The Theory of Impurity Conduction†

By N. F. MOTT and W. D. TWOSE‡
Department of Physics, University of Cambridge

CONTENTS

§ 1. INTRODUCTION

THE purpose of this article is to study one of the ways in which electricity can flow in a semiconductor containing impurities. The current in an impure semiconductor is due to two competing conduction processes which act in parallel. The first process is responsible for the current usually observed which (in for example n-type material) is carried by electrons in the conduction band in thermal equilibrium with electrons on donor impurities. The second process arises as follows: An electron occupying an isolated donor has a wave function localized about the impurity and an energy slightly below the conduction band minimum. Because there is a small but finite overlap of the wave function of an electron on one donor with neighbouring donors, a conduction process is possible in certain circumstances in which the electron moves between centres by tunnel effect without activation into the conduction band. This we call *impurity*

---

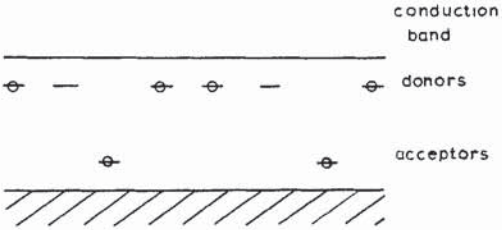
† Supported in part by the U.S. Air Force.

‡ Now at the Institute for the Study of Metals, University of Chicago.

conduction. The electrons of high mobility in the conduction band completely dominate the conductivity at higher temperatures. However, although the mobility of an electron moving in the impurity levels is very small since it depends on interaction between widely spaced impurities, at low temperatures impurity conduction will dominate due to the absence of electrons in the conduction band.

The circumstance in which impurity conduction is possible is the presence of 'compensation', by which we mean the presence of some minority centres, acceptors in an n-type conductor. These accept electrons from a certain proportion of the donors, thus allowing the movement of electrons from an occupied donor into an occupied one (fig. 1). Without compensation impurity conduction is not possible, unless the overlap between the centres is very large; when this is large enough, corresponding to a critical concentration N_c, another form of conductivity sets in, in which the electrons behave like a degenerate electron gas. This we shall treat in part II of this report and shall describe as metallic impurity conduction.

Fig. 1



Energy diagram of an n-type semiconductor containing donors and acceptors. The horizontal lines represent centres, the circles electrons in them.

Impurity conduction† was first observed by Busch and Labhart (1946) in silicon carbide, and has since been observed in a large number of both n- and p-type semiconductors; references are given in § 4. The concept of an electron bound to a donor centre is complementary to that of a hole bound to an acceptor. Hence our discussion, which for convenience is based on n-type, can readily be carried over to p-type material. Most experimental work has been done on the valence semiconductors, germanium and silicon, and this article is mainly about these.

A feature of impurity conduction, which distinguishes it from the usual semiconduction, is its extreme sensitivity to impurity concentration. For example, a change by a factor of 30 in the density of impurities in germanium can alter the conductivity by the impurity process by a factor of 10^7 (fig. 4), while the corresponding change in the conductivity in the conduction band in the exhaustion range of temperatures (i.e. at those when the electrons are nearly all in the conduction band) is only of order 28. Another

---

† The possibility of impurity conduction was suggested on theoretical grounds by Schottky in 1935 (see Gudden and Schottky 1935).

feature is that, when the impurity concentration is small, the curves plotting $\ln \rho$ against $1/T$ exhibit a finite slope in the temperature range where impurity conduction predominates, suggesting that the charge transfer between impurity centres must itself be thermally activated. Above the critical concentration N_c mentioned above, the resistivity becomes independent of temperature; the conductivity is then apparently metallic and carriers move freely without thermal activation. There is a small transition region (a factor of order 4 in the impurity concentration in germanium) just below N_c, in which the conductivity is non-metallic but the slopes of the curves ($\ln \rho$ vs. $1/T$) decrease and finally vanish at the critical concentration. A complicated temperature dependence of the Hall effect is also observed in this region (fig. 5).

The theoretical interest of these phenomena is two-fold. First they give the opportunity of studying the transition from metallic to non-metallic conduction which occurs as the concentration of carriers is decreased. It has been postulated by one of us in a number of papers (Mott 1949, 1952, 1956, 1957, 1961) that as the lattice spacing of an *ordered* array of atoms is increased there should be a sharp transition from a metallic to a non-metallic state of the valence electrons. The theoretical treatment of this transition is a many-body problem, involving the interactions between the electrons. No way is known, except possibly the use of high pressures, of changing the interatomic distance of a crystalline array of atoms over a large enough range, so that to test this hypothesis we are driven back to a study of a disordered array of centres such as occurs in doped germanium or silicon.

The second point of interest is the mobility of an electron in a disordered lattice, considered as a one-body problem. For high concentrations giving metallic conductivity, we have a problem like that of a liquid metal, but with a greater degree of disorder. For low concentrations, it appears that the electron moves by a hopping process from one centre to another, interaction with phonons being essential and the concept of a mean free path not appropriate.

The article therefore divides naturally into two parts. In the first we shall be concerned with the experimental observation and calculation of impurity conductivity in the region of low concentration. We discuss in § 5 the reasons for using localized states and a phonon-activated hopping process in this low concentration region. The effect of compensating impurities and disorder is considered, and the theory of the interaction of localized carriers with lattice vibrations is traced through from the limits of strong coupling (polar semiconductors) to weak coupling (valence semiconductors). Although we describe this process as a one-body problem, we must in our applications of Fermi statistics introduce the interaction between electrons in the sense that an electron cannot move into an impurity centre that is already occupied. The second part of this article will have as its theme the interaction between carriers in the impurity centres when this becomes large enough to lead to a transition to a metallic form of conductivity. This second section will start from an assumption of what is to be expected from a crystalline arrangement of centres and then discuss the effect of disorder and compensation.

PART I

§ 2. THE MODEL AND GENERAL DISCUSSION OF IMPURITY CONDUCTION

If we neglect the dependence on direction of the effective mass, an electron occupying an isolated donor can be taken to move in a hydrogen-like orbit in the Coulomb field of the donor ion, with a Bohr radius

$$a_0 = \kappa\,(m/m^*)\,a_{\mathrm{H}}$$

where m^* is the effective mass and a_{H} is the radius of a hydrogen atom (0·54 Å). Because of the large dielectric constant κ and small effective mass ratio m^*/m, the orbit may extend over several hundred of the host lattice sites. The energy of this state lies slightly below the lowest state of the conduction band. A similar description applies to a hole bound to a negative acceptor impurity; in this case the energy of the vacant state lies slightly above the top of the valence band. A more exact description of the impurity states is given in § 3, taking into account the dependence of m^* on direction.

In an n-type semiconductor (one in which the donor concentration N_{D} exceeds the accept or concentration N_{A}), at the absolute zero of temperature all the acceptors will be occupied and consequently negatively charged. The number of donors occupied and therefore neutral is $N_{\mathrm{D}}-N_{\mathrm{A}}$ (fig. 1). Overlap between wave functions corresponding to neighbouring sites allows movement from an occupied to an empty donor without activation into the conduction band. Our study of impurity conduction will therefore be a study of transport of electrons in a *random* lattice from one positively charged donor to another and in the field of fixed negatively charged acceptors. The host crystal is regarded as a dielectric medium in which this random impurity lattice is imbedded; thermal energy is supplied by vibrations of the host crystal.

As already emphasized in the introduction, at high concentrations of impurity the resistivity and Hall coefficient become independent of temperature at low temperatures, the electrons behaving like a degenerate electron gas. This behaviour, illustrated in figs. 4 and 5, is discussed in part II. In this section we consider low concentrations. There are two particularly simple cases which we may dicuss. If N_{D} is the concentration of majority centres (say donors) and N_{A} the concentration of minority centres (acceptors), these are:

Case (*a*). $N_{\mathrm{D}} \gg N_{\mathrm{D}}-N_{\mathrm{A}}$

There will then be a *small* number $(N_{\mathrm{D}}-N_{\mathrm{A}})$ of electrons in the donor states, and, owing to the random arrangement of immobile positively and negatively charged centres, a random fluctuation in the potential energy from one centre to another. The question then arises, as in all these considerations (§ 5), whether in the absence of lattice vibrations the characteristic wave functions (solutions of the Schrödinger equation for a single electron) are localized or whether they spread through the lattice. By a localized wave function we mean one that decays exponentially to zero at large enough distances from a given point in space. Considerations set out in § 7 show that in a one-dimensional lattice they are always localized. In a three-dimensional lattice they are localized if the degree of disorder, or the ratio of the energy in the random field to the band width, are great enough (§ 5). We believe this to be the case in the range of concentration for which experimental measurements are made; the electron can then jump from one centre to another only with the help of *phonons*. The process by which it does so is a main theme of this report. But it is possible that, as the concentration of donors increases, there may be formed unbound states so that the activation energy for motion would be zero. It is possible that a 'crystallization' of electrons, as envisaged by Wigner (1938) may occur, and a transition to a metallic state (condensed electron gas) only for higher concentrations of electrons. This is discussed in part II and by Mott (1961).

Case (*b*). $N_D \gg N_A$

In this case most of the donors are occupied and a small number vacant. The donor states which are unoccupied are to be thought of as carriers; as electrons jump from occupied to unoccupied donors, the positive vacancy moves through the lattice. As in case (*a*) when the centres are still localized, we think of this as a hopping process from centre to centre.

In case (*b*) it is possible to discuss the activation energy for charge transfer in terms of a simple model first suggested by one of us (Mott 1956). The carrier (the positive charge vacancy) will in its state of lowest energy lie as closely as possible to a negatively charged acceptor. Before conduction can occur the carrier must be thermally activated from this bound state. An energy of order

$$E = (e^2/\kappa)\langle 1/r_{AD}\rangle = 1{\cdot}46\,(e^2/\kappa)\,N_D^{1/3} \quad . \quad . \quad . \quad . \quad (1)$$

will be required to remove the carrier from the neighbourhood of the acceptor. Here r_{AD} is the nearest neighbour separation between a donor and an acceptor, and $\langle 1/r_{AD}\rangle$ denotes an average, assuming a Poisson distribution of centres. Price (1957) suggested as a better approximation that E should be of order

$$E \sim (e^2/\kappa)(N_D^{1/3} - 2N_A^{1/3}); \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.1)$$

he supposed that one should not consider removing the carrier to infinity but to a distance at which it is effectively outside the field of the particular acceptor concerned. The further movement of the 'free' carriers through the lattice may well require further activation, but the energies will be considerably smaller. Price has examined the statistics of this model.

We give here a simple derivation of the number n of free carriers. Assuming there is only one trap site associated with each acceptor, and a constant trapping energy E, the free energy of the carriers is:

$$nE - kT\,[\ln\{N_D!/n!(N_D-n)!\} - \ln\{N_A!/n!(N_A-n)!\}].$$

Minimizing this with respect to n, we obtain

$$\left.\begin{aligned} n^2/(N_D-n)(N_A-n) &= \exp(-E/kT) \\ \text{or} \qquad n &\simeq (N_D N_A)^{1/2}\exp(-E/2kT) \end{aligned}\right\} \quad . \; . \; . \quad (2)$$

at low temperatures. Thus we expect the slope ϵ_3 of the curve plotting $\ln\rho$ against $1/kT$ to be approximately $\frac{1}{2}E$ at low temperatures. Saturation in the carrier concentration should set in at temperatures given by

$$\exp(-E/2kT) \sim K^{1/2}, \quad (K = N_A/N_D). \quad . \; . \; . \; . \quad (3)$$

In, for example, p-type germanium containing $6{\cdot}5 \times 10^{15}$ majority impurities per cm^3 and a compensation ratio $K = 0{\cdot}05$, the observed activation energy at low temperatures is $1{\cdot}6 \times 10^{-3}$ ev. Therefore the saturation temperature T_s should be approximately 14°. A slight flattening of the resistivity curve is observed at around 5°, the high temperature end of the impurity conduction range (fig. 9, specimen (*a*)). However, this saturation effect is prominent only in specimens in which both the compensation and impurity concentration are small. Also, T_s does not follow in detail the prediction of (3). For example, in specimen (*b*) of fig. 9 (for which $K \sim 0{\cdot}4$) T_s should be 11°, lower than in specimen (*a*), whereas the observed slope is practically constant. This is one example of the limitations of the trapping model. We emphasize again that the model is valid only for very small K, when the acceptors are separated in general by a large number of donors.

The theory of Miller and Abrahams (1960), outlined in § 8, is not limited to small K and thus cannot use the assumption of trap sites. Moreover it is essential in their work to remember that some sites are occupied and some not, so that the electron or 'hole' cannot move except where there is a site ready for it. These authors find that the resistivity ρ is given by

$$\rho(T) \propto \exp(\epsilon_3/kT)$$

where

$$\epsilon_3 = \zeta - 1{\cdot}35\,\epsilon_A$$

and

$$\epsilon_A = (e^2/\kappa)\,(4\pi N_A/3)^{1/3}$$

and ζ is the Fermi energy. This is defined, if

$$f_i = 1/[1 + \exp\{(\epsilon_i - \zeta)/kT\}]$$

and ϵ_i is the energy, due to the random field, of the donor site i, by

$$\sum f_i = N_D - N_A.$$

The summation is approximated by an integration using a density of states function $p(\epsilon)\,d\epsilon$. This is obtained by assuming the energy spread to arise from nearest neighbour negatively charged acceptors. The probability that an acceptor is at a distance r from a donor, and is the nearest one, is

$$p(r)\,dr=\frac{3r^2}{{r_{\mathrm{A}}}^3}\exp\left\{-\left(\frac{r}{r_{\mathrm{A}}}\right)^3\right\}dr,$$

where the mean acceptor separation is given by

$$r_{\mathrm{A}}=(3/4\pi N_{\mathrm{A}})^{1/3}.$$

Then $\epsilon=e^2/\kappa r$, so we have

$$p(\epsilon)\,d\epsilon=(3{\epsilon_{\mathrm{A}}}^3/\epsilon^4)\exp\{-(\epsilon_{\mathrm{A}}/\epsilon)^3\}\,d\epsilon,$$

where

$$\epsilon_{\mathrm{A}}=\epsilon^2/\kappa r_{\mathrm{A}}.$$

On performing the integration of f_i over ϵ, we find

$$1-K=\exp\{-(\epsilon_{\mathrm{A}}/\zeta)^3\}\{1+\exp(-\zeta/kT)\}]^{-1}$$

which determines ζ. Unless K is extremely small or very close to unity,

$$\zeta=-\epsilon_{\mathrm{A}}\{\ln(1-K)\}^{-1/3}.$$

As we shall see in § 4, the magnitude of the resulting activation energy ϵ_3 agrees well with experimental values (cf. fig. 6). For $K\lesssim 0{\cdot}2$,

$$\epsilon_3=\epsilon_{\mathrm{D}}-1{\cdot}35\,\epsilon_{\mathrm{A}}=1{\cdot}61\,(e^2/\kappa)({N_{\mathrm{D}}}^{1/3}-1{\cdot}35{N_{\mathrm{A}}}^{1/3}), \quad . \quad . \quad . \quad (4)$$

which is similar in form to that predicted by Price (1957), formula (1.1).

Finally we may ask whether, in case (*b*), we may expect for holes a phenomenon similar to that suggested in case (*a*) where the overlap becomes large, namely a state of affairs when the states for the hole are not localized. It is clear that, if $K\ll 1$, the lowest state must be localized with an energy given by (1), because a Coulomb field *always* leads to bound states. However, when the carrier has escaped from the field of the nearest charged acceptor, then for a high concentration of centres a hopping process may no longer be an appropriate description of the motion.

In the remainder of this section we shall suppose that K is less than $\frac{1}{2}$. The temperatures at which impurity conduction can be observed is thus determined by the following factors. The mobility of a carrier moving in the impurity levels is much smaller than in the conduction band, since the former is determined by interactions between widely spaced impurities. On the other hand, the number of carriers in the conduction band is determined by an activation energy ϵ_1 ($\sim 10^{-2}$ ev for germanium), while the energy ϵ_3 regulating impurity conduction is at least an order of magnitude smaller, varying from zero to 10^{-3} ev, depending on the concentration of centres. Thus at higher temperatures the conductivity is determined by carriers in the conduction band, at low temperatures by those in the impurity levels, and the transition between the two regions is quite sharp (fig. 4). In the whole region we may write formally

$$\sigma=n_{\mathrm{c}}e\mu_{\mathrm{c}}+ne\mu$$

where n_{c}, μ_{c} are the number of carriers and the drift mobility in the conduction band and n, μ the same quantities for the impurity carriers.

The Hall coefficient R, in the same notation, should depend on these quantities according to the expression

$$R=(n_c\mu_c\mu_{Hc}+n\mu\mu_H)/ec(n_c\mu_c+n\mu)^2,$$

where μ_{Hc}, μ_H is the Hall mobility of the conduction band and impurity electrons respectively. Neglecting any temperature variation in the mobilities, this expression has a maximum when

$$n_c e\mu_c = ne\mu. \qquad \qquad (5)$$

The Hall curves for high concentration (samples 14, 15, fig. 5) can be qualitatively explained by assuming that a Hall effect exists for carriers in impurity levels, with Hall mobility comparable to the drift mobility. At high temperatures, for which $n_c\mu_c \gg n\mu$, so that the normal conduction band Hall effect is observed,

$$R_c=\frac{1}{ec}\left(\frac{\mu_{Hc}}{\mu_c}\right)\frac{1}{n_c}.$$

At temperatures below the Hall maximum, where $n_c\mu_c \ll n\mu$

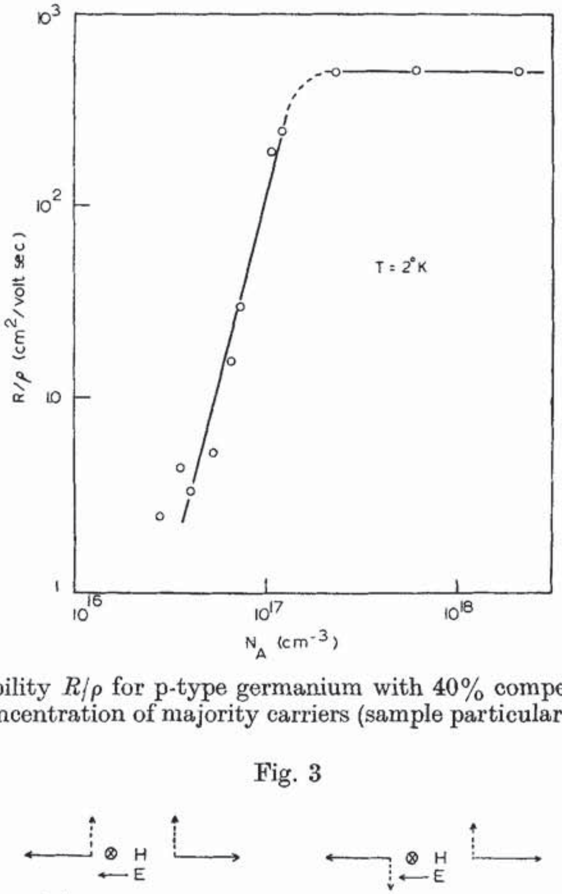
$$R_{imp}=\frac{1}{ec}\left(\frac{\mu_H}{\mu}\right)\frac{1}{n}. \qquad \qquad (6)$$

The number of carriers is practically independent of temperature in this range ($n\simeq N_D-N_A$), giving a temperature-independent Hall curve, as observed (fig. 5). The Hall coefficient R_{exh} in the exhaustion range (temperatures between 77° to 300° K) is due to the same number (N_D-N_A) of carriers, in the conduction band, but measured values of R_{exh} are about eight times larger than R_{imp}. This suggests that μ_H/μ is anomalously low. This is at present unexplained.

For the samples of low concentration it is not known whether a Hall effect exists at the impurity conduction temperatures. The drop in the Hall curves beyond the maximum can be attributed quantitatively to the rapidly decreasing contribution of the conduction band current to the total current. A flattening of the curves beyond the maximum (which would indicate a finite Hall mobility) is not observed down to the lowest temperatures at which it has been possible to make measurements (fig. 5). It is not clear on theoretical grounds whether a Hall effect is to be expected when the conduction process involves jumps of bound carriers to neighbouring sites. The experimental evidence suggests that R may exist, but be too small to be measured in the low concentration range. This is shown by the concentration dependence of R/ρ (fig. 2), measured at 2° K for the p-type samples of table 1. We see that R/ρ (which from (6) is proportional to μ_H) becomes small as the concentration falls through the transition region. This is discussed further in part II.

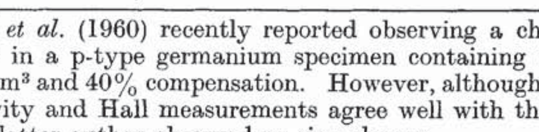
When charge carriers are positive vacancies (in n-type material with both impurity concentration and degree of compensation small), it might be thought that if a Hall effect exists it should be of opposite sign to that due to electrons in the conduction band. However, in a magnetic field H the vacancies do not behave as true particles with positive effective mass and charge. For example, if a positive and a negative particle are initially drifting in opposite directions, a magnetic field will deflect both particles in the same direction. However, a charge vacancy always moves in the opposite direction to an electron, even in a magnetic field (fig. 3). Hence we expect no change in sign in the Hall effect when the conduction is by charge transfer†.

Fig. 2



The Hall mobility R/ρ for p-type germanium with 40% compensation plotted against concentration of majority carriers (sample particulars in table 1).

Fig. 3



The motion in a crossed electric field E and magnetic field H of (*a*) an electron and a hole, (*b*) an electron and a charge vacancy (in a weakly compensated semiconductor).

---

† Yonemitsu *et al.* (1960) recently reported observing a change in sign of the Hall effect in a p-type germanium specimen containing 2×10^{16} gallium impurities per cm^3 and 40% compensation. However, although the magnitudes of their resistivity and Hall measurements agree well with those of Fritzsche (figs. 4, 5), the latter author observed no sign change.

In the band theory of periodic lattices the concept of a 'hole' which behaves as a positive charge with positive effective mass depends in part on the negative effective mass of electrons occupying states near the band maximum. This latter is a consequence of the increasing importance of Bragg reflection of electrons as their energies approach the band maximum. It is unlikely that the phases of electrons scattered by a completely disordered lattice will match sufficiently well for electrons of any energy to show a negative effective mass. Hence, in the metallic region of impurity conduction in n-type material, we would not expect to observe hole conduction or a positive Hall effect, for any concentration on impurity electrons. It is for this reason among others that we have avoided use of the misleading term 'impurity-band' when speaking of the metallic region.

§ 3. THE IMPURITY WAVE FUNCTIONS

We have assumed in the last section a hydrogen atom model for an isolated impurity centre. The departures from this model are briefly outlined here. It has been shown by Kohn and Luttinger (see Kohn 1957) that the donor electron wave functions in germanium and silicon have the form

$$\psi^{(i)}=\sum_{j=1}^{N}\alpha_j^{(i)}F_j(\mathbf{r})\,\phi_j(\mathbf{r}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where ϕ_j is the Bloch wave function at one of the conduction band minimum denoted by j, and the sum is over the N equivalent minima ($N=4$ for germanium, $N=6$ for silicon). The α_j are coefficients which are determined by the symmetry of the state. $F_j(\mathbf{r})$ is an envelope function satisfying a Schrödinger equation for the potential due to the impurity, but with the free electron mass replaced by the effective mass appropriate to the jth minimum:

$$\left\{-\frac{\hbar^2}{2m_l}\frac{\partial^2}{\partial z^2}-\frac{\hbar^2}{2m_t}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-\frac{e^2}{\kappa r}\right\}F_j(\mathbf{r})=\epsilon_1F_j(\mathbf{r}).$$

The z-axis lies along the direction of the k-vector of the jth minimum, and m_l, m_t are the longitudinal and transverse masses respectively. For the ground state, F_j has the hydrogen-like form

$$F_j(\mathbf{r})=(\pi a^2b)^{-1/2}\exp\left\{-\left(\frac{x^2+y^2}{a^2}+\frac{z^2}{b^2}\right)^{1/2}\right\}. \quad . \quad . \quad . \quad (8)$$

For germanium, $a=64{\cdot}5$ Å, $b=22{\cdot}7$ Å. Corrections to this effective-mass approximation partially lift the N-fold degeneracy of the ground state. These corrections arise chiefly from departures from a simple Coulomb potential in the immediate neighbourhood of the impurity ion, and because the concept of a uniform dielectric constant breaks down in that region. In germanium, the ground state splits into a non-degenerate lower level and a 3-fold degenerate upper level (Price 1956); the coefficients of the wave functions (7) corresponding to these levels are:

$$\begin{aligned} \alpha^{(1)} &= \tfrac{1}{2}(1,\ \ 1,\ \ 1,\ \ 1) && \text{lower level;} \\ \left.\begin{aligned} \alpha^{(2)} &= \tfrac{1}{2}(1, -1,\ \ 1, -1) \\ \alpha^{(3)} &= \tfrac{1}{2}(1,\ \ 1, -1, -1) \\ \alpha^{(4)} &= \tfrac{1}{2}(1, -1, -1,\ \ 1) \end{aligned}\right\} & && \text{upper level.} \end{aligned}$$

The splitting is $0{\cdot}57 \times 10^{-3}$ ev for antimony impurities, but is of order ten times larger for arsenic and phosphorus donors (Fritzsche 1960 a); hence the upper states can be neglected except in antimony.

The acceptor states can similarly be constructed from Bloch orbitals at the valence band maximum. In this case the envelope functions F satisfy a set of six coupled effective-mass equations (Kohn and Luttinger 1955, Kohn and Schechter 1959); the solution of these is difficult and hence only approximate solutions of the acceptor wave functions are known.

§ 4. OBSERVATIONS OF IMPURITY CONDUCTION

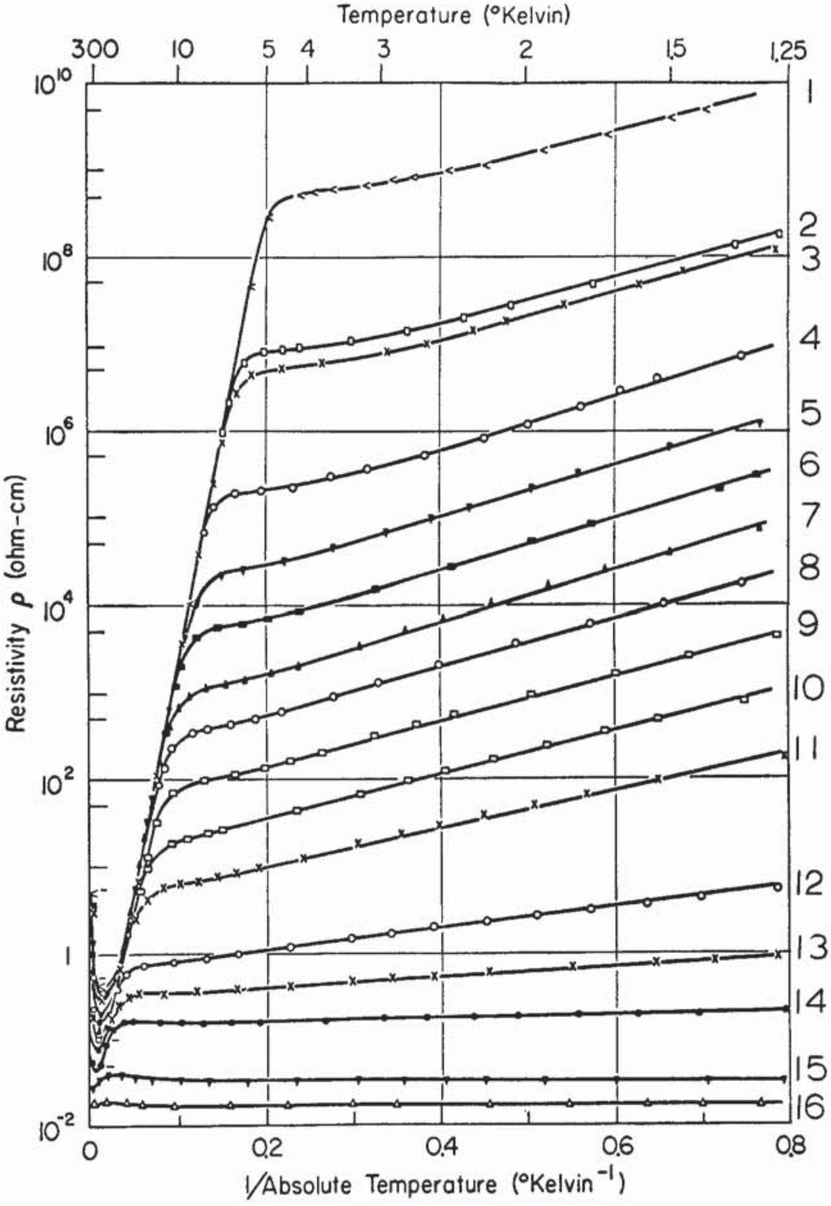
Impurity conduction has been observed in many semiconductors at low temperatures†. We do not attempt to give here an exhaustive review of experimental results but confine ourselves to the case of germanium, which has been studied most extensively. Results for other materials are qualitatively similar.

Prior to 1960, measurements in germanium and silicon were made in samples grown from melts doped with suitable impurities. For a quantitative comparison of theory and experiment it is desirable to have a range of measurements on samples where either (*a*) the majority impurity concentration is varied, but the degree of compensation K is kept constant, or (*b*) K is varied but the majority concentration is kept constant. It was practically impossible to achieve these conditions in the early measurements. Consequently, since the resistivity and Hall coefficient vary extremely rapidly with impurity concentration, comparison of theory and experiment was uncertain. Recently, however, Fritzsche and Cuevas (1960 a) has published measurements on p-type germanium samples in which the acceptor concentration (gallium) ranges between $8 \times 10^{14}\,\text{cm}^{-3}$ to $1{\cdot}3 \times 10^{18}\,\text{cm}^{-3}$ and the compensation ratio (arsenic and selenium) is kept fixed at $K = 0{\cdot}4$. The impurities were introduced into pure germanium by slow neutron bombardment (Cleland *et al.* 1950), causing transmutation of germanium atoms. The proportion of different impurities produced is determined by the cross sections for neutron capture and the decay schemes of the various germanium isotopes; therefore the compensation ratio is constant. The magnitude of the impurity concentration can be varied by the neutron flux and exposure times of different samples. Figures 4 and 5 show results of measurements of the resistivity ρ and Hall coefficient R; table 1 gives information about the specimens.

---

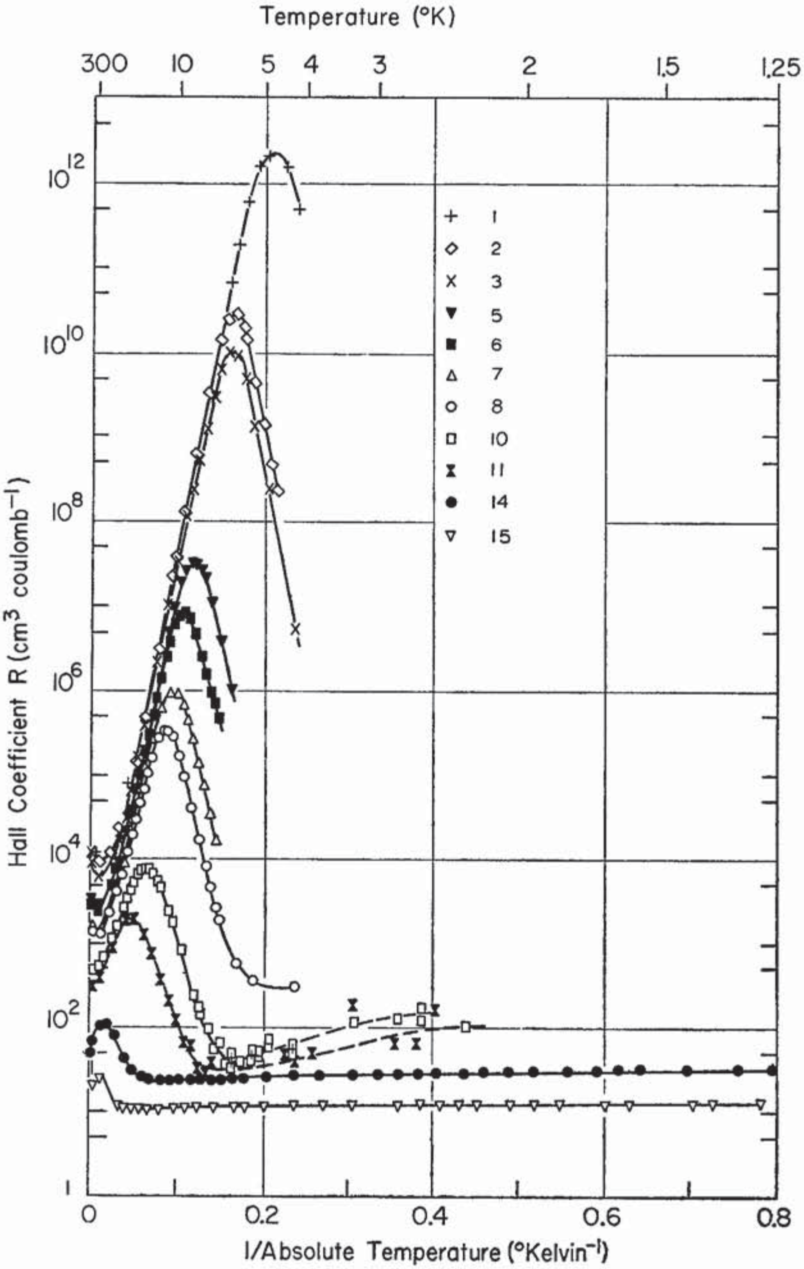
† SiC: Busch and Labhart (1946); Ge: Hung and Gliessman (1950, 1954); CdS: Kroger *et al.* (1954); Si: Morin and Maita (1954), Carlson (1955); p-InSb: Fritzsche and Lark-Horovitz (1955); n-InSb: Sladek (1958); Te: Fukuroi *et al.* (1954).

Fig. 4



Resistivity as a function of temperature of p-type germanium with compensation $K=0{\cdot}4$; particulars in table 1 (Fritzsche and Cuevas 1960 a).

Fig. 5



Hall constants of p-type germanium as functions of temperature (particulars in table 1).

Table 1

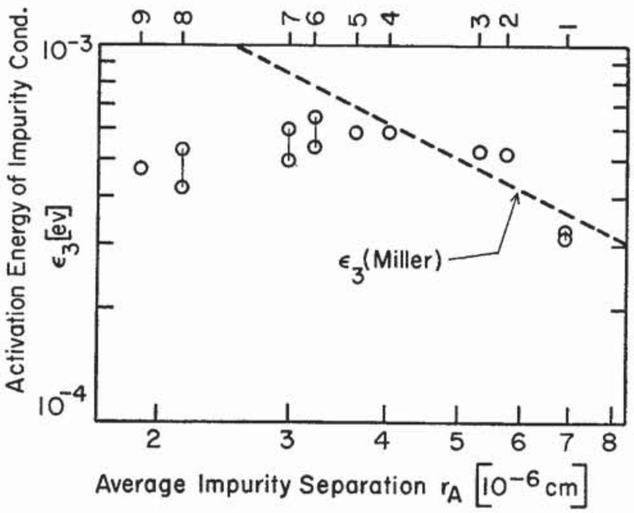
Sample	N_A (cm^{-3})	$\rho(T=2{\cdot}5°\text{K})$ Ω cm	μ(cm^2/volt sec)
1	$7{\cdot}5\times10^{14}$	$8{\cdot}9\times10^{9}$	$4{\cdot}3\times10^{-5}$
2	$1{\cdot}4\times10^{15}$	$1{\cdot}8\times10^{7}$	$1{\cdot}2\times10^{-3}$
3	$1{\cdot}5\times10^{15}$	$1{\cdot}0\times10^{7}$	$2{\cdot}3\times10^{-3}$
4	$2{\cdot}66\times10^{15}$	$5{\cdot}6\times10^{5}$	$3{\cdot}0\times10^{-2}$
5	$3{\cdot}6\times10^{15}$	$1{\cdot}0\times10^{5}$	0·11
6	$4{\cdot}9\times10^{15}$	$2{\cdot}5\times10^{4}$	0·53
7	$7{\cdot}2\times10^{15}$	$6{\cdot}3\times10^{3}$	2·4
8	$9{\cdot}0\times10^{15}$	$2{\cdot}0\times10^{3}$	
9	$1{\cdot}4\times10^{16}$	$4{\cdot}5\times10^{2}$	
10	$2{\cdot}4\times10^{16}$	$1{\cdot}0\times10^{2}$	1·6
11	$3{\cdot}5\times10^{16}$	28	4·0
12	$7{\cdot}3\times10^{16}$	20	
13	$1{\cdot}0\times10^{17}$	0·50	
14	$1{\cdot}5\times10^{17}$	0·18	180
15	$5{\cdot}0\times10^{17}$	$3{\cdot}2\times10^{-2}$	250
16	$1{\cdot}35\times10^{18}$	$1{\cdot}8\times10^{-2}$	

Estimated mobilities μ are shown in column 4 of table 1.

For specimens (1–7) with low impurity content, μ is obtained from values of the resistivity at the temperature of the Hall maximum by making use of eqn. (5):

$$\mu=\mu_c n_c/n=(N_A K e \rho_c)^{-1}.$$

Fig. 6



The activation energy ϵ_3 of impurity conduction for the samples in table 1. The dashed curve represents the calculation of Miller and Abrahams (1960).

We have assumed the number of impurity carriers n to be $N_D (= KN_A)$, since K is less than $\frac{1}{2}$. At higher concentrations, we use simply

$$\mu = R/\rho.$$

This is really the Hall mobility, which as we have seen in § 2 may be appreciably smaller than the drift mobility.

Values of the activation energy ϵ_3 in the low temperature region are plotted in fig. 6, against the average acceptor separation. The magnitude of ϵ_3 agrees well with the values calculated from the theory of Miller and Abrahams (1960), shown by the dotted line, in the region of low concentration.

Measurements of the conductivities $1/\rho$ of n-type samples can be fitted by a sum of three exponentials (see, for example, Fritzsche 1958),

$$1/\rho = c_1 \exp(-\epsilon_1/kT) + c_2 \exp(-\epsilon_2/kT) + c_3 \exp(-\epsilon_3/kT).$$

Here ϵ_1 is the activation energy for exciting an electron into the conduction band, and ϵ_3 that for impurity conduction. The role of ϵ_2, which occurs only for samples in the transition region $(2 \times 10^{16} < N < 8 \times 10^{16})$, is not clear. ϵ_2 is observed in weakly compensated n-type and p-type samples, not however in p-type samples having $K = 0{\cdot}4$. A fourth activation energy was observed at temperatures below 1°K (Zaravstikaya 1956), in low concentration specimens. This however has since been shown to be due to stray light quanta exciting electrons into the conduction band. The resistivity is so large at these temperatures (of order $10^{10}\,\Omega$. cm) that a very small fraction of excited electrons can lower the resistance appreciably.

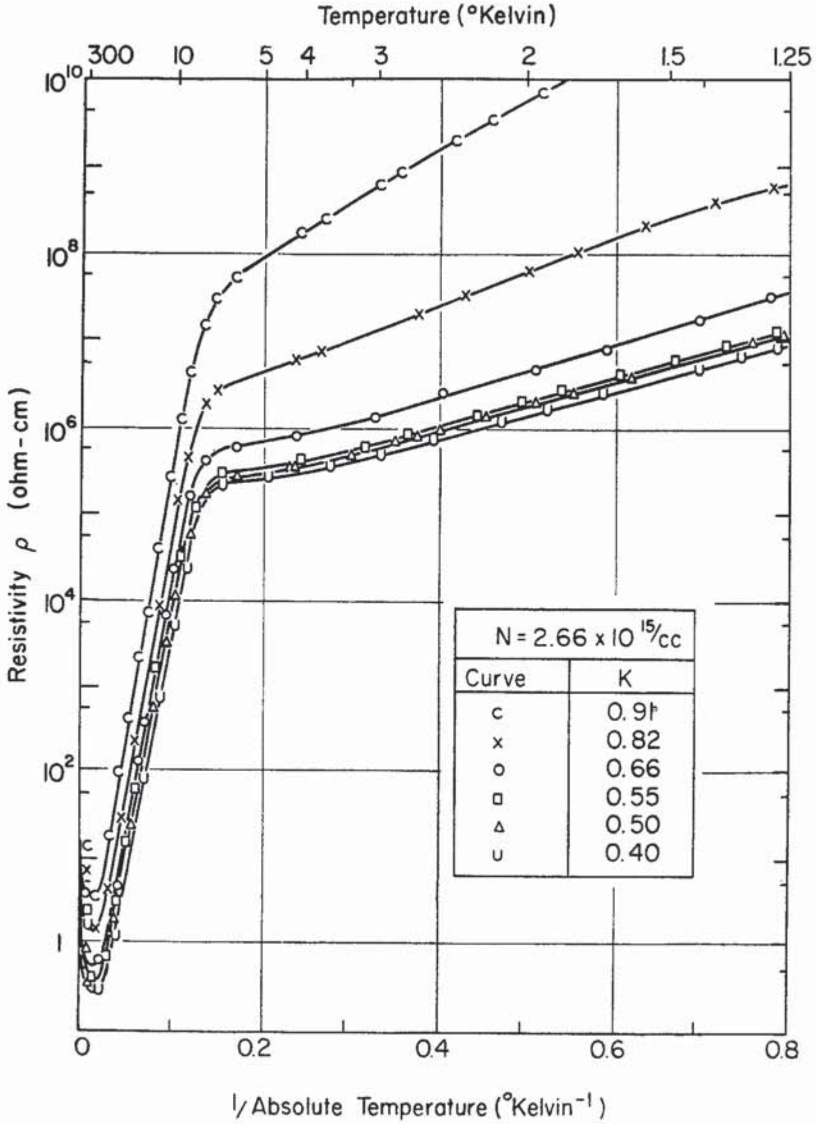
Some measurements of the resistivity have recently been reported on germanium samples of constant acceptor concentration $(2{\cdot}66 \times 10^{15}\ \mathrm{cm}^{-3})$ in which the degree of compensation was varied from 0·4 to 0·9 (Fritzsche and Cuevas 1960 b). These are shown in fig. 7.

The samples were produced by bombarding specimens of n-type germanium with slow neutrons. The same neutron flux and exposure time was used, so that the same acceptor (gallium) concentration was produced in each specimen. The degree of compensation however depends on the initial donor concentration. Figure 8 shows the observed activation energy as a function of K. The solid line is that predicted by Miller and Abrahams (1960) and is in good agreement, except at very high values of K.

The variation of the resistivity with the degree of compensaton should depend on whether the specimen shows metallic or non-metallic conduction. Measurements (in addition to those above) in which K is varied but the majority concentration is kept constant, have been made by Fritzsche (1955), Fritzsche and Lark-Horovitz (1959), in n- and p-type germanium, and by Ray and Longo (1959) in n- and p-type silicon. The results can be summarized as follows.

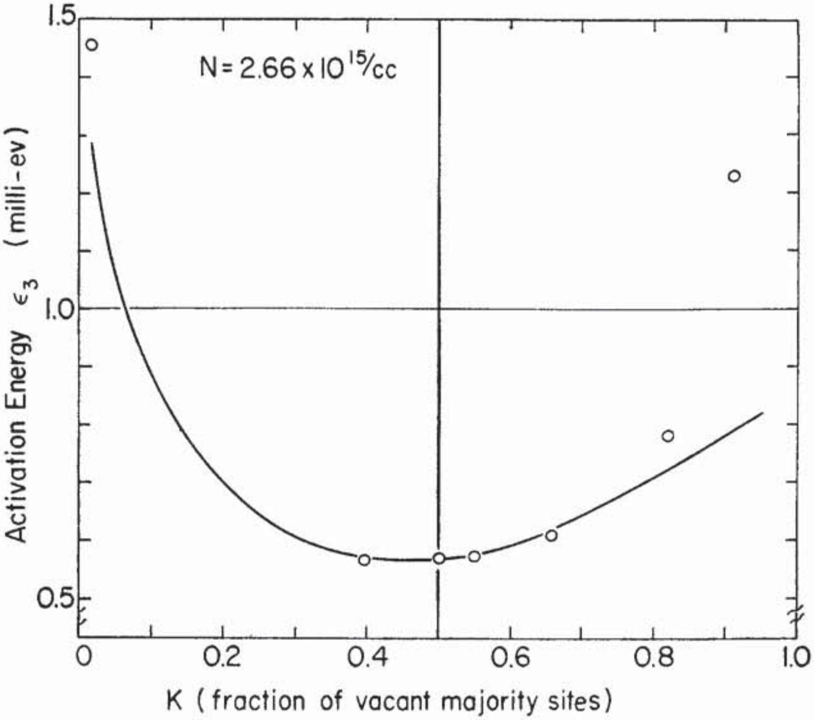
(i) In specimens showing metallic conductivity when $K \sim 0$, both the resistivity and Hall coefficient increase with K, the activation energy ϵ_3 remaining zero provided K is not too large. In p-type germanium samples

Fig. 7



The resistivity of p-type germanium with constant acceptor concentration N_A and compensation K varying from 0·4 to 0·9.

Fig. 8



with $N_A = 2{\cdot}5 \times 10^{17}\,\mathrm{cm}^{-3}$, Fritzsche and Lark-Horovitz (1959) observed that the conductivity becomes non-metallic as K increased between 0·4 and 0·7, the Hall and resistivity curves showing behaviour types of the transition region (figs. 9, 10).

(ii) In specimens showing non-metallic conductivity when weakly compensated, the resistivity at a constant temperature decreases to a minimum as K is increased to about 0·4, and thereafter increases.

Writing

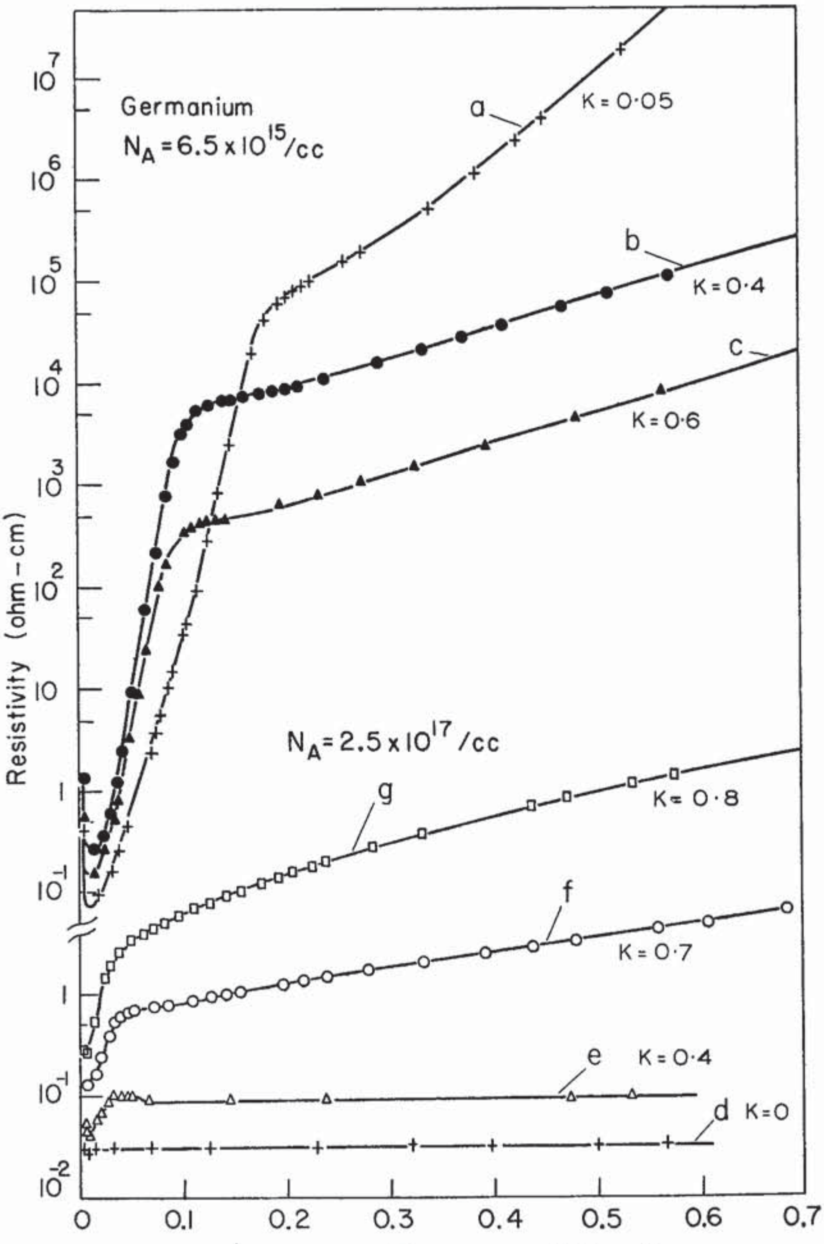
$$\rho = \rho_0 \exp(-\epsilon_3/kT),$$

ρ_0 is found to increase with K (fig. 11). However, in samples with $N_D = 10^{16}\,\mathrm{cm}^{-3}$, ρ_0 remains approximately constant in the range $10^{-3} < K < 10^{-2}$ (Fritzsche 1960, private communication); these samples have a non-vanishing Hall coefficient and are in the transition range.

These results can be understood in part by considering the effect of compensation on the carrier concentration. In the metallic range the number n of carriers is $N_{\mathrm{maj}}(1-K)$. In the non-metallic range $n = N_{\mathrm{maj}} K$ when $K < \frac{1}{2}$ (the carriers are vacancies on impurity sites) and $n = N_{\mathrm{maj}}$ $(1-K)$ when $K > \frac{1}{2}$. Hence, since

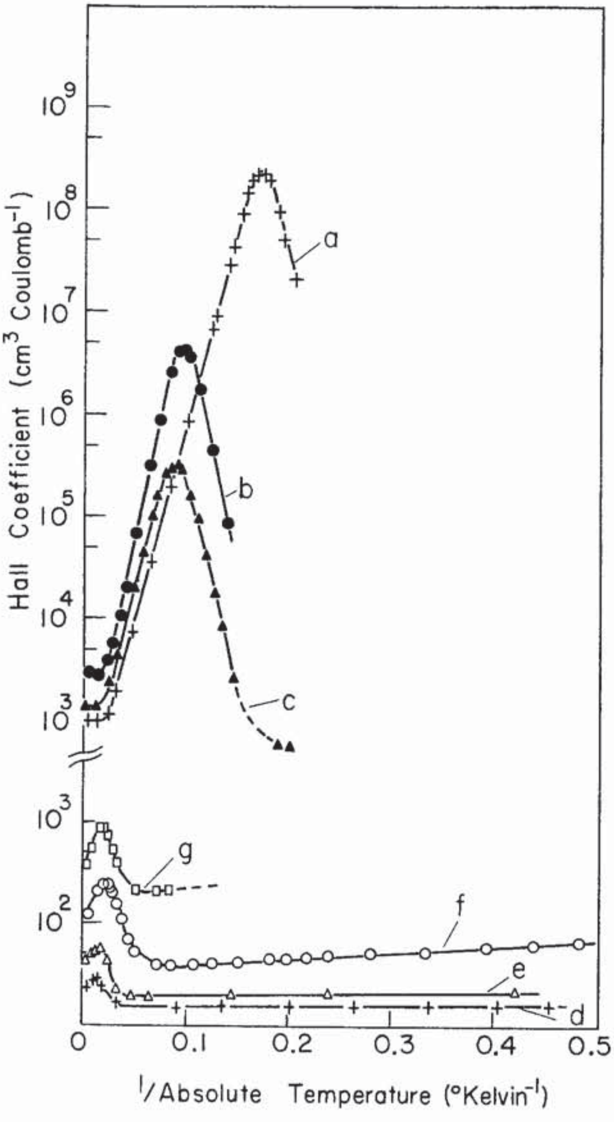
$$\rho = 1/ne\mu,$$

Fig. 9



Resistivity of p-type germanium with variable degree of compensation (Fritzsche and Lark-Horovitz 1959).

Fig. 10



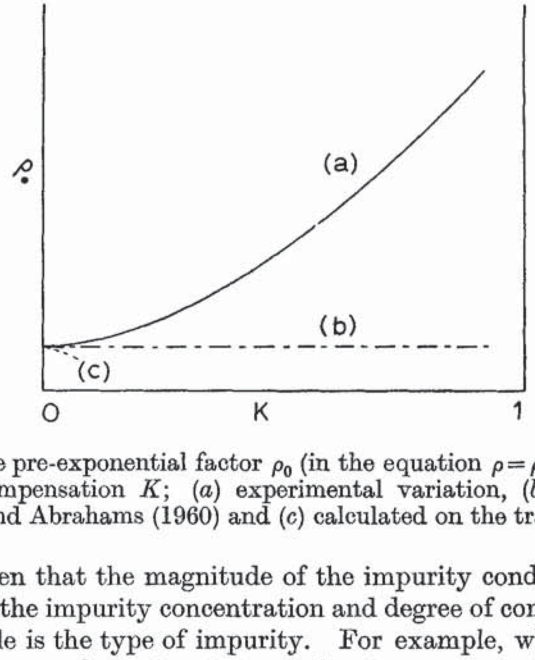
The Hall constant of the samples of fig. 9.

if the conduction is metallic ρ should always increase with K, assuming the carrier mobility μ to depend only weakly on K. If the conduction is non-metallic, ρ should decrease as K increases to $\frac{1}{2}$, and increase as K goes from $\frac{1}{2}$ to 1.

The dependence of ρ on K in the non-metallic range has not however been adequately explained. Since the term $\exp(\epsilon_3/kT)$ dominates ρ at low temperatures, and ϵ_3 is observed to have a minimum value for $K \simeq 0{\cdot}4$

(rather than $K = 0{\cdot}5$), a minimum is also observed in ρ at $K = 0{\cdot}4$. Although Miller and Abrahams (1960) have been able to calculate the observed dependence of ϵ_3 on K (fig. 8), in their theory the pre-exponential factor ρ_0 is independent of K. This independence is observed only for $K < 10^{-2}$, and then for specimens in the transition range, to which their theory does not apply. On the basis of the trapping model (§ 2), $\rho_0(K)$ is again predicted incorrectly. For in this model the number of free carriers is proportional to $K^{1/2}$ (eqn. (2)). When K is very small we can assume that the mobility of free carriers between sites at large distances from compensating centres is independent of K. Hence ρ_0 should be proportional to $K^{-1/2}$, whereas ρ_0 is observed to increase with K (fig. 11).

Fig. 11



Variation of the pre-exponential factor ρ_0 (in the equation $\rho = \rho_0 \exp(-\epsilon_3/kT)$) with compensation K; (*a*) experimental variation, (*b*) calculated by Miller and Abrahams (1960) and (*c*) calculated on the trap model.

We have seen that the magnitude of the impurity conduction depends sensitively on the impurity concentration and degree of compensation. A further variable is the type of impurity. For example, when corrections to the effective-mass formalism (§ 3) are taken into account, it is found that a donor electron is localized in a smaller volume around an arsenic impurity atom than around antimony. Hence the resistivity is larger for arsenic than for antimony impurities, if the same number of both are present, due to the smaller overlap of the arsenic wave functions with neighbouring states.

The Bohr radius of the impurity wave functions in silicon is of order $\frac{1}{3}$ that in germanium. Hence the impurity conductivity in silicon is always very much smaller than in germanium, for comparable impurity concentrations. Also, the energy ϵ_1 required to activate electrons into the conduction band is of order 5×10^{-2} ev in silicon and 1×10^{-2} ev in germanium. Therefore the onset of impurity conduction is observed at higher temperatures in silicon than in germanium (again for comparable impurity concentrations) due the faster freeze-out of electrons from the conduction band of silicon.

The impurity concentration at which the transition from non-metallic to metallic conduction is observed depends on the overlap of neighbouring donor states, and on the degree of compensation. Because the impurity states are more localized in silicon than in germanium, the transition is observed at higher impurity concentrations in silicon. Similarly, because the impurity Bohr radius is different for different impurities, the transition concentration is also a function of the type of impurity. Fritzsche and Cuevas (1960 b) finds that the activation energy ϵ_3 in p-type (gallium doped) germanium disappears at $N_A = 1{\cdot}09 \times 10^{17}\,\mathrm{cm}^{-3}$ when $K \simeq 0{\cdot}04$, and at $N_A = 1{\cdot}80 \times 10^{17}\,\mathrm{cm}^{-3}$ when $K = 0{\cdot}4$. When $N_A = 2{\cdot}5 \times 10^{17}\,\mathrm{cm}^{-3}$, the activation energy disappears at a value of K in the range 0·4 to 0·7 (Fritzsche and Lark-Horovitz 1959). These results are discussed in more detail in § 11.

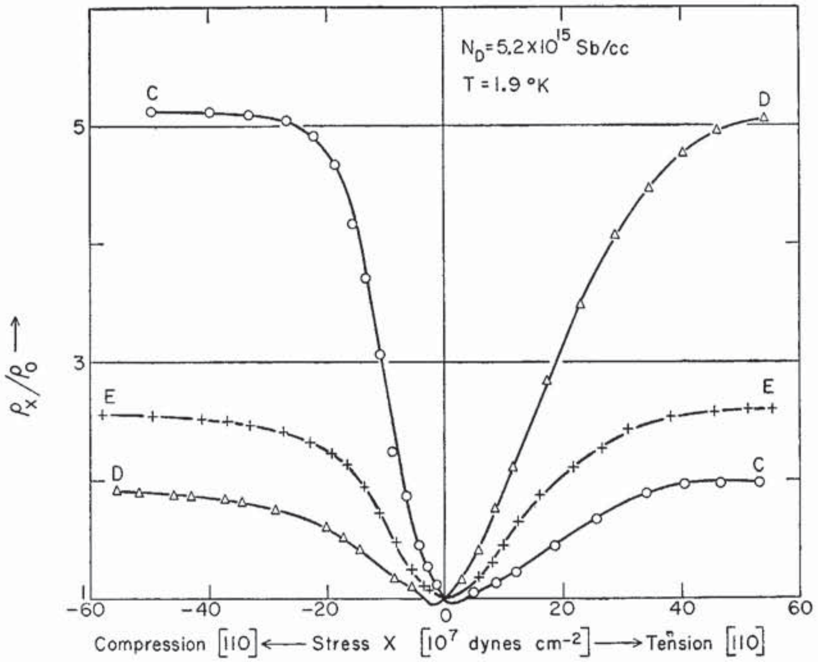
It is found that there is a large change in impurity resistivity when the overlap of neighbouring impurity states is altered by straining the crystal. We saw in § 3 that the wave function ψ of an isolated donor can be constructed from Bloch orbitals taken from the degenerate conduction band minima. Although the envelope functions F (eqn. (8)) are anisotropic, in an unstrained crystal ψ has the tetragonal symmetry of the lattice, and the overlap of neighbouring donor states leads to an isotopic resistivity. In a germanium crystal strained along, for example, the [110] direction, two conduction band minima are depressed in energy relative to the other two. If the strain is large enough, the donor electron ground state will be a sum of the $F_i \phi_i$ corresponding to the depressed minima only,

$$\psi_S = 2^{-1/2}(F_i \phi_i + F_j \phi_j).$$

The overlap of such functions will vary strongly with the direction to the neighbouring donor relative to the strain direction. Hence we expect that in a strained single crystal the impurity resistivity will be anisotropic, with a maximum value in the direction of least overlap. Fritzsche (1960 b) has measured the change in impurity conduction in germanium containing $5{\cdot}2 \times 10^{15}$ antimony atoms per unit volume at 1·9°K as a function of uniaxial tension and compression along the [110] direction. Figure 12 shows the resistivity ratios of the strained to the unstrained sample for the three principle directions of the resistivity tensor (labelled C, D and E). A 'saturation' is observed when strains are large enough for only the depressed minima to contribute to the donor wave function. The relative magnitudes of ρ_C, ρ_D and ρ_E are in agreement with predictions on the basis of the overlap of these functions.

Additional evidence that there is a conduction process in the impurity energy levels at low temperatures is provided by the work of Sladek (1956, 1958, 1959) on magnetically induced impurity banding in n-type indium antimonide. Because of the small effective mass ratio ($m^*/m = 0{\cdot}013$) and large dielectric constant ($\kappa=16$), the Bohr radius of a donor electron in indium antimonide is large, about 144 Å, and the ionisation energy of an isolated donor is small (0·0067 ev). For all obtainable purities the overlap of donor wave funtions is large and the levels are broadened and merge with the bottom of the conduction band. A strong magnetic field will shrink the donor wave funtions (Yafet *et al.* 1956), producing two effects.

Fig. 12

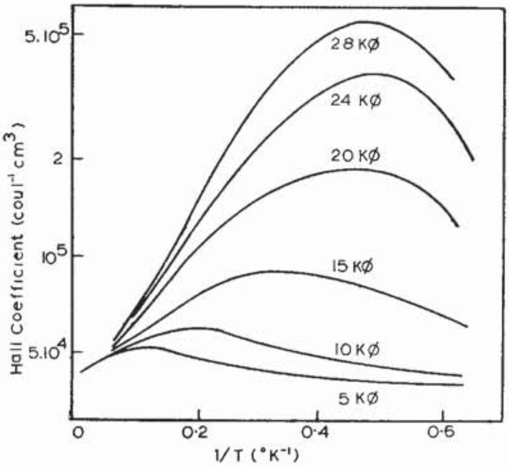


Ratio of the resistivity of stressed and unstressed n-type germanium, for the principle directions, as a function of stress (Fritzsche 1960 b).

(i) The donor ionization energy is increased due to the decrease in the Coulomb energy of the electron. Thus for a large enough magnetic field donor levels will be split off from the bottom of the conduction band.

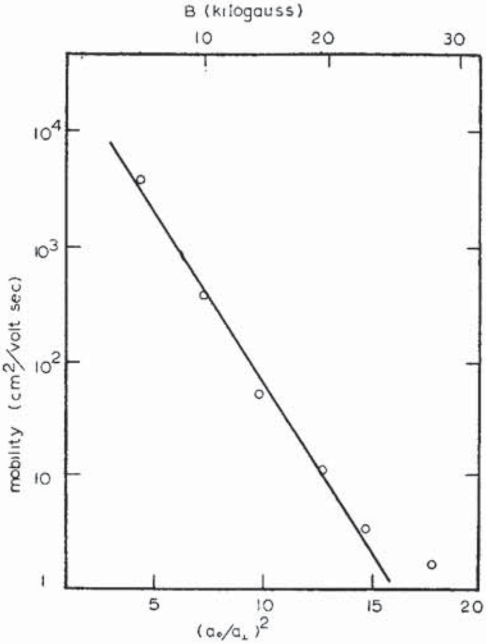
(ii) The overlap of the donor wave functions is decreased. This effect is large because of the small effective mass. Figure 13 shows measurements of the Hall coefficient R as a function of temperature and magnetic field strength, in a specimen containing $5\cdot3\times10^{14}$ donors per unit volume (Sladek 1958). For a small magnetic field R is approximately temperature-independent since the specimen is degenerate. At fields greater than about 10 k-gauss we see the onset of typical impurity conduction. R reaches a maximum value at around 20°K due to a competing conduction process in the split-off donor levels. The depression of the donor levels below the conduction band increases with the magnetic field B, as shown by the increasing slope of the Hall curve with B at temperatures above the Hall

Fig. 13



The Hall constant as a function of magnetic field in InSb containing $5{\cdot}3 \times 10^{14}$ donors/cm^3 (Sladek 1958).

Fig. 14



The effect of a magnetic field B transverse to the current direction on the mobility of impurity electrons in n-type InSb; $a_\perp$ is the Bohr radius of a donor wave function in a direction perpendicular to B (Sladek 1958).

maximum. The Hall mobility μ_H for the same specimen measured at a temperature below the Hall peak, is plotted in fig. 14 against $(a_0/a_\perp)^2$. The Bohr radius a_0 is the zero field value, while $a_\perp$ is the radius normal to the field H, calculated from the theory of Yafet *et al.* (1956). Thus as the magnetic field shrinks the donor electron's orbit and hence lowers the overlap of neighbouring donor states, the electron's mobility decreases, as is to be expected on the impurity conduction model.

§ 5. METHODS OF CALCULATING THE ELECTRICAL CONDUCTIVITY AT LOW CONCENTRATIONS

By low concentrations we mean here concentrations such that the electron gas is not 'metallic', so that the conductivity tends to zero with the temperature. In this range of concentration we have to do essentially with a 'one-body' problem, the movement of a single electron under circumstances in which the interaction between electrons is not important (except in so far as an occupied centre blocks the passage of an electron from another centre). However, even so, the problem is complicated, and far from a complete solution.

If we take as our first problem the movement of an electron in a disordered lattice in which the ions are assumed to be at rest, we have to ask whether the states are *localized* or unbounded in space. By disordered, we mean either in random positions or acted on by a random field (such as that from the charged minority centres) or both. By a 'localized' state we mean that each characteristic solution of the Schrodinger equation for an electron in this field decays exponentially to zero at sufficiently large distances from some point in space. The problem of the conditions under which states are localized or not is by no means solved. In § 7 we reach the conclusion that *all* states in a one-dimensional lattice may be localized. This may correspond to the theorem that in one dimension any potential hole, however small, leads to a bound state. In three dimensions the work of Anderson (1958) was the first to show that, if the separation between impurity centres is large enough, all states would be localized, but at some higher concentration one would go over to unbound states†. Anderson was concerned with spin diffusion; Twose (1959) extended his work to impurity conduction, finding as the criterion for bound states

$$N_t < 10^{-4}/a_0^3 \qquad (9)$$

where a_0 is the hydrogen radius of the centres $(\hbar^2\kappa/m^*e^2)$. Due to the approximations that it was necessary to make in this work, the result may be too small‡.

Suppose now that (9) is not satisfied, so that an electron has a wave function extending through the lattice. We shall now consider the

---

† This concentration has nothing to do with the concentration for metallic conduction discussed in part II.

‡ See Appendix for detailed discussion.

problem of its mobility, under conditions when the interaction with other electrons is not important (i.e. $1-K \ll 1$, or for a 'metallic' electron gas).

We shall have to calculate the time of relaxation τ and mean free path l. The question arises as to whether the usual transport theory can be used, developed as it is for the case where the scattering processes are weak compared with those interactions between lattice atoms which broaden the sharp electronic levels of the isolated atom into a band in the solid. If τ_{S} is an average relaxation time for a carrier scattered between states of different energy and momentum in the band, a suitable criterion that the band approach be a good one is (Joffe 1956, Herring 1959, Yamashita and Kurosawa 1960)

$$\tau_{\mathrm{S}} \gtrsim \hbar/W,$$

otherwise the uncertainty in energy of the scattered carrier would exceed the bandwidth. This leads to a lower limit on the carrier mobility μ below which the band model will not be an adequate starting point for a conductivity calculation. For

$$\mu = \frac{e\tau_{\mathrm{S}}}{m^*} \gtrsim \frac{e\hbar}{m^*W} \sim \frac{ea^2}{6\hbar}, \qquad \dots\dots \quad (10)$$

where m^* is the effective mass of the carrier and a the lattice spacing. Here we have used the Bloch tight-binding model to estimate m^*W; this should be a valid approximation in those substances for which the band approach is questionable, for example, the narrow 3d 'band' in transition metal oxides. With $a \sim 4$ Å, the mobility must be larger than of order

$$\mu_{\mathrm{L}} = 0{\cdot}5\ \mathrm{cm^2/volt\ sec}. \qquad \dots\dots \quad (11)$$

Similarly, the carrier mean free path $l \sim \tau_{\mathrm{S}}\bar{v}$ (where $\bar{v}$ is the mean velocity) must be larger than the order of the lattice spacing:

$$\left.\begin{aligned} &\bar{v} \sim \pi\hbar/m^*a, \\ &l \sim \frac{\pi\hbar\tau_{\mathrm{S}}}{m^*a} \gtrsim \frac{\pi\hbar^2}{am^*W} \sim \frac{\pi a}{6}. \end{aligned}\right\} \qquad \dots\dots \quad (12)$$

The numerical factors are of course approximate. It is possible to define the wave vector k of a Bloch electron only to within $1/l$, on a similar argument.

In metals, these conditions are usually easily satisfied. Electrons in copper, for example, have a mean free path of order 140 lattice spacings and mobilities 30 cm^2/volt sec at room temperature. In a large number of semiconductors however, the carrier mobility may be of order μ_{L} or lower. Taking the transition metal oxides as examples, in $\mathrm{L_{0\cdot1}i\,Ni_{0.9}\,O}$ carriers in the 3d (Ni) levels have mobilities ranging from 10^{-3} to 7×10^{-2} cm^2/volt sec from room temperature to 1000°K (Morin 1958). In titanium oxide (TiO) Morin estimates the 'd band' mobility as 0·4 cm^2/volt sec at room temperature. In the case of impurity conduction in germanium or silicon, the condition (10) for the mobility becomes

$$\mu_{\mathrm{L}} \gtrsim \frac{e}{6\hbar}\left(\frac{3}{4\pi N}\right)^{2/3} = 10^{14}N^{-2/3}\ \mathrm{cm^2/volt\ cm}. \qquad \dots \quad (13)$$

Here we have assumed that the majority impurities, of concentration N, lie on a lattice of average spacing $(3/4\pi N)^{1/3}$. Table 2 compares experimental estimates of the mobility, μ_{exp}, with μ_{L} for n-type germanium. Thus a treatment of impurity conduction which assumes that the impurity electrons can be described by modified Bloch-type functions would appear to break down as N decreases into the transition region.

Table 2

N (cm^{-3})	10^{15}	10^{16}	10^{17}	10^{18}
μ_{L} (cm^2/volt sec)	10^4	2×10^3	5×10^2	10^2
μ_{exp} (cm^2/volt sec)	5×10^{-6}	1·0	$4{\cdot}10^2$	10^3

Methods based on the density matrix have been developed to handle the conductivity of solids under these conditions; we have thought it worth while to include in the next section a simplified treatment, limited to one dimension, to show what these methods mean.

We turn now to the case when the states are localized. Then in general a finite amount of energy is necessary to move an electron from one localized state to the next. This can only come from phonons. We have thus to consider the interaction with phonons.

Phonons can act in two ways, which can be distinguished in the language of field theory as strong and weak interactions. In polar lattices it has been known for a long time that 'self-trapping' is a possibility (Landau 1933, Mott and Gurney 1940, p. 86). The electron or positive carrier polarises the lattice round it and can only move by carrying this polarization with it. A jump from one site to another is thus a multiphonon transition. This is believed to occur in nickel oxide, the carrier being a positive vacancy on a Ni^{2+} ion.

Single phonon transitions are probably of predominating importance in valence semiconductors, and these are discussed in § 7. We start here with the concept of a localized state; an electron is localized in a given centre and, receiving energy from a phonon, it makes a transition by tunnel effect to another centre where the energy is different. This is common to the treatment of Twose and of Miller and Abrahams: the latter are interested in the case where K is not small and so all centres are not available for any one moving electron or hole.

§ 6. THE PROPERTIES OF A ONE-DIMENSIONAL DISORDERED LATTICE

In this section we investigate some of the properties of a one-dimensional lattice, not with a view to applying them to the actual problem, but as an illustration of some of the principles involved.

We shall first (in § 6.1) investigate the mobility of an electron in a one-dimensional lattice, assuming that the wave functions are extended throughout the lattice. The purpose of the investigation is to give a method appropriate to the case when the mean free path is comparable with the electron's wavelength, so that the usual transport theory based on Boltzmann's equation is not applicable. The calculation for three dimensions has been given by Edwards (1958), but that in one dimension is so much simpler that it is worth reproducing.

In § 6.2 we shall examine the nature of the wave functions in a one-dimensional disordered lattice. There have been many investigations of the density of states in this case (for refs see Frisch and Lloyd 1960) but as far as we know none of the nature of the states. We shall show that in many cases (perhaps in all) the states are bound.

6.1. *The Conductivity of Electrons in a One-dimensional Disordered Lattice*

Throughout this section we discuss the conductivity of carriers on a one-dimensional 'wire' of length L, on which they have a non-periodic potential energy $V(x)$. We have to develop a transport theory appropriate to this case. Before doing this we shall set down the usual transport theory based on Boltzmann's equation in a form appropriate to a one-dimensional model.

In this case, where the state of the carrier can be defined by a wave number k and a time of relaxation τ can be defined, the current j is given by

$$j=e\int_{-\infty}^{\infty}N(k)f(k)\,u\,dk \qquad (14)$$

where $N(k)\,dk$ is the number of states in the range k to $k+dk$, $f(k)$ the probability that a state k is occupied and u the velocity ($u=dE/\hbar dk$). If a field F is applied, then in a steady state

$$f=f_0+(df/dk)eF\tau/\hbar,$$

where f_0 is the form of f in the absence of a field. Thus from (14) we obtain

$$j=\frac{e^2F}{\hbar}\int N(k)\frac{df}{dk}u\tau\,dk. \qquad (15)$$

We may write $u\tau=l$, where l is the mean free path, and

$$N(k)=L,$$

so that

$$j=\frac{e^2FL}{\hbar}\int l\frac{df}{dE}dE. \qquad (16)$$

This formula gives the total current in the wire. The current C at any point is obtained by dividing by L, so that

$$C=\frac{e^2F}{\hbar}\int l\frac{df}{dE}dE. \qquad (17)$$

If we wish to derive a mobility from these formulae, we may suppose a Boltzmann distribution of electrons and set

$$f=\text{const.}\exp(-E/kT),$$

so that, from (16)

$$j = \frac{e^2F}{\hbar}\int N(k)\frac{df}{dE}\,l\,\frac{dE}{dk}\,dk = e^2F\int N(k)flu/kT\,.\,dk,$$

from which we see that the mobility μ is given by

$$\mu = elu/kT \sim el/m^*u \sim e\tau/m^*,$$

where m^* is the effective mass.

We shall now derive by a simplified method a formula due to Greenwood (1958) for the current C, valid even when a mean free path cannot be defined. We assume as before that the electron moves on a circular wire of length L on which its potential energy $V(x)$ is that of some random (non-periodic) field. Then the Schrödinger equation of such an electron is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V)\psi = 0.$$

The solutions are subject to the boundary conditions

$$\psi(x) = \psi(x+L), \quad \psi'(x) = \psi'(x+L),$$

so that they are single-valued on the circular wire; they are real and in general non-degenerate. We denote the stationary wave functions and energy states by ψ_n, E_n. As in Greenwood's paper we imagine an electromotive force induced in the circuit by a magnetic field, threading the circuit and increasing uniformly with the time. The perturbing term in a time-dependent Hamiltonian is

$$(\hbar eF/mi)\,t\,\partial/\partial x,$$

where F is the induced electric field and t the time. If then an electron is initially in the state n, at a time t after the perturbation is switched on, the wave function Ψ is given by

$$\Psi = \psi_n \exp(-iE_n t/\hbar) + \sum_{n'} a_{nn'}(t)\,\psi_{n'}(x)\exp(-iE_{n'}t/\hbar),$$

where

$$(\hbar/i)\,a_{nn'}(t) = (\hbar eF/mi)D_{nn'}\int_0^t t\exp(i\omega_{nn'}t)\,dt.$$

Here

$$D_{nn'} = \int_0^L \psi_{n'}\frac{\partial}{\partial x}\psi_n\,dx,$$

and

$$\omega_{nn'} = (E_{n'} - E_n)/\hbar.$$

Integration by parts and neglect of oscillating terms as in Greenwood's paper gives

$$a_{n'} = (eF/m)\,D_{nn'}\{1 - \exp(i\omega_{nn'}t)\}/\omega^2_{nn'}.$$

The current due to an electron initially in state n is

$$\frac{e\hbar}{2mi}\int\left\{\Psi^*\frac{\partial}{\partial x}\Psi - \Psi\frac{\partial}{\partial x}\Psi^*\right\}dx = (e^2\hbar F/m^2)\sum|D_{nn'}|^2\sin\omega_{nn'}t/\omega^2_{nn'}.$$

If f_n is the occupation number of each state, the resultant current may be written

$$\frac{e^2\hbar^3 F}{m^2}\sum_n\sum_{n'}|D_{nn'}|^2\frac{f_n-f_{n'}}{E_n-E_{n'}}\frac{\sin\{(E_n-E_{n'})t/\hbar\}}{E_n-E_{n'}}. \qquad (18)$$

If $N(E)$ is the density of states, then for large values of the time t this becomes

$$\frac{\pi e^2\hbar^3 F}{m^2}\int\overline{|D_{nn'}|^2}\,\frac{\partial f}{\partial E_n}\{N(E_n)\}^2\,dE_n, \qquad (19)$$

where the bar gives an average over values of n' near to n. This formula is valid in general; it will be noted that m is the electronic mass, not the effective mass. Also to obtain the current C at any point we must divide by L.

Formula (19) is the basic expression for conductivity in one dimension. With any such formula, our first task is to show that for small perturbing energy V it leads to the formula given by the Boltzmann equation. This has been done, in the general three-dimensional case, by Edwards (1958). His proof uses advanced methods, and we shall now give a simple discussion to show how the Boltzmann expression (17) arises.

For any form of $V(x)$ we can define a wave number k such that $1/k$ is the mean distance between zeros of ψ_n. Then as before

$$N(E)=L/(dE/dk).$$

If the disordered field is small perturbation on a periodic field in which the electron has an effective mass m^*

$$N(E)=Lm^*/\hbar^2 k. \qquad (20)$$

We shall now estimate $|D_{nn'}|^2$ for this case. First we note that, while the diagonal element D_{nn} vanishes, the off-diagonal element $D_{nn'}$ is not small even if n and n' differ only by unity, since the phase of ψ_n may differ from that of $\psi_{n'}$ by a large amount. Thus the integral

$$\int\psi_n\,(\partial\psi_{n'}/\partial x)\,dx,$$

integrated over one mean free path l, a distance in which the wave functions are coherent, will be of order

$$lmk/m^*L; \qquad (21)$$

the quantity L in the denominator comes from the normalizing factor of the functions ψ_n. There are L/l mean free paths in the length in the wire, and the signs of the contribution of each to D will be random; thus, to obtain the root mean square of D, (21) should be multiplied by $(L/l)^{1/2}$, and

$$|D_{nn'}|\sim(l/L)^{1/2}\,mk/m^*. \qquad (22)$$

Substituting for $|D_{nn'}|^2$ and for $\{N(E)\}^2$ in eqn. (19) and dividing as before by L to get the curent at any point, we find

$$C=\frac{e^2F}{\hbar}\int l\,\frac{df}{dE}\,dE,$$

apart from numerical factors, which is identical with (17).

It is of interest to show directly that the mean free path as usually defined can be equated (apart from a numerical constant) with the distance in which any given phase relationship between two functions, ψ_n, $\psi_{n'}$ of very nearly the same energy is lost. For one dimension we shall treat the problem by considering a row of scattering centres at points $x_1, x_2, \ldots x_n$, each reflecting a *small* proportion of any incident wave. It is convenient to use at each of these centres the delta-function used by Lax and Phillips (1958) so that at each centre ψ is continuous and ψ' changes by $\Delta\psi$, where

$$\Delta\psi'/\psi = \eta$$

and η is small.

Suppose that such a centre is situated at $x = 0$ and we ask for the amplitude of the wave reflected from it. We set for an incident and reflected wave $(x < 0)$

$$\psi = \exp(ikx) + A\exp(-ikx)$$

and for a transmitted wave $(x > 0)$

$$\psi = B\exp(ikx).$$

The boundary conditions give

$$1 + A = B, \qquad 1 - A = B + q$$

where $q = \eta/k$. Hence

$$A = \tfrac{1}{2}q.$$

Thus the amplitude reflected by the first N of such centres is

$$\tfrac{1}{2}q\sum_{1}^{N}\exp(2ikx_{\mathrm{N}}).$$

An estimate of the number N of centres in one mean free path would be obtained by equating this to unity.

Now consider the phase of a *real* function ψ and let us ask how much it changes at a scattering centre. Let us write for $x < 0$

$$\psi = A_1\cos(kx + \zeta_1)$$

and for $x > 0$

$$\psi = A_2\cos(kx + \zeta_2).$$

The boundary conditions give

$$A_1\cos\zeta_1 = A_2\cos\zeta_2, \quad \tan\zeta_1 - \tan\zeta_2 = \eta/k = q,$$

and hence, since q is small

$$\Delta\zeta = \zeta_1 - \zeta_2 = q\cos\zeta_1\cos\zeta_2.$$

Since ζ_1 and ζ_2 are nearly equal, this may be written

$$\Delta\zeta = \tfrac{1}{2}q(\cos 2\zeta - 1).$$

The second term (-1) in the bracket obviously makes no contribution to the rate at which two waves get out of phase with each other. The total shift in the phase of two waves initially $\tfrac{1}{2}\pi$ out of phase with each other will be the difference between $\tfrac{1}{2}q\cos 2kx_n$ and $\tfrac{1}{2}q\sin 2kx_n$, which will be of order unity in one mean free path as defined above. Either method gives