REVIEWS OF MODERN PHYSICS VOLUME 40, NUMBER 4 OCTOBER 1968

SESSION I—INTRODUCTION

Metal–Insulator Transition

N. F. MOTT
Cavendish Laboratory, Cambridge, England

A general survey of the idea of the metal–nonmetal transition is given with particular emphasis to the following points: (a) is the transition of first order; (b) can one observe the transition; and (c) how is the transition affected when it occurs in ionic lattices or polar liquids.

I. INTRODUCTION

In 1931 Wilson[1] gave a description of the difference between metals and insulators in terms of a model of noninteracting electrons. This model has been very successful. However it does not always work, as de Boer and Verwey[2] were the first to point out; for instance, cubic nickel oxide should be a metal according to the model. These authors gave a qualitative explanation of why this is so. Wigner in 1938 specifically introduced the electron–electron interaction e^2/r_{12} into the problem and suggested that at low densities a free-electron gas should "crystallize" in a nonconducting state. In my paper[3] of 1949, I discussed a cubic crystalline array of one-electron atoms with lattice parameter d (Fig. 1), and gave reasons for supposing that for large values of d (but not so large as to prevent tunneling) the array must be an insulator. For small values of d, according to the Wilson model and common experience, the array will be a metal. If therefore one varies d, it must have a critical value which I call d_0 at which a transition occurs. If $d > d_0$, the array is an insulator at zero temperature; if $d < d_0$, it is a metal. This does not necessarily imply a phase transition since the change need be sharp only at $T=0$.

If $d > d_0$, an activation energy is necessary to form a pair of carriers. I denote this energy by 2ϵ, so that the conductivity varies as $\exp(-\epsilon/kT)$. If the atoms are far apart,

$$2\epsilon = I - E,$$

where I is the ionization energy and E the electron affinity of the atom. Note that these are unequal owing to the term e^2/r_{12}. As d becomes smaller, ϵ will decrease. Figure 2 shows two ways in which it may change. In my 1949 paper[3] and in later ones[4–6] I argued that the drop must be *discontinuous* (curve b), because an electron and hole can form a pair as a consequence of their Coulomb attraction; they have mutual potential energy $-e^2/\kappa r_{12}$ and therefore binding energy of order $me^4/\hbar^2\kappa^2$. κ is the background dielectric constant. The activation energy in the insulating state cannot be less than this. Pair formation will occur unless the other electrons screen the Coulomb attraction, so that $-e^2/r$ can be replaced by

$$-(e^2/r)\exp(-qr). \tag{1}$$

The constant q, on the Thomas–Fermi approximation, is given by

$$q^2 = 4me^2(3N_0/\pi)^{1/3}/\hbar^2,$$

where N is the density of free electrons. Putting in the condition that no bound states exist when $V(r)$ is given by (1), one obtains an estimate of the value of N for which the transition occurs, namely,

$$N^{1/3}a_H \simeq 0.2, \tag{2}$$

a value which agrees quite well with observations from doped germanium (Mott,[5] Mott and Twose[7]).

Li, Love, and Miller[8] have made essentially this estimate, in a more elaborate way, and have applied it to n-type InSb.

I remember that for almost ten years after my paper of 1949 the Wilson model was so well established in people's minds that the idea of the insulating state built up from one-electron atoms was received with complete incredulity by (at any rate) some of my colleagues. But now the situation has changed; there is a formal proof by Kohn[9] that, for large enough values of d, the array of Fig. 1 is nonconducting and there are developments by Hubbard,[10] Kemeny,[11] and Gutzwiller[12] which describe how the transition can occur. The time is certainly ripe for a conference on the subject and it gives me great pleasure to give the opening paper.

There are several points which I would like par-

---

[1] A. H. Wilson, Proc. Roy. Soc. (London) **A133**, 458 (1931).

[2] J. H. de Boer and E. J. W. Verwey, Proc. Phys. Soc. (London) **49**, 59 (1937).

[3] N. F. Mott, Proc. Phys. Soc. (London) **A62**, 416 (1949).

[4] N. F. Mott, Can. J. Phys. **34**, 1356 (1956).

[5] N. F. Mott, Phil. Mag. **6**, 287 (1961).

[6] N. F. Mott, Advan. Phys. **16**, 49 (1967).

[7] N. F. Mott and W. D. Twose, Advan. Phys. **10**, 107 (1961).

[8] S. P. Li, W. F. Love, and S. C. Miller, Phys. Rev. **162**, 728 (1967).

[9] W. Kohn, Phys. Rev. **133**, 171 (1964).

[10] J. Hubbard, Proc. Roy. Soc. (London) **A277**, 237 (1964).

[11] G. Kemeny, Ann. Phys. (N.Y.) **32**, 69 (1964).

[12] M. C. Gutzwiller, Phys. Rev. **137**, A1726 (1965).

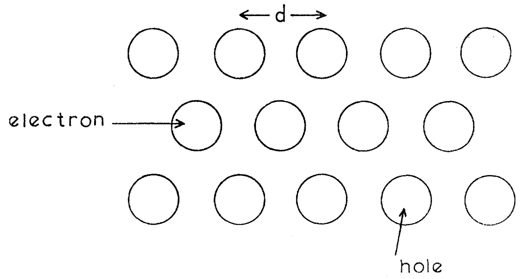

FIG. 1. A crystalline array of monovalent atoms.

ticularly to emphasize. These are:

(a) Is the transition of the first order? Kohn[13] has recently suggested that what should occur in a *rigid* lattice as d changes is an infinite series of transitions (Fig. 3) and this interests me particularly because I think that Kohn's model enables one to escape from the argument that ϵ must change discontinuously, because it allows κ to become infinite (Mott and Davis[14]). I shall develop this argument below.

(b) Can one observe the transition? It seems to me quite possible that one cannot observe it in crystalline arrays because it will always be accompanied by a lattice distortion. One is then limited to donors or acceptors in semiconductors. Here the force holding the donor in position may be very strong compared with that exerted by the electron gas. Since in this case the centers (donors or acceptors) are distributed at random, one has to ask how the transition is affected by this.

(c) How is the transition affected when it occurs in ionic lattices or polar liquids, for example V_2O_3 or doped titanates or sodium in ammonia?

II. TYPES OF METAL-INSULATOR TRANSITION

It is not of course the case that all metal–insulator transitions are due to electron–electron interaction and inexplicable in the model of noninteracting electrons. Naturally a change of crystal structure may lead to a band gap, and it has been suggested that this is what happens in the vanadium oxides (see papers at this conference by Adler[15] and by Hyland[16]). Moreover, as I have already stated and shall show below, a transition that *is* due to the term e^2/r_{12} can provoke a phase change. Also a metal–insulator transition can occur at a Néel point, as predicted by Slater[17] and observed in NiS (Sparks and Komoto[18,19]). Then of course in the Wilson model of noninteracting electrons any *divalent* metal will become an insulator if d is increased, and in some cases also if d is decreased. On the Wilson model ϵ should tend continuously to zero (Fig. 2a). In my early papers I predicted that the introduction of the Coulomb repulsion between electrons and holes would lead here too to a discontinuous change of ϵ; the interesting point here is whether—as I now believe—this prediction has to be replaced by the more complicated phenomenon of the "excitonic insulator" described in a number of papers recently (Keldysh and Kopaev,[20] Kohn,[21] Jérome, Rice, and Kohn,[22] and Halperin[23]).

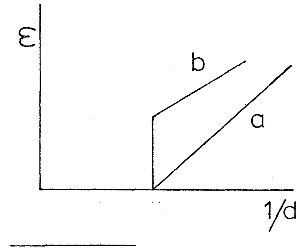

FIG 2. Different predictions about the way the activation energy changes at the transition.

I would like to mention another kind of transition, which can be described in the model of noninteraction electrons, and which I call the "Anderson transition."

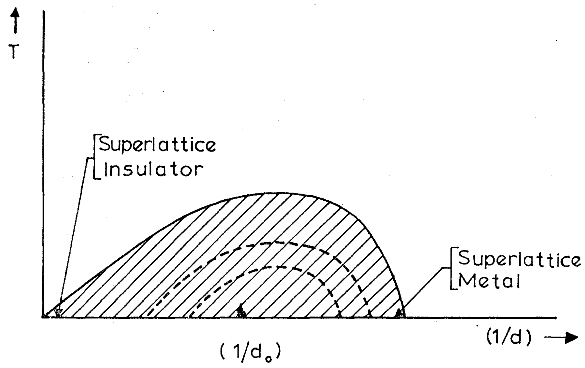

FIG. 3. Kohn's model for the behavior near the metal–insulator transition.

Anderson[24] in 1958 investigated the following model. The atoms of Fig. 1 are supposed now to be far enough apart for the tight binding approximation to be applicable and for the bandwidth J to depend on the interaction between nearest neighbors only. Anderson supposed that a potential energy V_n was applied at each atom n, with a random spread V_0. He showed that if V_0/J exceeded a critical value (about 6), no diffusion was possible. In other words the states are localized, and the conductivity $\sigma_E(0)$ averaged over all configurations (i.e., all values of V) is zero for all E. $\sigma_E(0)$ is the limit as $\omega \rightarrow 0$ of the conductivity at frequency ω given

---

[13] W. Kohn, Phys. Rev. Letters **19**, 789 (1967).

[14] N. F. Mott and E. A. Davis, Phil. Mag. **17**, 1269 (1968).

[15] D. Adler, Rev. Mod. Phys. **40**, 714 (1968), this issue.

[16] G. J. Hyland, Rev. Mod. Phys. **40**, 739 (1968), this issue.

[17] J. C. Slater, Phys. Rev. **82**, 538 (1951).

[18] J. T. Sparks and T. Komoto, Phys. Letters **25A**, 398 (1967).

[19] J. T. Sparks and T. Komoto, Rev. Mod. Phys. **40**, 752 (1968), this issue.

[20] L. V. Keldysh and Yu. V. Kopaev, Sov. Phys.—Solid State **6**, 2219 (1965).

[21] W. Kohn, Phys. Rev. Letters **19**, 439 (1967).

[22] D. Jérome, T. M. Rice and W. Kohn, Phys. Rev. **158**, 462 (1967).

[23] B. I. Halperin, Rev. Mod. Phys. **40**, 755 (1968), this issue.

[24] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).

by (Mott[25])

$$\sigma_E(\omega)=\frac{2\pi e^2\hbar^3}{m^2}\sum_{E'}\sum_{E''}\left|\int\psi_{E''}{}^*\frac{\partial}{\partial x}\psi_{E'}d^3r\right|^2 \times\delta(E-E')\,\delta(E+\hbar\omega-E''). \quad (3)$$

The situation is the same as in impurity conduction (Mott and Twose,[7] Miller and Abrahams,[26] Mott[6]). There is a continuous range of energies in the impurity band, and no energy gap at the Fermi energy; but none the less, since each state is localized and has its own quantized energy, a variable activation energy is needed for each "hop" to any other state for which the orbitals overlap appreciably.

A transition from a metallic to a nonmetallic state can occur, therefore, through a change of V_0, and I shall give later an example of this occurring.

III. THE NATURE OF THE TRANSITION

The first formal treatment of the transition in a crystalline lattice due to the term e^2/r_{12} is due to Hubbard.[10] He uses the Hamiltonian

$$\sum_k(n_{k\uparrow}+n_{k\downarrow})+I\sum n_{k\uparrow}n_{k\downarrow},$$

where $n_{k\sigma}$ is the number operator for an electron in the state k, σ and I is the average repulsive energy between two electrons in the same atom. Hubbard finds energy bands for his pseudoparticles as in Fig. 4. At a critical value d_0 of d an energy gap occurs, so that for $d>d_0$ the subband is occupied and the material is an insulator, with activation energy for conduction 2ϵ as shown. Hubbard's model does not give a discontinuous change in ϵ; but as it only considers short-range interactions, there is no reason why it should, as he himself remarks. His model contains no attraction between a hole on one atom and an electron on the next.

Kohn's model, which derives from recent work on excitonic insulators, uses the Hartree–Fock approximation and takes account of long-range interaction. In this model, as one approaches the transition from the side of small d (metallic conduction), the first thing that happens is that charge density waves (which may or not be different from the two spin directions) are set up. In this respect Kohn's model has similarities to that of Slater[17] in which a nonconducting state was supposed to be caused by the appearance of itinerant electron antiferromagnetism which would cause a splitting of the Brillouin zone and could separate the filled from the empty states. Kohn's waves are similar to the spin density waves which are known to exist in chromium. In any case, Kohn suggests that these are the first instabilities which occur as the density of the electron is decreased. When the value d_1 of d at which this first occurs is reached, a charge density wave of zero amplitude is set up of the form

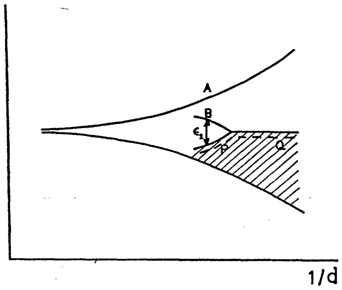

FIG. 4. Band gap in Hubbard's model. *PQ* shows the Fermi level if there is some compensation (Mott and Davis[14]).

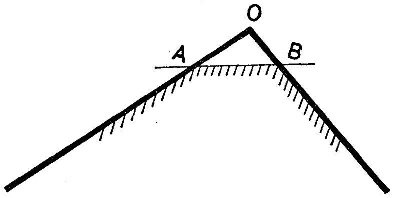

FIG. 5. Brillouin zone in Kohn's model; occupied states are shaded.

$$\Delta\rho=A_1(d)\sin(w_1x+\eta),$$

where w is some definite wave number and $A_1(d)$ an amplitude.

Kohn then goes on to argue that an infinite series of such second-order transitions, each setting up a charge density wave with a wave number w_n, occurs before the transition point d_0 is reached. The phase diagram is supposed to be as in Fig. 3. Naturally the A_n becomes small for large n.

The insulating state is thus described as a situation where the occupied states are separated from the empty states by planes in k space across which there is an energy gap caused by the charge density wave as illustrated in Fig. 5. Two predictions can be made from this model.

(a) It is possible that the dielectric constant κ *does* tend to infinity as d increases towards d_0. This is because, near d_0, some of the boundary planes will have very small energy gap, and allowed transitions can occur with very small energy (Mott and Davis[14]). Thus my former argument that there must be a discontinuous change in κ falls to the ground. In Fig. 3, a vanishingly small number of carriers at $T=0$ is possible just to the insulating side of d_0.

(b) If charge density waves are set up, a distortion of the lattice (a frozen phonon) with the same wave number w must occur. This will produce an energy gap, which cannot be distinguished from that set up by the charge density wave itself. This has been emphasized to the author particularly by Professor J. C. Phillips and R. G. Arkhipov.

For this reason it may be very difficult in principle to observe the Mott transition in a crystalline lattice.

---

[25] N. F. Mott, Phil. Mag. **17,** 1259 (1968).

[26] A. Miller and E. Abrahams, Phys. Rev. **120,** 745 (1960).

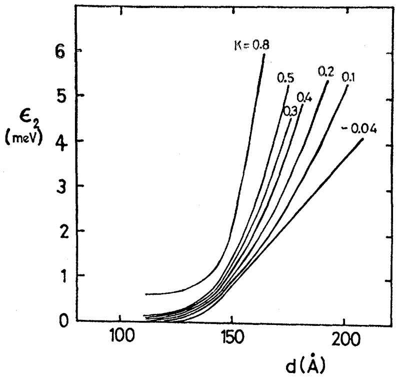

FIG. 6. Activation energy ϵ_2 as a function of the mean distance between centers and compensation K (Davis and Compton[28]).

One can of course observe it for n-type (or p-type) centers in a semiconductor of high dielectric constant such as germanium, so that (2) can be satisfied without unreasonably high doping. Here the covalent forces holding the atoms in position are certainly very strong compared with any forces exerted by the electron gas, and these substances provide a very good approximation to a rigid lattice. But the centers are distributed at random, not in a crystalline array, and the effect of this is what we have next to consider.

IV. THE IMPURITY BAND IN N-TYPE SEMICONDUCTORS

In doped and compensated germanium, three activation energies can be separated in the formula for the conductivity (cf. Fritzsche,[27] Davis and Compton,[28] Mott and Davis[14]), so that

$$\sigma=\sigma_1 \exp(-\epsilon_1/kT)+\sigma_2 \exp(-\epsilon_2/kT)+\sigma_3 \exp(-\epsilon_3/kT),$$

where $\epsilon_1>\epsilon_2>\epsilon_3$ and $\sigma_1\gg\sigma_2\gg\sigma_3$. Here $2\epsilon_2$ is the energy required to remove an electron from a center and put it on one of the neutral centers, where of course it will have a band of energies. Fritzsche[27] was the first to discuss the term in ϵ_2, observed in an intermediate region, in this way. The energy ϵ_2 is the quantity of main interest to us which must be compared with the ϵ sketched in Fig. 2. The term in ϵ_3 occurs *only* in the presence of compensation (i.e., of minority centers) and ϵ_3 is the activation energy for hopping explained satisfactorily by Miller and Abrahams[26]; except to emphasize that it depends essentially on compensation, I shall not discuss it further here.

Figure 6 shows the dependence of ϵ_2 on concentration of impurities for different degrees of compensation; the results are replotted from Davis and Compton[28] (Mott and Davis[14]). Elementary interpretations have been considered by Mott and Twose[7] and by Nishimura,[29] who consider the band of levels available to an extra electron on one of the neutral donor centers, and calculate how it broadens with decreasing d until ϵ_2 disappears. A more sophisticated treatment taking account of the facts of disorder has been given by Mott and Davis[14] with the following conclusions. The transition is a Mott transition, in the sense that it is due to e^2/r_{12}. But with a random array of atoms one would not expect charge-density waves to be set up. One would expect rather that *random* fluctuations of charge density will occur. These will become stronger as d decreases towards the value where the Mott transition is to be expected, and the field thus produced can localize the one-electron states at the Fermi energy in the sense of Anderson.[24] So the transition to the nonmetallic state can, as in Kohn's model, be described in a one-electron approximation, the states at the Fermi energy being localized so that $\langle\sigma(0)\rangle=0$. Mott and Davis show that $\kappa\to\infty$ at the transition and that ϵ_2 shows no discontinuity.

It must be emphasized that no phase change occurs, but simply a sharp value of d where ϵ_2 vanishes. To one side $\langle\sigma(0)\rangle$ vanishes; on the other side $\langle\sigma(0)\rangle\neq 0$. The present author[25] has given reasons for believing that there is a discontinuity in $\langle\sigma(0)\rangle$ at the critical value of d.

It is particularly interesting that at high values of the compensation the activation energy does not entirely disappear at concentrations which ought to be metallic. Mott and Davis ascribe this to the very strong scattering by the charged acceptors, and show that it is strong enough to produce localization. The activation energy which occurs when the compensation is above some limiting value is an example of an "Anderson transition," due to scattering by the charged acceptors rather than to the term e^2/r_{12}.

V. THE TRANSITION IN METALLIC LIQUIDS AT HIGH TEMPERATURES

The work of Franck and Hensel[30] and of Kikoin *et al.*[31] on mercury (see also Franck[32]) and on cesium are reported at this conference. By raising the temperature one can decrease the density and plot ln σ against T at various low values of the density. In the case of a monovalent metal, the interpretation must be similar to that of a doped n-type semiconductor with two differences:

(a) The issue is not complicated by compensation due to acceptors, giving rise to impurity (hopping) conduction.

(b) One does not know how the distribution of atoms changes with volume or temperature. This applies

---

[27] H. Fritzsche, J. Phys. Chem. Solids **6**, 69 (1958).

[28] E. A. Davis and W. D. Compton, Phys. Rev. **140**, A2183 (1965).

[29] H. Nishimura, Phys. Rev. **138**, A815 (1965).

[30] E. U. Franck and F. Hensel, Phys. Rev. **147**, 109 (1966).

[31] I. K. Kikoin, A. P. Senchenkov, É. V. Gel'man, M. M. Korsunskiĭ, and S. P. Naurzakov, Zh. Eksperim. i. Teor. Fiz. **49**, 124 (1965) [English transl. Sov. Phys.—JETP **22**, 89 (1966)].

[32] F. Hensel and E. U. Franck, Rev. Mod. Phys. **40**, 697 (1968), this issue.

equally to alkalis in ammonia (Thompson[33]) where a main difficulty of interpretation is to determine the arrangement of atoms in the nonmetallic region.

For divalent metals (Hg) a tentative description has been given by Mott.[34] For a crystalline material a sharp transition is expected on the one-electron (Wilson) model, and for the liquid at high temperatures one ought to be able to describe the whole phenomenon in terms of noninteracting electrons. I have suggested that the energy gap must be replaced by a "pseudogap," that is to say, a minimum in $N(E)$, and have given reasons for believing that, when the minimum is deep enough $[N(E)/N(E)_{\text{free}}\approx\frac{1}{2}]$, the states in the pseudogap become localized in Anderson's sense. This estimate of the depth of the gap, based on the idea that localization occurs when the electron wavelength and mean free path become comparable, is of course very crude; but I believe the basic idea is correct. If so, during the steady drop in the conductivity down to about 200 $\Omega^{-1}\ \text{cm}^{-1}$ the material remains "metallic," the resistivity reaching the low value due to a short mean free path and a drop in the density of states, and thereafter the material behaves as a semiconductor, electrons hopping from one localized state to another and with a rapidly increasing activation energy.

In the neighborhood of the critical point Franck's resistivity shows a very abnormal dependence on T at constant volume. I would like to speculate that this is a two-phase region, of the same nature as is observed for sodium in ammonia,[33] vapor and liquid being present in equilibrium.

VI. THE METAL–INSULATOR TRANSITION IN COMPOUNDS

A great many observations of the transition have been made in compounds, particularly oxides, and in liquids (alkalis in ammonia or in molten alkali halides). In such cases the problem is complicated by the existence of two dielectric constants, κ the static dielectric constant, and κ_0 the high-frequency dielectric constant. One has here the possibility of polaron formation, particularly when, as in many cases, conduction is in a comparatively narrow d band. In most of the cases considered in this section the electrons are contributed by donors distributed at random (for instance, titanates, tungsten bronzes, or alkali metals dissolved in ammonia). The particular case of VO_2 and similar crystals will be considered in the next section.

It seems to me that one can distinguish three cases:

(i) Weak coupling with phonons or, in other words, the situation when the electron moves so quickly that the ions cannot follow it. In this case, for instance, for tungsten bronzes or alkalis in ammonia, the onset of metallic conduction seems to be given reasonably well by formula (2) if, instead of κ, one writes κ_{eff} where

$$\kappa_{\text{eff}}^{-1}=\kappa^{-1}+\tfrac{5}{16}(\kappa_{\infty}^{-1}-\kappa^{-1}),\tag{4}$$

a formula given originally by Simpson.[35] In liquids such as Na in NH_3 the issue is complicated by uncertainty about the species (Na^+, Na, $Na_2\cdots$) that may form on the nonmetallic side (Thompson[33]). Materials such as GaAs and InSb are probably in this category. In the former, electrons appear to be free at the lowest temperatures for concentrations of impurities greater than 6×10^{13} cm^{-3} (cf. Li *et al.*[8]). Since $m_{\text{eff}}=0.013m$, $\kappa=15.7$, and κ_0 must be nearly as large as κ, the Bohr radius must be very large and this result is not surprising. The treatment given by Li *et al.*[8] does not make any difference between κ and κ_{eff}.

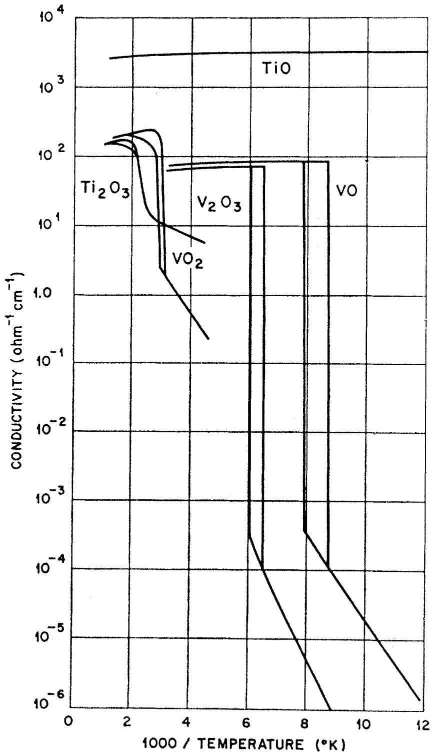

FIG. 7. Conductivities of oxides (Morin[36]).

Here I must emphasize that in crystalline undoped materials metallic conduction undoubtedly can occur in a d band. TiO is a case in point, as shown in Fig. 7, reproduced from Morin's[36] paper. Another case is $NbSe_2$, where the Nb atoms are 3 Å apart and are nearest neighbors in a layer lattice. The resistivity at room temperature is 0.35×10^{-3} Ω cm and drops to about 10^{-5} at 7°K where it becomes superconducting. These results (Yoffe[37]) are for pressed pellets, not single crystals, but there can be no doubt that these materials are metallic. Adler[15] in his review to this

---

[33] J. C. Thompson, Rev. Mod. Phys. **40**, 704 (1968), this issue.

[34] N. F. Mott, Phil. Mag. **13**, 989 (1966).

[35] J. H. Simpson, Proc. Roy. Soc. (London) **A197**, 269 (1949).

[36] F. J. Morin, Bell Syst. Tech. J. **37**, 1047 (1958).

[37] A. D. Yoffe (private communication).

conference gives many other examples. The criterion (2) should doubtless be modified for *d* bands to take into account the degeneracy, *n* being replaced by *n*/3 or *n*/5. If this is done, one is led to the conclusion that metallic conduction in a stoichiometric compound can occur if the band width is greater than about 2 eV or the effective mass less than about 5*m*. But the criterion (2) takes into account only the long-range forces between electron and hole. If metallic conduction occurs in materials with *d* bands as narrow as this, the short-range Hubbard interaction must be very small. In other words, in TiO to form Ti^{3+} and Ti^{+} the interaction between the electrons on Ti^{+} must give rise to negligible energy, presumably due to screening by the oxide ions which are strongly hybridized with the 3*d* orbitals. I assume this to be the case.

(ii) Cases where a free electron forms a small polaron. This will normally occur when the conduction band is a narrow *d* band, say less than 0.5 eV wide, and the static dielectric constant is high. At low temperatures the electron will behave as if it had high effective mass, and no "Mott transition" is therefore possible. On the other hand, the attractive force between the positively charged centers and the electron may be so small that escape of an electron from a center is possible at very low temperatures, so a first assessment of the facts might lead one to guess that the concentration of carriers was above that for the Mott transition. Examples where this happens may include $BaTiO_3$ (Gerthsen, Groth, and Härdtl[38]) and TiO_2 (Bogomolov and Zhuze[39]).

As is shown in the next section, it is possible that this is happening also in some of the crystalline vanadium oxides.

(iii) The case of $KTaO_3$ and similar ferroelectrics which show superconductivity. The puzzling situation presented by these materials was emphasized by Mott[6] (p. 123). They have a comparatively wide band gap and κ_∞ is not large; κ, on the other hand, is very large and ferroelectricity may occur. But unlike WO_3 doped with Na (tungsten bronzes), at low temperatures the electrons are free at very low concentrations of donors (2.7×10^{17} cm^{-3} according to Tufte and Chapman[40]) and unlike $BaTiO_4$ the electrons have very high mobility; clearly no small polarons are formed. I see no obvious way of explaining this behavior.

VII. THE VANADIUM OXIDES

The theoretical description given by Adler and others at Harvard is reviewed by Adler[15] at this conference. They point out that a lattice distortion occurs at the transition point, and suggest that the energy gap set up by this distortion separates occupied from empty states. As the temperature is increased, electrons are excited across the gap, and the entropy of these electrons leads to the disappearance of the distortion and hence of the energy gap. The material is then a metal. But in order to get enough entropy they have to assume (for V_2O_3) a very narrow band with $m_{\mathrm{eff}}\sim 50m$, so that the electron gas is only partially degenerate. My worry about this model is that so narrow a band must, if my speculations are not wildly wrong, be on the nonmetal side of the Mott transition, probably even more so than NiO which I think has a wider band than this.

It seems to me that a hypothesis worth considering is that the so-called metallic phase is not metallic at all, but that what we have here is a nondegenerate gas of small polarons. This would be consistent with the comparatively low conductivity (Fig. 7) and the temperature-dependent paramagnetism of the high temperature phase in VO_2.

I do not think this is at all impossible, if I assume that, even in the high-temperature phase, a Mott–Hubbard band gap as in Fig. 4 exists, but a small one (some tenths of an electron volt). Since transitions across this band gap are optically forbidden, the small band gap does not imply a large value of κ_∞. So if (in VO_2 for example) a pair of carriers V^{4+} and V^{2+} is formed, they may both form small polarons, and since the energy of a small polaron is expected to be of order

$$-\tfrac{1}{2}(e^2/r_0)(\kappa_\infty^{-1}-\kappa^{-1}), \qquad r_0\sim 1\ \text{Å},$$

the energy released may be *greater* than the energy required to form the pair of carriers in a lattice where the ions are held rigidly in place.

If the high-temperature phase, then, can be envisaged as a sort of Debye–Hückel gas of small polarons, it is essential to assume that in the high-temperature phase the static dielectric constant is very large, so that the attraction or repulsion between the carriers is small. The static dielectric constant of VO_2 lies in the range 26–40 (Barker *et al.*[41]) in the low-temperature phase and may well be much higher in the high-temperature phase. Thus I postulate that the high-temperature phase may be interpreted as a nondegenerate gas of small polarons interacting through very weak Coulomb forces, but which would condense in some way at a low enough temperature.

Discussion of Mott's Paper

H. Fritzsche (University of Chicago): When we talk about metal–nonmetal transitions we consider the material nonmetallic when it shows an activation energy and thereby is an insulator when the resistance is extrapolated to low temperatures. It is

---

[38] P. Gerthsen, R. Groth and K. H. Härdtl, Phys. Status Solidi **11**, 303 (1965).

[39] V. N. Bogomolov and Z. P. Zhuze, Sov. Phys.—Solid State **8**, 1904 (1967).

[40] O. N. Tufte and P. W. Chapman, Phys. Rev. **155**, 796 (1967).

[41] A. S. Barker, H. W. Verleur, and H. J. Guggenheim, Phys. Rev. Letters **17**, 1286 (1966).

considered to be metallic when the temperature resistance coefficient is zero and positive, and hence the low-temperature extrapolation yields a finite resistance. In semiconductors we usually consider the concentration at which the activation energy ϵ_2 of the resistivity vanishes the critical value N_c for the metal–nonmetal transition. The steepness of the ϵ_2 versus impurity concentration curve is often used as a measure of abruptness of the metal–nonmetal transition. I wish to point out the danger in calling the concentration range $N>N_c$ simply "metallic." There are several phenomena like the anomalous magnetoresistance effect which exist above as well as below N_c and which do not pay any attention to the transition. The Hall mobility continues to increase, for example, up to concentrations of 10^{18} cm^{-3} in Ge, although $N_c=10^{17}$ cm^{-3} for Sb-doped Ge and 3×10^{17} cm^{-3} for As-doped Ge [see M. Cuevas and H. Fritzsche, Phys. Rev. **137**, A1847 (1965); **139**, A1628 (1965)]. Only above $N=10^{18}$ cm^{-3} does the mobility decrease with increasing N as is expected for an impure metal. My question to Professor Mott is whether it might not be important for the understanding of the nature of the transition and its degree of abruptness to study both concentration ranges adjacent to N_c.

N. F. MOTT: It is a pity we have not got Professor Holstein at this meeting, because he and Friedman have made calculations of the Hall coefficient when conduction is due to hopping in impurity bands, and under his inspiration Amitay and Pollak have tried unsuccessfully to measure it. In a disordered system, on the metallic side of the transition, I believe that the mean free path will be very short, of the order of the de Broglie wavelength, which is as short as it can be. As far as I know, no theoretical derivation of the Hall coefficient has been given when the mean free path is as short as this. It would be very useful to have one for the interpretation of many observations, for instance the work of Cutler and of Enderby on liquid TeTl alloys.

H. FREDERIKSE (National Bureau of Standards): One of the parameters that appears in the activation energy, about which we haven't heard much this morning, is the dielectric constant κ. One often talks about the regime in which the optical dielectric constant is being used as compared to the one that is governed by the static dielectric constant. (I can imagine—by the way—if one goes to a *very* localized situation, that $\kappa=1$ might be a better approximation.) However, the whole transition between localized and delocalized states is very closely associated with the possibility of interpolating between the optical and the static dielectric constant. A brave attempt in this direction was made some time ago by Haken [H. Haken, in *Polarons and Excitons*, G. G. Kuper and G. D. Whitfield, Eds. (Plenum Press, New York, 1963), p. 295] which does not seem to fit in the few cases that it could be tested. I wonder if Dr. Mott could comment on this?

N. F. MOTT: Certainly I could talk about the static and dynamic dielectric constants and would have put this in my lecture, but would then have greatly exceeded the time allowed. In germanium and silicon the high-frequency and static dielectric constants are about the same and no problem arises. In the oxides, however, it seems that if the conditions for large polaron behavior are satisfied, the static dielectric constant does not play a major role and the Mott transition occurs at a concentration determined primarily by the high-frequency value. This must be the case, for instance, in the ammonia–metal solutions described by Thompson and also in the tungsten bronzes. If, however, the coupling constant between phonons and electrons is large and small polarons are formed, there will be quite a different situation and I do not think that the Mott transition, as I have described it, can occur. There is a point here about which I feel completely in the dark. Why is it that in the tungsten bronzes, materials with a high static dielectric constant and a moderate refractive index, show a transition at about 20% sodium, so that the static dielectric constant is not much affecting the issue, while on the other hand, in strontium tantalate and in some titanates, electrons being introduced by slight reduction, a metallic behavior just like that of doped germanium occurs for as low a concentration of donors as 3×10^{17} cm^{-3}? These materials have properties very similar to WO_3 and I do not see why this difference should occur. In the tantalates and titanates somehow the high dielectric constant seems able to free the electrons from the donors without turning them into small polarons and so giving them a low mobility.

R. W. KEYES (IBM): I'd like to come back to the point about the possibility of donors being displaced during a phase transition. I think that the electronic forces are large enough so that one shouldn't throw this possibility out although I don't know of any case where there is evidence that it occurs. The point is that the ground state of a donor in a multivalley semiconductor like silicon or germanium is degenerate or nearly degenerate [C. Kittel and A. H. Mitchell, Phys. Rev. **96**, 1488 (1954); M. A. Lampert, Phys. Rev. **97**, 352 (1955); W. Kohn and J. Luttinger, Phys. Rev. **97**, 1721 (1955); **98**, 915 (1955)], so that Jahn–Teller effects are possible. Although the numbers are such that such effects are not likely to be important in silicon and germanium [R. W. Keyes, Solid State Phys. **20**, 37 (1968)], it is not possible to exclude them for multivalley semiconductors in general. Another view of these effects looks at the energy of a donor in germanium as worked out by Peter Price [P. J. Price, Phys. Rev. **104**, 1223 (1956)], for example, as a function of strain. It is seen that there is a negative term in the energy of the lowest state of the donor which is proportional to the square of a strain, in other words, a negative contribution to an elastic constant [R. W. Keyes, IBM J. Res. Develop. **5**, 266 (1961)]. If this term actually became greater than the elastic constant, one would get some kind of a transition. Now in germanium, it doesn't because before one gets to the state where the number of electrons is large enough to overcome the elastic energy, the overlap of the donors is so large that the effect is wiped out; the overlap integral becomes greater than the chemical shift. [See previous reference.] The change in elastic constants is large, however. My colleague John Hall has measured the elastic constant of germanium containing as few as 3×10^{15} antimony donors, which is only perhaps about a part of 10^7 doping, and he finds something like a part in 10^3 change in the elastic constant. The electronic effects in elastic properties may affect the nature of metal–nonmetal transitions in multivalley semiconductors at low temperatures and also constitute another tool for study of the transition.

G. J. HYLAND (University of Liverpool): At one stage in his lecture Professor Mott mentioned long- and short-range effects in connection with Hubbard's work. Now I would just like to point out that while the long-range part of the Coulomb interaction can always be treated in terms of plasma oscillations, the description of *short*-range correlations requires not only the wave functions of a single band but, in addition, those from all higher bands; these short-range effects must be considered in cases where there is appreciable population of a band.