Theory of the Insulating State*

WALTER KOHN
University of California, San Diego, La Jolla, California
(Received 30 August 1963)

In this paper a new and more comprehensive characterization of the insulating state of matter is developed. This characterization includes the conventional insulators with energy gap as well as systems discussed by Mott which, in band theory, would be metals. The essential property is this: Every low-lying wave function Φ of an insulating ring breaks up into a sum of functions, $\Phi=\sum_{-\infty}^{\infty}\Phi_M$, which are localized in disconnected regions of the many-particle configuration space and have essentially vanishing overlap. This property is the analog of localization for a single particle and leads directly to the electrical properties characteristic of insulators. An Appendix deals with a soluble model exhibiting a transition between an insulating and a conducting state.

1. INTRODUCTION

MOST of the recent developments of many-body theory take as their starting point a gas of free particles. This approach is appropriate for a wide class of systems including nuclear matter, electrons in metals and liquid helium. However, as has been emphasized especially by Wigner[1] and Mott,[2] a collection of electrons at low temperatures and *low densities* must be expected to occupy states which are quite unrelated to a gas-like phase and in which the electrons are "localized."

Just what is the precise nature of this localization, in view of the fact that certainly there are finite overlap integrals between the single-particle electron wave functions? This is one of the questions to which a sharp answer is proposed in this paper.

Another closely related question is concerned with the electrical properties of such localized electrons. If their ground state is isolated from the excited states by a finite energy gap, as in a "conventional" insulator like NaCl or Ge, the low-frequency conductivity at $T=0°$ is immediately seen to vanish, by well-known elementary considerations. However, Mott has conjectured the existence of electronic systems *without an energy gap* which are also strict insulators. In this paper we show that insulating characteristics are a strict consequence of electronic localization (in an appropriate sense) and do not require an energy gap.

To make the issues more precise, consider a regular cubic lattice of hydrogen atoms with lattice parameter a at $T=0°$. We begin by considering the high-density regime, $a\ll a_0$, where a_0 is the Bohr radius.

Here the kinetic and potential energies per particle are, respectively,

$$t\approx\gamma_1\times\mathrm{Ry}\times(a_0/a)^2, \tag{1.1}$$

$$v\approx\gamma_2\times\mathrm{Ry}\times(a_0/a), \tag{1.2}$$

where γ_1 and γ_2 are of the order of unity. Hence

$$v\ll t. \tag{1.3}$$

Under these circumstances we expect a state resembling a Fermi gas and that the system will have a metallic character. If we denote the conductivity by

$$\sigma(\omega)=\sigma'(\omega)+i\sigma''(\omega), \tag{1.4}$$

we expect at low frequencies the behavior characteristic of free acceleration,

$$\omega\to 0:\sigma''(\omega)=-ne^2/m^*\omega, \tag{1.5}$$

where n is the density of the electrons and m^* is an effective mass parameter, of the order of the free electron mass.

Now consider $\sigma''(\omega)$ as the ratio (a/a_0) grows and the mean density of the electrons tends to zero. The electronic density between the hydrogen nuclei will approach zero and hence free acceleration will certainly become more difficult. However, an interesting question is now this: As (a/a_0) grows to infinity, does

(a) the conductivity $\sigma''(\omega)$ for low ω maintain the form (1.5), but with m^*/m presumably growing to infinity, i.e., the system remains a metal but with a larger and larger effective mass; or

(b) does the nature of the wave function change abruptly for some critical value of a/a_0, beyond which the electrons are localized and free acceleration ceases *entirely* in the sense that

$$\lim_{\omega\to 0}\omega\sigma''(\omega)=0. \tag{1.6}$$

Mott has given qualitative arguments in favor of the second answer and adduced experimental evidence to support this conclusion. In this paper we aim to place this conclusion on a more precise and firm theoretical basis.

2. CURRENT AND GAUGE TRANSFORMATION

We consider a system in the shape of a ring. For our purposes this is sufficiently characterized by imposing periodic boundary conditions in the x direction

$$\Phi(x_1y_1z_1;\cdots;x_i+L,y_i,z_i;\cdots)=\Phi(x_1y_1z_1;\cdots;x_i,y_i,z_i;\cdots). \tag{2.1}$$

* Supported in part by the Office of Naval Research.

[1] E. Wigner, Trans. Faraday Soc. **34**, 678 (1938).

[2] N. F. Mott, Proc. Phys. Soc. (London) **62**, 416 (1949); Progr. Metal Phys. **3**, 76 (1952); Can. J. Phys. **34**, 1356 (1956); Nuovo Cimento Suppl. **7**, 318 (1958); Phil. Mag. **6**, 287 (1961).

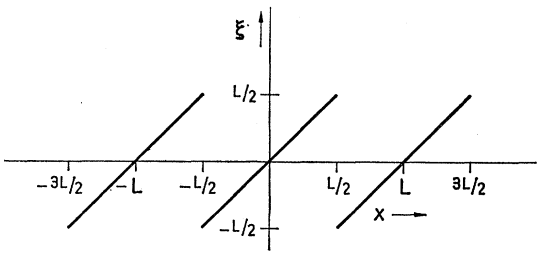

FIG. 1. The function $\xi(x)$.

We find it very useful to introduce, as a formal device, a constant vector potential in the x direction

$$\mathbf{A}\equiv-(c/e)\mathbf{k}, \tag{2.2}$$

$$\mathbf{k}=(k,0,0). \tag{2.3}$$

(This may be thought of as arising from a magnetic flux through the center of the ring.) It gives rise to no electric or magnetic fields inside the system. Our Hamiltonian is then, in atomic units,

$$H(k)\equiv\sum_i[\tfrac{1}{2}(\mathbf{p}_i+\mathbf{k})^2+V_i]+U, \tag{2.4}$$

where V_i is the external potential and U is the potential energy of interaction.

We denote the eigenfunction and eigenvalues of (2.4) and (2.1) by $\Phi_\alpha(k)$ and $E_\alpha(k)$, where the α's are state labels. Then

$$E_\alpha(k)=[\Phi_\alpha(k),H(k)\Phi_\alpha(k)] \tag{2.5}$$

and, because of the stationary nature of (2.5),

$$\begin{aligned} dE_\alpha(k)/dk&=\{\Phi_\alpha(k),[dH(k)/dk]\Phi_\alpha(k)\},\\ &=[\Phi_\alpha(k),\textstyle\sum_i(\mathbf{p}_i+\mathbf{k})_x\Phi_\alpha(k)], \qquad (2.6)\\ &=-\Omega J_\alpha(k), \end{aligned}$$

where Ω is the volume and $J_\alpha(k)$ is the current density in the x direction carried by the state α.

Another result which follows directly from (2.4) is

$$\left.\frac{d^2E_\alpha(k)}{dk^2}\right|_{k=0}=N-2\sum_{\alpha'\neq\alpha}\frac{|(\alpha'|P|\alpha)|^2}{E_{\alpha'}-E_\alpha}, \tag{2.7}$$

where

$$P\equiv(\textstyle\sum_i p_i)_x. \tag{2.8}$$

Next we consider the response of the system to a uniform time-dependent electric field in the x direction, $\mathcal{E}e^{i\omega t}e^{st}$. Here s is a positive infinitesimal which describes a field whose amplitude is slowly being turned on. The appropriate Hamiltonian has again the form (2.4) with

$$\mathbf{k}=(1/i\omega)(\mathcal{E}e^{(i\omega+s)t},0,0). \tag{2.9}$$

(Here, and in what follows, s is dropped where it plays no significant role.)

A straightforward calculation gives for the induced current density

$$J_\alpha=\sigma_\alpha(\omega)\mathcal{E}e^{(i\omega+s)t}, \tag{2.10}$$

where

$$\sigma_\alpha(\omega)=\frac{1}{\Omega}\left[\frac{N}{i\omega}-\frac{1}{i\omega}\sum_{\alpha'}'|(\alpha'|P|\alpha)|^2\left(\frac{1}{E_{\alpha'}-E_\alpha-\omega-is}+\frac{1}{E_{\alpha'}-E_\alpha+\omega+is}\right)\right]. \tag{2.11}$$

In particular for the imaginary part of $\sigma_\alpha(\omega)$ we find

$$\sigma_\alpha''(\omega)=-\frac{1}{\Omega}\frac{1}{\omega}\left[N-\sum_{\alpha'}'|(\alpha'|P|\alpha)|^2\left(\frac{E_{\alpha'}-E_\alpha-\omega}{(E_{\alpha'}-E_\alpha-\omega)^2+s^2}+\frac{E_{\alpha'}-E_\alpha+\omega}{(E_{\alpha'}-E_\alpha+\omega)^2+s^2}\right)\right]. \tag{2.12}$$

Comparison with (2.7) gives the important result

$$\lim_{\omega\to0}\omega\sigma_\alpha''(\omega)=-\frac{1}{\Omega}\frac{d^2E}{dk^2}. \tag{2.13}$$

For orientation we now consider two especially simple systems.

Free Particle

Here

$$H(k)=\tfrac{1}{2}(\mathbf{p}+\mathbf{k})^2 \tag{2.14}$$

whose eigenfunctions are, for *all* k

$$\varphi(\mathbf{r})=(1/\Omega^{1/2})e^{i\mathbf{q}\cdot\mathbf{r}}, \tag{2.15}$$

where, for simplicity, we assume periodic boundary conditions also in the y and z directions, which is however of no further consequence. Here

$$q_i=(2\pi/L)xl, \quad l=0,\pm1,\pm2,\cdots \tag{2.16}$$

to satisfy the ring conditions (2.1). Clearly

$$E(k)=\tfrac{1}{2}(q_1+k)^2+\tfrac{1}{2}(q_2^2+q_3^2). \tag{2.17}$$

Hence by (2.6), (2.7), and (2.13)

$$J(k)=-(1/\Omega)(q_1+k), \tag{2.18}$$

$$d^2E(k)/dk^2=1, \tag{2.19}$$

$$\lim_{\omega\to0}\omega\sigma''(\omega)=-\frac{1}{\Omega}. \tag{2.20}$$

These are all very familiar facts. Note, however, that in the present context they depend on a failure of gauge invariance in the conventional sense.[3] One is accustomed to assuming that the spectrum of (2.4) is independent of k, since the latter may be removed by a

[3] Y. Aharanoff and D. Bohm, Phys. Rev. **115**, 485 (1959).

simple gauge transformation

$$S=\exp(-ik\sum_i x_i). \tag{2.21}$$

However, the new wave function

$$\Phi'=\Phi\exp(-ik\sum_i x_i) \tag{2.22}$$

no longer satisfies the ring condition (2.1) if Φ did, and hence is not admissible. Thus it is not surprising that the correct E does have a k dependence.

Localized Electron

As a second example consider an electron bound near the origin in a region of linear dimensions $d \ll L$.

Define now $\xi(x)$, for any value of x, as the unique member of the set

$$x+\nu L \quad \nu=0, \pm 1, \pm 2, \cdots \tag{2.23}$$

such that

$$-L/2<\xi(x)\leqslant L/2. \tag{2.24}$$

(See Fig. 1.) Then if we denote the bound state of the electron corresponding to $k=0$ by $\Phi_\alpha(x,y,z;0)$, we have

$$\Phi_\alpha(x,y,z;0)\doteq 0 \quad \text{for} \quad \xi(x)\approx \pm L/2. \tag{2.25}$$

Here we have introduced the symbol $\doteq$ for equality apart from terms which tend exponentially to zero as $L\to\infty$ in a manner such as $e^{-L/b}$ where b is of atomic dimension and independent of L.

We now see that

$$\Phi_\alpha(x,y,z;k)\doteq\Phi_\alpha(x,y,z;0)e^{-ik\xi(x)}. \tag{2.26}$$

For the right-hand side of (2.26) satisfies the Schroedinger equation everywhere (except at the isolated points $x=\pm L/2, \pm 3L/2, \cdots$), and also obeys the ring boundary conditions. It does have impermissible discontinuities at the points $x=\pm L/2, \pm 3/L2, \cdots$, but since the wave function is near vanishing (in the sense $\doteq 0$) at these points, these discontinuities are of no consequence.[4] We therefore have

$$E_\alpha(k)\doteq E_\alpha(0), \tag{2.29}$$

and hence

$$J_\alpha(k)\doteq 0, \tag{2.30}$$

$$\lim_{\omega\to 0}\omega\sigma_\alpha''(\omega)=0. \tag{2.31}$$

Equations (2.30) and (2.31) reflect the nonconducting nature of the state. We see that it is intimately related to its localized character which permits the application of a conventional gauge transformation.

[4] If we consider L to be the x dimension of the unit cell of a periodic lattice, and remember that the introduction of k into H is equivalent to a change of boundary conditions, we see that these facts are familiar results of the extreme tight binding limit of band theory.

3. GAUGE INVARIANCE—SECOND-ORDER PERTURBATION THEORY

We now return to the lattice of hydrogen atoms described in the introduction. We demonstrate in this and the following section that for sufficiently large values of a/a_0 (low density), the energies of all low-lying levels are gauge invariant in the sense of (2.29), from which the insulator properties (2.30) and (2.31) follow.[5]

A major difficulty is that the wave functions for the lowest lying states of this system, which are presumably of an antiferromagnetic character, are not known even in zeroth order of approximation. For this reason we do not discuss the wave functions themselves but rather generate a k-independent effective Hamiltonian from which the gauge invariance of the entire low-lying part of the spectrum follows.

The external potential energy of our hydrogen lattice is

$$V(\mathbf{r})=\sum_{\nu=1}^{N} v(\mathbf{r}-\mathbf{R}_\nu), \tag{3.1}$$

where we assume $v(\mathbf{r})$ to fall off rapidly,[6] but where, since x and $x+L$ are physically identical, we have

$$v(x+L,y,z)=v(x,y,z). \tag{3.2}$$

In this lattice there are N particles interacting with a short-range[7] repulsion, giving rise to the additional potential energy

$$U=\sum_{i>j} u(\mathbf{r}_i-\mathbf{r}_j), \tag{3.3}$$

where the u's satisfy periodicity conditions analogous to (3.2). The total Hamiltonian, for $k=0$, is then

$$H=\sum_i [\tfrac{1}{2}p_i^2+V(\mathbf{r}_i)]+U. \tag{3.4}$$

We shall work in a representation of Wannier functions $w_n(\mathbf{r}-\mathbf{R}_\nu)$ associated with the single-particle Hamiltonian

$$h\equiv\tfrac{1}{2}p^2+V(\mathbf{r}), \tag{3.5}$$

We denote by $\varphi_n(\mathbf{r},\mathbf{q})$ the normalized Bloch-like eigenfunction of band index n and wave vector $\mathbf{q}$ associated with h:

$$\varphi_n(\mathbf{r}+\boldsymbol{\tau};\mathbf{q})=\exp(i\mathbf{q}\cdot\boldsymbol{\tau})\varphi_n(\mathbf{r};\mathbf{q}), \tag{3.6}$$
$$q_x=(2\pi/L)\times\text{integer},$$

where $\boldsymbol{\tau}$ is a lattice translation vector. Then

$$w_n(\mathbf{r}-\mathbf{R}_\nu)=(1/N^{1/2})\sum_{\mathbf{q}}\varphi_n(\mathbf{r},\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{R}_\nu}. \tag{3.7}$$

[5] In Sec. 5 we add to these the result $\lim_{\omega\to 0}\sigma'(\omega)=0$.

[6] The true $v(\mathbf{r})$ is of course Coulombic and a satisfactory treatment of such long-range forces would, as usual, require a more subtle discussion. We cannot, however, see any physical reason why our conclusions should be altered by the long-range character of the true interactions.

[7] See remarks of footnote 6.

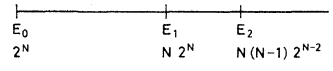

FIG. 2. Schematic spectrum of H_0 and degeneracies.

Notice that, in view of (3.6), our Wannier functions satisfy periodic bondary conditions

$$w_n(x+L, y, z)=w_n(x,y,z). \tag{3.8}$$

Denote by $A_{nm\nu}$ the destruction operator for an electron described by the Wannier function $w_n(\mathbf{r}-\mathbf{R}_\nu)$ and spin $m=\pm\frac{1}{2}$. We then use as a basis the set of states

$$(A_{n_N m_N \nu_N}{}^*)\cdots\quad(A_{n_1 m_1 \nu_1}{}^*)\Psi^{(0)}, \tag{3.9}$$

where $\Psi^{(0)}$ is the vacuum state.

In this representation we write the Hamiltonian in second quantized form as the sum of three terms:

$$H=H_0+H_1+H'. \tag{3.10}$$

H_0 is diagonal and describes (a) the expectation value of the one-particle part of H and (b) that part of the expectation value of U arising from pairs of Wannier functions on the same site $\mathbf{R}_\nu$. If we introduce the number operator

$$N_{nm\nu}\equiv A_{nm\nu}{}^* A_{nm\nu}, \tag{3.11}$$

then H_0 has the form

$$H_0=\sum_{n,m,\nu}\epsilon_n N_{n,m,\nu}+\sum_{nm}\sum_{n'm'}\sum_{\nu}U_{nn'}N_{nm\nu}N_{n'm'\nu}. \tag{3.12}$$

The spectrum of H_0 is schematically shown in Fig. 2. E_0 corresponds to each site being singly occupied by a particle in the lowest band Wannier function, spin up or down (degeneracy$=2^N$); E_1 to one site having an electron in the Wannier function of the first excited band (degeneracy$=N.2^N$); E_2 to one unoccupied site and doubly occupied with electrons in the lowest Wannier function, etc. Notice that the spectrum is discrete. We denote the general eigenvalue of H_0 by E_J, and its eigenstates of the form (3.9) by $\Psi_{J,j}$, where j is a degeneracy index.

H_1 is defined in such a way that H_0+H_1 is identical with H within each subspace $\Psi_{J,j}$, $j=1, 2, \cdots$, but has no matrix elements connecting states of different J; i.e.,

$$(J',j'|H_1|J,j)=\delta_{J'J}[(J',j'|H|J,j)-E_J\delta_{j'j}]. \tag{3.13}$$

Finally H' is that part of H which connects states of different J

$$(J',j'|H'|J,j)=(1-\delta_{J'J})(J',j'|H|J,j). \tag{3.14}$$

We are interested in the low-lying states which arise out of E_0 when the effects of H_1 and H' are taken into account, which will partly remove the degeneracy. In the limit of large separation the energy of these states is almost wholly determined by H_0. Furthermore E_0 remains separated from the first excited level by a finite gap. For these reasons we feel justified in assuming that the effects of H_1 and H' may be treated by the appropriate degenerate perturbation theory, although a formal proof of its convergence is not attempted. We therefore write

$$H=H_0+\lambda H_1+\lambda H' \tag{3.15}$$

and treat λ formally as small, although of course in fact $\lambda=1$. We now look for a canonical transformation

$$\tilde{H}=e^{-iS}He^{iS}=H+i[H,S]+\frac{i^2}{2}[[H,S],S]+\cdots \tag{3.16}$$

with

$$S=\lambda S_1+\lambda^2 S_2+\cdots \tag{3.17}$$

which eliminates matrix elements connecting states with $J=0$ to states with $J\neq 0$, to all orders in λ. Combining (3.15), (3.16), and (3.17) we obtain

$$\begin{aligned}\tilde{H}=&H_0+\lambda\{H_1+H'+i[H_0,S_1]\}\\&+\lambda^2\{i[H_0,S_2]+i[H_1+H', S_1]+(i^2/2)[[H_0,S_1],S_1]\}\\&+\lambda^3\cdots.\end{aligned} \tag{3.18}$$

It is evidently possible to choose S to all orders such that

$$(J',j'|S|J,j)=0 \quad \text{unless} \quad J=0, J'\neq 0; \quad \text{or} \quad J\neq 0, J'=0, \tag{3.19}$$

To first order in λ

$$(J',j'|S_1|0,j)=(1/i)[1/(E_0-E_{J'})](J',j'|H'|0,j), \quad J'\neq 0 \tag{3.20}$$

and hence up to second order in λ

$$(0,j_2|\tilde{H}|0,j_1)=(0,j_1|H_0+\lambda H_1|0,j_2)+\lambda^2\sum_{J',j'}\frac{(0,j_2|H'|J',j')(J',j'|H'|0,j_1)}{E_0-E_{J'}}. \tag{3.21}$$

$\tilde{H}$ is the new effective Hamiltonian in the 2^N dimensional space, $J=0$, $j=1, \cdots 2^N$.

Next we construct in a similar way the effective Hamiltonian corresponding to

$$H(k)=\sum_i[\tfrac{1}{2}(\mathbf{p}_i+\mathbf{k})^2+V(\mathbf{r}_i)]+U \tag{3.22}$$

with $k\neq 0$. For this purpose we use the fact that $w_n(\mathbf{r})$ is an exponentially decreasing function of $\mathbf{r}$,[8] the characteristic length being an atomic dimension.[9] We now introduce a new representation in which the Wannier functions $w_n(\mathbf{r})$ are replaced by

$$w_n(\mathbf{r}; k)=w_n(\mathbf{r})e^{-ik\xi(x)}. \tag{3.23}$$

[8] W. Kohn, Phys. Rev. **115**, 809 (1959). Actually this paper establishes the exponential drop of Wannier functions in one dimension only. We assume here that three-dimensional Wannier functions have the same property.

[9] It must, however, be remarked that, as the band index n approaches infinity, the spread of the Wannier function also increases without limit. But one may see, in simple model calculations, that this difficulty is spurious and that one arrives at the correct conclusions by regarding the spread of w_n as generally of the order of an atomic dimension rather than L.

Like $w_n(\mathbf{r})$ these functions satisfy the ring condition (3.8). The lack of continuity at $x=L/2$ is of no consequence in view of the just mentioned near vanishing of $w_n(\mathbf{r})$ at that point.

Consider now some typical matrix elements of $H(k)$. For example

$$M_1\equiv[w_n'(\mathbf{r}-\mathbf{R}_1';k),\tfrac{1}{2}(\mathbf{p}+\mathbf{k})^2w_n(\mathbf{r}-\mathbf{R}_1;k)]. \quad (3.24)$$

This is clearly a function only of $\mathbf{R}_1'-\mathbf{R}_1$, and we may therefore take

$$\xi(X_1)=0. \quad (3.25)$$

Because of the localization of the Wannier functions it then follows that unless

$$\xi(X_1')\ll L/2$$

the matrix element vanishes. In view of (3.24) and (3.25) we obtain

$$M_1\doteq e^{-ik\xi(X_1'-X_1)}[w_{n'}(\mathbf{r}-\mathbf{R}_1'),\tfrac{1}{2}p^2w_n(\mathbf{r}-\mathbf{R}_1)]. \quad (3.26)$$

Quite similarly

$$M_2\equiv[w_{n'}(\mathbf{r}-\mathbf{R}_1',k)\ V(\mathbf{r})w_n(\mathbf{r}-\mathbf{R}_1;k)] \quad (3.27)$$
$$\doteq e^{-ik\xi(X_1'-X_1)}[w_{n'}(\mathbf{r}-\mathbf{R}_1'),\ V(\mathbf{r})w_n(\mathbf{r}-\mathbf{R}_1)]$$

and

$$M_3\equiv[w_{m'}(\mathbf{r}_1-\mathbf{R}_1',k)w_{n'}(\mathbf{r}_2-\mathbf{R}_2'k),$$
$$\times u(\mathbf{r}_1-\mathbf{r}_2)w_m(\mathbf{r}_1-\mathbf{R}_1,k)w_n(\mathbf{r}_2-\mathbf{R}_2,k)]$$
$$\doteq e^{-ik\xi(X_1'+X_2'-X_1-X_2)}[w_{m'}(\mathbf{r}_1-\mathbf{R}_1')w_{n'}(\mathbf{r}_2-\mathbf{R}_2'),$$
$$\times u(\mathbf{r}_1-\mathbf{r}_2)w_m(\mathbf{r}_1-\mathbf{R}_1)w_n(\mathbf{r}_2-\mathbf{R}_2)], \quad (3.28)$$

where we have used the assumed short-range properties of v and u.

The construction of the effective Hamiltonian proceeds now exactly as before and leads again to an expression of the form (3.21). Clearly by (3.26)–(3.28) all diagonal matrix elements E_J are the same as before. Off-diagonal elements differ by phase factors which lead to such combinations as

$$e^{ik\xi(X_1'+X_2'-X_1-X_2)}e^{-ik\xi(X_1'+X_2'-X_1-X_2)}=1. \quad (3.29)$$

Therefore

$$\tilde{H}(k)\doteq\tilde{H}(0) \quad (3.30)$$

and consequently, up to second order in perturbation theory, the entire low-lying part of the energy spectrum of $H(k)$ is independent of k.

4. GAUGE INVARIANCE—ARBITRARY ORDER PERTURBATION THEORY

All higher order terms of $\tilde{H}(k)$ will contain products of matrix elements of the form

$$(0,j_2|\bar{H}|J^{(m)},j^{(m)})\cdots(J^{(2)},j^{(2)}|\bar{H}|J^{(1)},j^{(1)})$$
$$\times(J^{(1)},j^{(1)}|\bar{H}|0j_1), \quad (4.1)$$

where $\bar{H}$ is either H_1 or H'. Such an expression may be partly characterized by the following kind of graph to be read from the bottom up (Fig. 3). This graph states

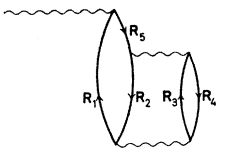

FIG. 3. A typical interaction diagram.

that $J^{(1)}$, $j^{(1)}$ has one additional occupancy of sites $\mathbf{R}_1$ and $\mathbf{R}_3$, and no occupancy of $\mathbf{R}_2$ and $\mathbf{R}_4$ (holes); and that state $J^{(2)}$, $j^{(2)}$ has one additional occupancy of $\mathbf{R}_1$ and a hole at $\mathbf{R}_5$. The initial and final states each have exactly one electron on each site. The phase factor associated with Fig. 3 is

$$f=e^{-ik\xi(X_1+X_3-X_2-X_4)}e^{-ik\xi(X_2+X_4-X_3-X_5)}$$
$$\times e^{-ik\xi(X_5-X_1)}. \quad (4.2)$$

As in Sec. 3 we may assume that all X_i are close to the origin, which permits us to drop the ξ's and gives $f=1$. Thus, to any finite order m (i.e., m fixed and independent of L) we see $\tilde{H}(k)$ to be indpendent of k.

A new feature arises in very high order $[O(L/a)]$ perturbation theory. Consider for example the following graph (Fig. 4), where $X_\nu=\nu a$ and $Ma=L$. The phase factor associated with this graph is

$$f=e^{-ik\xi(X_2-X_1)}e^{-i\xi(X_3-X_4)}e^{-i\xi(X_M-X_1)}$$
$$=e^{-ika}\cdots e^{-ika}=e^{-ikL}. \quad (4.3)$$

Such a term *does* introduce a k dependence into $\tilde{H}(k)$. However, since we may assume that the procedure of successive canonical transformations converges for large L at a rate independent of L, terms of this very high order are exponentially negligible.

This concludes the demonstration that to all orders in perturbation theory the low-lying levels are described by an effective Hamiltonian $\tilde{H}$ which is independent of k, apart from terms which are exponentially small in L and hence negligible. Thus, all these levels $E_\alpha(k)$ are in fact independent of k, from which it follows by (2.6) and (2.13) that none of them carry a current or exhibit a free acceleration in an external electric field.[10]

5. NATURE OF THE LOCALIZED MANY-PARTICLE WAVE FUNCTION

When we are dealing with a system which is strictly localized in the usual three-dimensional space, the independence of the spectrum of k is immediately demon-

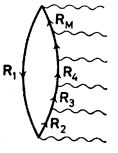

FIG. 4. A pathological high-order diagram.

[10] In particular this means that spin-wave states of an insulator carry no current and exhibit no free acceleration in a dc electric field. This is not an entirely trivial result, since it cannot be derived from any selection rules. In fact, for a ring of finite L, spin waves do carry a finite though small current.

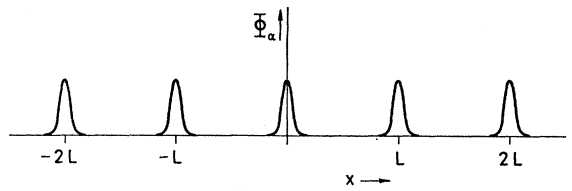

FIG. 5. Wave function of well-localized particles on a ring.

strated: If $\Phi_\alpha(0)$ is eigenfunction of $H(0)$, with eigenvalue E_α, then it is well known that the function

$$\Phi_\alpha(k)\equiv e^{-ik\Sigma x_i}\Phi_\alpha(0) \tag{5.1}$$

is an eigenfunction of $H(k)$ with the same eigenvalue E_α. On a "ring" this argument in general breaks down because, e.g.,

$$\Phi_\alpha(x_1+L, y_1, z_1; \cdots; k) = e^{-ikL}\Phi_\alpha(x_1,y_1,z_1; \cdots; k) \tag{5.2}$$

and therefore violates the ring boundary condition. Nevertheless we say in Sec. 2 that for a particle localized near the origin we could define

$$\Phi_\alpha(x,y,z; k)\doteq e^{-ik\xi(x)}\Phi_\alpha(x,y,z; 0), \tag{5.3}$$

which is single valued and gives rise to an eigenvalue $E_\alpha(k)$ which, apart from terms vanishing exponentially with L, is independent of k. The essential feature in the demonstration of this fact was that in going around the ring the wave function became exponentially small. The discontinuity in the phase factor occurred in the region where the function was exponentially small and thus introduced a negligible error in the energy. A plot of $\Phi_\alpha(x,y,z;0)$ as function of x has the following general appearance (Fig. 5). Note the essential characteristic that it consists of a sequence of practically disconnected parts.[11]

We show that a similar disconnectedness exists also for our many-particle system and is responsible for its insulating properties.

The essential features may be seen from an examination of the zeroth-order wave function. Denote one of the eigenfunctions of $(0,j_2|\tilde{H}|0,j_1)$, e.g., (3.21), by $(0,j_1|\alpha)$. Then the full eigenstate of H is given by

$$\Psi_\alpha=\sum\Psi_{J_1,j_1}(J_1,j_1|e^{iS}|0,j_2)(0,j_2|\alpha) = \Psi_\alpha^{(0)}+\Psi_\alpha^{(1)}+\cdots, \tag{5.4}$$

where

$$\Psi_\alpha^{(0)}=\sum\Psi_{0,j_2}(0,j_2|\alpha) \tag{5.5}$$
$$\Psi_\alpha^{(1)}=\sum\Psi_{J_1,j_1}(J_1j_1|iS_1|0,j_2)(0,j_2|\alpha),$$

etc. The states $\Psi_{0,j}$ all have one electron on each site in the ground-state Wannier function with spin up or down, and $\Psi_\alpha^{(0)}$ is a superposition of such states. Evidently the spatial wave function corresponding to $\Psi_\alpha^{(0)}$ will be large whenever

$$\mathbf{r}_i\approx\mathbf{R}_{\nu(i)}+m_i\mathbf{L}, \tag{5.6}$$

where $\mathbf{R}_{\nu(i)}$ is one of the lattice vectors which we may restrict to the interval

$$-L/2<X_\nu\leqslant L/2 \tag{5.7}$$

with

$$\sum_1^N X_\nu=0; \tag{5.8}$$

the set $\nu(i)$ exhaust all ν; m_i is an integer; and $\mathbf{L}$ is a vector of length L in the x direction.

We may write the periodic Wannier functions in the form

$$w_n(\mathbf{r}-\mathbf{R}_\nu)\doteq\sum_{m=-\infty}^{\infty}\bar{w}_n(r-\mathbf{R}_\nu-m\mathbf{L}), \tag{5.9}$$

where $\bar{w}_n$ is the Wannier function for the infinite interval. Then the wave function $\Phi_\alpha^{(0)}$ can be broken up correspondingly into an infinite sum

$$\Phi_\alpha^{(0)}\doteq\sum_{m_i=-\infty}^{\infty}\Phi_\alpha^{(0)}(m_1,m_2,\cdots m_N), \tag{5.10}$$

where $\Phi_\alpha^{(0)}(m_1,\cdots)$ is obtained from $\Phi_\alpha^{(0)}(0,0,\cdots)$ by shifting the locations of the Wannier functions from $\mathbf{R}_1, \mathbf{R}_2, \cdots$ to $\mathbf{R}_1+m_1\mathbf{L}$, $\mathbf{R}_2+m_2\mathbf{L}$, etc. We now show that each $\Phi_\alpha^{(0)}(m_1,m_2,\cdots m_N)$ is spatially localized in the $3-N$ dimensional space and has negligible overlap with all other $\Phi_\alpha^{(0)}$ $(m_1',m_2',\cdots m_N')$, for which $\sum m_i\neq\sum m_i'$.

The localization is evident. Thus $\Psi_\alpha^{(0)}(0,0,\cdots 0)$ has an electron localized near each lattice point in the volume $A\times L$ where A is the cross sectional area of our ring, so that this function extends only slightly beyond the boundaries of a hypervolume of dimension $(AL)^N$. To estimate the overlap, we consider the integral

$$I\equiv\int[\bar{w}_0(\mathbf{r}_1-\mathbf{R}_1-m_1\mathbf{L})\cdots\bar{w}_0(\mathbf{r}_N-\mathbf{R}_N-m_N\mathbf{L})]^2 \times[\bar{w}_0(\mathbf{r}_1-\mathbf{R}_1'-m_1'\mathbf{L})\cdots\bar{w}_0(\mathbf{r}_N-\mathbf{R}_N'-m_N'\mathbf{L})] \times d\mathbf{r}_1\cdots d\mathbf{r}_N. \tag{5.11}$$

We have chosen an integral with nonnegative integrand, so that accidental cancellations cannot occur.

The $\mathbf{R}_\nu'$ are some permutation of the $\mathbf{R}_\nu$; they arise from the antisymmetrization of the wave function. For $m_1=\cdots m_N'=0$ and $\mathbf{R}_i'=\mathbf{R}_i$ (maximum overlap) I has the value

$$I_0=Q^N, \tag{5.12}$$

where

$$Q\equiv\int\bar{w}_0^4(\mathbf{r})d\mathbf{r}.$$

[11] At this point we make contact with an important recent paper by C. N. Yang, Rev. Mod. Phys. 34, 694 (1962). Yang considers the behavior of *density matrices* in going around the ring and notes that for normal (i.e., nonsuperfluid) systems, they are similarly localized. From this point of view there is no basic distinction between a normal metal and an insulator. In the present work, where we consider the behavior of the *wave functions* in going around the ring, this distinction becomes apparent.

For the general case we use an upper bound of the form

$$\int \bar{w}_0^2(\mathbf{r})\bar{w}_0^2(\mathbf{r}-\mathbf{S})d\mathbf{r}\leqslant Qe^{-|S_1|/b}, \qquad (5.12')$$

where $\mathbf{S}$ is an arbitrary lattice vector and b is a characteristic length related to the range of the Wannier function. Then we may write

$$\begin{aligned} I &\leqslant I_0 \exp[-\sum_\nu |(X_\nu'-X_\nu)+(m_\nu'-m_\nu)L|/b] \\ &\leqslant I_0 \exp[-|\sum_\nu (X_\nu'-X_\nu)+(m_\nu'-m_\nu)L|/b] \\ &= I_0 \exp[-|\sum_\nu (m_\nu'-m_\nu)|L/b] \\ &= I_0 \exp[-|M'-M|L/b], \qquad (5.13) \end{aligned}$$

where

$$M\equiv\sum m_\nu. \qquad (5.14)$$

In quite the same way we can derive an upper bound for the unintegrated product of two wave functions,

$$\begin{aligned} &|\bar{w}_0(\mathbf{r}_1-\mathbf{R}_1-m_1\mathbf{L})\cdots\bar{w}_0(\mathbf{r}_N-\mathbf{R}_N-m_N\mathbf{L})| \\ &\times|\bar{w}_0(\mathbf{r}_1'-\mathbf{R}_1'-m_1'\mathbf{L})\cdots\bar{w}_0(\mathbf{r}_N-\mathbf{R}_N'-m_N'\mathbf{L})| \\ &\leqslant Ce^{-|M'-M|L/c}, \qquad (5.15) \end{aligned}$$

where C is a constant of order a^{-3N} and c is a length similar to b.

Thus we see that each zeroth-order wave function of the system breaks up into an infinite sequence of disconnected equivalent pieces, each of which is characterized by a common value of M:

$$\Phi_\alpha^{(0)}=\sum_{M=-\infty}^{\infty}\Phi_{\alpha,M}^{(0)}. \qquad (5.16)$$

Consider now an arbitrary one-particle position operator $\sum_i G(\mathbf{r}_i)$. It is clear from (5.15), and may be simply verified, that

$$\int_{-\infty}^{\infty}\Phi_{\alpha,M}^{(0)*}\sum G(\mathbf{r}_i)\Phi_{\alpha',M'}^{(0)}d\mathbf{r}_1\cdots d\mathbf{r}_N\doteq 0,$$
$$M\neq M',\ \alpha,\alpha' \text{ arbitrary.} \qquad (5.17)$$

Furthermore, since the perturbation corrections $\Psi_{\alpha,M}^{(1)}$, etc., differ from the unperturbed function by having a small number (of order 1) of Wannier functions shifted through distances of the order of atomic dimensions and possibly excited, it is clear that if the perturbation series converges, then we may write for the full wave functions

$$\Phi_\alpha=\sum_{M=-\infty}^{\infty}\Phi_{\alpha,M}, \qquad (5.18)$$

where

$$\int_{-\infty}^{\infty}\Phi_{\alpha,M}^{*}\sum G(\mathbf{r}_i)\Phi_{\alpha',M'}d\mathbf{r}_1\cdots d\mathbf{r}_N\doteq 0,$$
$$M\neq M',\ \alpha,\alpha' \text{ arbitrary.} \qquad (5.19)$$

This result implies that the entire configuration space $-\infty<x_i,y_i,z_i<+\infty$ may be divided into a sequence of geometrically similar nonoverlapping regions, $\mathfrak{R}_M$, and their outside $\mathfrak{R}'$ such that for every α the partial wave function $\Phi_{\alpha,M}$ is confined to $\mathfrak{R}_M$.[12] We shall see that from this property all the characteristics of an electrical insulator follow.

6. ELECTRICAL CONDUCTIVITY OF LOCALIZED ELECTRON SYSTEM

It is clear from the above discussion that if $\Phi_\alpha(0)$ is a perturbation eigenfunction of $H(0)$, then

$$\Phi_\alpha(k)\equiv\sum_{M=-\infty}^{\infty}\Phi_{\alpha,M}\exp[-ik(\sum x_i-ML)] \qquad (6.1)$$

is an equally good perturbation expansion of $\Phi_\alpha(k)$ which, because of the nonoverlap, has precisely the same energy

$$[\Phi_\alpha(k),H(k)\Phi_\alpha(k)]\doteq[\Phi_\alpha(0),H(0)\Phi_\alpha(0)], \qquad (6.2)$$

or

$$E_\alpha(k)\doteq E_\alpha(0). \qquad (6.3)$$

It can be directly verified that the procedure of Secs. 3 and 4 leads exactly to the function (6.1).

From (5.18) follow at once the insulating properties

$$j_\alpha\doteq 0, \quad \lim_{\omega\to 0}\omega\sigma_\alpha''(\omega)=0. \qquad (6.4)$$

These properties are, however, not yet sufficient to assure that the system is an insulator; they are also shared by metallic alloys. We need still to show that

$$\lim_{\omega\to 0}\sigma_\alpha'(\omega)=0, \qquad (6.5)$$

where σ_α' is the real (absorptive) part of the conductivity, given by

$$\begin{aligned} \sigma_\alpha'(\omega)=&\frac{1}{\Omega}\frac{e^2\pi}{\omega}\sum_{\alpha'}|(\alpha'|P|\alpha)|^2 \\ &\times[\delta(E_{\alpha'}-E_\alpha-\omega)-\delta(E_{\alpha'}-E_\alpha+\omega)]. \qquad (6.6) \end{aligned}$$

Here

$$(\alpha'|P|\alpha)\equiv\int_{\Omega^N}\Phi_{\alpha'}{}^{*}P\Phi_\alpha d\mathbf{r}_1\cdots d\mathbf{r}_N \qquad (6.7)$$

$$=\int_{\Omega^N}\Big(\sum_{M'=-\infty}^{\infty}\Phi_{\alpha',M'}{}^{*}P\sum_{M=-\infty}^{\infty}\Phi_{\alpha,M}\Big)\times d\mathbf{r}_1\cdots d\mathbf{r}_N, \qquad (6.8)$$

where the integration goes over the standard region of volume Ω^N, bounded in the x directions by $-L/2\leqslant x_i<L/2$.

[12] If we start at a point $(\mathbf{r}_1,\cdots\mathbf{r}_N)$ in $\mathfrak{R}_M$ and move one of the x coordinates around the ring, we come to a corresponding point in $\mathfrak{R}_{M+1}$. We may remark however that for our purposes it is not sufficient that the wave function vanishes when a single x_1 is taken around the ring, but it must vanish on every path leading from $\mathfrak{R}_M$ to $\mathfrak{R}_{M'}$, $M'\neq M$.

Now define a periodic operator $Q(\mathbf{r}_1,\cdots\mathbf{r}_N)$ as follows: As before we denote by $\mathfrak{R}_M$ a set of similar non-overlapping regions in the unbounded $3N$-dimensional space of the $\mathbf{r}_i$ such that $\Phi_{\alpha,M}$ is near-vanishing outside $\mathfrak{R}_M$ for all α; and we call $\mathfrak{R}'$ the region outside of all $\mathfrak{R}_M$. Then Q is given by

$$\begin{aligned} Q(\mathbf{r}_1,\cdots\mathbf{r}_N)&=x_1+x_2+\cdots x_N-ML \text{ in } \mathfrak{R}_M \\ &=F(\mathbf{r}_1,\cdots\mathbf{r}_N) \qquad \text{in } \mathfrak{R}', \end{aligned} \tag{6.9}$$

where F is a largely arbitrary, periodic function but chosen such that Q is everywhere twice differentiable. Q may be Fourier expanded as

$$Q=\sum_{\mathbf{q}} Q(\mathbf{q})\exp[i\mathbf{q}\cdot(\mathbf{r}_1+\mathbf{r}_2+\cdots\mathbf{r}_N)] \tag{6.10}$$

and is a well-behaved operator in our Hilbert space of periodic functions. It satisfies the following commutation relations:

$$[Q,v(\mathbf{r}_i)]=[Q,u(\mathbf{r}_i-\mathbf{r}_j)]=0 \tag{6.11}$$

and

$$[\tfrac{1}{2}\textstyle\sum \mathbf{p}_i^2,Q]=\dfrac{1}{-i}\displaystyle\sum_i p_{ix}=\dfrac{1}{i}P \text{ in all } \mathfrak{R}_M, \tag{6.12}$$

which may be combined to yield

$$[H,Q]=\frac{1}{i}P \text{ in all } \mathfrak{R}_M. \tag{6.13}$$

In view of the fact that $\Psi_{\alpha'}$ is near vanishing in $\mathfrak{R}'$, (6.13) may be substituted in (6.8) and yields

$$(\alpha'|P|\alpha)\doteq i(E_{\alpha'}-E_\alpha)(\alpha'|Q|\alpha). \tag{6.14}$$

Now the order of magnitude of $(\alpha'|Q|\alpha)$ may be estimated from the low-order perturbation expression (5.4) and is found to be a length independent of the dimension of the system, say

$$(\alpha'|Q|\alpha)=O(d). \tag{6.15}$$

When we substitute (5.28) into (5.21) we obtain

$$\begin{aligned} \sigma'(\omega)=\left(\frac{e^2\pi}{\Omega}\right)\omega\sum_{\alpha'}|(\alpha'|Q|\alpha)|^2 \\ \times[\delta(E_{\alpha'}-E_\alpha-\omega)-\delta(E_{\alpha'}-E_\alpha+\omega)]. \end{aligned} \tag{6.16}$$

In view of (6.15) and the difference of the two δ functions appearing in (6.16) we see that

$$\omega\to 0:\sigma_\alpha'(\omega)\sim\omega^n, \quad n\geqslant 2. \tag{6.17}$$

Thus the dc conductivity does indeed vanish.

7. CONCLUSIONS

We have seen that the essential property of the array of hydrogen atoms which assured its insulating characteristics was the fact that each of its low-lying wave functions Ψ_α consisted of a sum of disconnected functions $\Psi_{\alpha,M}$. It is a simple matter to verify that also the ground states and exciton states of a conventional insulator with an energy gap (like an array of He atoms) have this property. Furthermore preliminary work indicates that the same criterion applies also to disordered insulators. Thus, it would seem that this disconnectedness of the wave function, rather than more special conditions like an energy gap, is the essential characteristic of insulators.

APPENDIX 1. LOCALIZATION OF THE CENTER OF MASS

Consider a system of N one-dimensional particles, of coordinates $x_1,\cdots x_N$ on a periodic interval of length L. There are N potential wells and, in the sense of Secs. 3 and 4, there is in zeroth-order one particle in each well. We wish to explore the distribution of an appropriately defined center of mass.

For this purpose we define the operator

$$P(\bar{X})\equiv\delta_P(x_1+x_2+\cdots x_N-\bar{X}), \tag{A1.1}$$

where δ_P is the periodic δ function

$$\delta_P(x)\equiv(1/L)\sum_q e^{iqx}, \quad q=(2\pi/L)(0,\pm1,\cdots). \tag{A1.2}$$

Clearly the quantity $\langle P(\bar{X})\rangle d\bar{X}$, where $\langle\ \rangle$ denotes expectation value, is the probability that the sum $x_1+\cdots x_N$ has a value in the range $\bar{X}$, $\bar{X}+d\bar{X}$, modulo L.

We begin by evaluating $\langle P(\bar{X})\rangle$ in the zeroth-order wave function

$$w_0(x_1-X_1)w_0(x_2-X_2)\cdots w_0(x_N-X_N). \tag{A1.3}$$

This gives

$$P(\bar{X})=\frac{1}{L}\sum_q\exp[iq(X_1+\cdots X_N-\bar{X})]F^N(q), \tag{A1.4}$$

where

$$F(q)\equiv\int_{-L/2}^{L/2}w_0^2(x)e^{iqx}dx. \tag{A1.5}$$

For small q we may write

$$F(q)=1-b^2q^2+\cdots, \tag{A1.6}$$

where b is a length of the order of the range of the Wannier function. Hence we can write

$$\log F^N(q)=N\log F(q)=Nb^2q^2 \tag{A1.7}$$

and

$$F^N(q)=e^{-(N^{1/2}bq)^2}. \tag{A1.8}$$

This falls off so rapidly with q that the small q approximation is *a posteriori* justified. Substituting into (A1.4), and recalling the convention $\sum X_\nu=0$ gives

$$\begin{aligned} P(\bar{X})=\frac{1}{2\pi^{1/2}N^{1/2}b} \\ \times\sum_m\exp\{-[(\bar{X}-mL)/2N^{1/2}b]^2\}. \end{aligned} \tag{A1.9}$$

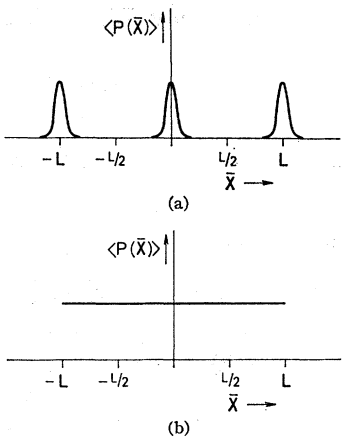

FIG. 6. (a) Localization of the center of mass in a one-dimensional insulator. (b) Nonlocalization of the center of a mass in a one-dimensional free electron gas.

Thus, we see that $\bar{X}$, which may be regarded as N times the center of mass, is localized—on the interval—$L/2 \leqslant \bar{X} < L/2$—in a region of width $(ba^{-1/2})L^{1/2}$ much narrower than L (see Fig. 6). It is clear that if exchange terms are included in the evaluation of $\langle P(\bar{X})\rangle$ or if the expectation value of $P(\bar{X})$ is evaluated in the true eigenfunctions, $\Phi_\alpha = e^{iS}\Phi_\alpha^{(0)}$, the general nature of the result is not altered. Thus, while the individual electron coordinates are of course spread all over the interval $-L/2 \leqslant x < L/2$, the center of mass, $\bar{X}/N$, is localized in an interval of width $\bar{a}/N^{1/2}$, where $\bar{a}$ is of the order of atomic dimensions.

This state of affairs may be contrasted with that of a free electron gas. Here

$$P(\bar{X}) = (1/L)\sum_q e^{-iq\bar{X}}\langle e^{iq(X_1+\cdots X_N)}\rangle \quad \text{(A1.10)}$$
$$= (1/L),$$

i.e., $\bar{X}$ is entirely unlocalized [see Fig. 6(b)].

When we go from one to three dimensions we encounter a somewhat strange situation. If we call the cross-sectional area of our ring A, we find, in exact analogy with the one dimensional case that the width of $P(\bar{X})$ is given by

$$w \sim (AL)^{1/2}a^{-3/2}b, \quad \text{(A1.11)}$$

which is much smaller than L, only for an extremely thin ring, for which

$$A^{1/2}/L \ll a^{3/2}b^{-1}L^{-1/2}. \quad \text{(A1.12)}$$

If all dimensions of the ring are of comparable magnitude ($A^{1/2} \sim L$), then $w \gg L$ and the center of mass becomes delocalized. Referring to the regions $\mathfrak{R}_M$, defined in Sec. 5, we may say that for the one and "thin" three-dimensional ring, $\bar{X}$ is restricted in each $\mathfrak{R}_M$ to a narrow range of values, which do not overlap, while for the thick ring the values of $\bar{X}$ in different $\mathfrak{R}_M$ overlap considerably.

APPENDIX 2. ONE-PARTICLE DENSITY MATRIX

The localization of the wave function is also reflected in the one-particle density matrix,

$$\rho_{mm'}(\mathbf{r},\mathbf{r}') \equiv (\Psi_\alpha, \psi_{m'}{}^*(\mathbf{r}')\psi_m(\mathbf{r})\Psi_\alpha). \quad \text{(A2.1)}$$

To zeroth order, we find, e.g., in the ferromagnetic state,

$$\rho_{mm'}(\mathbf{r},\mathbf{r}') = \delta_{m-1/2}\delta_{m'-1/2} \times \sum_\nu w_0(\mathbf{r}'-\mathbf{R}_\nu)w_0(\mathbf{r}-\mathbf{R}_\nu). \quad \text{(A2.2)}$$

Because of the exponential decay of the Wannier functions, we can write, for $|x'-x| \ll L$,

$$\rho_{mm'}(\mathbf{r},\mathbf{r}') < Ce^{-|\mathbf{r}-\mathbf{r}'|/c}, \quad \text{(A2.3)}$$

where C is some constant and c is of the order of an atomic dimension. A result of this form persists to all orders of perturbation theory.

It is tempting to suppose that this exponential behavior of ρ is characteristic only of insulators. This is not so. The same behavior is found for metallic alloys,[13] and even for free particles ρ vanishes for $|r-r'| \to \infty$, although not exponentially.

APPENDIX 3. MODEL EXHIBITING TRANSITION BETWEEN INSULATING AND CONDUCTING STATES

The main body of this paper has been concerned with a characterization of the insulating state. In particular it was shown that under certain conditions (e.g., large lattice parameter) a system with one electron per atom, which in band theory would be a monovalent metal, will be an insulator. Presumably as a is decreased, the system eventually becomes a metal but the nature of this transition is not at present known. This Appendix deals first with a soluble model exhibiting a transition of this general nature.

We consider a cubic lattice of N fixed nuclei and N interacting electrons and will be concerned with diagonalizing the Hamiltonian H within the Hilbert space spanned by the Bloch wave functions, $\varphi_\mathbf{k}$, of the lowest band. We note that the total z spin, S_z, and total wave vector $\mathbf{q}$, which characterizes the behavior under over-all translation by a lattice vector, are good quantum numbers.

When S_z has its maximum value $N/2$, our Hilbert space contains only one state vector, $\bar{\Psi}$, with all electron spins aligned and $\mathbf{q}=0$. The diagonalization of H is then trivial:

$$E_{N/2,0} = \sum_\mathbf{k} \epsilon_\mathbf{k} + \tfrac{1}{2}\sum_{\mathbf{k}_1,\mathbf{k}_2} [(\mathbf{k}_1\mathbf{k}_2|w|\mathbf{k}_1\mathbf{k}_2) - (\mathbf{k}_1\mathbf{k}_2|u|\mathbf{k}_2\mathbf{k}_1)], \quad \text{(A3.1)}$$

where all $\mathbf{k}$ run over the fundamental Brillouin zone.

For $S_z = N/2-1$, and given $\mathbf{q}$ our Hilbert space contains N state vectors, $\Psi_\mathbf{k}$, which differ from $\bar{\Psi}$ by having a spin-up electron missing in $\varphi_\mathbf{k}$, and having a spin-down electron present in state $\varphi_{\mathbf{k}+\mathbf{q}}$:

$$\Psi_\mathbf{k} \equiv a_{\mathbf{k}+\mathbf{q}-}{}^* a_{\mathbf{k}+}\bar{\Psi}. \quad \text{(A3.2)}$$

The eigenvectors of H must then be of the form

$$\Psi = \sum_\mathbf{k} A(\mathbf{k})\Psi_\mathbf{k}, \quad \text{(A3.3)}$$

[13] D. C. Mattis and J. Bardeen, Phys. Rev. 111, 412 (1958).

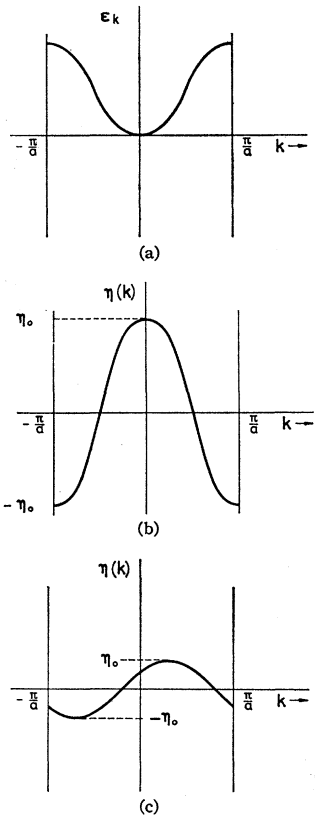

FIG. 7. (a) The function ϵ_k (schematic). (b) The function $\eta(k)$ for $q=(\pi/a)$ (schematic). (c) The function $\eta(k)$ for $q>0$ but $q\ll(\pi/a)$ (schematic).

and the diagonalization of the Hamiltonian leads to the equation

$$[\eta(\mathbf{k})-E]A(\mathbf{k})+\sum_{\mathbf{k}'} u(\mathbf{k},\mathbf{k}')A(\mathbf{k}')=0, \quad \text{(A3.4)}$$

where

$$\eta(\mathbf{k})\equiv\epsilon_{\mathbf{k}+\mathbf{q}}-\epsilon_{\mathbf{k}}, \quad \text{(A3.5)}$$

and

$$u(\mathbf{k},\mathbf{k}')\equiv(\Psi_{\mathbf{k}},U\Psi_{\mathbf{k}'}). \quad \text{(A3.6)}$$

It should be noted that (A3.4) has a structure analogous to a one-particle Schroedinger equation. We now study its solution in two especially simple cases.

Case I. Contact Potential

We first consider the case in which $u(\mathbf{k},\mathbf{k}')$ is assumed to have the form

$$u(\mathbf{k},\mathbf{k}')=\lambda[\gamma N\delta_{\mathbf{k}\mathbf{k}'}+(1/N)], \quad \text{(A3.7)}$$

where γ is a constant and λ is a strength parameter. Then defining

$$E'\equiv E-\lambda\gamma N, \quad \text{(A3.8)}$$

Eq. (A3.4) becomes

$$[\eta(\mathbf{k})-E']A(\mathbf{k})+(\lambda/N)\sum_{\mathbf{k}'} A(\mathbf{k}')=0. \quad \text{(A3.9)}$$

This type of equation is well known from the theory of superconductivity. We first introduce

$$\sigma\equiv\frac{1}{N}\sum_{\mathbf{k}'} A(\mathbf{k}') \quad \text{(A3.10)}$$

and then solve (A3.9) in terms of σ:

$$A(\mathbf{k})=-\lambda\sigma/[\eta(\mathbf{k})-E']. \quad \text{(A3.11)}$$

Summation on k results in the relation

$$1+\lambda(1/N)\sum\{1/[\eta(\mathbf{k})-E']\}=0 \quad \text{(A3.12)}$$

which is an implicit equation for E'.

Now $\eta(\mathbf{k})$, Eq. (A3.5), is a periodic function of $\mathbf{k}$, shown schematically in Figs. 7(b) and 7(c), whose minimum and maximum values we denote by $\mp\eta_0$. We now define

$$F(E')\equiv(1/N)\sum\{1/[\eta(\mathbf{k})-E']\} \quad \text{(A3.13)}$$

and study the solutions of (A3.12) by graphically equating

$$F(E')=-1/\lambda, \quad \text{(A3.14)}$$

(see Fig. 8). For $|\eta|\geqslant\eta_0$ it is useful to introduce the limiting function

$$\bar{F}(E)=\lim_{N\to\infty}F(E)=\frac{1}{(2\pi)^3}\left(\frac{\Omega}{N}\right)\int\frac{1}{\eta(\mathbf{k})-E}d\mathbf{k}. \quad \text{(A3.15)}$$

Notice in particular that because of the volume element in $d\mathbf{k}$, $F(\pm\eta_0)$ are *finite*, and call

$$\bar{F}(-\eta_0)\equiv F_0. \quad \text{(A3.16)}$$

Then clearly for attractive interaction, $\lambda<0$, we have the following possible situations:

(a) if $|\lambda|>1/F_0$, there is an isolated ground state plus a continuum;

(b) if $|\lambda|<1/F_0$, there is only a continuum.

In situation (a) the ground state is a spin-wave state, in which the spin-up hole and spin-down electron are bound together. It is straightforward to verify that the corresponding wave function has the disconnectedness property typical of insulating states. The continuum states however describe free electron hole pairs which conduct. In situation (b) we have only conducting states. Thus, as $|\lambda|$ decreases from values $>1/F_0$ to 0, we get a sharp transition from an insulating to a con-

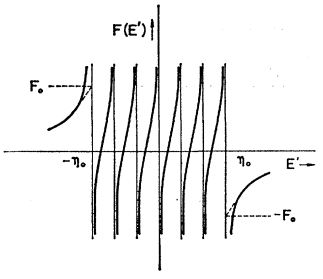

FIG. 8. The function $F(E')$, Eq. (A3.13).

ducting ground state. This is our soluble model exhibiting a sharp transition.

Another case, of great physical interest, which does not exhibit a transition is Case II.

Case II. Coulomb Potential

$$u(\mathbf{k},\mathbf{k}')=\lambda[\gamma N\delta_{\mathbf{k}\mathbf{k}'}+(1/N)(1-\delta_{\mathbf{k}\mathbf{k}'})f(\mathbf{k},\mathbf{k}')] \quad \text{(A3.17)}$$

where for $\mathbf{k}\approx\mathbf{k}'$,

$$f(\mathbf{k},\mathbf{k}')\approx(1/a^2|\mathbf{k}-\mathbf{k}'|^2). \quad \text{(A3.18)}$$

Just as in the ordinary two-particle problem with Coulomb interaction one obtains an isolated bound ground state, no matter how weak the interaction is, so also in the present case we obtain an insulating spin-wave ground state, describing a bound electron hole pair, for all negative values of λ.

Speculations about the Transition between Insulating and Metallic States

We conclude this Appendix with speculations about the nature of the ground state as S_z decreases from its maximum value $N/2$ to 0.

The ground state with $S_z=N/2$ is insulating. With $S_z=N/2-1$ the ground state is a spin-wave state with that wave number $\mathbf{q}$ which results in the lowest energy: For strong Coulomb interaction, $\mathbf{q}=0$, maximizing the effect of the interaction (ferromagnet); for weak Coulomb interaction $\mathbf{q}=(\pi/a,0,0)$ which gives the lowest band energy [see Fig. 7(b)]; in intermediate cases the lowest state may have some other value of $\mathbf{q}$ (spiral magnetization).

When $S_z=N/2-n$, with $n\ll N$, the ground state will have n of the lowest energy spin waves and still be an insulator.

However, when n becomes comparable to N, interactions between spin waves must be taken into account. For sufficiently large n, the spin-wave excitons may overlap sufficiently so as to substantially screen the Coulomb attraction between electrons and holes and thus finally lead to a dissolution of the bound pairs. This would then result in a conducting ground state.[14] On the other hand, the overlap may not sufficiently weaken the attraction, so that an insulating ground state can result even for $S_z=0$. This is the case which has been the main subject of this paper.

[14] Compare an analogous discussion in Sec. 2 of N. F. Mott, Phil. Mag. **6**, 287 (1961).