IMPURITY BAND CONDUCTION. EXPERIMENT AND THEORY: THE METAL-INSULATOR TRANSITION IN AN IMPURITY BAND

N. Mott

To cite this version:
N. Mott. IMPURITY BAND CONDUCTION. EXPERIMENT AND THEORY: THE METAL-INSULATOR TRANSITION IN AN IMPURITY BAND. Journal de Physique Colloques, 1976, 37 (C4), pp.C4-301-C4-306. 10.1051/jphyscol:1976453. jpa-00216567

HAL Id: jpa-00216567
https://hal.archives-ouvertes.fr/jpa-00216567

Submitted on 1 Jan 1976

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
THE METAL-INSULATOR TRANSITION IN AN IMPURITY BAND

N. F. MOTT

Cavendish Laboratory, Cambridge, U. K.

Résumé. — Nous présentons un résumé du phénomène de la conduction dans les bandes d'impuretés et de la transition métal-isolant dans les semiconducteurs dopés. La transition semble être du type Anderson, sans changement discontinu du « gap » ni de rayon des états localisés. Nous discutons l'effet du « U » de Hubbard et nous proposons que cela a un effet moins important que ce que nous avons pensé au sujet de la concentration déterminant un comportement métallique. L'effet Hall et le pouvoir thermoélectrique sont discutés et nous faisons une comparaison avec les propriétés des bronzes de tungstène.

Abstract. — An outline is given of the phenomenon of impurity conduction and of the metal-insulator transition in impurity bands. The transition appears to be of Anderson type, with no discontinuous change in the gap or in the radius of localized states. The effect of the Hubbard $U$ is discussed and it is suggested that it may have less effect in determining the concentration for non-metallic behaviour than previously thought. The Hall effect and thermopower are discussed, and a comparison is made with some properties of tungsten bronzes.

The phenomenon now known as impurity-conduction was introduced by Hung and Gleissman [1] as a new conduction mechanism predominant at low temperatures in doped silicon and germanium, the conductivity following the law

$$\sigma = \sigma_3 \exp(-\epsilon_3/kT).$$

(1)

Here $\sigma_3$ is strongly dependent on the concentration of majority centres and $\epsilon_3$ is an order of magnitude smaller than the energy $\epsilon_4$ required to ionize the centre. In the early days it was hotly disputed whether this phenomenon depended on compensation, but it is now clear that it does. For an uncompensated sample in the limit as $T \to 0$, the Fermi energy lies at an energy half way between the donor and the conduction bands, but for a compensated sample it is pinned at the level of the impurities, and only then can impurity conduction occur. Miller and Abrahams [2] first described the hopping process by which conduction occurs; for n-type materials the donor levels are distributed in energy through a range $\sim e^2/kR_0$ due to the charged acceptors, $R_0$ being the mean distance between the latter, and an electron jumps to an empty donor from an occupied one, obtaining energy from phonons. The energy $\epsilon_3$ is proportional to $e^2/kR_0$, and the factor $\sigma_3$ to $\exp(-2aR)$, $R$ being the distance between majority carriers (donors) and $a^{-1}$ the spatial extent of localized wave function.

We can therefore describe the density of states as in figure 1. The states in the impurity band are localized, in the sense of Anderson [3]; the breadth of the band is due both to the random field of the charged acceptors and to the variation of the energy overlap integral between localized states, which contains the factor $\exp(-aR)$. The Fermi energy $E_F$ at $T = 0$ lies in the impurity band and its position depends on the degree of compensation.

The analysis of Miller and Abrahams supposed, that the factor $\exp(-2aR)$ was so small that carriers always jumped to nearest neighbours. Nonetheless, one would always expect variable-range hopping [4] with $\sigma = A \exp(-B/T^{1/4})$, at low enough temperatures. For lightly doped and compensated material, it would be instructive to look for it at very low temperatures, though it is certainly observed some way from the transition both in doped Ge, as shown by Shlimak and Nikulin [5], and in Si : P as shown by Wallis [6], and in Ga : P down to 0.13 K (Emel’yanenko et al. [7]). Some unpublished results of Wallis are reproduced in figure 2.

The thermopower $S$ in this and other systems when conduction is due to hopping is not fully understood. It certainly increases with $T$ as for a metal and should
have the same sign as $d\ln N(E_F)/dE$. According to Overhof \cite{8} it increases as $T^{1/2}$ at low $T$, when charge transport is by variable-range hopping; other earlier work is quoted by Overhof. Here it must be remarked that \text{Si:P} below the metal-insulator transition should be an amorphous antiferromagnet, and should have a Neél temperature above which the directions of the spin moments are not frozen \cite{9}. According to Butcher \cite{10}, above the Neél temperature there is an additional term in $S$ equal to $(k/e)\ln S$. Choosing a concentration where the Neél temperature is not too low, it would be interesting to look for a sudden increase in $S$ as $T$ goes through the Neél temperature.

Figure 1 also shows the upper Hubbard band, sometimes called the $e_2$ band. Conduction in this band occurs because an electron is excited to form a negatively charged donor ($D^-$) and then moves from one neutral donor ($D$) to another. The idea of the existence of negatively charged donors goes back more than 20 years (Ansel'm \cite{11}, Sellar \cite{12}), the first optical evidence for their existence being that of Dean, Hayes and Flood \cite{12} who observed recombination from shallow donors in silicon and germanium. Fritzche \cite{14} and Davis and Compton \cite{15} identified conduction by this mechanism; according to Pollak \cite{16} this form of conduction is not by hopping, so it is reasonable to assume a mobility edge in the upper Hubbard band, marked in figure 1 as $E_c$. Thus

$$e_2 = E_c - E_F.$$  

(2)

Recently Norton \cite{17} has excited electrons into (presumably) localized states in the tail of the $D^-$ band, where at low temperatures they are frozen-in, and then observed photoconduction by exciting these electrons, which requires a wavelength $\sim 300$ $\mu$m contrasted with $\sim 30$ $\mu$m to ionize the D. It is very significant that the former energy increases when the donor concentration increases above $\sim 10^{16}$ $\text{cm}^{-3}$. This shows that the radius of $D^-$ is above $2 \times 10^{-5}$ cm (about five times that of D), so that overlap between $D^-$ and D occurs, producing a tail to $N(E)$; we have here an experimental method of observing the density of states in a random array of s-like centres, calculated by Cyrot-Lackmann and Gaspard \cite{18}, yielding the kind of shape illustrated in figure 1. It would be extremely useful if calculations of the position of the mobility edge could be made, but this has not been done for positional disorder \cite{9}.

We now turn to the metal-insulator transition. In the author’s opinion one can distinguish two main types.

1. Band crossing transitions in crystalline materials, including Hubbard bands (the so-called Mott transition). For a rigid lattice such transitions are always discontinuous \cite{9}, the number of free electrons $n$ at $T = 0$ changing, as some parameter $x$ varies, from zero to a finite value. This property depends on electron-electron interaction. In a real material, moreover, the neighbourhood of the transition always lies in a two-phase region of $x$, and could only be observed in quenched alloys.

2. Anderson transitions in disordered systems. These can be described by a model of non-interacting electrons, though interaction may affect the concentration at which the transition occurs. I believe that we have good understanding of the Anderson transition, perhaps particularly through the experiments on the silicon inversion layer to be described by Dr Adkins in this volume. Its characteristics are as follows. Consider a (disordered) band as in figure 3a, which could be an impurity band, and suppose that the number of electrons and hence $E_F$ can be changed, for instance by changing the degree of compensation. $E_c$ denotes the mobility edge. Thus if $E_F$ lies below $E_c$, we make the following predictions, amply born out by experiment.

If $E_F$ lies below $E_c$, then at low $T$ conduction is by variable-range hopping and at high $T$ by excitation of carriers to the mobility edge, when the conductivity is of the form

$$\sigma = \sigma_{\text{min}} \exp\left\{ - (E_c - E_F)/kT \right\}$$  

(3)

and $\sigma_{\text{min}} \approx 0.05 \, e^2/\hbar a$, $a$ being the hopping distance. A metal-insulator transition occurs when $E_c = E_F$, and

(1) For disorder produced by a random field see Abou-Chacra and Thouless \cite{19}.
THE METAL-INSULATOR TRANSITION IN AN IMPURITY BAND

Fig. 3. — Density of states in an impurity band just on insulator side of transition (a) if correlation unimportant, (b) if correlation is strong.

Then $\sigma = \sigma_{\text{min}}$. As we approach the transition from below, $\alpha$ behaves like

$$\alpha = \text{const}(E_c - E_F)^{\sigma},$$

where $\sigma$ is variously estimated to be 0.6 and 2/3, and for some systems can be investigated experimentally by an analysis of variable-range hopping [20, 21, 22, 23]. What I want to stress is that there is no discontinuity in $\alpha$.

Now, I know of no system that certainly shows the discontinuity in $n$ expected for a Mott or band-crossing transition, without change of structure. Copper or zinc atomically dispersed in some insulator might show it (3); we await experiments. What is certain, in my view, is that systems like Si : P do not. Thus Davis and Compton [15] showed clearly that, as the concentration of donors increases, $e_2$ decreases linearly with increasing concentrations of donors and eventually disappears. $T^{1/4}$ hopping is observed near the transition, whether the sample is uncompensated or not. In fact, the transition is certainly of Anderson type, and the only question we have to ask is whether correlation (the Hubbard intra-donor $U$, equal [9] to $(5/8) e^2/\hbar a$ where $a$ is the donor radius) has any major effect on it.

Assuming that it does, the present author has interpreted the transition in the following way. As the donor concentration increases, the two Hubbard bands overlap (Fig. 3a) and the transition, as in others of Anderson type, occurs when $E_c - E_F$ vanishes. A prediction is, then, that variable range hopping should be observed, near the transition on the non-metal side, for uncompensated as well as compensated samples, as appears to be the case. There are some other consequences of the model to be described below. Figure 3 shows the two possibilities for the density of states.

Before discussing the role of the Hubbard $U$, I would like to give my assessment of how controversial this description is. First, we may ask if there is any evidence that the transition in Si : P is not an Anderson transition, but a discontinuous transition. The only evidence for the latter, in my view, is the discontinuous disappearance of the Knight shift at the transition found by Sasaki [25]. This has been discussed by the author in a previous paper [26]. The suggestion made there is that when states at the Fermi energy become Anderson localized, there is no discontinuity in the Pauli susceptibility, but since states are localized, the region in which the magnetic field is changed is confined to a small volume proportional to $H$. The Knight shift thus disappears for the spins of most nuclei.

The other controversial point is whether for this or any other system, the quantity $\sigma_{\text{min}}$ exists for transitions of Anderson type. The evidence is now very strong [20], though Cohen and co-workers [27] in a series of papers have argued otherwise, and also many workers, e.g. Mitzushiba [28], have supposed that, in a doped semiconductor in a certain range of concentration, metallic regions are separated by non-metallic. I believe this to be wrong; these considerations only apply if fluctuations exist over regions wide enough to prevent tunnelling.

As regards the magnitude of the minimum metallic conductivity, $\sigma_{\text{min}} = 0.05 e^2/\hbar a$, experimental evidence now exists over two orders of magnitude of $a$ (Pepper [29]). For ordinary amorphous materials in which $E_F$ crosses $E_c$ due to a change of composition, $a$ is a few angstrom and $\sigma_{\text{min}}$ of the order 600-1000 $\Omega^{-1}$ cm$^{-1}$. For impurity conduction in doped Ge with $a \sim 2 \times 10^{-6}$ cm, values of about 10 $\Omega^{-1}$ cm$^{-1}$ are found. Ferre, Dubois and Biskubski [30] have observed an approach to a metal-insulator transition in n-type InSb with donor concentration near $10^{14}$ cm$^{-3}$ induced by a magnetic field; since $a$ is not varied, the $\sigma = 1/T$ curves extrapolate to the same $\sigma_{\text{min}}$. The value of $\sigma_{\text{min}}$ obtained corresponds to 0.05 $e^2/\hbar a$ with $a^3 \sim 10^{-16}$ cm$^3$.

I now want to turn to the question of whether, when the transition occurs, the two Hubbard bands have merged. In my earlier papers [30, 31] I gave a criterion for the transition

$$n^{1/3} a_H \simeq 0.2$$

based on considerations of screening. With the Hubbard Hamiltonian, assuming that both Hubbard bands

\[ \text{(3) Compare Shanfield et al. [25] who find an Anderson transition for Xe-Fe mixtures; the variable valency of Fe could mitigate against a discontinuity.} \]
have the same width (B), the condition for overlap is
\( B = U \), and putting in a calculated value for these
quantities one gets approximately the same result (2). 
Certainly the upper Hubbard band must be broader
than the lower, but the width occurs in a logarithm
and this will make little difference. An Anderson
transition in a pseudogap occurs when \( N(E)/N(E)_{\text{free}} \sim 1/3 \),
so it was thought that this condition would be near
enough to that for the disordered case.

I turn now to the Hall coefficient. Friedman [32]
has shown that, for conduction in which carriers are
excited to a mobility edge, \( \mu_\text{H} \) is independent of \( T \)
and of order
\[ 0.1 e a^2/\hbar. \tag{5} \]

There is extensive evidence that this is correct.
Unfortunately he also predicts that \( R_\text{H} \) should always
be negative, and this is not so in some amorphous
semiconductors, a fact which is unexplained [32], but
I shall nonetheless assume the validity of the formula.
It was also suggested by Nagels, Callears and
Denayer [34] that the Hall coefficient for hopping
is small ; the reason is not fully understood (for a
discussion see Mott, Davis and Street [35]). It follows
that in the temperature range where the conductivity
mechanism goes from excitation to \( E_\text{c} \) to hopping,
behaviour like that of figure 4 is expected. This is
observed in two quite different systems, Si : P near
the transition (Tomotomi [36] ; reproduced by Mott
et al. [20], Kamimura and Mott [37]), and As₂Te₃.

Evidence for the behaviour (5) hitherto has come
from metal-ammonia [38] and expanded fluid mercury [39, 9]. Both cases should correspond to figure 5a
(formation of a pseudogap).

Recently results from Cornell [40] show that for
tungsten bronze \( \text{Na}_x\text{WO}_3 \), as \( x \) approaches the lower
limit of the metallic region (\( x = 0.2 \)), a similar anomaly
is observed (fig. 6). It is not thought that in this material
the gas is highly correlated ; the anomaly is simply
due to a spreading out of the band, as in figure 5.
This is confirmed by the electronic specific heat, which
lies \( \sim 1/3 \) below the free electron value (fig. 7).

The results of Yamanouchi et al. [35] on Si : P do not
show any appreciable evidence Friedman anomaly
in \( R_\text{H} \) denoted by (6), though whether a small effect
occurs depends on where one puts the metal-insulator
transition. The drop in \( 1/R_\text{H} \) just on the insulating side
is of course due to excitation to \( E_\text{c} \), with the Hall
mobility given by (4). To explain this I suggest the
following hypothesis. The metal-insulator transition

\[ g \text{ is here defined as} \]
\[ N(E)/N(E)_{\text{free}}. \]

Two ways in which this can occur are shown in figure 5.
In the first a pseudogap occurs between two bands, in
the second the width of the band is spread out by
disorder. The value of \( g \) for localization is found by
writing
\[ S_{\text{F}} e^2 a^2 g^2/12 \pi^3 \hbar = \sigma_{\text{min}}, \]

and since \( \sigma_{\text{min}} \) depends on the square of the Anderson
localization criterion \( (B/V_0)^2 \), so does \( g \). The values
1/3 is obtained by setting \( B/V_0 \sim 1/2 \).
occurs when the two Hubbard bands have merged; in other words it is not a Mott-Hubbard transition at all and the Hubbard $U$ plays little role. If this is so, one has to show that pure Anderson localization can lead to the condition (1), and thus the localization criterion should not be $V/B \sim 2$ but $V/B \sim 1$, which would give little Friedman anomaly.

Mott and Davis [42] gave a rough criterion for Anderson localization in this case (a random distribution of centres), which we now modify. Following Lifshitz [43], we pair each centre with its nearest neighbour, which we suppose to be at a distance $r_1$ from it, given by

$$ (4\pi/3) r_1^3 = 1/N . $$

The energy of an electron located on a pair of centres is written $I_0 \exp(-\alpha r_1)$, where $I_0$ is of order $e^2/ka$, and since this quantity varies from pair to pair, as a very rough approximation we take it as the Anderson localization parameter $V_0$. The mean distance between pairs is

$$ r_2 = 1/(\frac{4}{3} N)^{1/3} $$

and we take the band-width to be

$$ B = 2 z I_0 \exp(-\alpha r_2) $$

where $z$ is a co-ordination number. Thus the condition for Anderson localization at the centre of the band is

$$ \exp(-\alpha r_1) = 2 z (V_0/B) \exp(-\alpha r_2) $$

where $V_0/B$ is the Anderson localisation ratio. As the centres are similar to $H_2$ molecules, it is unlikely that any one would overlap more than two others, and so we take $z = 2$. Probably $V_0/B \sim 1$ would be appropriate for so small a number. This would account for the small Friedman effect.

Taking this value and $z = 2$ the Anderson transition occurs where

$$ a N^{-1/3} \left \{ 2^{1/3} - (3/4 \pi)^{1/3} \right \} = \ln 4 $$

which yields

$$ N^{1/3} a_H \simeq 0.4 , $$

which, in view of the extremely rough approximation, seems satisfactory agreement.

A test of these ideas would be an examination of the thermopower ($S$). When the Hubbard bands are separated (low concentration), the thermopower for a lightly compensated specimen will have the opposite sign to that in the $\delta$ region; this was observed to be the case in $Si : P$ at $5 \times 10^{18}$ by Geballe and Hull [44] (see Mott and Davis [42]). When the bands have merged, the opposite is the case. Allen [45] found no positive values in GaAs near the transition, and deduced that the gas is not highly correlated. One needs experiments to see at what concentration $S$ changes sign.

Next we consider the magnetic properties. The susceptibility at low $T$ rises above the free electron value ($U_e$ and Maekawa [46]) and this can be partly due to the narrowing of the impurity band, partly to correlation in the sense of Brinkman and Rice [47]. The latter has been discussed by Berggren [48] and Kamimura and Kanehisa [49]. If however the electron gas is not highly correlated, this is not admissible. We may, however, following Toyozawa [50] consider that statistical fluctuations in density may give regions where a local moment exists, but must insist that the spin for metallic concentrations must flip from one position to the other through exchange with the metallic electrons, in the same sense as the Kondo effect. In fact, the model still has similarities with that of the highly correlated gas. In the latter, most sites are singly occupied at any one moment, the spin flipping in the Kondo sense. In the uncorrelated gas doubly, singly and unoccupied sites must have comparable weight, but we postulate that there are some (perhaps $10\%$) of sites where the distance to nearest neighbours is large, the site is normally singly occupied, and spin flips with a (small) frequency $\omega_K$ when a conduction electron moves into it, the original electron jumping
away. $\omega_K$ is the Kondo frequency, and for the various sites $\omega_K$ will spread over a large range of values, perhaps from zero upwards. This assumption has already been made by the present author to explain the negative magnetoresistance linear in $H$ (Mott [9] p. 228). It seems likely that a quantitative explanation of the susceptibility could be given along these lines.

If metal-insulator transitions in undoped Si : P are pure Anderson transitions, with the Hubbard $U$ playing no major role, it becomes much easier to understand why quite strong compensation (up to 80%) has only a minor effect on the concentration at which the transition occurs. This was observed by Davis and Compton [15]; the results are displayed in a way that shows this by Mott and Davis [42] and an attempt to explain it given in terms of the authors’ original theory. Of course it has been realized for some time that for compensated samples the transition is of Anderson type. If now we say that for uncompensated specimens there is no pseudogap and that Anderson localization is due to lateral disorder, then the effect of compensation will be two fold.

1. To reduce the number of electrons. The transition will occur when states are localized some way down into the band. The results of Abou-Chacra and Thouless [19] show that only a very small change in the disorder parameter is required to shift the mobility edge from the centre to quite deep in the band (see their figure 2), so changing the number of carriers will affect very little the concentration of centres required for localization, unless the compensation is high enough to shift the Fermi energy into the tail.

2. The random field will add some disorder in the energies of sites to the effect of positional disorder. In ordinary impurity conduction, as in the theory of Miller and Abrahams [2], it is of course the former which is responsible for the hopping energy $(\varepsilon_{\text{h}})$. But in the neighbourhood of the transition, it is by no means clear that random field of the charged centres is more important than the variation of the overlap integral; in fact a rough calculation suggests that the latter may be larger. If so, reasonable levels of compensation should not affect the concentration of the transition.

References

[29] Pepper, M., Communications on Physics 1 (1976) 147.
[34] Nagels, P., Calloerts, R. and Denayer, M., Fifth Int. Conf. on Amorphous and Liquid Semiconductors (Taylor & Francis) 1975, p. 867.
[38] Acrivos, J. V., Phil. Mag. 25 (1972) 757.