the field strength, our second assumption (23) makes the knowledge of E_0 unnecessary. For E_1 we assume

$$E_1=2C_1\cosh[\alpha(x-l_1)], \quad l_1=l/2, \tag{24}$$

with α given by

$$\alpha^2=(8\pi\epsilon/K)(b'/D')c_0. \tag{25}$$

This form of the field is suggested by the fact that it represents a first-order approximation in the static case studied in the author's older theory.[15] Furthermore, it can be proven that it is an "adequate solution" in the dc case, even when a current flows.[10] This, of course, does not prove yet that Eq. (24) is a good approximation in the ac case, except for low frequencies. As a matter of fact, it can be foreseen that E_1 becomes more and more homogeneous as the frequency increases.[16] From this point of view it might appear that the simpler assumption $E_1=\text{const}$ represents as good a zero approximation as (24). However, if $E_1=\text{const}$ is used, the possibility of determining the thickness of the polarization layer (see below) is lost. For this reason we retain the form (24).

Finally, we assume that there is only one depletion layer, at $x=0$, but no polarization layer at the other electrode. This we realize by using the solution (24) only from $x=0$ to $x=l/2=l_1$. Then the constant C_1 becomes

$$2C_1=\alpha V_1/\sinh(\alpha l_1) \tag{26}$$

if V_1 is the voltage applied to the layer of thickness l_1.

In calculating the current, we may still use Eq. (21). Owing to the complete symmetry of our assumptions with regard to $x=l/2$, Eq. (21), when divided by $2V_1$, yields the impedance of two polarization layers in series, and therefore, half of the conductance and capacitance of each of them.

If we substitute (23) to (26) into Eq. (21) the integrations become quite simple and yield

$$\begin{aligned}(j_1)_{\text{ion}}=\epsilon b'c_0(V_1/l_1)\{1-(1/(\alpha^2-\mu_1'^2))\\ \times[\alpha^2-\alpha\mu_1'\nu\tanh(\mu_1'l_1)]\}.\end{aligned} \tag{27}$$

Here the abbreviation

$$\nu=\coth(\alpha l_1) \tag{28}$$

has been introduced. Since μ_1' [see Eq. (15b)] is complex, Eq. (27) defines a complex admittance. If the real and imaginary parts are separated, the admittance for the layer extending from $x=0$ *to* $x=l_1$ assumes the form

$$Y_1=1/R_{\text{el}}+(1/R_\infty)\Gamma_2(\lambda',\beta)+i\omega[C_{\text{el}}+C_0\Gamma_1(\lambda',\beta)]. \tag{29}$$

The two functions of two arguments, Γ_1 and Γ_2, are defined by

$$\begin{aligned}\Gamma_1(\lambda',\beta)=(1+4(\lambda'/\beta)^4)^{-1}[(\nu/\lambda')g_1(\lambda')\\ -2/\beta+(2\nu\lambda'/\beta^2)g_2(\lambda')]\end{aligned} \tag{30}$$

and

$$\begin{aligned}\Gamma_2(\lambda',\beta)=(1+4(\lambda'/\beta)^4)^{-1}[(\nu\lambda'/\beta)g_2(\lambda')\\ -2\nu(\lambda'/\beta)^3g_1(\lambda')+4(\lambda'/\beta)^4],\end{aligned} \tag{31}$$

and the two functions of one argument, g_1 and g_2, are

$$g_1(\lambda')=(\sinh\lambda'+\sin\lambda')/(\cosh\lambda'+\cos\lambda') \tag{32}$$

and

$$g_2(\lambda')=(\sinh\lambda'-\sin\lambda')/(\cosh\lambda'+\cos\lambda'). \tag{33}$$

---

[14] Using the static solution and the constants obtained in Sec. III, we find for the ratio of n_1 at the anode to the undisturbed value about 1.5 for a potential difference of 10^{-2} volt, whereas the same ratio for E is of the order of 100.

[15] See reference 3, pp. 228 to 229.

[16] H. Chang has carried the calculation of E_1 to the second order and verified the statements in the text.

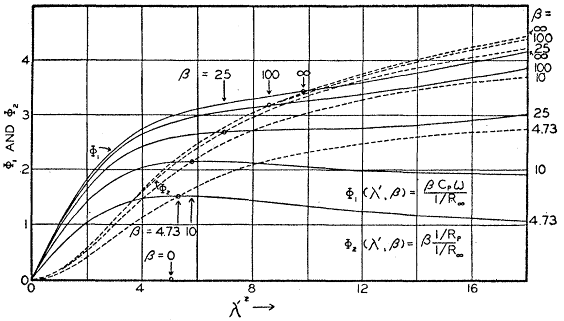

FIG. 1. Theoretical curves representing the reduced susceptance, $\Phi_1(\lambda', \beta)$, (full lines), and the reduced conductance, $\Phi_2(\lambda', \beta)$ (broken lines), as functions of the reduced frequency, λ'^2, for various values of the reduced plate distance β.

The variable λ' is given by

$$\lambda' = 2l_1(\omega/2D')^{\frac{1}{2}}, \tag{34}$$

and the dimensionless parameter β by

$$\beta = 2\alpha l_1. \tag{35}$$

The variable λ' is dimensionless, but the combination $(2l_1)^2/D' = \tau'$ has the dimension of time and may be taken as the "time of relaxation" for the ions of diffusivity D'. It represents the average time such ions take to travel over the distance $2l_1$ by diffusion. Consequently, $(\lambda')^2$ itself is proportional to the ratio of τ' to the length of the period. It is seen that the time-dependence of Y_1 is expressed by this ratio exclusively. This specifies quantitatively the statements made in the introduction.

As for the dimensionless parameter β, it was introduced in the author's older theory by the name of "reduced plate distance" and designated by δ (see reference 3, pp. 226 ff.). It characterizes the behavior of the polarized layer. If β is sufficiently large, and that means from about $\beta = 10$ on, the thickness of the polarization layer will be small compared with the total thickness of the dielectric.

Since Γ_1 tends towards unity as $\lambda' \rightarrow 0$ and $\beta \rightarrow \infty$, C_0 is the equivalent parallel capacitance (of ionic origin) in that limiting case. Its value is

$$C_0 = \epsilon b' c_0/(\alpha D'). \tag{36}$$

On the other hand, Γ_2 tends to unity as $\lambda' \rightarrow \infty$. Hence

$$1/R_\infty = \epsilon b' c_0 / l_1, \tag{37}$$

defines the equivalent parallel conductance (of ionic origin) in the limiting case of very high frequencies. All constants refer to one cm^2 area.

If we designate by R_p and C_p the equivalent parallel resistance and capacitance, respectively, their frequency dependence will be expressed by

$$C_p = C_0 \Gamma_1(\lambda', \beta), \quad 1/R_p = (1/R_\infty)\Gamma_2(\lambda', \beta). \tag{38}$$

In order to illustrate this dependence, we have plotted in Fig. 1 the two dimensionless functions

$$\begin{aligned} \Phi_1(\lambda', \beta) &= \beta \omega C_p / (1/R_\infty), \\ \Phi_2(\lambda', \beta) &= \beta (1/R_p)/(1/R_\infty). \end{aligned} \tag{39}$$

They represent reduced values of the susceptance and conductance. As abscissas we have chosen λ'^2 which is proportional to the frequency, and representative values of β, from $\beta = 4.73$ to $\beta = \infty$, have been selected.

The curves for Φ_1 all pass through a maximum and ultimately approach zero with $\omega \rightarrow \infty$. These features are not exhibited by the diagram as they occur for higher values of λ'^2. On the other hand, the curves representing Φ_2 (broken lines) increase to the asymptotic value β. Thus, with increasing frequency the ionic conductance increases from zero to the maximum value given by (37). Conversely the capacitance decreases from its initial value at $\omega = 0$, $C_0\ (1 - 2/\beta)$ to zero.

For a given value of β the corresponding curves for Φ_1 and Φ_2 intersect at a characteristic frequency $f = f^*$, and this point of intersection moves from $\lambda'^2 = 5.1$ to $\lambda'^2 = 9.8$ as β varies from 0 to ∞. It will be seen in the next section that such points of intersection have been observed and play an important part in the evaluation of the ionic constants.

The formulas so far developed in this section refer to defect semiconductors. For excess semiconductors analogous results are obtained by replacing in (21), and correspondingly in all subsequent equations, the symbols p, n, b', D', μ_1' by n, p, b, D, μ_1, respectively.

The thickness of the layer l_1 over which the solution is to be applied, remains to be determined. This we do by taking into account the bulk resistance of the entire semiconductor, and by requiring the current and the field to be continuous at $x = l_1$. By this procedure it can be shown[10] that

$$\beta = 2\alpha l_1 = 4 \ln\alpha - 2 \ln(\omega/2D'), \tag{40}$$

if λ' is small and β not small (say $\beta > 10$).

Thus l_1 is, for a given concentration of the ions, a slowly variable function of the frequency. As λ' increases, the approximate solution (40) ceases to be correct, and graphical methods have to be applied (see Sec. III); however, l_1 continues to decrease slowly with increasing λ'. On the other hand, with decreasing λ' the thickness l_1 will increase until it becomes equal to the total thickness of the semiconductor.

In the applications it will become imperative to assume the presence of more than one species of moveable ions. The resulting expression for the admittance is of the same form as (29); only there are as many terms involving functions $\Gamma_{1i}(\lambda_i', \beta)$ and $\Gamma_{2i}(\lambda_i', \beta)$ as there are different species of mobile carriers (positive or negative). Naturally, all of them move in the same field, and this necessity finds its expression in the fact that β is the same throughout. Provided all ions have the same valence, the expression for α has to be changed from (25) to

$$\alpha^2=(8\pi\epsilon/K)(b/D)\textstyle\sum_i c_i, \qquad (41)$$

where c_i is the concentration of the i-species of carriers and where the summation extends over all carriers of one sign.[17]

III. COMPARISON OF THE THEORY WITH EXPERIMENTAL RESULTS

It is well known that the ac characteristic of semiconductors shows frequency dependence at much lower frequencies than might be expected from the "internal time of relaxation" of electronic carriers. This was one of the main reasons for developing the present theory. At the same time an extensive series of bridge measurements was carried through.[18] These measurements cover a range of frequencies from 10 cps to 200,000 cps and were taken on locally fabricated selenium disks between nickel and cadmium electrodes.

It has been customary, particularly in the work of Schottky and his associates,[19] to discuss the ac behavior of semiconductors in terms of equivalent circuits. Valuable as this method is for a rapid survey of results, it cannot be considered adequate for the description of the finer details. If the fundamental concepts of the present investigation are accepted, Eq. (29) and its generalization for several species of carriers indicate clearly enough that the equivalent parallel susceptance and conductance are rather complicated functions of frequency. Furthermore, the more general treatment indicated at the beginning of Sec. II(b) proves that a simple harmonic impressed emf will cause overtones of all orders in the current. Therefore, in our opinion, the correct description of a semiconductor is that by numbers and mobilities of the relevant carriers. An attempt at determining these from the measurements mentioned above is given in what follows.

The experimental curves representing ωC_p and $1/R_p$ as functions of frequency invariably show the character of the theoretical curves given in Fig. 1, though the numerical values may vary considerably from one disk to another and for one and the same disk with the previous treatment. It was pointed out in Sec. II(b) that there is a point of intersection between the ωC_p and the $1/R_p$ curves at a determined frequency f^*. The determination of ionic constants starts most conveniently from these points where they have been observed.

In Table I we have given all cases where such points of intersection have been observed between about 100 cps and 200,000 cps. In some cases a point of intersection was observed before and after formation, where the formation consisted in the application of 40 to 160 volts in the blocking direction over a period of 3 to 6 hours. Many other disks indicated the existence of a point of intersection at higher frequencies than were observed.

The effective resistances of the locally fabricated disks at the lowest observed frequencies, i.e., 10 cps, were considerably (from 30 to 100 times) larger than the dc resistance of commercial disks. Thus, it is evident that the locally fabricated disks must have contained much less accidental impurities than the commercial ones.

Furthermore, it is characteristic of the locally fabricated disks that the electronic component may be disregarded altogether for the interval of frequencies over which a match with the theory was attempted. This is justified by the fact that, at lowest frequencies, where only the electronic component can contribute to the observed conductances, the conductance is negligibly small compared to that in the matched interval (see Figs. 2 and 3). Therefore, in the following computations the electronic component of the current has been disregarded.

It will be noticed from Table I that disk 74, in its two states, represents extreme conditions. (The forming in this case was performed by application of 40 volts for 4 hours.) In all cases the frequencies f^* and the corresponding values of ωC_p and $1/R_p$ are contained within the limits set by the disk 74 (with one unimportant exception in the case of disk 61). For this reason the laborious calculations were carried through only for the two states of disk 74, which both refer to zero bias (i.e., $U_0=0$). If the very considerable changes occurring in disk 74 during formation are represented in a satisfactory way, it seems reasonable to assume that also the other, intermediate, cases can be matched.

A great simplification of the numerical evaluations arises from the fact that the values of λ' and β which belong to f^*, say λ'^* and β^* are completely determined by the theory, as long as only one species of ions is being considered. The intersection occurs for

$$\Phi_1(\lambda',\beta)=\Phi_2(\lambda',\beta). \qquad (42)$$

If this equation is combined with one which arises from the boundary conditions for $x=l_1$, there remain two equations for the determination of λ'^* and β^*. We have

TABLE I. Observed points of intersection between the curves representing susceptance (ωC_p) and conductance ($1/R_p$).

	Virgin		Formed	
Disk	f^* (cps)	$1/R_p=C_p\omega$ (μmhos)	f^* (cps)	$1/R_p=C_p\omega$ (μmhos)
30	1.18×10^5	2.29×10^4	not observed	
47	0.312×10^5	2.03×10^4	not observed	
61	0.652×10^5	2.10×10^4	0.472×10^5	1.07×10^4
67	1.38×10^5	4.12×10^4	not observed	
74	2.05×10^5	4.12×10^4	0.416×10^5	1.29×10^4
84	1.02×10^5	2.98×10^4	0.647×10^5	1.38×10^4

---

[17] These results are obtained by first generalizing the stationary case of our older theory (see reference 3, paragraph 1) so as to include several species of carriers. This leads in a simple way to a first-order field distribution given by Eq. (24) with the modified value of α just defined. All other calculations remain the same as indicated in this section, and the final result for the current, and thereby for the admittance, is obtained by summation.

[18] The bridge measurements were performed by Mr. Allan McDonald. The details of the method and of the procedure in fabricating the disks are to be found in Report 11 and 3rd annual report (bridge measurements), and 2nd annual report (preparation of disks), Signal Corps Project No. 152 B.

[19] W. Schottky and W. Deutschmann, Physik Z. 30, 839 (1929).

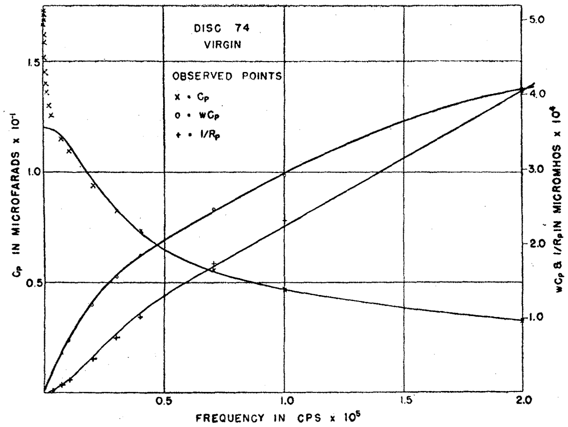

FIG. 2. Comparison of the theoretical curves for C_p, ωC_p, and $1/R_p$ with observed values. Selenium disk No. 74, virgin.

carried through the evaluation by graphical procedures; they lead to the result

$$\lambda'^* = 2.30, \quad \beta^* = 4.73. \tag{43}$$

With λ'^* and β^* known, C_0 and $1/R_\infty$ can be determined from the curves of Fig. 1. Then, Eqs. (25) and (34) to (37), together with Townsend's relation

$$b'/D' = \epsilon/kT, \tag{44}$$

represent a system of six simple equations for the determination of α, β, l_1, b', D', and c_0.

Unfortunately, the attempt to match experimental curves by our theoretical formulas invariably leads to the conclusion that more than one species of ions, at least two, have to be assumed as present in the semiconductor if the curves are to be matched over a wide range of frequencies. This might have been anticipated if the ions in question are of accidental origin; a whole "spectrum" of ions of different mobilities would appear probable.

In the following computations two species of ions are assumed to contribute essentially. Furthermore, β is treated as a constant, though it really is a slowly varying function of the frequency. Naturally β may change more markedly by the formation process which affects the number of ions present, and possibly their mobilities, and this circumstance has to be taken into account.

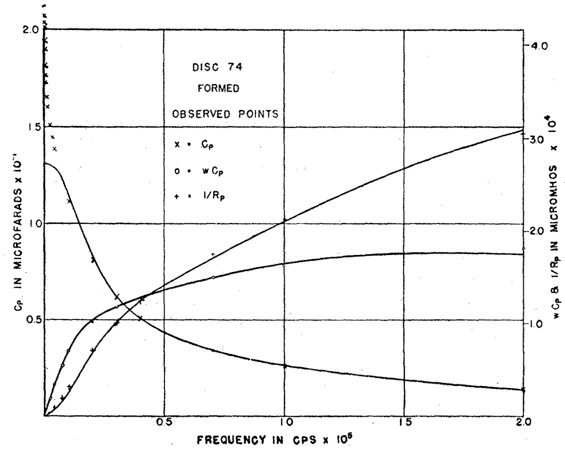

FIG. 3. Comparison of the theoretical curves for C_p, ωC_p, and $1/R_p$ with observed values. Selenium disk No. 74, formed.

If there are two species of ions, it follows from the last paragraph of Sec. II(b) that the effective values of ωC_p and $1/R_p$ will be given by

$$\omega C_p = A_1\Phi_1(\lambda_1', \beta) + A_2\Phi_1(\lambda_2', \beta) \tag{45}$$

and

$$1/R_p = A_1\Phi_2(\lambda_1', \beta) + A_2\Phi_2(\lambda_2', \beta), \tag{46}$$

TABLE II. Experimental constants used in the calculation of the curves for C_p, ωC_p and $1/R_p$ drawn in Figs. 2 and 3.

Disk	β	A_1 (μmhos)	A_2 (μmhos)	τ_1 (sec)	τ_2 (sec)
74, virgin	3.80	2.96×10^4	5.92×10^3	2.78×10^{-6}	2.52×10^{-5}
74, formed	4.73	0.87×10^4	4.96×10^3	4.80×10^{-6}	3.60×10^{-5}

where Φ_1 and Φ_2 are the functions defined by Eq. (39) and plotted in Fig. 1.

The experimental curves have to be matched with (45) and (46). It should be stressed that the curves for ωC_p and $1/R_p$, being mutually related, should be matched by the same constants, and if this is feasible it represents a strong argument in favor of the interpretation given here.

There are five constants available in Eqs. (45) and (46), namely A_1, A_2, β, and two constants which reduce the frequency scale to the λ'^2-scale for the two species of ions. Since λ'^2 is connected with the frequency by its definition (34), the "internal time constants"

$$\tau_i = l_1^2/D_i', \quad i = 1, 2, \tag{47}$$

may be used as such constants.

We have determined the five constants by trial and error using the following procedure. First one species of ions is considered only. By making the theoretical point of intersection coincide with the observed one, the three relevant constants (say A_1, β, and τ_1) are found without ambiguity, but the curves do not match sufficiently well. Then, a second (slower) species of ions is introduced and the constants first determined are modified until the matching is as good as feasible with the five constants available.

TABLE III. Limiting values of specific conductivity, σ_i, and capacitance, C_{0i}, as derived from the experimental data of Table II.

Disk	σ_1 (μmhos/cm)	σ_2 (μmhos/cm)	C_{01} (μfd/cm²)	C_{02} (μfd/cm²)
74, virgin	3.24×10^4	0.649×10^4	1.23×10^{-2}	2.23×10^{-2}
74, formed	1.19×10^4	0.677×10^4	0.72×10^{-2}	3.07×10^{-2}

The choice of the parameter β should be restricted to narrow limits. It will be seen from Fig. 1 that the shape of the theoretical curves changes but slowly with the value of β; hence, β cannot be determined with any degree of accuracy. On the other hand, the mobilities (and diffusivities) change strongly with β (being propor-

TABLE IV. Ionic constants, and thickness of the depletion layer l_i, derived from the data of Tables II and III.

Disk	l_1 (cm)	D_1' (cm²/sec)	b_1' (cm²/volt ×sec)	c_1 (cm⁻³)	D_2' (cm²/sec)	b_2' (cm²/volt ×sec)	c_2 (cm⁻³)
74, virgin	0.79×10^{-5}	4.5×10^{-5}	1.71×10^{-3}	0.94×10^{15}	5.0×10^{-6}	1.90×10^{-4}	1.70×10^{15}
74, formed	1.04×10^{-5}	4.5×10^{-5}	1.71×10^{-3}	0.45×10^{15}	6.0×10^{-6}	2.27×10^{-4}	1.92×10^{15}

tional to the square of β). Hence, β should not vary very much. In order to introduce as little ambiguity as possible into the numerical evaluations, the following procedure for the determination of β was adopted. For the formed disk the theoretical value $\beta=4.73$, which is valid for a unique group of ions, was chosen because the calculations showed that for the formed disk the one group of ions (the slower one) predominates rather strongly. To represent, then, the virgin disk, β had to be decreased, but this was done as little as possible.

The degree of agreement which could be realized over the range of frequencies from about 2×10^3 to 2×10^5 cps will be evident from Figs. 2 and 3 where the curves are calculated from theory. The constants used in these computations are given in Table II, and Table III contains the limiting values (per cm² area) of capacitance (as $f\rightarrow 0$), C_{0i}, and of specific conductance (as $f\rightarrow\infty$), σ_i. These limiting values have been obtained by extrapolation on the basis of the theoretical curves. The calculations involve the area of the disks utilized, which was 3.47-cm², and of the dielectric constant of selenium, which was found in the literature to be $K=6.3$.

The agreement over the indicated range of frequencies is about as good as might be expected in consideration of the various approximations which had to be introduced, as well into the theory, as into the numerical evaluations. At lower frequencies systematic deviations set in. The curves for C_p have been added because they show more clearly than those for ωC_p that for lower frequencies ($f<2\times 10^3$ cps) a new group of ions, of still lower mobility, makes its appearance. It would not be difficult to represent this behavior by adding a third group to the theoretical representation.

Finally, the experimental constants of Table II, together with the extrapolated values of Table III, permit of determining all ionic constants if the reasonable assumption $p_{\text{II}}=c_0=c_1+c_2$ is made. It means physically that the number of mobile negative ions (in the undisturbed state) is equal to the number of positive holes. The following equations are available: two equations of form

$$\sigma_i=\epsilon b_i' c_i, \quad i=1, 2, \tag{48}$$

(from Eq. (37)), two equations from (44), two equations of form (47), and the one equation which results from the substitution of (41) into the definition of β, (35). These are seven equations for the unknowns c_i, b_i', D_i' ($i=1, 2$), and l_1.

The values of the ionic constants, including l_1, obtained in this manner are contained in Table IV. The concentrations of the mobile ions, c_1 and c_2, appear quite reasonable, and the thickness of the depletion layer, l_1, is of the same order of magnitude found by other observers. The mobilities and diffusivities are of particular interest since the latter may be considered as determined directly from a comparison with known frequencies. The mobilities are distinctly of electrolytic order of magnitude, the swifter ions having a mobility slightly above that of simple electrolytic ions (the mobility of the K^+-ion is $b=6.8\times 10^{-4}$ cm²/(volt×sec) at 18°C), and the slower ions having somewhat smaller mobilities. In all events, these mobilities are far too small to be attributed to the positive holes. On the other hand, the values found here appear somewhat high for ions in a solid medium. This might very well be due to the lack of accuracy in the determination of β (see above).

With regard to the formation process the only conclusion which can be drawn with some certainty from the numbers of Table IV is that the importance of the swifter group of ions decreases by the application of the forming potential. The ratio c_1/c_2 decreases from 0.54 to 0.23, and this result persists even when the experimental constants A_1 and A_2 are determined in a slightly different way (within the limits which will not seriously affect the agreement between experiment and theory). The decrease of c_1 is easily interpreted; it indicates that the fastest ions are electrolyzed out first by the formation process.

On the whole, we can summarize the analysis of the experimental data by stating that the present theory coordinates the observed facts in a satisfactory way. In all events it shows the importance of the ionic component in the conductance of semiconductors like selenium, since the facts correlated here cannot be explained on the basis of a purely electronic theory.

The usefulness of the theory developed in the present paper is not limited to semiconductors. Either directly or with slight modifications, it is applicable to all cases where polarization effects are caused by the motion of ions. Work on such applications to electrolytes and to poorly conducting liquids is in progress and will be reported on separately.

ACKNOWLEDGMENTS

The author wishes to express his sincere obligation to the project engineer, Dr. L. W. Morris, and to Mr. A. McDonald for letting him utilize and publish here results obtained by them in the course of their work on the Signal Corps Project No. 152 (B). He also is indebted to Dr. L. W. Morris for many clarifying discussions, and to Dr. K. Lehovec for valuable suggestions.