JULY, 1940 REVIEWS OF MODERN PHYSICS VOLUME 12

Electrical Conduction and Related Phenomena in Solid Dielectrics

MILLARD F. MANNING, *Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania*

AND

MAURICE E. BELL, *Westinghouse Research Laboratory, East Pittsburgh, Pennsylvania*

TABLE OF CONTENTS

I. INTRODUCTION 215
II. EXPERIMENTAL METHODS 216
  A. Preparation of samples 216
  B. Current measurements 217
    1. Older methods 217
    2. Modern methods 218
III. METHODS OF DISTINGUISHING BETWEEN ELECTRONIC AND IONIC CONDUCTION 220
  A. General considerations 220
  B. Faraday's law and transference numbers 221
IV. IONIC CONDUCTIVITY 224
  A. High temperatures 224
  B. Low temperatures 225
    1. Polarization phenomena and dielectric anomalies 226
    2. Temperature dependence 230
V. DIELECTRICS IN ALTERNATING FIELDS 232
VI. LATTICE DEFECTS AND IONIC CONDUCTIVITY . . 235
VII. QUANTUM-MECHANICAL IDEAS 241
VIII. SPECIAL TOPICS 244
  A. Semiconductors 244
  B. Color centers and photoconductivity . . . 248
  C. Breakdown 252
BIBLIOGRAPHY AND REFERENCES 254

I. INTRODUCTION

DURING the past decade, considerable progress has been made in understanding the mechanism of electrical conduction in solid nonmetals. It is the purpose of this article to review the fundamental experimental results and the theoretical interpretations which have been suggested.

Substances are grouped roughly into metallic conductors, semiconductors, and insulators. At room temperatures metallic conductors usually have a conductivity of 10^3 ohm^{-1} cm^{-1} or greater. If the room-temperature conductivity of a substance is less than 10^{-10} ohm^{-1} cm^{-1}, it is classed as an insulator. The substances in the intermediate group are called semiconductors.

The mechanism of electrical conduction in solids may be either electronic or ionic, and various methods of distinguishing between the two types have been developed. In metals the conductivity is, of course, electronic, and may be, at room temperatures, as high as 10^6 ohm^{-1} cm^{-1}. There are also insulators and semiconductors which owe their conductivity to electronic transfer. Because of their symmetrical structures, pure samples of diamond and sulphur must owe their feeble conductivity ($\sim 10^{-18}$ ohm^{-1} cm^{-1}, or less) to electrons. In heteropolar or ionic compounds there is the possibility of ionic transfer. At ordinary temperatures the conductivity of rocksalt is about 10^{-16} ohm^{-1} cm^{-1} and is due to ionic transfer. There are also semiconductors whose conductivity is ionic. The highest ionic conductivity which has been observed is that of the high temperature form of AgI which has a conductivity of about 2 ohm^{-1} cm^{-1}.

The reason why some electronic conductors are metals and others are semiconductors or insulators has received a satisfactory explanation in terms of the quantum-mechanical theory of energy levels within a crystal. The exponential variation with temperature of the conductivity of semiconductors and insulators can also be interpreted in terms of this theory.

Ionic conductors show an exponential variation of conductivity with temperature. This has been satisfactorily interpreted in terms of a mechanism of ionic conductivity which postulates the existence of vacant lattice points or lattice defects. These lattice defects are also important in determining the properties of electronic semiconductors.

For most insulators, the observed current for a given temperature and field intensity does not stay constant but decreases for some time after

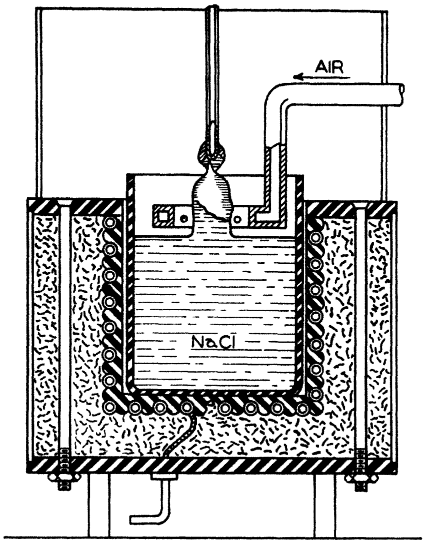

FIG. 1. Diagram showing formation of crystals from the molten salt.

the field has been applied. This behavior is responsible for the fact that under alternating fields the phenomena are dependent upon the frequency.

Besides the topics which have been mentioned in the previous paragraphs, we include a brief section on the conductivity of alkali halides colored by impurities, a section on photoconductivity, and a few remarks about electrical breakdown phenomena.

II. EXPERIMENTAL METHODS

In order to measure the conductivity of an insulator, one must have a suitable sample of the material to which electrodes have been applied in a manner which enables the experimenter to interpret his results as a property of the material. Any consistent group of data from which the specific conductivity can be calculated will serve. Having prepared the sample and electrodes, it is necessary to have either a sensitive instrument for measuring current or a resistance bridge of suitable range.

A. Preparation of samples

Many substances are available as powders which can be pressed into disks at pressures of a few thousand pounds per square inch. This method yields the size and shape desired, but has the disadvantage of allowing the formation of internal cracks. Under very high pressures some of the disks are translucent, or occasionally transparent. Ceramic materials are pressed into disks of the proper shape and fired in the usual way. Glass is cast in the form of a plate after which the surface may be ground.

Electrodes are applied to the sample in order to produce a uniform field over a region of known area. There are three general types of electrodes. Foils and plates of lead, platinum and other metals are frequently used at low temperatures. The use of petrolatum or other high resistance material to hold the electrode in place should be avoided. Air layers between electrode and specimen should likewise be carefully eliminated. Better contact, at the expense of somewhat greater effort is obtained with electrodes which are painted, evaporated or sputtered on the material. This method can be used only for nonporous specimens. A suspension of colloidal graphite in water may be applied with a brush in one or more coats. Such an electrode has a high lateral resistance and should be backed with a sprayed metal coating or a metal plate. A study of the properties of this type of electrode has been made by Hoch.[1]

Platinum can be applied either in the form of a paint or by evaporation or sputtering. Commercial paints are available consisting of a platinum compound in a suitable solvent.* The paint is applied with a brush, dried, then heated to a few hundred degrees in air to burn out the organic material. The specimen is then fired, producing a bright, firmly-adhering surface. Sometimes two or three coats are required to get sufficiently low surface resistance. This can be improved by applying sprayed copper or lead over the platinum. This type of electrode is excellent for measurements at high temperature, but can be used only for substances which will stand the necessary heat treatment. Platinum or other metals can be evaporated or sputtered to form a good electrode without heat-treating the sample.

Liquids have been used for electrodes since the earliest times. A description of the method

---

* One such paint is supplied by the Hanovia Chemical and Manufacturing Company, Newark, N. J., under the name "Liquid Bright Platinum, No. 05."

of employing mercury is given in the handbook of the American Society for Testing Materials.[2] Mercury, because of its vapor pressure, is suitable only for temperatures below 100°C. For certain higher temperature ranges, Wood's metal or molten cadmium can be used. In one of the historic measurements on glass, J. Hopkinson[3] used a solution of H_2SO_4 in water to make contact with the inside and outside of a flask.

The alkali halides, magnesium oxide, and many other substances can be prepared artificially by growing single crystals from the melt. This method enables one to avoid internal flaws and control the purity of the material, the crystal orientation, and the size. Kyropoulos[4] has grown large single crystals of sodium chloride by fusing the salt in an electrically-heated platinum crucible. The temperature is held constant just above the melting point while the crystal grows around a platinum tube closed at one end and cooled internally by a stream of air. As the crystal grows, the tube is raised slowly from the melt by means of a micrometer screw. Korth[5] and Hilsch[6] used a modification of the Kyropoulos method to produce single crystals of the alkali halides as large as 12 cm high and 10 cm in diameter. Single crystals of cylindrical shape, 30 cm long and 2 cm in diameter have been grown by H. Walther.[7] The crystal was cooled by streams of air from a perforated annular tube and simultaneously raised and rotated by a clockwork device. A schematic diagram of Walther's apparatus is shown in Fig. 1.

In studying a wax or other easily fusible substance, a cell described by Voglis[8] can be used. The arrangement consists of a condenser whose plates are supported by insulators which maintain the spacing accurately whether the wax is present or not. The supports are guarded in order that they do not affect the measured resistance. The wax is melted and poured into the condenser, the whole being cooled very slowly until the material solidifies. A similar cell can be used to measure the conductivity of liquids.

B. Current measurements

1. *Older methods.*—Early investigations of the conductivity of insulators were made by Jacques and Pierre Curie[9] and published in 1888–1889. These thorough and capable workers established the basic method for conductivity measurements in which the sample, in the form of a plate or slab, is placed between two metal electrodes. One electrode, which should be provided with a guard ring to eliminate surface conduction effects, is connected through an arrangement of shunts and capacitances to ground. The current through the sample may be measured either by the drop across a standard resistance or by the time required to charge a condenser of known capacity to a given potential. The potential drop is usually measured by an electrometer.

The other electrode is connected to a source of constant potential of known value, the other side of which is grounded. A switch is usually provided in order that the sample may be discharged when desired. The circuit is shown in Fig. 2.

For measuring relatively high currents the resistance method is used. The current I produces a drop IR which is measured by the electrometer E.

The current through many good insulating materials under ordinary conditions is so small that a measurable drop is not produced across available standard resistances. For example, the conductivity of ordinary glass at room temperature is approximately 10^{-14} ohm^{-1} cm^{-1}. This means that a specimen of 10 cm^2 area and 1 mm thickness with 10 volts potential difference would pass a current of only 10^{-11} ampere, producing a drop of only 0.1 volt across a 10^{10}-ohm resistor, which is the largest reliable resistance that can be made easily. In this case it is often better to charge a small high quality condenser and measure the time rate of increase of its potential. The current is equal to $C(dV/dt)$, where C includes the capacity of the connections and electrometer as well as that of the condenser itself.

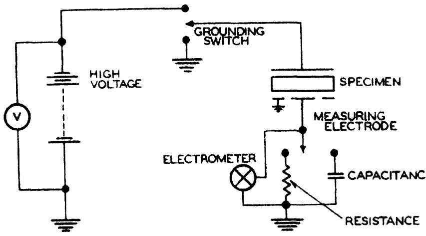

Fig. 2. Simple circuit for measuring the conductivity of dielectrics.

The development of the high sensitivity D'Arsonval galvanometer provided a very direct means of measuring the conduction current in an insulator. The galvanometer, combined with an Ayrton shunt to provide proper damping and changes in sensitivity, is connected between the guarded electrode and ground. In this way currents as small as 10^{-10} ampere can be measured accurately, but the method has the disadvantages of limited sensitivity and inherently slow response.

These two methods and their variations have been the basis for most insulation-resistance measurements in the past.

2. *Modern methods.*—The quadrant electrometer as modified by K. T. and A. H. Compton[10] has a much higher sensitivity than the older form. Values of 5000 mm per volt are not unusual. The extremely small currents in dielectrics at low voltages can be measured by such an electrometer in combination with a high resistance or a condenser of small capacity. A null method is ordinarily employed. The drop across the standard resistance is balanced by an e.m.f. applied between the standard and ground. The electrometer is used merely to indicate when the high side of the resistor is at ground potential and thus need not be linear in response nor even calibrated. The current equals the balancing e.m.f. divided by the resistance of the standard. The specimen and insulated parts of the circuit can be very effectively guarded because they are at ground potential.

The two apparent disadvantages of this method are the slow response of the electrometer and the manual effort required to apply the balancing e.m.f. to the resistor. Both objections are very serious when we are attempting to measure the conduction current at short time intervals after closing the circuit. Recently, however, vacuum-tube electrometer circuits have been developed which, when used in conjunction with rapid-response instruments, such as a string galvanometer or cathode-ray oscillograph, can measure small currents in 1/1000 second or less. The second difficulty has also been solved by a vacuum-tube circuit employing the new inverse feedback principle. In this circuit the signal is amplified and applied with opposite polarity to the input side. The amplifier thus automatically provides its own balancing e.m.f. which may be measured and recorded directly.

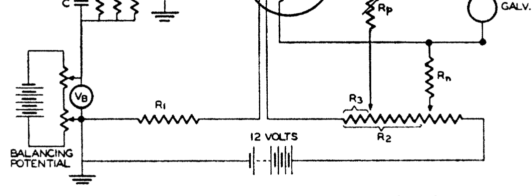

FIG. 3. The Barth electrometer circuit.

In the past decade, two special electrometer tubes have been developed; the G.E. Type F.P. 54,[11–13] and the Western Electric Type D-96475.[14] In these the object is to reduce the grid current to the lowest possible value. The grid current in these tubes under normal operating conditions is less than 10^{-15} ampere, compared with about 10^{-10} in an ordinary radio tube. Since currents of the same order as the grid current can be measured, these special tubes extend the range of measurements by a factor of 10^4 beyond that of a high sensitivity galvanometer. The response time is determined principally by the time constant of the grid circuit, RC, where R is the shunt value, and C will be 50×10^{-12} farad. When measuring, for example, 10^{-12} ampere, a resistance of at least 10^8 ohms should be used. For this case, the time constant would be 0.005 second. For the higher currents a much more rapid response can be realized. Compared with the period of 10 seconds or more for either an electrometer or a galvanometer of sufficient sensitivity, the period of the electrometer tube permits us to gain a factor of the order of a thousand in rapidity of response.

In the simplest circuit for an electrometer tube, the change in plate current is used to indicate a change in grid voltage. This arrangement suffers from instability, nonlinearity, and low amplification. The most prominent causes of instability are drift of filament and plate supply voltages and changes in tube characteristics with time. These difficulties can be diminished by connecting two identical tubes in a balanced circuit so that changes in battery voltages cancel in the output.[8] The signal is applied to one tube only, resulting in an unbalance between the two plate currents.

A change in the characteristics of only one of the tubes is not balanced, however, and appears as an instability in the output. This circuit has the added disadvantage of requiring two tubes.

A more satisfactory arrangement is a balanced circuit employing only one tube and using a resistance network to compensate for changes in battery voltage and tube characteristics. A good comparative discussion of the various balanced circuits has been given by Penick.[14] He shows stability characteristics for the Barth circuit, which is considered most generally useful of those he investigated. As battery voltage is varied, deflection of the galvanometer in the plate circuit decreases, and reaches a minimum at a particular battery voltage. When the circuit is operated in the neighborhood of this minimum, first-order effects of battery voltage fluctuations and changes in filament emission are eliminated. A very stable circuit results whose zero-point drift is negligible. Figure 3 shows the Barth circuit as employed for measuring the very small conduction currents in insulators.

The unguarded electrode of the specimen is connected to a source of known potential *HV*. The measuring electrode is provided with a guard ring which is grounded to eliminate surface leakage. A group of high resistances is arranged so that any one may be connected to the measuring electrode. The amplifier is used to indicate when the measuring electrode is at ground potential.

In measuring a current, the amplifier is first adjusted to read zero when the grid circuit is grounded. A resistor which will produce a convenient potential drop is then placed in series with the specimen, the ground is removed and a balancing e.m.f. is applied to restore the measuring electrode to ground potential. The current is equal to the balancing e.m.f. divided by the standard resistance in series with the specimen. This method is quite accurate and capable of high sensitivity—measurements of 10^{-15} ampere to within five or ten percent are possible. It is tedious, however, and not readily adapted to measuring rapidly-varying currents such as we find in a dielectric immediately after application of voltage.

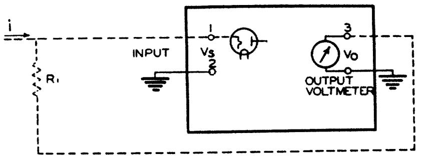

FIG. 4. Feed-back amplifier for measuring current.

In 1936, Vance[15] described a circuit characterized by good stability, rapid response, and high inherent accuracy. The principle employed is called inverse or negative feedback because the output signal is used as a balancing e.m.f. applied to the input resistor, so that the signal voltage V_s is equal to $V_i - V_0$, where V_i is the voltage to be measured and V_0 is the output voltage (see Fig. 4). For an amplifier with voltage gain G, a signal V_s produces an output V_0 equal to GV_s.

Hence,

$$V_0 = V_i\{1 - 1/(1+G)\}. \qquad (1)$$

By making G large (a value of 1000 is readily obtainable), V_0 can be made as nearly equal to V_i as desired.

This amplifier may be used to measure current by connecting a resistor, R_i, across terminals 1 and 3. The current to be measured produces a drop V_i across R_i which is practically equal to V_0.

The response of the circuit is essentially linear as long as the input resistance obeys Ohm's law. By Eq. (1) we see that a change in G due to a nonlinearity of the amplifier itself or to a change in the amplifier characteristics with time will have a negligible effect on V_0 as long as G remains large.

Let us consider the effect of inverse feedback upon the time constant of the input circuit. For an ordinary electrometer circuit the time constant is equal to RC, where C is the capacity of the measuring electrode, connections and input circuit of the electrometer, and R is the standard resistance. In the feed-back amplifier the potential of the input circuit changes only by V_s which is equal to $(1/G)\,V_0$. This is equivalent to decreasing the input resistance by a factor G. Therefore, for the inverse feed-back amplifier, the time constant will be approximately RC/G, which is much smaller than for an ordinary electrometer or straight amplifier circuit. At first this result seems paradoxical, but it is less surprising, perhaps, if we realize that most of the charge necessary to establish the voltage across R_i comes from the amplifier, actuated by the signal V_s.

Besides the small time constant there are two other advantages of the inverse feed-back circuit. The power output can be made large enough to operate a recorder and it is possible to guard the electrometer circuit effectively because the difference in potential from ground is only $(1/G)\,V_i$.

III. Methods of Distinguishing Between Electronic and Ionic Conduction

A. General considerations

One of the important questions to be settled before the phenomena of electrical conduction in solid salts can be understood is whether the conductivity is due to the motion of electrons, positive ions, negative ions, or to two or more types simultaneously. Actually, it is probable that there will be some contribution from all three types, but for a given material at a particular temperature it is usually found that one type predominates.

There are certain phenomena which usually accompany electrolytic conduction and which can usually be taken as evidence that a portion, at least, of the current is carried by ions. One of these phenomena is polarization. As was mentioned in the introduction, when a potential difference is applied to a substance like rocksalt, the current does not stay constant, but decreases with time. At ordinary temperatures, the change continues for minutes or even hours and the steady-state current may be less than 1/1000 of that which flowed when the potential difference was first applied. Another phenomenon which is sometimes difficult to investigate, but which is usually definite evidence for electrolytic conduction, is the actual appearance of material at an electrode. Since the resistivities of the materials investigated are usually very high, the current densities are usually so low that the amount deposited will be small unless the time is long. The difficulty with using a long time is that an appreciable amount of material may be lost by vaporization. In most cases when a metallic element is deposited from a solid by electrolysis it does not form a smooth film on the electrode, but instead forms local deposits. At these points the field intensity is greater, resulting in further deposition at these same places and eventually dendritic bridges are formed from one electrode to the other. These bridges then carry the current and there is no further electrolysis. In transparent substances the progress of the dendrite formation can be followed visually, but in opaque substances this phenomenon manifests itself only by a decrease in resistance. Dendrites can also be formed at the anode where the nonmetal is deposited. For instance, in Cu_2S, S is evolved at the anode and then reacts with the Cu_2S to form CuS which is a good conductor and can form dendritic bridges much as a metal does at the cathode.

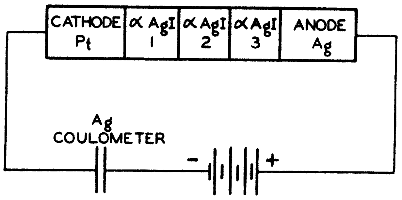

FIG. 5. Cell for transference number measurements in α AgI.

Before the development of the quantum-mechanical theories of the solid state, it was believed that a large increase in conductivity upon melting was definite evidence for ionic conduction. Many ionic crystals, for instance NaCl and KCl, show over a thousand-fold increase in conductivity upon melting. However, large changes in conductivity upon melting have also been observed by Norbury[16] in a group of substances in which the conductivity is electronic. These substances (Mg_2Sn, Mg_2Pb, Mg_2Si, Mg_2Ge, Li_2S, Na_2S, Cu_2S, Cu_2Se, Be_2C, and others) crystallize in the CaF_2 (fluorspar) type of structure. In the solid state, when the compositions are exact, they are almost insulators. In the molten state, however, they show an electronic conductivity which is about the same as that of molten tin. The explanation of this behavior in terms of their electronic structure has been given by Mott and Jones[17] and will be referred to again in Chapter VII.

One of the best indicators that a portion of the conductivity is electronic is obtained from studies of the Hall effect. If a magnetic field is applied perpendicular to an electronic current, a potential difference appears perpendicular to the plane of the magnetic field and the current. Even for an electronic conductor the Hall effect sometimes vanishes in a certain temperature range,[132] but it is never found for a purely ionic conductor.

When a solid is raised to a high temperature, charged particles are emitted. The nature of these particles, electrons, positive ions or negative ions, gives indirect evidence concerning the predominant current-carrier at a given temperature.[19]

However, the best method of determining what proportion of the total charge is transferred by each type of carrier is by studies of the applicability of Faraday's law of electrolysis, which should, of course, be valid for solid ionic conductors as well as for fused salts and water solutions. This method was developed by Tubandt[22] and has been applied by him and his associates to a large number of substances. The methods used and the results obtained previous to 1932 are described in detail in Tubandt's article in the *Handbuch der Experimentalphysik*, Vol. XII.

Since the results obtained are significant for an understanding of conduction in solids and since certain difficulties have developed in the application of the method, it will be described in the following section.

B. Faraday's law and transference numbers

The first evidence for ionic conduction in solids was obtained by Warburg[20] who worked with glass. The first quantitative test was made by Tubandt[21] on the halides of lead and silver. Measurements of electrolytic deposition from solids are difficult because dendritic deposits frequently form at the cathode and cannot be separated from the material and weighed. Fortunately, it turns out that some of the halides, particularly α AgI, form a plate of cohering material at the cathode. Tubandt has made use of this property by inserting a block of α AgI, or similar material, next to the cathode to prevent dendritic formation. Sometimes it is necessary to use protecting layers at both electrodes and occasionally a large number are required.

As an illustration, the application of the method to α AgI is described. The specimens used are cylinders about a centimeter in diameter and a centimeter high, formed by subjecting the powder to a high pressure. The schematic arrangement[22] is shown in Fig. 5.

Fig. 6. Cell for PbI_2.

The Ag anode loses a weight equal to that deposited in the silver coulometer. The same amount is deposited as a separate layer at the cathode, and the weight of each of the AgI cylinders is unchanged. There is a straightforward interpretation of these data. There can be no electronic conduction through the cylinders or the amount of Ag transferred through them would be less than that deposited in the coulometer in the external circuit. If there were any charge transferred by motion of I^- ions, the AgI cylinder adjacent to the cathode would lose weight and that adjacent to the anode would gain weight. The transference number for a particular type of ion is defined as the ratio of the quantity of electricity transferred by that type of ion to the total quantity transferred. In this case it is unity for the Ag^+ ion and negligible for the I^- ion.

In determining the transference numbers for PbI_2 a more elaborate set-up had to be used. This can be represented schematically, as shown in Fig. 6.

The data for this case as given by Tubandt[23] are:

change in weight of cathode+(1)	−0.0035	gram
" " " " (2)	0.0000	"
" " " " (3)+(4)	0.0012	"
" " " " (5)	0.0000	"
" " " " (6)+(7)	−0.0012	"
" " " " (8)+anode	0.0035	"
Ag in coulometer	0.0106	"

The first item and the next to the last item are the chlorine equivalents of the silver deposited in the coulometer. The third and fifth items are the ones which give us the information desired. If only the I^- ions were mobile, the combination of cylinders listed as the third item would show a decrease in weight of 0.01247 gram (the I^- ion equivalent of 0.0106 gram of Ag). If only the Pb^{++} ions were mobile, this same combination of cylinders would show an increase in weight of 0.01017 gram (the Pb^{++} ion equivalent of 0.0106 gram of Ag). Since neither of these is observed, we must have both types of ion mobile and the proportion of the current carried by each can be determined by setting up a relation for the weight gained by (3)±(4). If x represents the proportion of the current carried by the Pb^{++} ions and $1-x$ the proportion of the current carried by the I^- ions, we then have:

$$x(0.01017)+0.0035-(1-x)(0.01247)=0.0012.$$

The first term represents the weight carried into the block considered from the right by Pb^{++} ions. The second term represents the weight carried into the block from the left by Cl^- ions. The third term represents the weight carried out of the block toward the right by the I^- ions. If this equation is solved for x, the result is

$$x=0.45=\text{transfer No. for } Pb^{++},$$
$$1-x=0.55=\text{transfer No. for } I^-.$$

This analysis is valid only if there is supplementary evidence indicating that there is no electronic conduction.

An example of a mixed conductor is given by Tubandt.[24] The arrangement of the test samples for β Ag_2S can be represented diagramatically as shown in Fig. 7.

The data obtained at 170°C as reported by Tubandt were:

change in weight of cathode +(1)	0.3475	gram
" " " " (2)	0.0000	"
" " " " (3)+(4)	−0.0750	"
" " " " (5)	0.0000	"
" " " " (6)	0.0000	"
" " " " (7)	0.0680	"
" " " " (8)+anode	−0.3474	"
Ag in coulometer	0.3475	"

If there were pure ionic conduction throughout with only the Ag^+ ion mobile, the AgI cylinder (7) would not experience any net change in weight. Since there is an increase in weight observed, it must mean that while positive charge is being carried into it from the right, this must be balanced partially by negative charge flowing in from the left. The fact that the weight of (6) remains unchanged indicates that there is no transfer of charge by S^{--} ions but only by Ag^+ ions and electrons. The third item and the sixth should be numerically equal. Assuming that the 0.0680 is the correct value, we conclude that 0.0680/0.3474 or 19 percent of the quantity of electricity passing through β Ag_2S is carried by electrons and the remainder by Ag^+ ions.

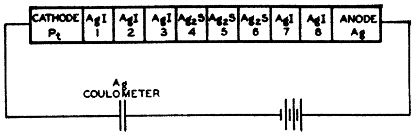

FIG. 7. Cell for β Ag_2S.

For substances in which both types of ion are mobile (PbI_2, and some of the alkali halides), the transference numbers change with temperature. The cuprous halides are mixed conductors with electronic conductivity at low temperatures and ionic conductivity at high temperatures. The low temperature or β form of Ag_2S is a mixed conductor and, according to Tubandt and Reinhold,[24] the portion of the current carried by electrons increases with the current density. The high temperature or α form of Ag_2S was originally believed to be an ionic conductor, but more recently it has been concluded that the conductivity is electronic.

Transference measurements on α Ag_2S (stable above 179°C) made by Tubandt, Eggert and Schibbe[25] indicated that all of the charge was transferred by Ag^+ ions. As Jost[31] has pointed out, it is difficult to reconcile this interpretation with the high numerical value of the conductivity and the results of diffusion measurements. The observed value of the conductivity is about 500 $\text{ohm}^{-1}\,\text{cm}^{-1}$ and even with very liberal estimates of the number of mobile ions, a conductivity as high as this could be obtained only if the mobility exceeded that of ions in water solutions.

Measurements of the diffusion coefficient of the mobile ion also give information about ionic conduction. There is a relation between the coefficient of diffusion and the mobility of an ion which is known as the Einstein relation:

$$D=(RT/N)B, \tag{2}$$

where D is the coefficient of diffusion, R is the universal gas constant, T is the absolute temperature, N is Loschmidt's number, and B is the mobility (constant of proportionality between the drift velocity and the electric force). Until recently it has been possible to measure the diffusion constant directly only for Pb salts[26] where there is a radioactive isotope (ThB) which can be used. Artificially radioactive atoms[27] have recently been used in the study of self-diffusion in

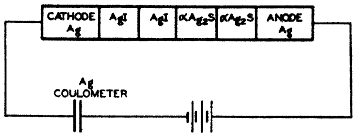

FIG. 8. Arrangement for α Ag_2S.

metals and they should be equally helpful in studying ionic conduction.

Tubandt, Reinhold, and Jost[28] and others used indirect method of measuring diffusion constants. They added a small amount of another salt with one element in common and followed by chemical means, or by weighing, the diffusion of the foreign element. Wagner[30] has carried out a thermodynamic analysis which indicates how the data so obtained can be used to obtain the coefficient of self-diffusion. When Tubandt, Reinhold, and Jost used their data on the diffusion of Cu^+ ions in α Ag_2S, to compute the ionic conductivity of α Ag_2S, they obtained only about 2 ohm^{-1} cm^{-1} which is very much less than the observed value.

In a review of the evidence concerning the nature of the conductivity, Jost[31] suggested the following considerations: (1) The numerical magnitude of the conductivity is at least fifty times greater than the value calculated when reasonable values of the ionic mobility are assumed; (2) The conductivity is several hundred times that calculated from diffusion data; (3) The material shows a definite Hall effect;[29] (4) The conductivity decreases with temperature. Jost pointed out that if it were not for the transference data, these considerations would suggest that the conductivity was mainly electronic. Since he saw no reason to doubt the transference data, he sought some anomaly in the diffusion data. To meet the objection raised by item (3), Tubandt[32] even suggested that a good ionic conductor might show a Hall effect. The final answer to the puzzle came from Wagner[18] who suggested that the conductivity was primarily electronic and that the discrepancy was in the transference measurements. The reason which he suggested for the false results of the transference measurements was as follows.

The experimental arrangement[25] used in measuring the transference numbers is shown in Fig. 8. Suppose that the conductivity in α Ag_2S is entirely electronic. During the passage of one faraday, one equivalent of Ag is carried through the AgI and deposited as a compact layer at the surface of the cathode. At the interface between the α Ag_2S and the AgI there will be a transfer of Ag^+ ions to the AgI. If there is only electronic conduction in the Ag_2S, this will leave free S at the interface. There is thus free S on one surface of the Ag_2S and free Ag at the other surface. Since Ag and S form a stable compound, the Ag at the anode will combine with the S if there is any mechanism by which they can be brought into contact. Such a mechanism is provided by the fact that the diffusion coefficient of Ag in Ag_2S is high. This problem has been discussed by Wagner[33] in connection with a somewhat different problem—that of a metal in equilibrium with the vapor of a nonmetal, such as O, S or a halogen. In many cases of this sort a layer of the compound will form at the surface and its continued increase in thickness must be due to some sort of diffusion process. Wagner has investigated the conditions under which this transfer will take place and the factors affecting the rate of transfer. It is obvious that one of these is the thickness of the layer. At least two methods suggest themselves for testing Wagner's hypothesis that diffusion is responsible for the transfer of Ag through the Ag_2S. Since the rate of transfer of Ag by diffusion for a definite specimen at a definite temperature has an upper limit, departures from Faraday's law might be produced by working at higher current densities. This experiment was performed by Jost and Rüter[34] who found that in all cases the amount of Ag lost by the anode was less than that plated out in the coulometer and that the discrepancy was greater at higher current densities. Another test of Wagner's hypothesis is to increase the thickness of the Ag_2S layer, or more conveniently to increase the number of Ag_2S cylinders. This experiment was performed by Wagner[35] who used fifteen cylinders (total length 8.4 cm) of Ag_2S in series and found that the weight lost by the Ag anode was only one-fourth of that plated out in the coulometer.

Further evidence brought forward by Wagner in favor of his hypothesis is the great decrease in conductivity produced by excess S. When the conductivity of an ionic conductor is below 10^{-6} or 10^{-8} ohm^{-1} cm^{-1} it is increased appreciably by

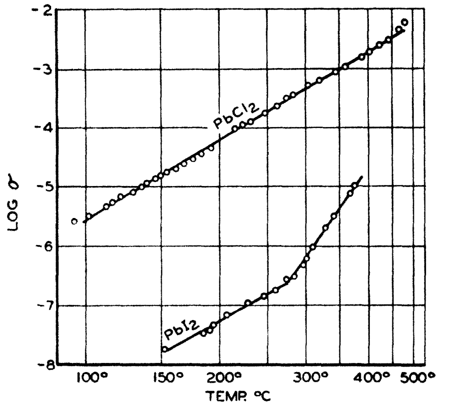

FIG. 9. Conductivity of $PbCl_2$ and PbI_2.

the presence of impurities.[36] However, ionic conductivities of the order of 1 ohm^{-1} cm^{-1} are little affected by impurities, but electronic semiconductors may change their conductivity by several powers of 10 upon departures from exact stoichiometric proportions. Other evidence which argues for predominantly electronic conduction is the fact that a cell with Ag and S electrodes and α Ag_2S for "electrolyte" shows a much smaller e.m.f. than is computed from the Gibbs-Helmholz equation, and the heat of formation of Ag_2S.*

The anomalies found for α Ag_2S have been discussed in detail because they indicate that the data obtained from transference numbers is not always dependable. In particular, in the light of modern knowledge of electronic semiconductors and their great sensitivity to exceedingly small amounts of impurity, the proportion of the current carried by electrons at a definite temperature should be regarded as a characteristic of the particular specimen rather than of the material. It is probable that a systematic study of transference numbers using artificially radioactive atoms as tracers would help to clarify some of the less-understood features of ionic conduction.

IV. IONIC CONDUCTIVITY

A. High temperatures

At temperatures near the melting point the conductivity of an ionic crystal is much less ambiguous than at lower temperatures. The conductivity may vary somewhat from one sample to another, but usually by much less than an order of magnitude. At these temperatures polarization phenomena are negligible. The conductivity is usually high enough so that it can be measured by the galvanometer method previously described.

The temperature dependence of the conductivity of a number of ionic conductors has been studied by Phipps and co-workers,[37, 51–53] Seith,[40] Lehfeldt,[41] Kassel,[68] Rochow,[38] Brennecke[39] and others. It is found that over a restricted temperature range the conductivity can be expressed by a formula of the Van't Hoff[42] form

$$\sigma = A \exp(-B/T), \tag{3}$$

where σ is the conductivity, T is the absolute temperature, and in the high temperature region A and B are constants for a given sample. B is usually the same for different specimens of the same material, but A may vary slightly from sample to sample. The values of A vary from about 10^4 to 10^7 ohm^{-1} cm^{-1} for different materials. The values of B are about 10^4 $(°C)^{-1}$. A formula such as (3) predicts a very rapid variation with temperature at low temperatures. For instance, if B equals 10^4, the conductivity changes by a factor of $e^5(\simeq 150)$ between 400°K and 500°K. This means that in any experimental determination of conductivities, it is necessary to be very careful about temperature control.

The validity of Eq. (3) is tested by making a plot of log σ *vs.* $1/T$. For some salts ($PbCl_2$, TlCl, TlBr, TlI, α AgI, for instance[40, 41]) the slope of the log σ *vs.* $1/T$ curve is constant over a considerable range of temperature, showing that the Van't Hoff relation is satisfactory for these substances. For other salts, particularly PbI_2, the log σ *vs.* $1/T$ curve can be approximated by two straight lines and a transition region[40] (see Fig. 9). This suggests representing the conductivity by an expression consisting of two Van't Hoff terms

$$\sigma = A_1 \exp(-B_1/T) + A_2 \exp(-B_2/T). \tag{4}$$

On a logarithmic plot the first term will determine the slope in the low temperature region and the second term will determine the slope in the high temperature region; the plot of

---

* As an illustration of the higher electronic conductivity of α Ag_2S than of β Ag_2S, Hochberg found that there is a corresponding increase in the thermal conductivity of α Ag_2S compared with β Ag_2S. It is well known that mobile electrons contribute greatly to the thermal conductivity.

log σ *vs.* $1/T$ consists of two straight lines and a transition region.

Phipps, Lansing and Cook[37] offered an interesting suggestion concerning the mechanism which is responsible for the two terms in the formula for the conductivity. They suggested that one member represented the contribution of the cation to the current and the other the contribution of the anion. Seith[40] pointed out that if this were true, $PbCl_2$ which is known from transference measurements[23] to be a unipolar conductor should show a straight line on a log σ *vs.* $1/T$ curve, and PbI_2, which is known to be a bipolar conductor should exhibit a kink in the log σ *vs.* $1/T$ curve. The graphs reproduced in Fig. 9 indicate that this is definitely true. Transference numbers calculated from the relative magnitudes of the two terms in Eq. (4) checked very closely those found by transference measurements. The transference numbers calculated by Smekal[43] from Seith's[40] data were 0.65 and 0.35 at 290°C, against 0.67 and 0.33 as found by Tubandt[23] for the same temperature.

However, PbI_2 is the only case for which the suggestion of Phipps, Lansing, and Cook[37] has been checked. In some of the alkali halides the transference numbers have been measured as a function of the temperature by Tubandt, Reinhold, and Liebold.[23] Their values do not check those computed by using a formula like (4) in which the constants have been adjusted to fit the experimental data on conductivities. A reason for this is given in Chapter VI.

An interesting study of diffusion in lead salts has been carried out by Hevesy and Seith.[26] They made use of the fact that ThB, which emits α-particles,* has the same electronic structure as Pb and hence should show the same behavior in diffusion measurements as Pb. The method of studying the diffusion of ThB in a Pb salt was to coat the surface of the Pb salt with the corresponding salt of the ThB and then increase the temperature. The emitted particles can penetrate only a thin layer of the salt and hence as the thorium atoms diffuse inward the observed activity decreases much more rapidly than that of ThB under normal conditions. Since both the conductivity and diffusion are given by an expression of the form $\exp(-B/T)$ we can obtain information about the B from the diffusion data. In $PbCl_2$ they found that B was so large compared with that for the Cl^- ion that no appreciable contribution of the Pb^{++} ion to the conductivity was to be expected. For PbI_2 they obtained a value of B for the Pb^{++} ion which agreed very closely with that found by conductivity measurements.

A table of values of A_1, B_1, A_2, B_2, etc. as given by Tubandt[44] is given in Table I.

B. Low temperatures

At lower temperatures (say below about half of the melting temperature on an absolute scale) the behavior is much more complicated than in the high temperature region. The effect of small amounts of impurities is much more marked and the conductivity depends upon the previous thermal history of the particular specimen.

TABLE I. *Constants for the conductivity of solid salts. The first four columns are from calculations by Smekal; the last two from calculations by Joffé.*

	A_1	B_1	A_2	B_2	B_1	B_2
NaCl	≤ 0.8	10,300	3.5×10^6	23,600	10,500	30,000
NaBr	0.5	9430	1.1×10^6	21,200	9200	22,000
NaI	0.023	6440	0.13×10^6	16,600	6300	16,000
PbI_2	2.95×10^{-4}	4140	8.0×10^4	14,720	4700	13,900
$PbCl_2$	6.55	5535	—	—	5600	—
AgCl	(1.3×10^{-4})	(2140)	0.51×10^6	11,100	12,500	—
AgBr	(0.049)	(3100)	1.5×10^6	10,300	10,000	—
β-AgI	—	(≤ 2000)	3×10^6	9550	12,000	—
α-AgI	5.48	593	—	—	—	—
TlCl	—	—	1.6×10^4	10,300	10,500	—
TlBr	—	—	5.1×10^3	9900	9800	—
TlI	—	—	5.0×10^2	8100	8100	—
$NaNO_3$	3.6×10^5	15,200	8.1×10^{16}	29,000	10,000	—
KNO_3	—	—	—	—	10,000	—

---

* Actually the α-particles are emitted by a disintegration product of ThB.

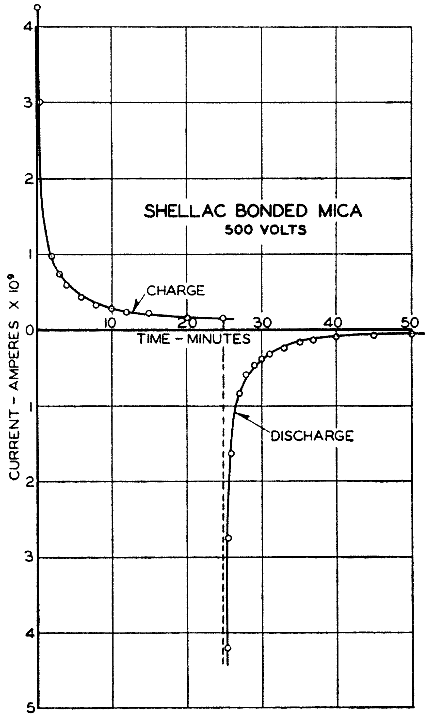

FIG. 10. Charging and discharging currents *vs.* time.

Properties of a specimen which show marked dependence upon the previous history are known as structure-sensitive properties, while properties which depend upon the chemical composition, temperature, pressure, etc., are known as structure-insensitive properties. Thus in the low temperature region, ionic conductivity is a structure-sensitive property, but at higher temperatures it is a structure-insensitive property. Smekal[45] has been particularly active in calling the attention of physicists to the structure-sensitive properties of matter.

In the low temperature region the current for a given specimen at a given temperature and for a particular applied field, changes with time. The rapid variation of the current, combined with the fact that substances like rocksalt have a conductivity of about 10^{-14} ohm^{-1} cm^{-1}, or less, at room temperatures, imposes great difficulties upon the measurements. It is only recently that current-measuring devices having proper sensitivity and speed of response have been available.

1. *Polarization phenomena and dielectric anomalies.*—The variation of the current in a dielectric with time was studied by Hopkinson[3, 81] and the Curies[9] more than fifty years ago. It was later studied by Tank[83] and Richardson,[46–48] and still later by Joffé.[49, 50] Recently a number of Russian and German investigators have worked on the problem.

The general phenomena are as follows: When a potential difference is applied to a dielectric the current changes with time as shown in Fig. 10.* It has not been possible to determine the current at very small time intervals after application of voltage, but all results indicate that it rises very rapidly toward $t=0$. At ordinary temperatures the steady-state value of the current may be reached only after several hours or even days and may be as low as 10^{-4} of that observed initially. If the specimen is short-circuited, it is found that a current, known as the discharging current, flows in the reverse direction. If the steady state has been reached before short-circuiting, the discharging current shows essentially the same variation with time as the charging current. At room temperature, for a poor conductor like rocksalt, the quantity of electricity which flows in the reverse direction is almost as large as that which flowed in the positive direction. If the steady state has not been reached, the quantity which flows during discharge is still approximately equal to the quantity which flowed during charge, but the initial current for discharge is equal to the difference of the initial and final currents during charge. (See Fig. 11.)

The discharging current manifests itself in the familiar "absorption of charge" experiment. If an ordinary condenser is charged and then discharged the time that the discharge key is closed is usually less than the time required for all of the charge to flow out of the condenser, so that repeated discharges without intervening charges will produce successively decreasing throws of a ballistic galvanometer. It was from this point of view that the problem was first studied by Richardson.[47, 48] He found that the dielectric constant depended upon the time of charging and the time of discharging in a manner which could be correlated with a variation of current with time like that indicated by Figs. 10 and 11.

Richardson assumed that the decrease of current with time was due to the building up of a

---

* The data for Figs. 10 and 11 were kindly supplied by Dr. L. J. Berberich of the Westinghouse staff.

counter-e.m.f. of polarization within the crystal. This polarization e.m.f. could be detected by measuring the potential difference across the crystal after the external field had been removed, or by decreasing the applied field and noting that there was a much greater decrease in the current. Richardson believed that he could determine the polarization e.m.f. by the following process: The charging current was allowed to flow for a certain length of time and the external voltage was suddenly decreased to a new value. If the new voltage were just equal (and opposite) to the polarization e.m.f., the current would be zero for a short time after switching. The values of the polarization voltage P were found to increase with time very rapidly immediately after application of an external voltage, but later increased less rapidly and eventually became constant. Richardson also found that the limiting polarization was proportional to the applied voltage. Later experiments have indicated deviations from this result for sufficiently high fields.

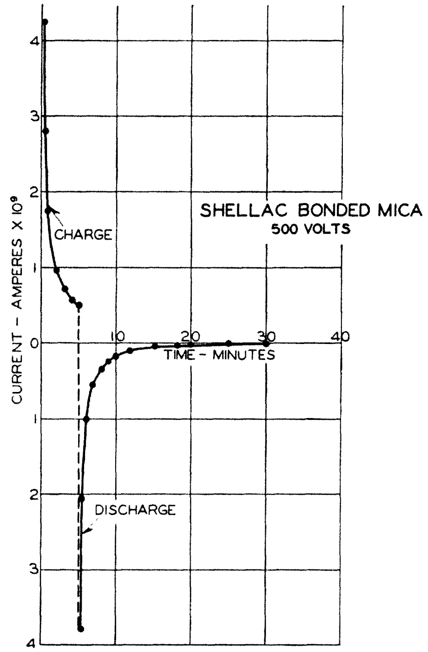

FIG. 11. Partial charging and discharging current *vs.* time.

The behavior depicted in Figs. 10 and 11 is a special case of a more general relation, known as the superposition principle,† which states that the change in current produced by a given change in applied voltage depends only upon the change in voltage. It does not depend upon the voltage previously applied or the current already flowing. Goldhammer[56] reports that the superposition principle was not applicable in studies which he made upon the conduction of electricity in quartz. This is probably because the conductivity of his sample was influenced by the presence of impurities which were preferentially removed by the passage of the current.

The question of whether or not insulators obey Ohm's law is complicated by the difficulty of knowing what current to use in the equation $i=\sigma VA/d$. Frequently the steady-state current has been used. Joffé[49, 50] has proposed that the conductivity should be found from the relation

$$i=\sigma(V-P)A/d, \tag{5}$$

where V is the applied voltage and P is the e.m.f. of polarization. According to Joffé there are three possible methods of finding σ from this equation. They are:

(1) At a very short interval of time after the external voltage is applied, the polarization e.m.f. is negligible and the ratio of the initial current to the applied voltage determines the conductivity.

(2) If we produce a sudden change in the external voltage from V_1 to V_2 and the current changes from i_1 to i_2, then, if the superposition principle is valid, we have:

$$\sigma=\frac{i_2-i_1}{V_2-V_1}\frac{d}{A}.$$

(3) If we can measure P and i at some instant, we can use Eq. (5) as it stands. This method is particularly simple to apply when the steady state has been reached.

The disadvantage of the first method is the difficulty in measuring the current at short intervals after the switch is closed. The second method of determining the polarization and conductivity suffers from the same disadvantage as the first. The chief reason for its use is to study the conditions under which the superposition principle is valid. A number of methods of measuring the polarization have been developed. Before discussing these, we shall describe some of the attempts which have been made to measure the

---

† A more complete discussion of the superposition principle is given in Chapter V.

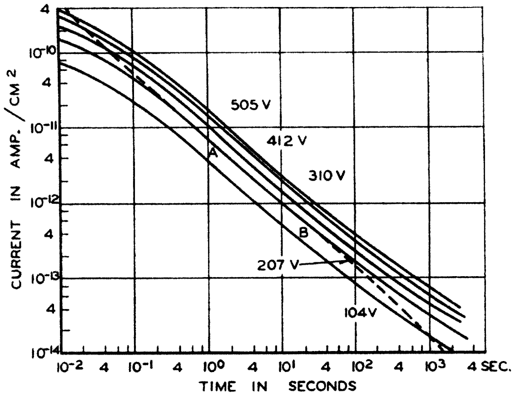

FIG. 12. Charging currents in mica.

current at short time intervals after closing the switch.

Goldhammer[57, 58] and Salessky[59] have developed ingenious methods of measuring the current at various time intervals after the external circuit is closed. The essential feature is a set of rotating switches which first close the external circuit; at an arbitrary time later, connect the electrometer across the standard resistance for a definite interval of time; and still later, open the main circuit. The methods of Goldhammer have been criticized by Quittner.[60] Methods have been developed by Hippauf and Stein[61] and by Voglis[8] which use electron-tube circuits and give satisfactory results. We shall discuss the work of Voglis in some detail.

He used a balanced-tube type of d.c. amplifier which actuated an electromagnetically-controlled string galvanometer. The deflection of the string was magnified optically and recorded photographically. Satisfactory records were obtained down to 10^{-2} second after closing the switch. The substances studied were wax, glass, mica and "Kerafar U." The results on mica are typical and will be outlined here.

Previous to Voglis, it had usually been assumed that the variation of current with time could be represented by the formula

$$i = At^{-n}, \tag{6}$$

where A is a constant depending upon the material and the voltage, and n is independent of the voltage and characteristic of the material. n is usually less than one. The applicability of this formula can be tested by making a plot of log i *vs.* log t. If formula (6) is applicable, the plot is a straight line with slop equal to $-n$. As Fig. 12 shows, the plot is a straight line between about 10^{-1} sec. and 100 sec., but departures are observed for values of t outside this range. Voglis suggested that the i *vs.* t curves could be represented, down to the smallest times observed, by the formula:

$$i = A(t+\tau)^{-n}. \tag{6a}$$

To find τ, prolong the straight portion of the curve (Fig. 13) and find the horizontal displacement necessary to shift any point on the observed curve to a position on the straight line. The corresponding time interval is equal to τ. To test the formula, determine the value of τ required to shift each point to the straight line. If Eq. (6a) is to represent the data, the values of τ so found must be sensibly constant. The fact that in Fig. 13, all the displaced points lie on the straight line, shows that τ is constant for a particular set of data and that the formula is satisfactory. For this case τ is 2.1×10^{-2} sec. and n has the value 0.87.

In Fig. 12, all of the log i *vs.* log t curves are plotted for the same specimen, but for different voltages. It is interesting to note that since the curves are parallel, only A and not n is a function of the applied field. The dotted curve marked 207 volts is for a *discharging* current which shows a departure from the t^{-n} formula at large values of t. The departure is in such a direction as to keep finite the total quantity of electricity which flows during discharge.

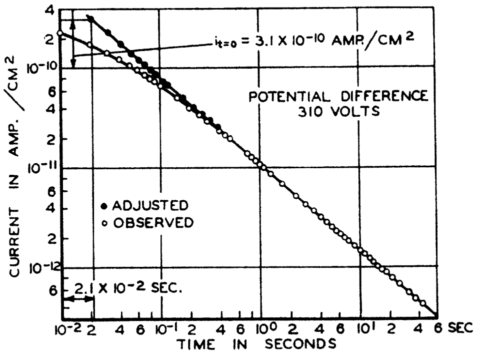

FIG. 13. Diagram illustrating method of finding τ.

Theoretical studies of the process of formation of the polarization e.m.f. and its effect on the current have been carried out by Schumann[62] and by Jaffé.[63] Schumann obtained a formula which is more complicated than that proposed by Voglis, but which shows many of the same general characteristics.

There are a number of methods of measuring the polarization e.m.f. Joffé and others made studies by inserting probes, and by a neat method of slicing off thin sheets and measuring the potential difference across each of them. When this is done, it is sometimes found that the polarization e.m.f. is distributed uniformly through the specimen, but sometimes it is concentrated near one of the electrodes. In the latter case Joffé[64] and co-workers have found that a drop of hundreds of volts can be concentrated in a region only a few thousand angstroms thick. Smekal[65] has claimed that the concentrated polarization e.m.f.'s are due to impurities and that the linear distribution is the normal case.

Another method of determining the polarization e.m.f. has been used by Beran and Quittner.[66] They allow the current to flow until a steady state has been reached and then suddenly decrease the applied voltage from V to V'. If V' is equal to the value of P before the potential was decreased, no current flows immediately after the switching operation. This method has the advantage of being essentially a null method. The method of Beran and Quittner has been criticized by Goldhammer.[57]

The steady-state polarization e.m.f. has been measured by a number of different authors[56, 57] for different materials as a function of applied voltage and temperature. Hochberg and Joffé[67] studied the polarization in $NaNO_3$ and found that it was concentrated near the electrodes. The dependence upon applied voltage and temperature is indicated by Fig. 14. The data obtained by other investigators for other materials are similar. In general, the steady-state polarization increases with potential but approaches a limit at higher values. The maximum value of P for a given applied voltage is smaller for higher temperatures. The general run of the data indicates that the polarization reaches its equilibrium value in a shorter time at higher temperatures. At sufficiently high temperatures no polarization has been found.

Kassel[68] determined the steady-state polarization e.m.f. for samples that had been aged at different temperatures. He measured the steady-state conductivity, and assumed that the true conductivity could be found by extrapolation from high temperatures. He assumed that the discrepancy was due to polarization. Wenderowitsch and Drisina[69] have claimed that this extrapolation is not justified. Later (Chapter VI) we shall present arguments which indicate that the objection of Wenderowitsch and Drisina is probably sound. They also investigated the steady-state polarization in rocksalt and found that tempering at a higher temperature (but for only four hours, whereas Kassel used 10 to 12 hours) caused a decrease in polarization. The highest temperature which they used was 350°C, and it may be that at this temperature the aging period was not long enough to establish thermal equilibrium.

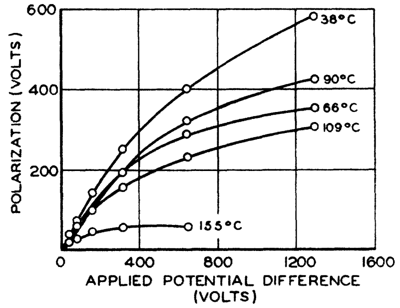

FIG. 14. Variation of polarization e.m.f. with temperature and voltage.

The effect of high field intensities on the conductivity was first investigated by Poole.[77] He found an increase of steady-state conductivity with field strength which could be represented by the formula

$$\sigma_d = a \exp(bE). \tag{7}$$

Joffé[78] has claimed that only the steady-state conductivity increases with field strength and that what he calls the "true" conductivity σ_w is independent of field strength. Joffé assigned the difference to polarization effects. It may be well to emphasize that a polarization which was proportional to the applied voltage would not yield any variation of steady-state conductivity with field strength. However, as indicated by Fig. 14, the polarization e.m.f. increases less rapidly than the applied voltage and the effective field will

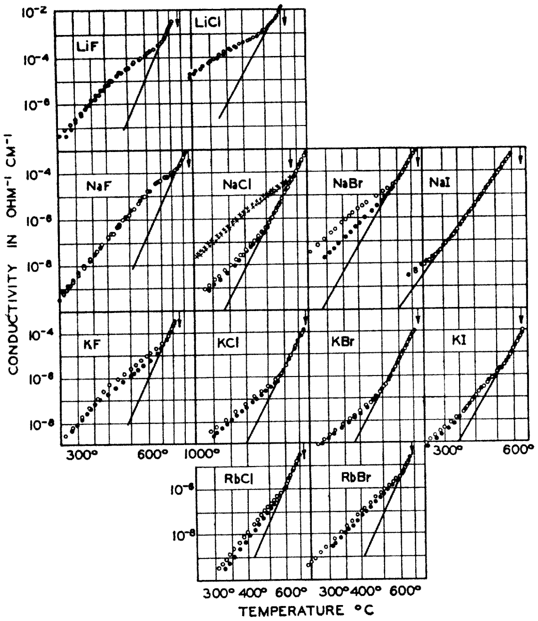

FIG. 15. Electrical conductivity of alkali halide crystals.

increase more rapidly. This would cause the steady-state conductivity to increase with applied voltage and approach the "true" conductivity at very high fields. Beran and Quittner[66] have emphasized this conclusion. They made a number of determinations of both the steady-state conductivity and the true conductivity of natural rocksalt crystals as a function of applied field. The true conductivity was found by determining the polarization e.m.f., using the method of a sudden shift of applied field as previously described. Beran and Quittner found that at all temperatures both the true and steady-state conductivities increased with field strength. In some cases a field intensity of only 10 kv/cm produced noticeable changes in the true conductivity.

Hochberg made studies of the effect of field intensity upon the conductivity of rocksalt crystals formed from the melt in a platinum crucible. He found interesting effects due to the previous thermal history. Crystals which had been tempered at 700°C showed much less dependence of conductivity upon field strength than untempered crystals. At high temperatures the tempered specimens showed negligible variation of conductivity with field strength. At 107°C the conductivity of the untempered specimens increased by a factor of 100, while the conductivity of the tempered specimens changed by only about 50 percent between 0 and 120 kv/cm. Similar results have been obtained by Wenderowitsch and Drisina,[69] who include a brief discussion of previous results and also a discussion of the controversy between Hochberg and Quittner. Wenderowitsch and Drisina found that crystals which showed small polarization effects showed a conductivity almost independent of field strength. For the most carefully prepared specimens, they found that the conductivity changed by less than a factor of two between low fields and 120 kv/cm. This small increase is in agreement with theoretical considerations which we shall present in Chapter VI.

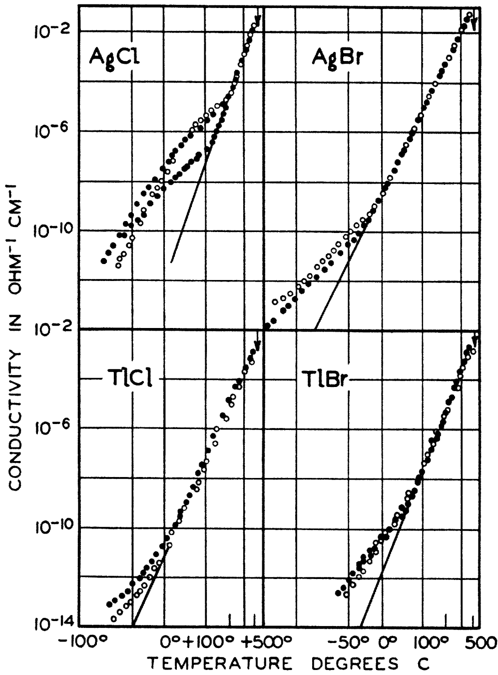

FIG. 16. Conductivity of AgCl, AgBr, TlCl and TlBr.

At much higher fields, an entirely new phenomenon known as breakdown appears. This will be discussed in Chapter VIII.

2. *Temperature dependence.*—The literature on temperature dependence of the conductivity is complicated by the fact that some authors talk about the "true" or initial conductivity denoted by σ_w and others talk about the steady-state conductivity denoted by σ_d.

The temperature dependence of the conductivity of a number of salts in both high and low temperature regions has been studied by Lehfeldt.[41] His results for the temperature variation are shown in Figs. 15 and 16. The important thing to notice about these results is that in all cases there is a straight line at high temperatures which is essentially the same for all of the specimens of a given material, but is different for the different specimens at low temperatures. Figure 17 shows the results for a number of different specimens of KCl having small amounts of impurities. At high temperatures the impurities have little effect, but at lower temperatures the effect is pronounced and increases with the amount of impurity. If the added cation is divalent when forming chlorides, the effect is more pronounced than is the case when a cation which forms monovalent chlorides is added. A similar effect has been found by Gyulai,[70] who investigated the conductivity of "pure" $PbCl_2$ and $PbCl_2$ with small amounts of added KCl and found that not only was A_1 changed, but also B_1. Ketzer[71] found that only 0.001 percent of NaCl added to $PbCl_2$ increased the conductivity by 50 percent. It is to be noted that in this case a divalent ion is replaced by a monovalent ion.

A striking example of the effect of added substances has been given by Koch and Wagner.[72] They added about 10 percent of $CdCl_2$ to AgCl and found that a solid solution was formed. This solid solution was found to have a conductivity about a thousand times that of pure AgCl. The fact that many of the phenomena connected with structure-sensitive behavior are due to very small amounts of impurity is shown by the results of some experiments with natural calcite. One specimen showed marked polarization phenomena with the polarization e.m.f. concentrated near the cathode. Another specimen of completely different origin had about one-thousandth of the conductivity and showed no detectable polarization effects.[64]

In the middle section of Fig. 17 the effect of electrolytic purification is shown. This electrolytic purification is frequently used to produce specimens that will give reproducible effects. A large single crystal (say a natural crystal of rocksalt) is placed between electrodes and heated to several hundred degrees centigrade and a convenient current caused to flow until it shows no change with time. The middle sections of such a crystal show nearly similar behavior from one specimen to another.

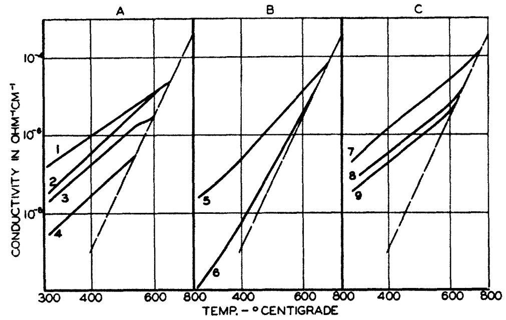

FIG. 17. Impurity conduction in KCl crystals. (1) With 0.07 mol. percent Pb; (2) With 0.04 mol. percent Pb; (3) Natural sylvite; (4) With 0.2 mol. percent each Cu and Ag; (5) With 0.07 mol. percent Pb before electrolytic purification (10 amp. sec.); (6) With 0.07 mol. percent Pb after electrolytic purification (10 amp. sec.); (7) Starting material; (8) ~0.3 mol. percent Cu added; (9) ~0.3 mol. percent Pb added to above.

The usual tables of data give varying values of the conductivity for different specimens. Joffé[74] was able by repeated purifications to obtain results that were reproducible from one sample of the material to another.

The suggestion of Phipps, Lansing, and Cook[37] that the two exponential terms in the expression for the conductivity correspond to the currents carried by the two different types of ion is definitely not valid in some cases. For instance, as pointed out by Smekal,[75] in the case of AgCl, there is a definite bend in the log σ *vs.* $1/T$ curve and yet Tubandt's experiments indicate that only the Ag^+ ion is mobile. The expression for the conductivity of AgCl as given by Smekal is

$$\sigma=(0.00015)\exp(-2160/T)+(3.2\times10^5)\exp(-11{,}060/T). \quad (8)$$

The first term represents the structure-sensitive portion of the conductivity and hence the coefficient of the exponential is only approximate.

The substance which has been studied most extensively is rocksalt. For this substance Smekal[76] gives as the expression for the conductivity, valid from room temperature almost to the melting point,

$$\sigma=A_1\exp(-10{,}300/T)+(1.4\times10^6)\exp(-23{,}000/T)+(3.6\times10^6)\exp(-25{,}700/T). \quad (9)$$

The second term represents cation conduction and the third, anion conduction. The value of A_1 varies from specimen to specimen between about 10^{-1} and unity.

An example of the variation of the constant A_1 with tempering temperature has been given by Kassel.[68] Some of his results are reproduced in Fig. 18. All of the curves have roughly the same slopes, indicating that B_1 is little affected by thermal treatment, but the displacement of the curves upwards definitely corresponds to an increase of A_1 with the tempering temperature. In the theoretical section we shall show that this behavior is consistent with the present interpretation of ionic conduction in solids.

Smekal[75, 76] has suggested that the first term is due to conduction along pores, cracks, or grain boundaries. This, according to Smekal, is the reason that B_1 is smaller than B_2. He interpreted A_1 as a measure of the number of ions set free at the cracks and used this as an explanation of the observed fact that A_1 is much less than A_2. In general, he assumed that the ratio of A_2 to A_1 was of the same order as the ratio of the number of "surface" ions to the total number of ions. There are undoubtedly special effects at grain boundaries, but it seems probable that they do not constitute the complete explanation of the structure-sensitive properties of ionic crystals. Jost[54] and Wagner[55] in particular have pointed out rather serious difficulties in Smekal's interpretation and have themselves proposed a more satisfactory mechanism.

V. Dielectrics in Alternating Fields

We have seen that when a constant voltage is applied to a dielectric, a current flows in the external circuit which usually decreases with time. In engineering applications the important feature is the dependence of the dielectric constant and power factor on the charge-discharge characteristics. A fairly satisfactory theory, applicable whenever the superposition principle is valid, has been developed to correlate the charge-discharge characteristics with the behavior of a dielectric in an alternating field. The attempts to identify the individual mechanisms responsible for dielectric behavior have met with smaller success. This is partly because different mechanisms lead to similar dielectric behavior, so that it is necessary to devise crucial experiments which will decide between different mechanisms. In this particular field the historical order of development has closely paralleled the logical order and therefore will be followed in the present discussion.

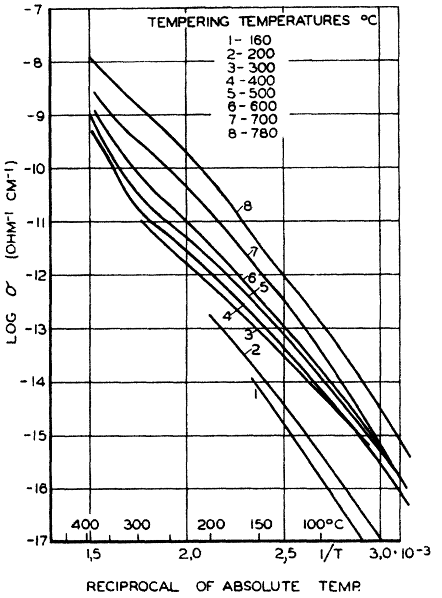

Fig. 18. Temperature dependence of the ionic conductivity of rocksalt.

The anomalous behavior of a dielectric in an alternating field was first reported by von Siemens[80] in 1864 who observed that considerable heat was produced in certain dielectrics in an alternating field. In the many following investigations, this phenomenon became known as the "Siemens heat."

A short time later, Hopkinson,[81] and Jacques and Pierre Curie[9] in France began investigating the charge and discharge currents in a dielectric as a function of time. Hopkinson studied the charge and discharge currents in a Leyden jar and found that (1) simultaneous voltages are superposable and (2) electric displacement through a dielectric may be supposed to depend not only on the voltage at the instant, but also in part on the voltage at all previous times.

It remained for the Curies, ten years later, to give clear expression to these ideas. They worked with glass, crystalline quartz, and other common materials. They expressed their results in three laws which have been approximately verified in modern work. The Curie laws regarding the charging current are as follows.

(1) Proportionality to the voltage. The ordinates of the charging-current curve as a function of the time are rigorously proportional to the voltage.

(2) Law of thickness. For the same voltage, the ordinates of the charging current-curve are in inverse proportion to the thickness of the dielectric.

(3) Law of superposition. Each variation of voltage between the two faces of the sample produces the same change in current as if it alone were acting. The law of superposition as expressed by the Curies is seen to be equivalent to the principle less clearly enunciated by Hopkinson. Accordingly, in present-day literature it is called the Curie-Hopkinson principle after its joint discoverers.

The superposition principle was generalized and put into a mathematical form by von Schweidler in 1907.[82] If $C_v V\varphi(t)$ is the absorption current which flows in a previously unstressed dielectric filling a condenser whose vacuum capacitance is C_v, the superposition principle states that a change in current Δi due to a change in voltage ΔV_T made at an instant T will be

$$\Delta i = C_v \Delta V_T \varphi(t-T). \tag{10}$$

If a series of changes is made at times T_1, T_2, $\cdots$ etc., the resulting absorption current will be

$$i = C_v \sum_k \Delta V_k \varphi(t-T_k). \tag{11}$$

If the voltage on the dielectric changes continuously, we replace the summation by an integral:

$$i = C_v \int_{-\infty}^{t} \frac{dV(T)}{dT} \varphi(t-T)dT. \tag{12}$$

The integration of Eq. (12) can be carried out more conveniently by means of a transformation which changes the limits. Let $\vartheta \equiv (t-T)$; then $d/dT = -d/d\vartheta$. Eq. (12) becomes

$$i = -C_v \int_0^{\infty} \frac{dV(t-\vartheta)}{d\vartheta} \varphi(\vartheta)d\vartheta. \tag{13}$$

The physical meaning of the variable ϑ is simply the time interval between t and a time in the past T. The integration is carried out over all such time intervals extending from zero to ∞. This equation is in a form suitable for use when an alternating voltage is applied to the dielectric. A sinusoidal voltage may be expressed by the real part of

$$V = V_0 \exp \{j\omega(t+\delta)\}, \tag{14}$$

where δ is an arbitrary phase, and $\omega = 2\pi \times$ frequency. At any time t, the voltage that has been applied to the dielectric an interval ϑ previously is

$$V = V_0 \exp \{j\omega(t-\vartheta+\delta)\}. \tag{15}$$

Differentiating with respect to ϑ, we have

$$\frac{dV}{d\vartheta} = -V_0 j\omega \exp \{j\omega(t-\vartheta+\delta)\}. \tag{16}$$

Combining Eqs. (13) and (16) gives the absorption current which flows under alternating voltage:

$$i = j\omega C_v V_0 \int_0^{\infty} \exp \{j\omega(t-\vartheta+\delta)\} \varphi(\vartheta)d\vartheta. \tag{17}$$

The exponential function may be split up to remove the part which does not depend on ϑ. If that part of the exponential which is left within the integral sign is written as

$$\exp(-j\omega\vartheta) = \cos(\omega\vartheta) - j\sin(\omega\vartheta),$$

Eq. (17) becomes

$$i = \omega C_v V_0 \exp \{j\omega(t+\delta)\} \left[j\int_0^{\infty} \cos(\omega\vartheta)\varphi(\vartheta)d\vartheta + \int_0^{\infty} \sin(\omega\vartheta)\varphi(\vartheta)d\vartheta \right]. \tag{18}$$

The absorption current thus consists of two parts. The first is in quadrature with the applied voltage, and contributes, as we shall show, to the dielectric constant. The second is in phase with the voltage and contributes to the dielectric loss.

When a steady voltage is applied to a condenser, the capacitance is defined as the total charge divided by the voltage. If absorption in the dielectric is present, we can regard the total charge as the sum of a portion due to the geometric capacitance* of the condenser and a portion absorbed in the dielectric.

When alternating voltage is applied, the current divides into at least three parts, according to the mechanisms producing it. The total current I in a capacitance C when an alternating voltage V is applied is given by

$$I = i_\infty + i_a + i_c. \tag{19}$$

i_∞ is due to the geometric capacitance C_∞ and is exactly 90° out of phase with the applied voltage. The part of the dielectric constant corresponding to C_∞ is ϵ_∞. i_a is the absorptlon current which we have been discussing. Since it is not exactly 90° out of phase with the voltage, it contributes to both the effective dielectric constant and the dielectric loss. The third part, i_c, is the ohmic conduction current and is exactly in phase with the applied voltage. It contributes only to the dielectric loss.

The properties of a dielectric at a given frequency and temperature can be described by two constants, ϵ' and ϵ'' which we shall call the dielectric constant and loss factor, respectively. They can be combined in a complex dielectric constant,

$$\epsilon = \epsilon' - j\epsilon'', \tag{20}$$

such that the current in a condenser containing the dielectric will be given both in magnitude and phase by the usual equation

$$I = j\omega\epsilon C_v V, \tag{21}$$

where C_v is the capacitance of the electrode system in a vacuum.

The dielectric constant can now be written as the sum of three terms,

$$\epsilon = \epsilon_\infty + \epsilon_a + \epsilon_c. \tag{22}$$

ϵ_∞ contributes only to ϵ' in Eq. (20) and ϵ_c contributes only to ϵ''. ϵ_a, in which we are primarily interested here is determined directly from the absorption Eq. (18),

$$\begin{aligned} i_a &= j\omega C_v V\epsilon_a = j\omega C_v V(\epsilon_a' - j\epsilon_a'') \\ &= j\omega C_v V\left[\int_0^\infty \cos(\omega\vartheta)\varphi(\vartheta)d\vartheta \right. \\ &\qquad \left. - j\int_0^\infty \sin(\omega\vartheta)\varphi(\vartheta)d\vartheta\right]. \end{aligned} \tag{23}$$

In the literature, ϵ_a' is often written as the difference between the observed dielectric constant, ϵ', and the geometric dielectric constant ϵ_∞, which is due to electronic polarization. Comparing the two sides of Eq. (23),

$$\epsilon_a' = \epsilon' - \epsilon_\infty = \int_0^\infty \cos(\omega\vartheta)\varphi(\vartheta)d\vartheta, \tag{24}$$

$$\epsilon_a'' = \int_0^\infty \sin(\omega\vartheta)\varphi(\vartheta)d\vartheta. \tag{25}$$

Equations (24) and (25), together with $i = C_v V\varphi(t)$ are often called the fundamental equations of dielectric absorption. They are general, depending only on Curie's laws and the Hopkinson principle of superposition. They relate the absorption current observed as a function of time in a dielectric under constant voltage, with the dielectric constant and dielectric loss in the same material under an alternating voltage.

In order to make use of these equations, we must have an approximate relaxation function $\varphi(t)$ which characterizes the material and which permits evaluation of the integrals. The empirical relaxation function which was observed by the Curies,

$$\varphi(t) = At^{-n}, \tag{26}$$

has been found to hold over the range of times observed for a large number of materials by Tank,[83] Benedict[84] and others. Von Schweidler[82] has evaluated the integrals in Eqs. (24) and (25) for this relaxation function and obtained for ϵ_a' and ϵ_a'':

$$\epsilon_a' = \omega^{n-1}A\Gamma(1-n)\cos\{(1-n)\pi\}/2, \quad 0<n<1, \tag{27}$$

$$\epsilon_a'' = \omega^{n-1}A\Gamma(1-n)\sin\{(1-n)\pi\}/2, \quad 0<n<2. \tag{28}$$

The ranges given for n are such that the integrals converge. Furthermore, the current is tacitly assumed to be infinite at the instant of application of voltage, which is probably incorrect.[8] Consequently, it is not surprising that Eq. (28) gives a frequency dependence for ϵ'' different from that observed. The power factor computed from Eq. (28) will decrease, remain constant, or

---

* By the geometric capacitance is meant the vacuum capacitance multiplied by the square of the refractive index.

increase with frequency, depending on whether the value of n is less than one, equal to one, or greater than one.[85] Most materials show a maximum power factor somewhere in the frequency range, but such behavior is not predicted from Eq. (28).

Whitehead[86] has shown that for several materials the t^{-n} expression holds, but with $n > 1$. An example is commercial abietic acid, for which the discharge current-time relation is of this type, with n having values from 0.1 to 2.5 at different temperatures.[87] This is another serious limitation to the use of the expressions given in Eqs. (27) and (28) for predetermination of the alternating-current behavior from the charge-discharge data.

Maxwell[88] and Pellat[89] have proposed a relaxation function of the form

$$\varphi(t) = k \exp(-t/\tau). \tag{29}$$

This function has a certain justification in theory, for it gives the time rate of decay of charge in a perfect condenser in parallel with a resistor. When a time τ has elapsed after the beginning of discharge, the condenser will contain only $1/e$ times its original charge. τ is called the relaxation time. The expressions for the increase in the dielectric constant, as given by Eq. (24), and for the loss factor, as given by Eq. (25), due to absorption become upon substituting in Eq. (29):

$$\epsilon_a' = k\int_0^\infty \cos(\omega t)\exp(-t/\tau)dt, \tag{30}$$

$$\epsilon_a'' = k\int_0^\infty \sin(\omega t)\exp(-t/\tau)dt, \tag{31}$$

which can be readily integrated with the help of Pierce's *Tables of Integrals*, formulae 506 and 507:

$$\epsilon_a' = k\tau/(1+\omega^2\tau^2), \tag{32}$$

$$\epsilon_a'' = k\omega\tau/(1+\omega^2\tau^2). \tag{33}$$

The simplification brought about by using a negative exponential for the relaxation function is very helpful. While the loss factor, according to Eq. (33), has a peak at the proper frequency; it covers a much more narrow frequency region than that observed for most materials. Furthermore, the charge-discharge curves usually cannot be fitted very well with a simple exponential relaxation function. For these reasons, Wagner[90] introduced the concept of a polydisperse system having a distribution of relaxation times determined from the empirical relaxation function. The equations for the contribution to the dielectric constant and loss factor were integrated numerically by Yager[91] and compared with the experimental results for several materials, including synthetic resins, glycerol and abietic acid. The experimental results are in good agreement with the theory for most of the materials, indicating that it is possible to predict the a.c. behavior of dielectrics from their charge-discharge curves, and that a small number of constants can be found which characterize the material satisfactorily.

VI. LATTICE DEFECTS AND IONIC CONDUCTIVITY

As Frenkel[92] was the first to point out, it seems impossible to conceive any mechanism which can account for ionic conduction through an ideal lattice. Frenkel suggested that the difficulty could be avoided by assuming that in any actual crystal a certain number of ions have been displaced to interstitial positions, leaving an equal number of vacant places or "holes" in the lattice. In this model conduction can take place either by motion of interstitial ions or by the jumping of ions from regular lattice points to vacant lattice points. This latter process is usually described as a migration of holes. It has been necessary to extend Frenkel's ideas, but his basic suggestion has proved to be correct and the concept of "holes" has become intrenched in the literature of physics. All modern theories of ionic conduction in solids postulate some type of lattice defect.

Vacant lattice points can be produced by thermal agitation, by the presence of certain types of impurity, and by departures from exact stoichiometric proportions. There are a few examples of a type of crystal structure in which one kind of ion forms a regular lattice with the ions of the other kind distributed at random over a number of possible positions which is much larger than the actual number of ions of this kind.

We shall first discuss the "holes" produced by thermal agitation. At first sight it seems that a perfectly-ordered lattice would have a minimum potential energy and hence would be more stable than a lattice which had been disordered by moving atoms into irregular positions. This reasoning is valid at absolute zero, but at higher temperatures it is the Gibbs free energy $E-TS+PV$, and not the potential energy which should be a minimum. Usually we can neglect the last term so that our condition for equilibrium is that the Helmholtz free energy $E-TS$ should be a minimum. Defects occur at finite temperatures because, for sufficiently small amounts of disorder, the entropy contribution to the free energy arising from the disorder is greater than the energy required to produce the disorder.

The entropy associated with the disorder can be found by the methods of statistical mechanics. Detailed discussions of a number of general cases have been given by Wagner and Schottky[93] and by Jost.[94] As an illustration we shall consider the case when n ions of a particular type are displaced from normal to irregular* points. If we let N_1 stand for the number of lattice points and N_2 for the number of irregular positions, the expression for the entropy is:†

$$S=kn\log(n/N_1)+kn\log(n/N_2), \tag{34}$$

where k is Boltzmann's constant. We can always consider n to be much less than either N_1 or N_2. The first term represents the contribution to the entropy which results from distributing n holes at random over N_1 lattice points and the second term represents the contribution to the entropy which results from distributing n ions over N_2 irregular positions. The expression for the free energy becomes:

$$F=nE+knT\log(n/N_1)+knT\log(n/N_2), \tag{35}$$

where E is the energy required to move a single ion from a lattice point to an irregular position. The number of ions in irregular positions is determined by the condition that the free energy shall be a minimum with respect to changes in n—that is, $(\partial F/\partial n)=0$. If we carry out the differentiation and solve for n on the assumption that log n is much greater than unity, we obtain

$$(n/N_1)=(N_2/N_1)^{\frac{1}{2}}\exp(-E/2kT). \tag{36}$$

If we assume still further than $N_1 \simeq N_2$, we obtain

$$(n/N)\simeq\exp(-E/2kT). \tag{37}$$

There are four different types of irregularity which can be regarded as fundamental. They are:[93] (1) Vacant anion lattice points with anions in interstitial positions; (2) vacant cation lattice points with cations in interstitial positions; (3) interstitial ions without vacant lattice points; (4) equal numbers of vacant lattice points for anions and cations and with the corresponding ions transferred to the surface.

In general, for stoichiometric crystals, we can represent any conceivable type of disorder as some combination of these different types. Actually, for any particular material the energies involved in the different cases are so different that only one type of disorder is present. No example of type 3 has been established. The fourth type was first discussed by Schottky.[95]

From a theoretical point of view, the next state after obtaining the expression for n is the computation of a numerical value of E for several typical salts. At first sight, it would seem that this would be equal to, or perhaps greater than, the lattice energy as computed by the Born-Madelung method. For most salts the lattice energy (binding energy per ion pair, which is also the work to remove a single ion to an infinite distance if secondary effects are neglected) is about 10 ev. This would give a conductivity proportional to exp $(-60{,}000/T)$, which gives far too small a conductivity and far too large a temperature coefficient.

The first explanation of the much smaller values of the disorder energy was given by Jost.[96] He considered the interstitial type which will hereafter be referred to as the Frenkel-Jost mechanism. In computing the energy required to transfer an ion from a normal position to an interstitial position, the following items must be taken into account. In the normal position the Coulomb potential energy due to the charges at all other lattice points is $-1.74e^2/a$,[97] where 1.74 is the Madelung number, e is the electronic

---

* It turns out that besides the interstitial ions there are other possibilities of disordered arrangement. We use here the term "irregular point" (German, Fehlordnung) as a general term to include all of the possibilities.

† This expression is a special case of the formula given by Jost, reference 94, page 55.

charge, and a is the lattice spacing. The alkali halides, AgCl, and AgBr, crystallize in a NaCl (simple cubic) type of lattice. For this type of lattice the most obvious interstitial position is the center of the cube. At this point the electrostatic potential is zero since the arrangement of positive and negative charges around such a point is symmetrical. In order to account for the stability of the crystal lattice, Born had to postulate not only the attractive potential $-e^2/r$ between adjacent ions, but also a repulsive potential $+A/r^n$. Compressibility data indicate that n should have a value of about nine. In the normal position this repulsive potential energy decreases the total potential energy by a factor $1-1/n$. If we assume that an ion in an interstitial position produces only a distortion and not a displacement of the neighboring ions, we arrive at a very large repulsive potential energy. Even if we allow for displacement of the neighboring ions, it still seems certain that the repulsive potential energy will be larger in this position than in the normal position. There are, however, two effects which favor the interstitial position. They are the van der Waals energy and the polarization energy of a charge buried in a dielectric. The van der Waals energy is a binding energy of mutual polarization of the type which is responsible for all of the binding in the case of solidified inert gases. This contribution to the energy increases rapidly as atoms are brought closer together ($\simeq r^{-6}$ at large values of r) so that it favors the interstitial position. If a charged sphere of radius r_0 is buried in a dielectric, we have an additional polarization energy in the medium above that when it is in a vacuum. If we compute this polarization energy, using the ordinary expression $\epsilon\mathcal{E}^2/8\pi$ (ϵ is the dielectric constant, and $\mathcal{E}$ is the field intensity) for the energy density, we obtain $-(e^2/r_0)(\epsilon-1)/\epsilon$. A vacant lattice point will produce approximately a polarization equal and opposite to that of the removed ion, so that we shall have an equal term for the polarization due to the hole. There is some doubt about the correct value to use for r_0, but $a/2$ seems reasonable. Using this value, we have for the total polarization energy

$$E_{\text{pol}}=\frac{2e^2}{a}\left(\frac{\epsilon-1}{\epsilon}\right). \tag{38}$$

For AgCl and AgBr the dielectric constant, ϵ, is approximately 10, so that the energy of polarization is practically equal to the lattice energy, leaving only the difference between the van der Waals and repulsion potentials to determine the disorder energy E. For NaCl and other alkali halides ϵ is only about 2 or 2.5; hence, the polarization energy is only about half of the lattice energy. In addition, the van der Waals energies will be smaller and the repulsive effects greater, so that interstitial ions are much less likely for NaCl than for AgCl.

In the case of Schottky disorder (type 4), the work to remove an ion from the normal position is also equal to the Madelung energy, but the polarization and van der Waals energies of the displaced ions are smaller than when they are removed to interstitial positions. However, the repulsive potential energy is less for the displaced ion than in the original position so that this effect decreases the disordering energy.

The qualitative conclusions presented here have been made more quantitative by Schottky,[95] Jost and Nehlep,[98, 99] and Mott and Littleton.[100] The results of these calculations indicate that for AgCl and AgBr the high dielectric constant and high van der Waals effects make the Frenkel-Jost mechanism almost certain. On the other hand, for NaCl, KCl, etc., the lower dielectric constant, the smaller van der Waals effects, and the greater repulsive effects, make the Schottky mechanism probable.

As Jost[96] has pointed out, the observed exponential variation of the conductivity with temperature is due not only to a change in the number of quasi-free ions, but is also due to a change in their mobility. Between one interstitial position and another, or between a lattice point and a vacant lattice point, there will, in general, be a potential barrier. For particles having such large mass there is no chance of a quantum-mechanical "tunnel" effect, and the only way that an ion can make a transition from one position of minimum potential energy to another is by obtaining sufficient energy of thermal agitation to pass over the potential barrier. If we denote the height of the barrier† by U, the probability that an ion have sufficient energy to jump

---

† For lack of a better term, we shall refer to this quantity as the activation energy.

from one stable position to another is proportional to exp $(-U/kT)$. The conductivity is thus given by the expression:

$$\sigma = (\text{const}) \exp\{-(U+E/2)/kT\}. \quad (39)$$

In addition to the motion of interstitial ions, we have for both the Frenkel and Schottky mechanisms the possibility of an ion jumping from a lattice position to one of the holes. This process is equivalent to the motion of the holes in the opposite direction and is usually so described. For the Schottky mechanism this is the only method of conduction, but since there are vacant lattice points of both types, there may be conduction due to both types of ion, although in general with different activation energies. For the Frenkel-Jost mechanism it is also likely that there will be two different values for the activation energy.

Calculations of the activation energy U have been attempted by Jost,[96] Jost and Nehlep,[101] and Mott and Littleton.[100] In the process of motion from one position to another, an ion must push neighboring ions aside. This means that the problem of computing the energy required encounters many of the same difficulties as the problem of coupled three-dimensional oscillations. Another feature which makes the calculations difficult is that a "squeezing-through" process is very sensitive to the nature of the interatomic forces. Since accurate expressions for the repulsive forces are not known, an accurate calculation of the value of U should not be expected. For NaCl the B_2 and B_3 values given by Smekal[102] correspond to values of 2.0 ev and 2.2 ev for the values of $U+E/2$ for the Na and Cl^- ions, respectively. Mott and Littleton obtained values of 1.5 and 1.6 ev. They assign most of the discrepancy to the calculation of U.

In the Schottky mechanism there are equal numbers of holes in both types of lattice so that the different contributions of the two types of ions to the conductivity are due only to the different activation energies. For most of the alkali halides the negative ions are larger than the positive ions, so it is not surprising that in NaCl, KCl and NaF Tubandt[23,103] has found that the negative ions are mobile only at high temperatures. The Frenkel-Jost mechanism is favored when one type of ion is much larger than the other because under these circumstances there is a smaller repulsive potential energy. For this type of conduction both the disorder energy E and the activation energy U will be greater for the larger ion in a diatomic salt. We should almost always expect unipolar conduction for this mechanism. The chlorides and bromides of silver have been proved to be interstitial conductors. These salts are also known from the work of Tubandt to be unipolar conductors. Similar statements are probably true for the corresponding thallium salts.

The empirical formula for the high temperature conductivity has the same form as that given in Eq. (39), but before we can identify the value of B found from the slope of a log σ *vs.* $1/T$ graph with $(U+E/2)/k$, we must be certain that this quantity is independent of temperature. One reason for believing that U in particular should depend upon the temperature is because its value is very sensitive to changes in the repulsive potential and therefore to changes in interatomic distances accompanying thermal expansion. The disorder energy E, particularly in the Frenkel mechanism, should also be expected to decrease with an increase in volume. The numerical value of A can be estimated from diffusion data and values of the order 10–100 are predicted.[96] Many of the observed values of A are in the neighborhood of 10^6, so that it is necessary to account for a factor of about 10^4. If we assume that $U+E/2$ decreases uniformly with temperature,[98] we have $U+E/2=(U_0+E_0/2)-KT$ and the expression for the conductivity becomes

$$\sigma = A'(\exp K/k) \exp\{-(U_0+E_0/2)\ kT\}. \quad (40)$$

The value of A ordinarily quoted would accordingly be equal to $A' \exp(K/k)$. In order to account for the factor of 10^4, it is necessary to assume that K/k has a value of about nine. Another method of expressing this result is to say that in a temperature range of 300°C the value of $U+E/2$ changes by about 0.2 ev. For the poorer conducting salts the values of $U+E/2$ are about 1.5 or 2.0 ev. This explanation is quite reasonable. For the better conducting salts like the silver halides, the values of $U+E/2$ are smaller, but the values of A are also smaller, so that the relative changes of $U+E/2$ are not unreasonably large.

Jost and Nehlep[101] also checked this argument from another point of view. If $U+E/2$ is sensitive to volume changes, it should be possible to produce a decrease in conductivity by increasing the pressure. Assuming that the change in U for a given change in interatomic distance is the same whether it is produced by thermal expansion or by application of pressure, they were able to calculate the effect of a change in pressure. They performed experiments on silver chloride and silver bromide and found that the experimental results checked the theoretical predictions within the tolerance which must be allowed due to the approximate values of A' used. Jost[98] has pointed out that when high pressures are applied, the expression for the number of atoms in irregular places is no longer correct because the effect of the pressure, particularly in the Schottky mechanism, is to squeeze the displaced ions back into the vacant places. Thermodynamically, this means that the Gibbs free energy which includes the pV term should be used instead of the Helmholtz free energy in setting up the equilibrium conditions.

For field strengths of 100 kv/cm or greater, the energy gained in jumping from one potential minimum to another is nearly a hundredth of an electron volt. Since the activation energy for most salts is less than half of an electron volt, the effect of such a field is to reduce U by at least one percent. At room temperatures this corresponds to a change of about 50 percent in the conductivity. At higher temperatures the change would be smaller. This is probably the explanation of the effects observed by Wenderowitsch and Drisina[69] and others.

Another anomaly in the behavior of the constants A and B is their considerable change between high and low temperatures. (See Table I.) In the high temperature region the value of A is usually about 10^6 and the value in the low temperature region is variable but usually somewhat less than unity. The value of B appropriate to the low temperature region is usually about half of that found at high temperatures. The explanation of these differences has been given by Jost.[104] Suppose that the temperature of a specimen of some salt is raised almost to the melting point and kept there for sufficient time to establish thermal equilibrium. The number of irregular ions will then be given by Eq. (36). If the specimen is cooled from this temperature, the process of establishing thermal equilibrium depends upon diffusion which is a comparatively slow process. In general, this means that at low temperatures the number of defects will be greater than the equilibrium value corresponding to the lower temperature, unless long tempering periods are allowed. At temperatures which are more than a few hundred degrees below the melting point, diffusion is so slow that we can regard the number of defects as independent of the temperature and the time. If T_c is a critical temperature below which the number of defects stays constant, we have as the expression for the conductivity in the low temperature region

$$\sigma = A \exp(-E/2kT_c) \exp(-U/kT). \quad (41)$$

For the alkali halides $E/2$ is about 1 ev. If we take for T_c a value of about 700°K, we obtain about $\exp(+12{,}000/T_c)$ as the ratio of the high temperature value of A to the low temperature value. This is about 10^7, in reasonable agreement with the values given in Table I. The low temperature value for B will be equal to U/k. Since U is estimated to be somewhat smaller than E, B_1 should be about equal to one-half of B_2 in agreement with the general run of the data in Table I. Jost[104] has pointed out that while, in the Shottky mechanism, it is necessary for holes to diffuse to the surface in order to produce thermal equilibrium, in the Frenkel mechanism it is only necessary for a hole and an interstitial ion to diffuse together. Therefore, it is to be expected than when the Schottky mechanism is prevalent, the establishment of thermal equilibrium should be slower, and the critical temperature nearer to the melting point. Jost has even expressed doubt whether the "freezing in" of the holes should occur at all for the Frenkel mechanism, and has suggested that a change in the slope of the log σ *vs.* $1/T$ curve was evidence for the Schottky mechanism. This suggestion is contradicted by the experimental results of Lehfeldt[41] who found this sort of behavior for AgCl and AgBr, although all other evidence indicates that these salts owe their conductivity to interstitial Ag^+ ions. Seitz[105] has used the data of Lehfeldt to separate the contributions of U and $E/2$ for AgBr. At high temperatures the slope of the log σ *vs.* $1/T$ graph

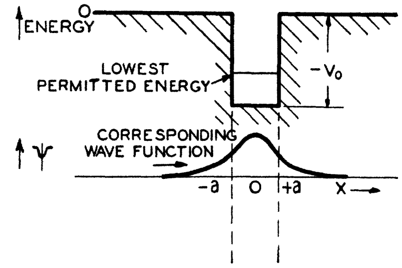

FIG. 19. Energy level and wave function for a particle in a box.

corresponds to a value of $U+E/2$ of 0.66 ev; at low temperatures the slope (see Fig. 16) is much smaller and corresponds to a value of U of about 0.30 ev. Values nearly equal to these have been found by Koch and Wagner[72] by a very clever method which we shall describe briefly later.

The value of T_c should depend upon the manner in which the specimen has been tempered. If a specimen has been kept at a temperature near the melting point for a long period and then cooled at the usual rate, the value of T_c should be very nearly equal to the temperature during the aging process. If the specimen was not tempered sufficiently or if it was cooled too slowly, the value of T_c might be either higher or lower. Kassel[68] has made conductivity studies on a number of specimens of NaCl which had been aged at different temperatures. He found (see Fig. 18) that the specimens which had been aged at the higher temperatures showed the larger values of A_1. This behavior is consistent with our picture that aging at a higher temperature causes a larger number of defects to be "frozen in" during the cooling. Further experimental work on this point would be desirable.

The model that has been suggested for the conductivity of AgCl and AgBr has been checked by Koch and Wagner[72] in a series of clever experiments. The basis for these experiments is the behavior of solid solutions of certain divalent halides in the corresponding Ag salts. For instance, in a favorable temperature region, as much as 10 mol percent of $CdCl_2$ forms a homogeneous mixture with AgCl, in which the Cd^{++} ions are distributed at random over the cation lattice points. However, for a given amount of Cl, there will not be enough Cd^{++} ions to fill the corresponding cation sites, so that we have vacant places in the cation lattice equal in number to half the number of added Cd^{++} ions. At 210°C the addition of 10 percent $CdCl_2$ to "pure" AgCl increased the conductivity from 8.6×10^{-5} ohm^{-1} cm^{-1} to 5.2×10^{-2} ohm^{-1} cm^{-1}. Similar results were obtained by adding $PbCl_2$ to AgCl and by adding $PbBr_2$ to AgBr. From an analysis of their data, Koch and Wagner conclude that in AgCl and AgBr without impurities, the conductivity is due about equally to the motion of interstitial ions and holes. Their results indicate that the disorder energies are 0.43 ev and 0.54 ev for AgBr and AgCl, respectively, in comparison with the values of 0.36 and 0.50 which Seitz has computed from Lehfeldt's[41] data.

The experiments of Koch and Wagner furnish the explanation of the effects observed by Lehfeldt,[41] Ketzer[71] and others of the effect of small amounts of impurity on the conductivity. NaCl crystals formed by precipitation from a supersaturated solution frequently contain $PbCl_2$; this offers an explanation of the variation of conductivity from one specimen to another.

There are a few salts which may be described as possessing natural disorder and therefore the exponential in the expression for the conductivity includes only the activation energy. One illustration of this type of lattice is α AgI (the form which is stable above 140°C). In this structure the unit cell contains two molecules. The I^- ions form a body-centered cubic lattice and the two Ag^+ ions per unit cell are distributed at random over the thirty largest holes resulting from this packing.[108] This structure is surprising at first sight. Wyckoff[106] remarks, "such an unusual structure needs further confirmation." At present the structure is accepted and a few other examples have been found. Another somewhat similar structure is that of α Ag_2HgI_4 in which there are four possible cation sites and only three cations per unit cell. This substance has been investigated by Ketelaar.[107] Other salts which owe their high ionic conductivity to natural disorder are the high temperature forms of CuBr and CuI. One interesting characteristic of these four salts is that their conductivity is not noticeably increased upon melting.

There is another type of lattice defect which also influences the ionic conductivity. Many of the compounds of Cu and Ag with S, Se and Te show a wide range of compositions with an excess of the electronegative component above that given by the formula R_2X. The compounds are sufficiently ionic to make a substitutional solid solution unlikely and the excess electronegative component is taken care of by holes in the cation lattice. Actually, the large anions form a lattice of their own with the cations in interstices somewhat as in α AgI. It is easy to see why there should be wide ranges of composition. Actually, most of these compounds show an electronic conductivity of about 10^3 ohm^{-1} cm^{-1}, but the diffusion and electrolytic transfer have been measured and found to be comparable with those in α AgI.[109]

VII. Quantum-Mechanical Ideas

During the past ten years quantum-mechanical studies of the electronic structure of semiconductors and insulators have added so much to our knowledge of their structure that a review article of this type would be incomplete without some reference to this topic. On the other hand, so many excellent books and review articles have appeared that any extended discussion seems inappropriate. In this section we shall present the essential ideas in nonmathematical form for those readers who have some acquaintance with quantum mechanics, and who are interested in the application to solids. For those who desire either a more introductory or a more elaborate discussion, a list of reference books is included in the bibliography.

As a starting point we shall consider a one-dimensional model in which the potential energy is represented by a series of rectangular wells. This model is admittedly artificial, but it gives a chance of illustrating in a simple manner certain features which are also characteristic of a more realistic model.

The behavior of a quantum-mechanical system is described by specifying the wave-functions and the energy. For a one-dimensional problem, the wave functions are found by a solution of the differential equation

$$\frac{d^2\psi}{dx^2}+\frac{8\pi^2 m}{h^2}[E-V(x)]\psi=0. \tag{42}$$

In this equation x is the coordinate, m is the mass of the particle, h is Planck's constant, $V(x)$ is the potential energy, and E is the total energy. In general, E can take on only certain values which permit solutions for ψ which show neither discontinuities nor infinities. The square of the function ψ is interpreted as being proportional to the probability of finding the particle in a region dx at x. If we consider a potential energy function like that represented graphically in Fig. 19 we can derive without mathematical analysis certain general characteristics of the wave functions.* The interesting values of E are those between 0 and $-V_0$. For energies within this range we can consider separately the parts for $x<-a$, $-a\leq x\leq a$, $x>a$. In the regions outside the well the solutions for ψ are exponentials which come down to the axis as indicated. Within the well the wave function is curved toward the axis. It can be seen that only for certain values of E can the sinusoidal portion inside the well be joined properly to both exponentials. The lowest energy and corresponding wave function are shown in Fig. 19. It might be pointed out that a ψ function which was at all points the negative of the one shown would also be a solution.

If there are two such potential wells at a large distance apart, we can consider them as two separate problems each having its own energy levels. If, however, they are close together, as shown in Fig. 20, the joining conditions become changed somewhat and we can have two different types of wave function and two different energy

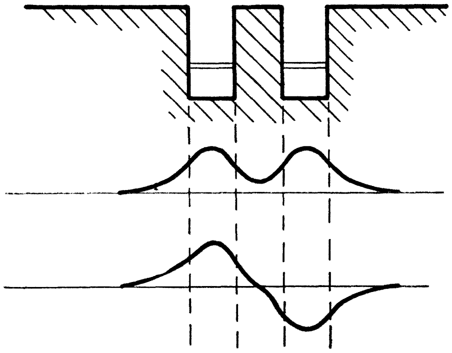

FIG. 20. Energy levels and wave functions for two similar boxes.

---

* This sort of argument has been reproduced in so many places (particularly Gurney's book on quantum mechanics) that the authors are apologetic about including it here. It is included because it seems a natural approach to the study of energy bands.

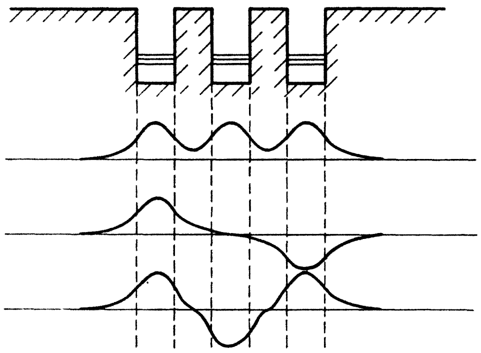

FIG. 21. Energy levels and wave functions for three similar boxes.

levels. If the boxes are far apart, the separation of the energy levels will be small; if the boxes are close together, the separation will be larger.*

One of the properties of this system results from the fact that in the left-hand well the wave functions are almost equal in magnitude and have the same sign; while in the other box, the wave functions are almost equal but have opposite signs. If we place a particle in one of the boxes and not in the other, we can represent this situation by a wave function which is half the sum of the two wave functions pictured in the diagram. This wave function would correspond to a very large probability of the particle being in the left-hand box, and a very small probability of its being in the right-hand box. However, since the wave function in the right-hand box is not quite zero, the probability of the particle being found in the right-hand box increases with time and the probability of its being found in the left-hand box decreases with time. In a certain length of time, the probability that the particle is in the right-hand box will be large, and the probability of its being in the left-hand box will be small. The process will then reverse. The particle will thus shift back and forth from one box to the other with a certain frequency. This frequency multiplied by Planck's constant is measured by the energy separation between the levels. This shifting back and forth of particles from one potential well to another is an important aspect of the quantum mechanics of solids. It turns out, for instance, that in a metal like Na the frequency with which electrons shift from atomic center to atomic center is so high that we should no longer think of a particular electron as belonging with a particular atom.

If we have three similar boxes, there will be three different types of wave functions and three closely-spaced energy levels. The important thing to note in this case is that the wave function for the lowest level is repeated without change in sign in going from one box to the next and that the wave function for the highest energy level changes sign in going from one box to the next. (See Fig. 21.)

If we have N separate and similar boxes, we will have N closely-spaced energy levels, the lowest corresponding to the function which repeats and the highest corresponding to the function which changes sign. In the cases that we have discussed, we cannot speak of the energy levels of N separate boxes, but only of the N energy levels of the system.

For the single potential well there will of course be other energy levels—for instance, the one whose wave function is indicated in Fig. 22. Corresponding to this energy level there will be, for the N symmetrically-placed boxes, N closely-spaced energy levels. In general, however, this "band" or group of energy levels will be widely separated in energy from the lower energy band.

For an actual atomic system, we will still have the same general features, except that we must, in accordance with the Pauli principle, allow two electrons to each energy level and must also take into account the different possible "orientations" of the wave functions. If we have a solid consisting of N atoms, each s level of the separate atoms will be broken up into a band containing $2N$ energy levels. Each p level (or group of levels) will be broken up into $6N$ energy levels. Each d level (or group of levels) will be broken up into a group of $10N$ energy levels.

Another method of understanding the separation of separate energy levels into bands of energy levels, is to start with a free-electron model. In this model we visualize a block of metal as a box in which the electrons are confined. The only function of the positive cores is to neutralize the electrostatic interactions of the electrons. The motion of the electrons in such a box can be described in terms of de Broglie's

---

* Interesting analogies have been drawn between the splitting of levels and the splitting of frequencies when oscillators are coupled together. See, for instance, W. Shockley, Bell Sys. Tech. J. **18**, 645–723 (1939).

hypothesis by standing-wave patterns. The lowest energy electron will have a de Broglie wave-length equal to twice the length of the box. The next-higher energy electron will have a wave-length equal to the length of the box, etc. Since the box is large compared with atomic dimensions, the lowest energy level corresponds for all practical purposes to an electron which is at rest and the higher energy levels correspond to electrons having kinetic energy of motion through the lattice. This free-electron model seems to predict a continuous distribution of energy levels without breaks. However, when reflections of the Davisson-Germer type from planes of atoms are taken into account, it turns out that there are certain energies for which propagation of electron waves through the lattice is forbidden, and we again obtain a system of allowed and forbidden ranges, or bands, of energy levels.

We are now in a position to understand why some substances are insulators and some are conductors. When an electric field is applied to a metal, some of the electrons have a drift-velocity superposed upon their normal motion. Quantum-mechanically, this is described as a transition to a higher energy level. However, if we have a solid in which the total number of valence electrons is exactly equal to the number of energy levels in the lowest energy band, there will be no vacant levels to which the electrons can be excited by the small amount of energy available from the field. Such a substance will be an insulator or perhaps an electronic semiconductor. One type of substance which is always a conductor is an element which crystallizes in a cubic system and which has an odd number of electrons. Many of the metals fall into this class. For these substances there will be a partly-filled band of energy levels and there is no difficulty about obtaining vacant levels for conduction. In substances like diamond, boron and silicon, there are an even number of electrons per unit cell. More detailed consideration of some of these has shown that the number of electrons available is exactly equal to the number of energy levels in a band. Some of the divalent metals like calcium, zinc, etc. might be expected[110] to have filled-energy bands, since each has an even number of electrons per unit cell. These metals are conductors because there is a small overlap in energy. Some energy levels in the higher band will be occupied and some of the energy levels in the lower band will be vacant. There will be conduction due to the electrons in the excited band and the "holes" in the lower band.

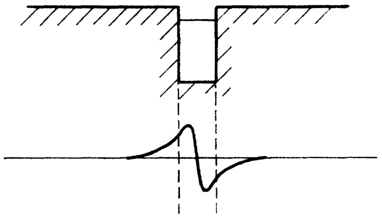

FIG. 22. Second energy level and corresponding wave function for a particle in a box.

Substances like NaCl, Cu_2O and most saturated-valence compounds also have an even number of electrons per unit cell, and hence the fact that they are insulators is explicable in terms of the band picture. The grouping of electrons into energy bands depends upon the crystal structure, so that it is understandable that marked changes in electrical properties may accompany polymorphic transformations or melting. An interesting example of the effect of crystal structure has been quoted by Mott and Jones.[17] There are a number of salts or intermetallic compounds (Mg_2Sn, Mg_2Pb, Mg_2Si, Mg_2Ge, Li_2S, Na_2S, Cu_2S, Cu_2Se, Be_2C, and others) that crystallize in the CaF_2 (fluorite) structure which has 12 atoms per unit cell. According to quantum-mechanical calculations, for this type of lattice there is an energy band with levels for exactly thirty-two electrons per unit cell. Since the compounds which are listed above all have thirty-two valence electrons per unit cell, we expect these substances to be insulators. The electrical resistance of these compounds has been measured by Norbury,[16] who has found that for exact stoichiometric compositions, they have a very high resistance. However, upon melting, their conductivity approaches that of molten metals such as tin. This behavior is in direct contrast with the viewpoint of von Hevesy[111] who has suggested that "from the standpoint of the electrical constitution of matter, the melting point is not a special point" and proceeds to argue that a large change in resistance upon melting is evidence for ionic conductivity. Von Hevesy's remarks were made before the development of the band theory and the only reason for repeating them here is to point out that melting corresponds to a change in structure and hence may cause a marked change in the energy bands.

There are a few compounds which the band theory indicates should be conductors, but which actually are almost insulators. CoO and NiO are examples. These salts crystallize in the NaCl type of lattice. CoO has an odd number of electrons per unit cell, so that in accordance with what we have said it should be a conductor. At room temperatures the observed conductivity is about 10^{-8} ohm^{-1} cm^{-1}. The value for NiO is about the same, so that we see that in this case some factor other than the odd number of electrons per unit cell must be involved. We can say that two of the electrons from the metallic element are lost to the oxygen. These electrons fill the *p* shell in oxygen, or, in other words, they form a filled band. The remaining *d* electrons, of which there will be eight in Ni and seven in Co, will only partly fill the *d* shells of their respective atoms. According to the band picture, the *d* band is only partly filled and hence should take part in the conductivity.[112] No detailed discussion of these compounds has been carried out, but some general considerations have been advanced by de Boer and Verwey,[113] and by Peierls[113a] in a discussion of their paper. The essential difference between the conductivity of a substance like Na and that of NiO involves one essential factor. In metallic Na the potential energy of an electron as a function of the distance along a line of centers is quite flat except near the nuclei. The wave functions of the valence electrons are large in this region. Since an electron in this region is only slightly under the influence of any particular nucleus, an electric field causes a drift of electrons without any piling up in the neighborhood of a particular nucleus. On the other hand, for NiO the wave functions of the *d* electrons in Ni are concentrated in regions near the nuclei; hence, a drift motion of electrons can take place only by creating a Ni^{+++} and a Ni^{+} ion instead of a pair of Ni^{++} ions. Such a process is unlikely.

If we have a solid with the highest, filled energy-levels separated by a forbidden energy region from the lowest vacant levels, this substance will be an insulator at absolute zero. However, at higher temperatures, particularly when irradiated with light in the proper frequency range, such substances may become conductors. The next chapter will be devoted to electronic conduction in insulators and semiconductors.

VIII. SPECIAL TOPICS

In the first six chapters we discussed principally the phenomena of ionic conduction. The present chapter will be devoted principally to electronic conduction. The phenomena of electronic conduction in nonmetals are best understood by considering cases where it is greatly enhanced. At absolute zero a substance with filled energy bands is an insulator. Conductivity can be produced by either thermal agitation, illumination with light of a suitable frequency, or by subjecting the material to very high field intensities. The first two types of conductivity are greatly affected by the presence of impurities and much of our attention will be devoted to crystals which depart in some way from an ideal arrangement. We shall consider first the case when the conductivity is due to thermal agitation.

A. Semiconductors

The boundary between so-called semiconductors and insulators is not definite, but is usually assumed to be between 10^{-8} and 10^{-10} ohm^{-1} cm^{-1} for the room temperature conductivity. The boundary between semiconductors and metals is about 10^{3} ohm^{-1} cm^{-1} for the room temperature conductivity. Here, however, we have more definite criteria. For a metal, the conductivity decreases with an increase in temperature, while for a semiconductor the conductivity increases with an increase in temperature. In this connection it might be remarked that some of the elements (Si, Ti, Zr and perhaps a few others) were formerly classed as electronic semiconductors because of the peculiar manner in which their resistance changed with temperature. Gudden[114] has explained these anomalies in terms of high resistance oxide layers. These oxide layers have been the cause of many difficulties in the experimental studies of semiconductors.

Semiconductors and metals are also effected differently by the presence of impurities. Small additions of another element lower the conductivity of a metal. For semiconductors, departures from stoichiometric proportions increase the conductivity. Some apparent exceptions to this rule are due to the fact that some substances as ordinarily prepared are not stoichiometric and so "impurities" may bring them closer to the ideal composition.

The sulphides, selenides and tellurides of copper and silver belong to a special group, which are sometimes called semimetallic alloys. The electronic conductivity increases with departures from stoichiometric proportions, but at the same time it decreases with temperature, although in a rather erratic manner. The place where these compounds fit in the general scheme can be seen by considering the "compounds" of Ag with the elements immediately following it in the periodic table; namely, Cd, In, Sn, Sb, Te, and I. The alloys of Ag with Cd, In, and Sn are primarily of the substitutional type with wide ranges of solid solution and typically metallic properties. The alloys with Sb, and to some extent with Sn, are still metallic, but are characterized by the occurrence of definite intermetallic compounds with high melting points and narrow composition ranges. AgI is definitely an ionic compound. The compound of Ag with Te (and similarly for the other salts mentioned above) is thus a sort of bridge between the definitely ionic and the definitely metallic compounds.

The explanation of the essential difference between a metal, a semiconductor, and an insulator was first given by Wilson.[115] According to his theory, for an insulator or semiconductor there will be a band of energy levels which is exactly filled at absolute zero, and, separated from it by a forbidden range of energy, a higher band which is vacant at absolute zero. The difference between an insulator and a semiconductor lies in the magnitude of the energy separation, which is large for an insulator and small for a semiconductor. When the temperature is above absolute zero, electrons are excited to the upper band and conduction can take place. According to the Fermi statistics, the number of electrons in the upper band at absolute temperature T is given by an expression of the form[117]

$$N \sim \exp\left(-\Delta E/2kT\right). \tag{43}$$

If ΔE is of the order of a few tenths of an electron-volt, the conductivity will be high enough for the material to be classed as a semiconductor. The substances to which this model is applicable are known as intrinsic semiconductors.

It seems definite, however, that this model does not apply to most semiconductors. Gudden[118] was the first to point out that the conductivity of a semiconductor could not be classed as a constant, characteristic of the material, but is markedly dependent upon the presence of small amounts of impurity and upon previous thermal treatment. This suggests a different model for a semiconductor. In the neighborhood of each impurity we will have a localized wave function with an energy level which may lie in the region between the filled and empty bands. Electrons can be excited from these localized levels to the conduction band. Substances in which the conductivity is due to impurities are known as extrinsic semiconductors. This case has also been worked out by Wilson.[116] The expression for the number of electrons has the same form as that given in Eq. (43), but the factor in front of the exponential is smaller. Other models have also been discussed by Fowler.[119] There is one other case we shall mention. It may happen that the impurities produce vacant levels in the region between the two bands. Electrons can be excited to these vacant levels from the filled band, thus leaving vacant places in the lower band and permitting conduction.

The easiest way of introducing impurities is to heat the material in the vapor of the electronegative component. One of the earliest experiments on the effect of excess electronegative component was performed by Baedeker,[120] who found that the conductivity of CuI was definitely dependent upon the iodine vapor pressure. Another method of showing this is to immerse the CuI in a solution of iodine in CS_2. In this experiment the conductivity was found to increase proportionally to the 5th root of the iodine concentration. Tubandt[121] and associates also found that the conductivity of cuprous halides changed from predominantly ionic to predominantly electronic with an excess of halide.

Most of the studies on semiconductors have been carried out on oxides. The excess oxygen required to produce large changes in the conductivity is so small that it is impossible to detect it chemically. Instead, the partial pressure of oxygen is varied and the resulting changes in conductivity are studied. Substances are classed in one of three different groups, depending upon the change of conductivity with the partial pressure of oxygen. We shall first list the different classifications and then discuss some of the special cases and such interpretations as are available.

(1) In oxidation semiconductors the conductivity increases with an increase in the partial pressure of oxygen. Typical examples are Cu_2O,[122] NiO,[123] CoO,[124] FeO,[124] and UO_2;[170]

(2) In reduction semiconductors the conductivity decreases with an increase in the partial pressure of oxygen. Typical examples are ZnO,[125] CdO,[125] Al_2O_3,[170] Ta_2O_5,[170] TiO_2,[170] and BaO;[126]

(3) For a few substances the conductivity is independent of the partial pressure of oxygen. Examples are CuO,[127] Fe_2O_3,[124] Fe_3O_4,[124] and Co_3O_4.[124]

The behavior of the first group depends upon the manner in which the excess oxygen is accommodated.[128] The oxygen is too large for an interstitial position and in an ionic crystal it is very unlikely that a simple substitutional lattice be formed, as this requires putting a negative ion at a point where its potential energy would be positive. Hence, it seems probable that the excess oxygen will be balanced by having vacant positions in the cation lattice. If, however, we add neutral oxygen to the lattice to form negative ions, the corresponding electrons must be removed from other copper ions. Each additional oxygen ion formed above stoichiometric proportions requires that there be two vacant places in the cation lattice and also that two Cu^{++} ions be formed. The mechanism by which the conductivity is produced may be described by saying that there are now vacant places in the *3d* band of copper, or by saying that the location of a doubly-charged ion shifts due to motion of an electron in the opposite direction. This type of conductivity is known as defect conductivity.

The mechanism by which vacant places in the cation lattice are formed is as follows: Cu^+ ions diffuse to the surface where they react with oxygen molecules to produce O^{--} ions and Cu^{++} ions. A Cu^{++} ion is equivalent to an electron defect in the copper conduction band. The law of mass action can be applied to this type of chemical reaction and yields the result that the conductivity should be proportional to the one-eighth power of the partial pressure of oxygen.[129] The experimental results of Wagner[130] indicate that between 800°C and 1000°C the conductivity is proportional to the one-seventh power of the partial pressure of oxygen. This is regarded as reasonable agreement for this sort of measurement. At lower temperatures the conductivity depends upon the previous history as well as upon the temperature and the oxygen partial pressure.

For the reduction semiconductors the situation is assumed to be quite different.[131] For instance, ZnO at 600°C is assumed to be partially dissociated—a portion of the oxygen being vaporized and leaving behind an excess of metal. This excess metal could be accommodated either by having holes in the anion lattice or by having zinc ions in interstitial positions. The relatively large size of the O^{--} ions compared with Zn^{++} ions makes the interstitial position more likely. The conductivity is due to the fact that when the oxygen evaporates, the excess electrons are left behind in the conduction band. This type of conductivity is known as excess conductivity. Increasing the partial pressure of oxygen in the surrounding atmosphere decreases the amount of oxygen which evaporates and thus decreases the number of excess electrons and hence the conductivity.

There is one essential difference between oxidation and reduction semiconductors. In the oxidation semiconductors ions of higher positive charge are created and this is likely only when dealing with a salt in which the metallic element is exhibiting its lower valence. For instance, Cu_2O is an oxidation semiconductor, but CuO is not. On the other hand, in reduction semiconductors the metal is usually exhibiting its highest valence.

One of the most satisfactory methods of distinguishing between excess semiconductors and defect semiconductors is by means of the Hall effect. A magnetic field is applied in a direction perpendicular to the flow of current and a potential difference appears between two points which were initially at the same potential. The experiments are rather complicated, particularly

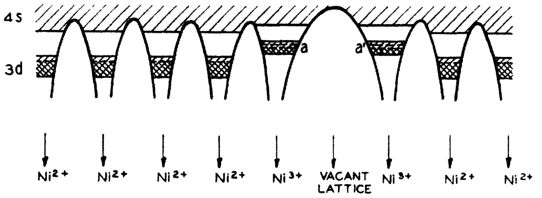

FIG. 23. Schematic representation of the *3d* and *4s* bands in NiO containing a small amount of oxygen in excess.

in the matter of contacts, but results have been obtained for a few substances. For magnetic fields of about 10,000 gauss and readily obtainable currents, the potential differences measured are usually of the order of millivolts. The Hall coefficient is defined by the relation $R=E_y/J_xH_z$, where E_y is the transverse potential difference, J_x is the current density in the x direction, and H_z is the magnetic field in the z direction. If there are only a few electrons in an otherwise empty band (ZnO, for instance), the Hall coefficient is equal to $3\pi/8nec$, where n is the number of electrons per unit volume in the upper band. In this case, the direction of the Hall e.m.f. is the same as that found by a simple classical model of a free electron moving across a magnetic field. This is referred to as the "normal" case. If there is a filled band with n vacancies per unit volume, we have the same expression, except that the algebraic sign is changed. This is referred to as the "anomalous" case. In some substances there may be conduction due to both excess electrons and holes, and the measured Hall effect will be the difference of the two. This probably explains the results obtained by Schottky and Waibel[132] for Cu_2O at different temperatures. At ordinary temperatures, the Hall effect is "anomalous" indicating, in accordance with the previous discussion, that the conductivity is due to the deficit of electrons in a nearly-filled band. At higher temperatures the Hall coefficient becomes zero and then changes sign, indicating that at higher temperatures "excess" conductivity predominates. Measurements of the Hall coefficient have sometimes been used to study the variation of n with temperature. If we have both types of conduction, this gives an upper limit for the number of free electrons (or holes) per unit volume.

The thermoelectric effect also furnishes information about the mechanism of conduction in a semiconductor.[129] The thermoelectric power of an excess conductor against a metal is negative, and that of a defect conductor against a metal is positive. The results obtained by this method have checked those obtained by other methods.

Of the oxides whose conductivity has been found independent of oxygen partial pressure, Fe_2O_3 and CuO are assumed to be intrinsic semiconductors according to the original model of Wilson. The conductivities of Fe_3O_4 and Co_3O_4 are intimately connected with their crystal structures. According to Wagner[131] in Co_3O_4 there are two different types of cobalt position in the lattice (spinel type) and these are normally occupied by ions of different net charge. At high temperatures there will be some irregularity in this distribution and the swapping around of electrons from divalent to trivalent cobalt accounts for the conduction. At low temperatures this process is unlikely and we have a low conductivity. The case of Fe_3O_4 is somewhat similar, but much more favorable for electrical conduction. According to de Boer and Verwey,[113] the lattice structure of this substance corresponds to a statistical distribution of Fe^{++} and Fe^{+++} over crystallographically identical lattice points. The effect of an electric field is to produce slight changes in this distribution so that it is easy to see why this substance should have a conductivity ($\backsimeq 10^{+2}$ at room temperatures) which is many times that of other oxides of the iron group ($\backsimeq 10^{-8}$ at room temperatures).

De Boer and Verwey have suggested a model to explain the dependence of the conductivity of NiO on oxygen pressure. When excess oxygen is absorbed into the lattice, it is assumed that holes occur in the cation lattice. Each added oxygen atom becomes an O^{--} ion by removing electrons from two nickel ions, so that we have in the anion lattice a regular distribution of oxygen ions, and in the cation lattice a number of vacant places and twice this number of Ni^{+++} ions. The vacant lattice site, as we have pointed out in another connection, has an external field qualitatively like that of an ion of opposite sign. This means that the Ni^{+++} ions and the holes will be associated to some extent. However, in the neighborhood of a vacant point, the potential energy of an electron is raised somewhat above the rest of the lattice so that we will have special

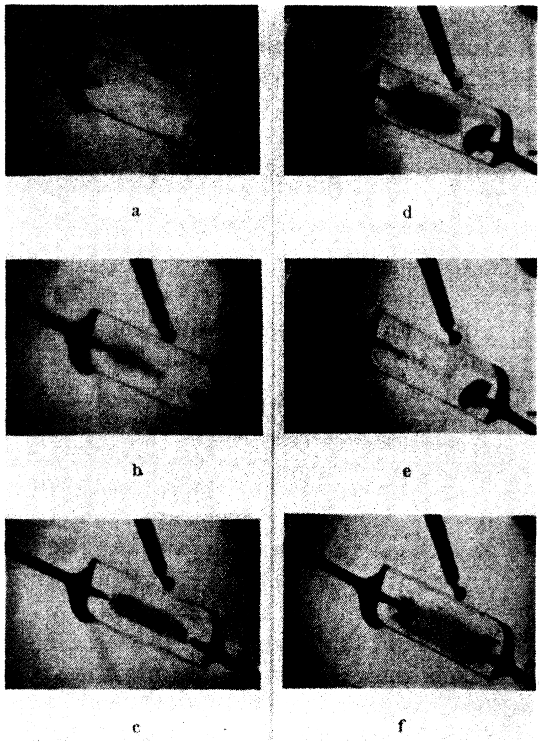

FIG. 24. Formation and movement of F centers in an NaCl crystal.

levels of somewhat higher energy as indicated in the diagram (Fig. 23). At absolute zero there can be a shift in charge only among the ten Ni^{++} and two Ni^{+++} ions which surround a vacant lattice point. At higher temperatures there will be some transitions of electrons to Ni^{+++} ions from Ni^{++} ions outside the ring which surrounds a vacant point. We shall then have a situation similar to that in Fe_2O_3. The number of "free" Ni^{+++} ions created will increase with temperature, showing that this is a model which explains satisfactorily the observed increase of conductivity with temperature.

Another indication that the excess oxygen creates special levels is afforded by the fact that ordinary NiO is colored a light yellowish-green, but NiO with an excess of oxygen is black. De Boer and Verwey suggested that the absorption is due to transitions from the levels marked a and a' to the $4s$ conduction band. In general, there is only a qualitative and not a quantitative connection between the long wave-length threshold and the energy difference ΔE in Eq. (43). The reason for this has been pointed out by a number of authors.[133] When an electron is released optically from the highest level of the lower band to the lowest level of the higher band, there is no time for polarization of the surrounding medium to take place; hence, the energy is greater in this case than in the slower thermal process when polarization can take place.

The theoretical discussion of the temperature variation of conductivity is actually more complicated than appeared in what was said previously. For an intrinsic semiconductor, the conductivity is proportional to exp $(-\Delta E/2kT)$. A similar expression is obtained if we assume that the number of impurities with which thermally-freed electrons can recombine is the same as the number of freed electrons. However, if we assume that the number of "traps" for a free electron is much greater than the number of free electrons, we obtain the result that the conductivity is proportional to exp $(-\Delta E/kT)$. This idea has been developed by Nijboer,[134] who shows that under certain circumstances the $\log \sigma$ *vs.* $1/T$ graph will have two different slopes in different temperature intervals—one corresponding to ΔE and the other to $\Delta E/2$.

One of the important practical applications of semiconductors is to the construction of rectifiers. The ordinary Cu_2O rectifier consists of a layer of Cu_2O on pure Cu.[135] Near the Cu there can be no excess oxygen and hence no defects. Therefore, a high resistance boundary layer is present. The Cu_2O further from the Cu is prepared in an excess of oxygen and hence is a defect conductor. Rectification is due to the fact that the resistance of the boundary layer is greater for current flowing from the oxide to the metal than for current flowing in the reverse direction. Older theories of rectification gave the wrong direction of easy flow. A recent paper by Mott[136] has proposed a different mechanism which is in agreement with the observed behavior.

B. Color centers and photoconductivity

In the past decade a new type of conduction in insulating crystals has been investigated—electronic conduction in alkali halide crystals containing coloring due to an excess of alkali, or of halogen, or to some special condition of the crystal. This discovery offers us a model of a metallic conductor in which the number of free electrons is only about one one-hundred-thousandth of the number in a typical metal. The material is transparent, enabling the observer to watch the progress of electricity through the crystal, to measure directly the density of charge carriers, the energy with which they are bound to the lattice, their mobility, and other properties. Valuable knowledge of the impurity type of semiconductor, of conduction in metals, and of contact rectifiers has been derived from a study of conduction in colored alkali halide crystals.[137]

It was discovered in 1932 by Stasiw[138] that electrons could apparently penetrate from a metallic cathode into an alkali halide crystal which was heated to a high temperature. A good description of this experiment has been given by von Hippel.[139] A crystal of NaCl is mounted in an electric furnace between two pointed copper wires fused into the ends to make good contact. A thermocouple is mounted in thermal contact with the crystal. An image of the crystal is projected upon a screen.

If the temperature is brought to about 550°C and an electric field of a few hundred volts per centimeter is applied between the electrodes, at first nothing visible takes place and an ammeter in the circuit registers the usual electrolytic current of about a milliampere. After a time a colored cloud suddenly emerges from the cathode (Fig. 24b) and moves toward the anode. All this time the current is increasing even though the voltage is constant (see Fig. 25, the interval between 20 and 60 seconds). As the cloud reaches the vicinity of the anode a deep red cap forms very quickly around the electrode at a distance which depends upon how much electrolysis took place before the colored cloud began to enter the crystal. (See Fig. 24c.)

Upon reversing the voltage, the edge of the colored cloud moves backward. There is one noticeable difference between the forward and the backward motion. During the forward motion the boundary is diffuse, but during the backward motion the boundary is much more definite. This is due to the fact that electrical conduction is much easier in the colored region, necessitating a smaller field to carry the current than is required in the clear section of the crystal. Thus, when the colored region is advancing toward the clear, any irregularity tends to be emphasized, for the path from cathode to anode lying to the greatest extent in the colored region offers the lowest over-all resistance and tends to carry more current. If, however, the boundary is moving toward the colored region, the condition for minimum resistance which decides the path of the current, tends to sharpen the boundary between clear and colored regions (Fig. 24d).

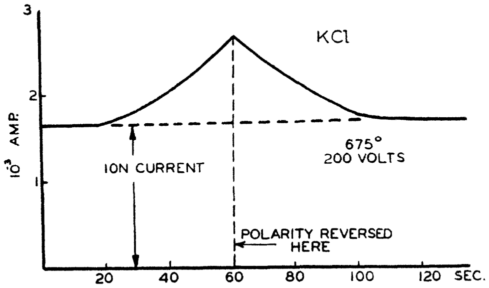

FIG. 25. Showing superposition of an electron current on the ion current.

While the crystal is being emptied of color, the current decreases steadily and reaches the original value for electrolytic conduction as the last of the colored cloud disappears at the electrode which was originally the cathode. This behavior is depicted in Fig. 25. When once formed, however, the deep red coloring about the anode remains and no further treatment will remove it. Analysis of that part of the crystal shows it to contain copper[140] from the anode which entered the crystal as copper ions to replace the sodium ions moved towards the cathode by electrolysis. As long as the copper was in ionic form it could be detected only by ultraviolet light. The copper ions have a greater affinity for electrons in the lattice than do the sodium ions, so that when they were touched by the colored cloud, the copper ions, due to their relatively high affinity for electrons, became neutralized to remain in the crystal as impurities.

If we continue to apply voltage in the reverse direction after the crystal has been swept clear, for a time the cap of copper impurity atoms over the new cathode suppresses the entrance of a new cloud from this end of the crystal and conduction is purely electrolytic. Sooner or later, a colored cloud breaks through, diffusing through the crystal very much as it had done previously from the opposite electrode. Upon approaching the new anode, a red-colored region appears here also, showing that copper ions have penetrated this end as well, while electrolysis was taking place in the crystal.

Electronic conduction in insulating crystals can also be produced by suitable illumination. This is known as photoconduction. The earlier work has been summarized in review articles in this journal[171, 172] and elsewhere, so we shall describe it only briefly. Some of the features which were not understood when Hughes' article was written have since been explained, and tentative explanations of some of the others have been suggested. We shall limit the discussion to cases where impurities have been intentionally introduced and where they are known to influence the conductivity greatly. Such crystals are known as allochromatic conductors.

The colored alkali halides have been extensively studied and some of their features satisfactorily explained. A pure alkali halide crystal shows absorption in the infra-red, which is due to ionic vibrations, and a very strong absorption in the ultraviolet. The interpretation of this absorption on the band picture (and its necessary refinements) has been given by Slater and Shockley.[143] Since no effect on the conductivity by illumination in this absorption band has been observed, we will not discuss it further.

Crystals which have been subject to various treatments show absorption in different regions of the spectrum. In alkali halides there are three types of absorption band which have been studied extensively. On the visible side of the characteristic ultraviolet absorption, there is a region of comparatively weak absorption known as the *U* band. (See Fig. 26.) In the visible region there is an absorption region known as the *F* band. Further toward the infra-red there is another absorption region known as the *F′* band.

The *F* band is the most interesting and easiest to observe. *F* centers can be produced in a number of ways. (1) In the experiment described previously, a cloud of *F* centers entered the crystal from a cathode point when the temperature and field were sufficiently high. (2) *F* centers are produced when light in the *U* band is absorbed by a crystal containing *U* centers. (3) Heating a crystal in the vapor of an alkali metal, not necessarily the same metal as the positive ions of the crystal, produces a uniform distribution of *F* centers whose density is a function of both the temperature and of the number of alkali atoms per cm^3 in the vapor. (4) Crystals are found in nature containing this coloration. (5) Irradiation with x-rays, or with ultraviolet light in the long wave-length foot of the characteristic absorption band, will produce *F* centers in an originally clear crystal.

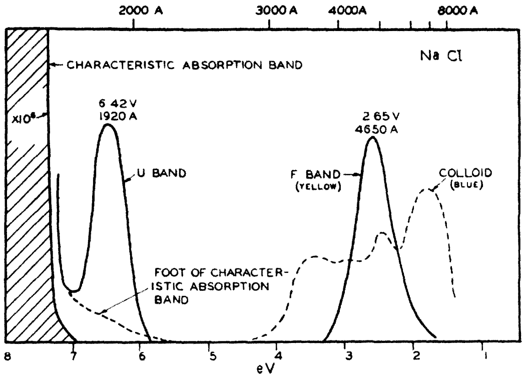

FIG. 26. Relative positions of the characteristic absorption band, the *U* band, the *F* band, and the band due to colloidal particles.

The first three methods, at least, are accompanied by destruction of the equality between the amount of metal and halogen. When dissolved in water, crystals colored by these methods show excess alkalinity.

When a crystal containing *F* centers is irradiated with light in the *F* band, *F′* centers are formed. The number of *F* centers decreases as shown by decreased absorption in the *F* band (Fig. 27). At the same time, the broad characteristic band due to *F′* centers appears at a longer wave-length. The crystal is said to be "excited" when containing *F′* centers; the transitions can be made to go the other way by flooding the crystal with red light in the *F′* band. At low temperatures the *F′* centers are stable, but at higher temperatures, they have only a short lifetime.

In measuring photoelectric currents in solid insulators, the sample is provided with electrodes and connected in series with a current indicator and a source of e.m.f. Illumination is provided

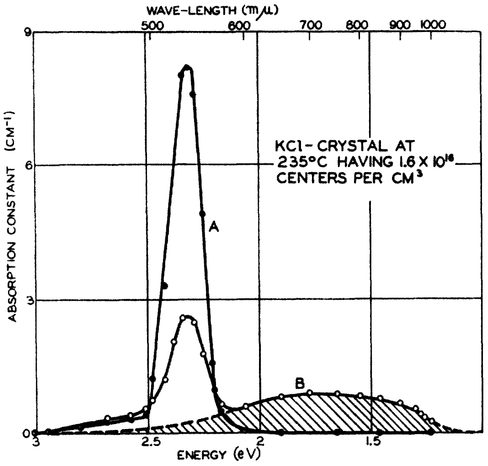

FIG. 27. Absorption in F centers and F' centers.

in either the transverse or longitudinal direction. For longitudinal illumination the electrodes are usually made of a transparent conducting liquid.

Gudden and Pohl[144] have shown that the total current which flows when the material is illuminated consists of two parts—a primary current and a secondary current. The primary current consists of photoelectrons moving toward the anode and a replacement current to keep the crystal neutral. This replacement current may be due to electrons from the cathode, or may be electrolytic. The secondary current usually exhibits a time lag and may persist a long time after illumination has been discontinued. It is observed as an increase in conductivity produced by the primary current. A secondary current is not observed in alkali halides, but in selenium it may be many times greater than the primary current.

In a nearly perfect lattice, or at low temperatures, insufficient replacement current can flow to keep the crystal neutral. Hence, a space-charge e.m.f. builds up which opposes the external voltage and causes a decrease in the observed current. This effect may be avoided by keeping the illumination low.

Figure 28 shows the primary photoelectric current in a NaCl crystal containing F centers for different temperatures. While at low temperatures, the dark current, which is due to ionic conduction, is negligible, at high temperatures it becomes much greater than the photoelectric current.

The data shown in Fig. 28 were obtained as follows: During the time-interval A, the crystal was illuminated in the F band. At 30°C, a current which is practically constant flows without time lag. If the temperature is higher, an additional current flows which depends upon the temperature and the time interval since illumination was begun. This is a replacement current due to ions and to thermally-released electrons, and is indicated by the shaded portion of the graph. The process may be pictured as follows: Electrons ejected from the F centers move a mean distance w under the influence of the field and become loosely bound in the crystal as F' centers. Measurements have shown that the mean value of w is proportional to the applied field E.

During the dark interval B, nothing happens if the temperature is low, but if the temperature is higher, the positive charges left behind by the photoelectrons are neutralized by ions and by electrons thermally released from F' centers.

During the time-interval C, the crystal was flooded with red light in the F' band. This releases electrons loosely bound to F' centers which find their way to the positive charges not already neutralized. The current produced in this way is shown by the cross-hatched parts of Fig. 28.

The nature of the color centers in alkali-halide crystals has been the subject of much speculation. Hilsch and Pohl[145] have suggested that U centers are due to hydrogen as an impurity and probably occur as H^- ions at otherwise vacant anion lattice points. Hilsch[146] made a lengthy study of the solution of hydrogen in KBr. After the crystal had been exposed to hydrogen for some time, it was exposed to potassium vapor and an optical method was used to measure the number of U centers. Another way of looking at the process of formation of U centers is that we have a small amount of alkali hydride in solid solution in the alkali halide. We should expect it to be less difficult to remove an electron from an H^- ion than from a Cl^- or other halogen ion. The observation that the U absorption is always on the visible side of the characteristic absorption is thus in agreement with the Hilsch-Pohl model of a U center.

Pohl's model of an F center was an electron trapped at an alkali ion to produce an alkali atom in the crystal. De Boer,[147] however, has suggested an alternate model which seems more reasonable. He suggested that an *F* center was formed when an electron was trapped at a vacant anion lattice point. In such a region there is a minimum potential energy for an electron, so that the hypothesis is reasonable. The *F* absorption band has been ascribed by Gurney and Mott to a transition from an *s*-like to a *p*-like state inside of the potential well.[148]

An *F′* center is said to be due to a second electron trapped at, or in the neighborhood of, an *F* center. One item of evidence for this is that conduction electrons are trapped by *F* centers to become *F′* centers.[173] The fact that a single light quantum in the *F* band destroys two *F* centers is also taken as evidence for the above hypothesis concerning an *F′* center.

C. Breakdown

The breakdown strength of a dielectric is measured by providing the specimen with electrodes, and applying a voltage which is raised at a uniform rate until puncture takes place. Field concentrations at the electrode edges must be eliminated, if possible, in order that the breakdown shall be at the center of the specimen. Von Hippel has developed a method for elimination of edge effects in breakdown tests.[149] He uses a dielectric guard ring of beeswax, or other material, under high pressure between the edge of the insulator and the electrode.

Materials of low conductivity, when tested at a low temperature, show a breakdown strength independent of the rate of voltage rise. If the circuit is opened before damage can be caused by the arcing current, the breakdown path will show no signs of melting or burning. This is called electronic or electrical breakdown. The other type, thermal breakdown, takes place in materials of higher conductivity, particularly if the conductivity increases rapidly with increasing temperature. Such a material breaks down at a much lower voltage for a slow rate of voltage rise than for an impulse. The breakdown path shows signs of fusion, or, perhaps carbonization in the case of an organic material. Thermal breakdown is very important commercially in high voltage oil-and-paper insulated cables and in transformers and rotating machines that operate at high temperatures. Fuoss[150] has demonstrated with polyvinyl chloride, a synthetic resin, that rapid increase of conductivity with temperature increases the likelihood of thermal breakdown. Wagner,[151] Hayden and Steinmetz,[152] and Güntherschulze,[153] discussed the thermal breakdown mechanism in 1922. Contributions were made by von Karman,[154] Dreyfus,[155] and Rogowski.[156] The conditions for thermal instability were worked out mathematically by Fock[157] and later, in a form more suitable for application, by Moon.[158] The work has been continued by Böning,[159] Gemant,[160] and others.

Electrical breakdown has received more attention from physicists recently because of its close connection with the electronic structure of matter. The electrical breakdown strength, obtained by impulse tests, may be several times as high as the thermal breakdown obtained when the voltage has been applied slowly. The breakdown mechanism depends also upon the temperature of the specimen. At low temperatures, breakdown is often electronic because the Joule heat created by conduction is removed before any hot spots can form. If we raise the temperature of the specimen, this may no longer be true and thermal breakdown sets in, as evidenced by the decrease in breakdown strength and the

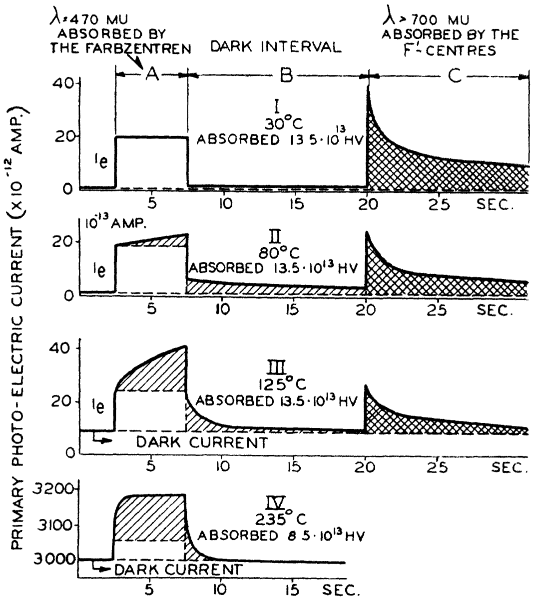

FIG. 28. Photoelectric primary current in sodium chloride crystal with *F* centers.

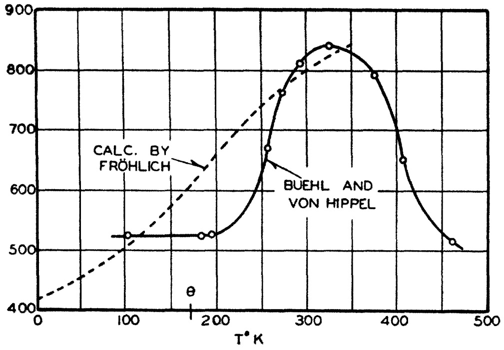

FIG. 29. The breakdown strength of KBr as $f(T)$. θ = Debye temperature.

dependence upon rate of voltage rise. Moon and Norcross have found an intermediate region where both mechanisms contribute.[161]

In the electrical breakdown region, according to Inge, Semenoff and Walther,[162] the breakdown strength is independent of temperature. According to them, this affords a convenient method for distinguishing between the mechanisms. Contradictory results have been obtained by Buehl and von Hippel,[163] who found that for pure ionic crystals, the electronic breakdown strength is not independent of temperature, but is very low at low temperatures, increasing to a maximum at the onset of thermal breakdown. Similar behavior was reported for KBr by Austen and Hackett.[164] Walther and Inge studied the breakdown of rocksalt in the temperature range of 300°–700°C and concluded that the breakdown mechanism is purely thermal at high temperatures.[165] They took exception to the belief expressed earlier by von Hippel that breakdown in the alkali halides is electrical in nature and that the decrease in strength at high temperatures is due to a decrease in the inherent electronic breakdown strength and should yield valuable information. Von Hippel replied to Walther and Inge, emphasizing that the type of breakdown can be established only by careful examination of the path taken by the discharge, provided it was arrested before subsequent melting. This, he says, they have failed to recognize.

Several theories of electronic breakdown, based on the quantum-mechanical picture of a solid dielectric, have been advanced. Zener,[166] and Cernuschi[167] studied the probability of transition of electrons from the filled band to the next unfilled band under the influence of an applied electric field. They take as their condition for breakdown that the probability of transition give an appreciable number of electrons in the empty band. We can see that this is similar to the assumption that, under high fields, an insulator begins to conduct like a metal. Postulating an energy gap of 5 volts between the bands, they obtained breakdown strengths of the order of those observed, $\backsimeq 10^7$ volts/cm. Their value is nearly independent of temperature and thus their theory is incapable of predicting the observed temperature variation.

Frölich,[168] and Seeger and Teller[169] have proposed theories of electronic breakdown based upon von Hippel's picture that breakdown occurs when thermally-released electrons gain energy from the field more rapidly than they lose it in collisions with the lattice. Fröhlich assumes that electrons which finally gain enough energy produce other electrons through ionization by collision. As is well known in the theory of metals, the mean free path of electrons increases greatly at low temperatures. This means that, at low temperatures the energy gained between collisions is greater than at higher temperatures and there is a greater probability of producing free electrons by collision. The variation of breakdown voltage with temperature as predicted by Frölich is in qualitative but not quantitative agreement with results obtained by Buehl and von Hippel (see Fig. 29). An indirect check of Frölich's analysis is the behavior of solid solutions of KCl—RbCl. As in alloys, the departure from perfect periodicity of the lattice decreases the mean free path. In Frölich's theory this would correspond to increases of the break-

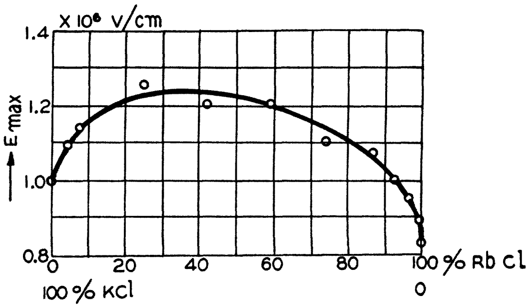

FIG. 30. Breakdown strength of mixed crystal.

down voltage for the solid solutions as compared with that for the components. Such results have actually been obtained by von Hippel[142,175] (see Fig. 30). Further studies of the theoretical aspects of the problem have been made by Seitz and Sampson,[176] but only an abstract has been published.

ACKNOWLEDGMENTS

The authors wish to thank Dr. C. F. Hill for many helpful suggestions during the writing of this report, and Dr. E. U. Condon for reading and criticizing the manuscript.

BIBLIOGRAPHY

Books and general articles

Chapter III

G. von Hevesy, Chapter 7 in *Handbuch der Physik*, Vol. XIII (1928).

C. Tubandt, Part III in *Handbuch der Experimentalphysik*, Vol. XII (1932).

Chapter IV

A. Joffé, *The Physics of Crystals* (McGraw-Hill, 1928).

A. Smekal, Chapter 5 in *Handbuch der Physik*, Vol. XXIV/2 (1933).

A. Smekal, Die Physik **4**, 17 (1936).

W. Seith, Zeits. f. Elektrochemie **42**, 635 (1936).

Chapter V

J. B. Whitehead, *Dielectric Theory and Insulation* (McGraw-Hill, 1928).

E. v. Schweidler, Ann. d. Physik **24**, 711 (1907).

W. A. Yager, J. App. Phys. **7**, 434 (1936).

Chapter VI

W. Jost, *Diffusion und Chemische Reaktion in Festen Stoffen* (Theodore Steinkopf, Dresden, 1937).

W. Jost, Physik. Zeits. **36**, 757 (1935).

C. Wagner, Physik. Zeits. **36**, 720 (1935).

B. Gudden and W. Schottky, Physik. Zeits. **36**, 715 (1935).

Chapter VII

The most recent and most complete book on both the experimental and theoretical aspects of all types of solids is *Modern Theory of Solids* by F. Seitz (McGraw-Hill, 1940). Other books and review articles on the electronic structure of solids are:

A. Sommerfeld and H. Bethe, Chapter 3, *Handbuch der Physik*, Vol. XXIV/2 (1933).

J. C. Slater, Rev. Mod. Phys. **6**, 208 (1934).

N. F. Mott and H. Jones, *The Properties of Metals and Alloys* (Oxford Univ. Press, 1936).

A. H. Wilson, *The Theory of Metals* (Camb. Univ. Press, 1936).

H. Fröhlich, *Elektronentheorie der Metalle* (Julius Springer, 1936).

Nonmathematical accounts have been published by F. Seitz and R. P. Johnson, J. App. Phys .2, 84 (1937), and W. Shockley, Bell Sys. Tech. J. **17**, 645 (1939).

Most of these books assume some knowledge of quantum mechanics. For an introductory account, see R. W. Gurney, *Elementary Quantum Mechanics* (Cambridge Univ. Press, 1936). For a more comprehensive treatment see L. Pauling and E. B. Wilson, *An Introduction to Quantum Mechanics* (McGraw-Hill, 1935).

Chapter VIII

For semiconductors

B. Gudden, Ergeb. d. exakt. Naturwiss. **13**, 223 (1934).

A. H. Wilson, *Semi-Conductors and Metals* (Cambridge University Press, 1939).

Photoconductivity

R. W. Pohl, "Photoconductivity and color-centers," Proc. Phys. Soc. **49** (extra part) (1937).

A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena* (McGraw-Hill, 1932), Chapter 8.

F. C. Nix, Rev. Mod. Phys. **4**, 723 (1932).

A. L. Hughes, Rev. Mod. Phys. **8**, 294 (1936).

Breakdown

A. v. Hippel, Ergeb. d. exakt. Naturwiss. **14**, 79 (1935).

REFERENCES

1. E. T. Hoch, Bell Sys. Tech. J. **5**, 555 (1926).
2. Am. Soc. Testing Materials, Committee D-9 Standards on Electrical Insulation Materials (1939), p. 299.
3. J. Hopkinson, Phil. Mag. [5] **2**, 314 (1876).
4. S. Kyropoulos, Zeits. f. anorg. allgem. Chemie **154**, 308 (1926).
5. K. Korth, Zeits. f. Physik **84**, 677 (1933).
6. R. Hilsch, Angew. Chemie **49**, 69 (1936).
7. H. Walther, Rev. Sci. Inst. **8**, 406 (1937).
8. G. M. Voglis, Zeits. f. Physik **109**, 52 (1938).
9. J. Curie, Ann. Chim. Phys. [6] **17**, 385 (1888); [6] **18**, 203 (1889).
10. A. H. and K. T. Compton, Phys. Rev. **14**, 85 (1919).
11. G. F. Metcalf and B. J. Thompson, Phys. Rev. **36**, 1489 (1930).
12. L. A. du Bridge, Phys. Rev. **37**, 392 (1931).
13. W. B. Nottingham, J. Frank. Inst. **209**, 287 (1930).
14. D. B. Penick, Rev. Sci. Inst. **6**, 115 (1935).
15. A. W. Vance, Rev. Sci. Inst. **7**, 489 (1936).
16. A. L. Norbury, quoted by Mott and Jones see next reference.
17. Mott and Jones, *Properties of Metals and Alloys* (Oxford, 1936), p. 169.
18. C. Wagner, Zeits. f. physik. Chemie, **B21**, 42 (1933).
19. G. C. Schmidt, Ann. d. Physik **82**, 664 (1927).
20. E. Warburg, Ann. d. Physik **21**, 622 (1884).
21. C. Tubandt, Ber. d. Natur. f. Geselbach Z. Halle **4**, 21 (1917).
22. C. Tubandt, *Handbuch der Experimentalphysik*, Vol. XII (1932), p. 396.

23. C. Tubandt, *Handbuch der Experimentalphysik*, Vol. XII (1932), p. 397. C. Tubandt, H. Reinhold and G. Liebold, Zeits. f. anorg. allgem. Chemie, **197**, 225 (1931).
24. C. Tubandt, *Handbuch der Experimentalphysik*, Vol. 12, p. 399. C. Tubandt and H. Reinhold, Zeits. f. Elektrochem. **37**, 589 (1931).
25. C. Tubandt, S. Eggert, and G. Schibbe, Zeits. f. anorg. allgem. Chemie, **117**, 1 (1921).
26. G. von Hevesy and W. Seith, Zeits. f. Physik **56**, 790 (1929).
27. J. Steigman, W. Shockley and F. C. Nix, Phys. Rev. **56**, 13 (1939).
28. C. Tubandt, H. Reinhold, and W. Jost, Zeits. f. physik Chemie **129**, 69 (1927).
29. F. Klaiber, Ann. d. Physik **3**, 229 (1929).
30. C. Wagner, Zeits. f. physik. Chemie **B11**, 139 (1930).
31. W. Jost, Zeits. f. physik. Chemie **B16**, 129 (1932).
32. C. Tubandt, *Handbuch der Experimentalphysik*, Vol. XII, p. 468.
33. C. Wagner, Zeits. f. physik. Chemie **B21**, 25 (1933).
34. W. Jost and H. Ruter, Zeits. f. physik. Chemie **B21**, 48 (1933).
35. C. Wagner, Zeits. f. physik. Chemie **B23**, 469 (1933).
36. C. Tubandt and H. Reinhold, Zeits. f. physik. Chemie **B24**, 22 (1934).
37. T. Phipps W. Lansing and T. Cooke, J. Am. Chem. Soc. **48**, 112 (1926).
38. E. G. Rochow, J. App. Phys. **9**, 664 (1938).
39. C. Brennecke, J. App. Phys. **11**, 202 (1940).
40. W. Seith, Zeits. f. Physik **56**, 802 (1929).
41. W. Lehfeldt, Zeits. f. Physik **85**, 717 (1933).
42. E. Rasch and F. Hinrichson, Zeits. f. Elektrochem. **14**, 41 (1908).
43. A. Smekal, Zeits. f. Physik **58**, 322 (1929).
44. C. Tubandt, *Handbuch der Experimentalphysik*, Vol. XII, p. 442.
45. A. Smekal, *Handbuch der Physik* (Vol. XXIV/2 (1933), p. 876.
46. S. W. Richardson, Proc. Roy. Soc. **A92**, 42 (1915).
47. S. W. Richardson, Proc. Roy. Soc. **A92**, 101 (1916).
48. S. W. Richardson, Proc. Roy. Soc. **A107**, 101 (1925).
49. A. Joffé, Ann. d. Physik, **72**, 461 (1923).
50. A. Joffé, *The Physics of Crystals* (McGraw-Hill Book Co., 1928), p. 79.
51. T. E. Phipps and R. T. Leslie, J. Am. Chem. Soc. **50**, 2412 (1928).
52. D. C. Ginnings and T. E. Phipps, J. Am. Chem. Soc. **52**, 1341 (1930).
53. T. Phipps and E. Partridge, J. Am. Chem. Soc. **51**, 1331 (1929).
54. W. Jost, Zeits. f. physik. Chemie **B7**, 234 (1930).
55. C. Wagner, Zeits. f. physik. Chemie **B22**, 181 (1933).
56. A. Goldhammer, Zeits. f. Physik. **52**, 708 (1928).
57. A. Goldhammer and N. Salessky, Physik. Zeits. Sowjetunion **5**, 553 (1934).
58. A. Goldhammer, Zeits. f. Physik **51**, 173 (1929).
59. N. Salessky, Zeits. f. Physik **52**, 695 (1928).
60. F. Quittner, Physik. Zeits. Sowjetunion **6**, 319 (1934).
61. E. Hippauf and R. Stein, Physik. Zeits. **39**, 90 (1938).
62. W. O. Shumann, Zeits. f. Physik **79**, 532 (1932).
63. G. Jaffé, Ann. d. Physik **16**, 217 (1933).
64. A. Joffé, D. Rojansky and K. Sinjelnikow, Zeits. f. Physik **66**, 143 (1930).
65. A. Smekal, Physik. Zeits. **36**, 742 (1935).
66. O. Beran and F. Quittner, Zeits. f. Physik **64**, 760 (1930).
67. B. Hochberg and V. Joffé, Zeits. f. Physik **66**, 172 (1930).
68. M. Kassel, Zeits. f. Physik **90**, 287 (1934).
69. A. Wenderowitsch and A. Drisina, Zeits. f. Physik **98**, 108 (1935).
70. Z. Gyulai, Zeits. f. Physik. **67**, 812 (1931).
71. R. Ketzer, Zeits. f. Elektrochem. **26**, 77 (1920).
72. E. Koch and C. Wagner, Zeits. f. physik. Chemie **B38**, 295 (1937).
73. J. Hopkinson, Phil. Mag. **2**, 314 (1876).
74. A. Joffé, *The Physics of Crystals* (McGraw-Hill, 1928), p. 75.
75. A. Smekal, Die Physik **4**, 14 (1936).
76. A. Smekal, *Handbuch der Physik*, Vol. XXIV/2, p. 878.
77. H. Poole, Phil. Mag. **32**, 112 (1916); **34**, 195 (1917); **42**, 488 (1921).
78. A. Joffé, Zeits. f. Physik **48**, 288 (1928).
79. B. Hochberg, Zeits. f. Physik. **70**, 635 (1931).
80. W. von Siemans, Pogg. Ann. **125**, 137 (1864).
81. J. Hopkinson, Phil. Trans. **166**, 489 (1876); **167**, 599 (1877); Proc. Roy. Soc. [6] **25**, 496 (1876).
82. E. von Schweidler, Ann. d. Physik **24**, 711 (1907).
83. F. Tank, Ann. d. Physik **48**, 307 (1915).
84. R. R. Benedict, Trans. A. I. E. E. **49**, 221 (1930).
85. J. B. Whitehead, *Impregnated Paper Insulation* (McGraw Hill, 1935), p. 190.
86. J. B. Whitehead, J. Frank. Inst. **208**, 453 (1929).
87. J. B. Whitehead, Physics **2**, 89 (1932).
88. J. C. Maxwell, *Electricity and Magnetism* (Oxford Press, 1892), Vol. I, p. 328.
89. H. Pellat, Ann. Chim. Phys. **18**, 150 (1899).
90. K. W. Wagner, Archiv. f. Elektrotech. **3**, 83 (1914).
91. W. A. Yager, Physics **7**, 434 (1936).
92. J. Frenkel, Zeits. f. Physik. **35**, 652 (1926).
93. C. Wagner and W. Schottky, Zeits. f. physik. Chemie **B11**, 163 (1930).
94. W. Jost, *Diffusion und Chemische Reaktion in Festen Stoffen* (Steinkopf, Dresden, 1937), Chap. II.
95. W. Schottky, Zeits. f. physik. Chemie **B29**, 335 (1935).
96. W. Jost, J. Chem. Phys. **1**, 466 (1933).
97. J. C. Slater, *Introduction to Chemical Physics* (McGraw-Hill, 1939), p. 388.
98. W. Jost and G. Nehlep, Zeits. f. physik. Chemie **B32**, 1 (1936).
99. W. Jost, Trans. Faraday Soc. **34**, 860 (1938).
100. N. F. Mott and M. J. Littleton, Trans. Faraday Soc. **34**, 485 (1938).
101. W. Jost and G. Nehlep, Zeits. f. physik. Chemie **B34**, 348 (1936).
102. A. Smekal, *Handbuch der Physik*, Vol. XXIV/2, p. 883.
103. C. Tubandt, *Handbuch der Experimentalphysik*, Vol. XII, p. 402.

104. W. Jost, Physik. Zeits. **36**, 757 (1935).
105. F. Seitz, Phys. Rev. **54**, 1111 (1938).
106. R. W. G. Wyckoff, *The Structure of Crystals* (Supplement) (Reinhold, New York, 1935).
107. J. A. A. Ketelaar, Zeits. f. Krist. **87**, 436 (1934); Trans. Faraday Soc. **34**, 874 (1938).
108. L. W. Strock, Zeits. f. physik. Chemie **B25**, 441 (1935).
109. H. Reinhold and H. Bräuniger, Zeits. f. physik. Chemie **B41**, 397 (1938) and other references there cited.
110. N. F. Mott and H. Jones, *The Properties of Metals and Alloys* (Oxford, 1936), p. 69.
111. G. von Hevesy, *Handbuch der Physik*, Vol. XIII, Chapter 7, p. 266.
112. F. Hund, Physik. Zeits. **36**, 725 (1935).
113. J. H. de Boer and E. J. Verwey, Proc. Phys. Soc. **49** (extra part), 59 (1937).
113a. R. Peierls, Proc. Phys. Soc. **49** (extra part), 72 (1937).
114. B. Gudden, Ergeb. d. exakt. Naturwiss. **13**, 224 (1934).
115. A. H. Wilson, Proc. Roy. Soc. **A133**, 458 (1931).
116. A. H. Wilson, Proc. Roy. Soc. **A134**, 277 (1931).
117. H. Fröhlich, *Elektronentheorie der Metalle* (Springer, 1936), p. 225.
118. B. Gudden, Erl. Ber. **62**, 289 (1930).
119. R. H. Fowler, *Statistical Mechanics* (Cambridge, 1936) p. 398, ff.
120. K. Baedecker, Ann. d. Physik **4**, 29, 566 (1909); Physik. Zeits. **13**, 1080 (1912).
121. C. Tubandt, E. Rindtorf and W. Jost, Zeits. f. anorg. allgem. Chemie **165**, 195 (1927).
122. J. Gundermann, K. Hauffe and C. Wagner, Zeits. f. physik. Chemie **B37**, 148 (1937).
123. H. von Baumbach and C. Wagner, Zeits. f. physik. Chemie **B24**, 59 (1934).
124. C. Wagner and E. Koch, Zeits. f. physik. Chemie **B32**, 439 (1936).
125. H. von Baumbach and C. Wagner, Zeits. f. physik. Chemie **B22**, 199 (1933).
126. W. Meyer and A. Schmidt, Zeits. f. tech. Physik **13**, 137 (1932).
127. H. von Baumbach, H. Dunwald and C. Wagner, Zeits. f. physik. Chemie **B22**, 226 (1933).
128. C. Wagner, Zeits. f. physik. Chemie **B22**, 181 (1933).
129. A. H. Wilson, *Semi-conductors and Metals* (Cambridge, 1939), Chapter IV.
130. C. Wagner, Zeits. f. physik. Chemie **B22**, 195 (1933).
131. C. Wagner, Physik. Zeits. **36**, 721 (1935).
132. W. Schottky and F. Waibel, Physik. Zeits. **34**, 858 (1933); **36**, 912 (1935).
133. N. F. Mott, Science Progress **32**, 657 (1938).
134. B. R. A. Nijboer, Proc. Phys. Soc. **51**, 578 (1939).
135. L. O. Grondahl, Rev. Mod. Phys. **5**, 141 (1933).
136. N. F. Mott, Proc. Roy. Soc. **A171**, 27 (1939).
137. R. Hilsch and R. W. Pohl, Zeits. f. Physik. **111**, 399 (1938).
138. O. Stasiw, Göttingen Nachrichten Math.-Phys. Kl. 261 (1932).
139. A. von Hippel, Zeits. f. Physik. **98**, 580 (1936).
140. S. Arzybysev *et al.*, Comptes rendus U.S.S.R. **1**, 177; **3**, 23; **4**, 28 (1934).
141. A. von Hippel, Zeits. f. Physik. **98**, 588 (1936).
142. A. von Hippel, Ergeb. d. exakt. Naturwiss. **14**, 79 (1935).
143. J. C. Slater and W. Shockley, Phys. Rev. **50**, 705 (1936).
144. B. Gudden and R. W. Pohl, Zeits. f. Physik **6**, 248 (1921).
145. R. Hilsch and R. W. Pohl, Göttingen Nachrichten Math.-Phys. Kl., N.S. **2**, 139 (1936).
146. R. Hilsch, Ann. d. Physik **29**, 407 (1937).
147. J. H. de Boer, Rec. Trav. Chim. Pays. Bas. **56**, 301 (1937).
148. R. W. Gurney and N. F. Mott, Trans. Faraday Soc. **34**, 506 (1938).
149. A. von Hippel, Zeits. f. Physik. **75**, 145 (1932).
150. R. M. Fuoss, J. Am. Chem. Soc. **60**, 456 (1938).
151. K. W. Wagner, J. A. I. E. E. **41**, 1034 (1922).
152. J. L. R. Hayden and C. P. Steinmetz, Elec. World **80**, 865 (1922).
153. A. Guntherschulze, Jahrb. d. Radioakt. u. Elektr. **19**, 92 (1922).
154. T. von Karman, Archiv f. Elektrotechnik **13**, 174 (1924).
155. L. Dreyfuss, Schweiz, Elektr. Verein Bull. **15**, 321, 577 (1924).
156. W. Rogowski, Archiv f. Elektrotechnik **13**, 153 (1924); **18**, 123 (1927).
157. V. Fock, Archiv f. Elektrotechnik **19**, 71 (1927).
158. P. H. Moon, Elec. Eng. **50**, 676 (1931).
159. P. Böning, Bull. Swiss Electrotech. Assn. (1938), p. 368.
160. A. Gemant, J. Frank. Inst. **228**, 79 (1939).
161. P. H. Moon and A. S. Norcross, J. A. I. E. E. **49**, 125 (1930).
162. L. Inge, N. Semenoff, and A. Walther, Archiv f. Elektrotechnik **17**, 433 (1926).
163. R. C. Buehl and A. von Hippel, Phys. Rev. **56**, 941 (1939).
164. A. E. W. Austen and W. Hackett, Nature **143**, 637 (1939).
165. A. Walther and L. Inge, Tech. Phys. U.S.S.R. **5**, 335 (1938).
166. C. Zener, Proc. Roy. Soc. **A145**, 523 (1934).
167. F. Cernuschi, Proc. Camb. Phil. Soc. **32**, 276 (1936).
168. H. Fröhlich, Proc. Roy. Soc. **A160**, 230 (1937); Phys. Rev. **56**, 349 (1939).
169. R. J. Seeger and E. Teller, Phys. Rev. **54**, 515 (1938); **56**, 352 (1939).
170. W. Hartman, Zeits. f. Physik **102**, 709 (1936).
171. F. C. Nix, Rev. Mod. Phys. **4**, 723 (1932).
172. A. L. Hughes, Rev. Mod. Phys. **8**, 294 (1936).
173. N. F. Mott, Proc. Phys. Soc. **50**, 196 (1938).
174. R. W. Pohl, Proc. Phys. Soc. **49** (extra part), 3 (1937).
175. A. von Hippel, J. App. Phys. **8**, 815 (1937).
176. F. Seitz and J. B. Sampson, Phys. Rev. **57**, 558A (1940).

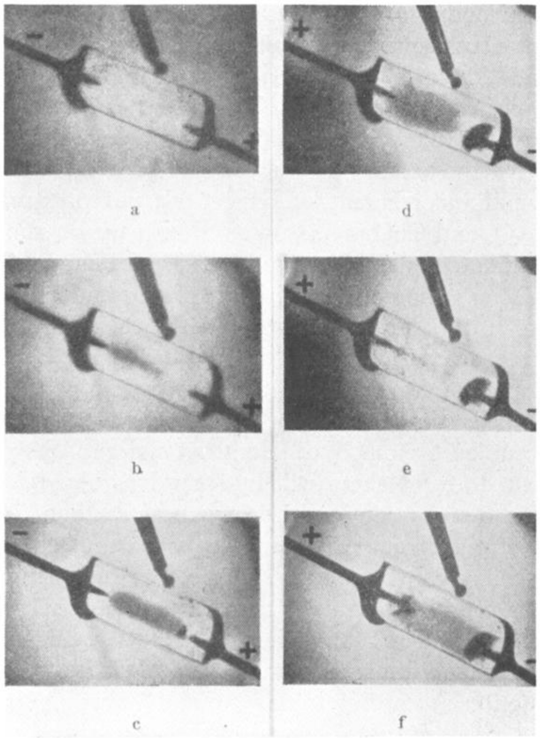

FIG. 24. Formation and movement of *F* centers in an NaCl crystal.