Localized versus Extended Holes in a Two-Band Model for Semiconductor-Metal Transitions*

T. A. Kaplan
Department of Physics, Michigan State University, East Lansing, Michigan 48823
(Received 25 March 1971)

We consider essentially the two-band Hubbard-type model of a nonmagnetic semiconductor studied recently by Falicov and Kimball (FK), but we allow the bandwidth Δ_v of the valence band to be nonzero and the Coulomb repulsion U_{11} between holes to be noninfinite. FK treated this model in an approximation based to some extent on the free-energy variational principle, and they used one-electron wave functions which are localized (Wannier) functions for the valence band, and extended (Bloch) functions for the conduction band. This problem is formulated within the framework of the recently introduced thermal single-determinant approximation. The limit $U_{11}\to\infty$, $\Delta_v\to 0$ then provides a strictly variational derivation of the FK results. We then show that for $\Delta_v\neq 0$, $U_{11}<\infty$, and temperature sufficiently small ($T<T_0$), a *lower* free energy is obtained when valence-band *Bloch* functions are substituted for the Wannier functions. That T_0 can be appreciable even when $U_{11}/\Delta_v\gg 1$ and the possibility of a transition to the localized picture at T_0 are pointed out. It is shown that magnetic inelastic neutron scattering (arising from the magnetic dipolar interaction between neutrons and electrons) can distinguish, at least in principle, between the extended and the localized pictures. The fact that band gaps occur in the thermal-neutron range for interesting materials is noted.

I. INTRODUCTION

Recently Falicov and co-workers[1,2] considered semiconductor-to-metal transitions in a system with a narrow valence band and a broad conduction band which, respectively, are filled and empty at low temperatures. Coulomb interactions U_{11} were considered to be very large between holes in the valence band—so much larger than the width Δ_v of the valence band that it was argued that the holes should be considered as spatially localized, the limit $U_{11}\to\infty$ being taken. Their further development of the properties of this system was based in part on the minimum-free-energy principle. In the present paper we shall show that if Δ_v is nonzero and U_{11} is finite, then a *lower* free energy is obtained when the localized hole states in the treatment of Falicov and Kimball (FK) are replaced by spatially extended states, at sufficiently low temperatures. Hence, by the motivation of minimizing the free energy, used by FK, *this extended picture is preferable* to that of FK in this temperature range. Furthermore, it is found that the latter range can be appreciable even when $U_{11}/\Delta_v\gg 1$; the possibility of a transition to the localized picture at higher temperature is also discussed.

The technique we use, which provides incidentally a strictly variational derivation of the FK results, is the recently introduced thermal single-determinant approximation.[3,4] The greater generality of this technique over the usual Hartree-Fock (HF) or mean-field approximation is essential here since in the latter the localized picture would never give a lower free energy than the extended state, as can be shown from simple symmetry considerations. The scattering cross section for inelastic magnetic neutron scattering is calculated and discussed as a means, at least in principle, of distinguishing between the localized-hole and extended-hole pictures.

The Hamiltonian we consider is

$$H=\sum_{\nu\vec{k}\sigma}\epsilon_{\nu\vec{k}}\,a^{\dagger}_{\nu\vec{k}\sigma}\,a_{\nu\vec{k}\sigma}+\sum_{i\,\nu}U_{\nu\nu}N_{\nu i\uparrow}N_{\nu i\downarrow}+U_{12}\sum_{i}N_{1i}N_{2i}\ . \tag{1}$$

Here ν runs over 1 and 2, $\nu=1$ labels the valence or hole band, and $\nu=2$ refers to the conduction band $a^{\dagger}_{\nu\vec{k}\sigma}$ creating an electron in a Bloch function of band ν, wave vector $\vec{k}$, and spin σ; $N_{\nu i\sigma}=b^{\dagger}_{\nu i\sigma}b_{\nu i\sigma}$ where $b^{\dagger}_{\nu i\sigma}$ creates an electron in a Wannier function for band ν at crystal site i, and $N_{\nu i}=N_{\nu i\uparrow}+N_{\nu i\downarrow}$. We consider the case of two electrons per site,

$$\langle N_{1i}\rangle+\langle N_{2i}\rangle=2\ , \tag{2}$$

the brackets indicating a thermal average, and choose, with FK, $U_{22}=0$. N_a will denote the number of sites i.

Equation (1) is identified, within a constant, to FK's Hamiltonian [their Eq. (1) plus the contributions from their (3) and (6)] provided we put $\epsilon_{2\vec{k}}\equiv\epsilon_2(\vec{k})-2U_{12}$, $\epsilon_{1\vec{k}}\equiv -E-U_{11}$, , consider only one conduction band ($\nu=2$), and make the identifications $U_{11}\equiv G_1$, $U_{12}\equiv -G_4(=G)$. At this stage FK take the limit $U_{11}\to\infty$, projecting out all states with two valence-band holes on the same site. They then treat

the remaining statistical problem in a self-consistent-field approximation, where the interband term ($\propto U_{12}$) is simply replaced by an average ("mean-field") value. They then calculate a "free energy" $\mathfrak{F}$ which they proceed to minimize to determine physical properties. But it was not shown that $\mathfrak{F}$ is an upper bound to the exact free energy corresponding to the model Hamiltonian being considered, so that FK's procedure of minimizing $\mathfrak{F}$ has no apparent justification. Furthermore, in the case we wish to consider (finite U_{11}, nonzero bandwidth of the valence band), the same "mean-field" approximation consistently applied to the U_{11} term in addition to the U_{12} term would lead to the standard thermal Hartree-Fock approximation,[5] no localization then occurring. Hence we shall follow another course, discussed in Sec. II.

II. GENERAL THEORY: REDUCTION TO FK RESULT

We will use explicitly the well-known variational principle[5,6]

$$A(\rho) \equiv \mathrm{tr}\rho H + kT\,\mathrm{tr}\rho\ln\rho \geq A_{\mathrm{exact}}(\rho)\ , \tag{3}$$

where A_{exact} is the exact Helmholtz free energy

$$A_{\mathrm{exact}} = \mu\bar{N} - kT\ln\mathrm{tr}e^{-\beta(H-\mu N)}\ , \tag{4}$$

ρ is any density matrix, the trace is taken over a complete set of eigenstates of H (so N varies from 0 to $4N_a$), $\bar{N}$ is the average N, and μ is the chemical potential.

A class of ρ's that conforms to FK's physical picture of localized valence electrons and spatially extended conduction electrons is defined by

$$\tilde{H}_{1\mathrm{oc}} = U_{11}\sum_i N_{1i}N_{1i} + \sum_i w_1 N_{1i} + \sum_{\vec{k}\sigma} w_{2\vec{k}} n_{2\vec{k}\sigma}\ , \tag{5}$$

with the density matrix defined thereby as

$$\rho_{1\mathrm{oc}} = e^{-\beta(\tilde{H}_{1\mathrm{oc}}-\mu N)}/\mathrm{tr}e^{-\beta(\tilde{H}_{1\mathrm{oc}}-\mu N)} \tag{6}$$

($n_{2\vec{k}\sigma} = a^\dagger_{2\vec{k}\sigma}a_{2\vec{k}\sigma}$). In (5) w_1 and $w_{2\vec{k}}$ are real "one-electron energy" parameters, to be varied to minimize $A(\rho_{1\mathrm{oc}})$ subject to

$$\bar{N} = \mathrm{tr}\rho_{1\mathrm{oc}}N = 2N_a\ . \tag{7}$$

The first two terms of (5) lead to the localized description of band 1, the last term giving the extended description of band 2. This is true because a complete set of eigenstates of (5) (the approximate energy eigenstates in this picture) is the set of single Slater determinants in which valence-band Wannier functions and conduction-band Bloch functions are occupied in all possible ways. The choice (5) is a special case of the thermal single-determinant approximation[3,4]; there, $\tilde{H}$ is allowed to be a general function of occupation numbers N_ν corresponding to some complete orthonormal set of one-electron states ψ_ν, the general variational equations determining $\tilde{H}$ and the ψ_ν having been obtained.[3,4] But for our present purposes it is sufficient to consider the choice (5) and, below, one other choice leading to an extended description of both bands.

Using (5) and (6), (3) gives (with $\langle N_{1i}N_{2i}\rangle \equiv \mathrm{tr}\rho_{1\mathrm{oc}}N_{1i}N_{2i}$)

$$A(\rho_{1\mathrm{oc}}) = \sum_{\vec{k}\sigma}(\epsilon_{2\vec{k}} - w_{2\vec{k}} + \mu)\bar{n}_{2\vec{k}\sigma} + \sum_{i\sigma}(b_{ii} - w_1 + \mu)\bar{N}_{1i\sigma} + U_{12}\sum_i\langle N_{1i}N_{2i}\rangle - kT\ln Z\ , \tag{8}$$

where

$$b_{ij} \equiv (1/N_a)\textstyle\sum_{\vec{k}} e^{i\vec{k}\cdot\vec{R}_{ij}}\epsilon_{1\vec{k}}\ , \tag{9}$$

$\vec{R}_{ij}$ connects sites i and j,

$$\bar{Q} \equiv \mathrm{tr}\rho_{1\mathrm{oc}}Q\ , \tag{10}$$

and Z is the denominator in (6). With the help of (5), we see that

$$\langle N_{1i}N_{2i}\rangle = \bar{N}_{1i}\bar{N}_{2i}\ , \tag{11}$$

$$\bar{n}_{2\vec{k}\sigma} = (e^{\beta(w_{2\vec{k}}-\mu)}+1)^{-1} \equiv f(w_{2\vec{k}})\ , \tag{12}$$

$$\bar{N}_{1i\sigma} = (x + x^2e^{-\beta U_{11}})/(1 + 2x + x^2e^{-\beta U_{11}})\ , \tag{13}$$

where

$$x = e^{-\beta(w_1-\mu)}\ . \tag{14}$$

We also have

$$Z = z^{N_a}\prod_{\vec{k}\sigma} z_{\vec{k}}\ , \tag{15}$$

where

$$z = 1 + 2x + x^2e^{-\beta U_{11}}\ , \tag{16}$$

$$z_{\vec{k}} = 1 + e^{-\beta(w_{2\vec{k}}-\mu)}\ . \tag{17}$$

Using (2) plus the fact that $\bar{N}_{1i\sigma}$ is independent of $i\sigma$ and $\bar{n}_{2\vec{k}\sigma}$ is independent of σ, we have

$$\bar{N}_{1i\sigma} = 1 - \tfrac{1}{2}n\ , \tag{18}$$

where

$$n = (1/N_a)\textstyle\sum_{\vec{k}}\bar{n}_{2\vec{k}\sigma}\ , \tag{19}$$

the average number of conduction electrons per site. Thus (13) gives

$$x = \frac{[(1-n)^2 + (2-n)ne^{-\beta U_{11}}]^{1/2} + 1 - n}{ne^{-\beta U_{11}}}\ . \tag{20}$$

We now show that for $n < 1$, (8) reduces to FK's expression for the free energy when, with FK, we put $\epsilon_{2\vec{k}} = \epsilon_2$, let $U_{11} \to \infty$, and choose $w_{2\vec{k}} = w_2$. (For simplicity, we consider only the case presented in the original FK paper.[1]) We readily see that (8) is, then,

$$A(\rho_{1\mathrm{oc}}) = N_a n\{\epsilon_2 - kT\ln[(2-n)/n]\}$$

$$+N_a(2-n)(b_1+kT\ln x)$$
$$+U_{12}N_a(2-n)n-kTN_a\ln(1+2x+x^2e^{-\beta U_{11}})$$
$$+kT(2N_a)\ln(1-\tfrac{1}{2}n)\,,$$

where $b_1=b_{ii}$. From (20), we see that for $n<1$,

$$x\simeq 2n^{-1}(1-n)e^{\beta U_{11}} \quad \text{as } U_{11}\to\infty\,,$$

so that

$$\frac{1}{N_a}A(\rho_{\rm loc})\simeq n\left(\epsilon_2-kT\ln\frac{2-n}{n}\right)+(2-n)\left[b_1+kT\ln\left(2\frac{1-n}{n}\right)+U_1\right]+U_{12}n(2-n)-kT\ln\left(4\frac{1-n}{n^2}\right)-U_{11}+2kT\ln\frac{2-n}{2}\,.$$

As pointed out above, for FK's original Hamiltonian and ours to be the same we need here $\epsilon_2=\epsilon-2U_{12}$ and $b_1=-E-U_{11}$. Hence, dropping a constant (independent of n), we have

$$\frac{1}{N_a}A(\rho_{\rm loc})\simeq n(\epsilon+E)-U_{12}n^2$$
$$+kT[2n\ln n+(2-n)\ln(2-n)+(1-n)\ln(1-n)$$
$$-n\ln 2-2\ln 2]\,.$$

Noting FK's definition $\Delta\equiv\epsilon+E$ and that the proper correspondence is $U_{12}=G$ (as pointed out above), we see that this expression is identical to FK's equation (11) with their $m=1$, $J=\frac{1}{2}$, $J'=0$ (the values appropriate to the case we are considering).[7] The case $n>1$ clearly cannot give minimum free energy in this limit.

Returning to the more general expression (8), we vary the parameters $w_{2\vec{k}}$, w_1 subject to (2); requiring stationarity of A then gives, after a straightforward calculation,

$$w_{2\vec{k}}=\epsilon_{2\vec{k}}+U_{12}\bar{N}_1\,,\quad w_1=b_1+U_{12}n\,, \tag{21}$$

where $\bar{N}_1\equiv\bar{N}_{1i}(=2-n)$ and $n=\bar{N}_{2i}$. Using this result plus (2) and (11)–(19), Eq. (8) gives

$$\mathfrak{a}_{\rm loc}\equiv\frac{1}{N_a}A(\rho_{\rm loc})$$
$$=2\mu-U_{12}n(2-n)-kT\ln(1+2x+x^2e^{-\beta U_{11}})$$
$$+\frac{kT}{N_a}\sum_{\vec{k}\sigma}\ln(1-\bar{n}_{2\vec{k}\sigma})\,. \tag{22}$$

We shall treat the extended-hole case in the well-known thermal HF approximation.[5] In this, one takes the density matrix to be defined, through $\rho\propto e^{-\beta(\tilde{H}-\mu N)}$, by $\tilde{H}=\sum\tilde{\epsilon}_i n_i$, the n_i being the occupation numbers for some complete set of one-electron states. In conformity with the extended picture, we take the latter to be the Bloch functions for the two bands of the model of Eq. (1). The free energy (3), after requiring stationarity under variation of the $\tilde{\epsilon}_i$, in this case can easily be shown to be

$$A(\rho_{\rm ext})=2N_a\mu+kT\sum_{\nu\vec{k}\sigma}\ln(1-\bar{n}_{\nu\vec{k}\sigma})$$
$$-\tfrac{1}{2}\sum_{\nu\vec{k}\sigma}(\tilde{\epsilon}_{\nu\vec{k}\sigma}-\epsilon_{\nu\vec{k}\sigma})\bar{n}_{\nu\vec{k}\sigma}\,, \tag{23}$$

with $\bar{n}_{\nu\vec{k}\sigma}=f(\tilde{\epsilon}_{\nu\vec{k}\sigma})$ the Fermi-Dirac distribution of (12) and

$$\begin{aligned}\tilde{\epsilon}_{1\vec{k}\sigma}&=\epsilon_{1\vec{k}}+U_{11}\bar{N}_{1i\sigma}+U_{12}\bar{N}_{2i}\\&=\epsilon_{1\vec{k}}+U_{11}(1-\tfrac{1}{2}n)+U_{12}n\,,\\\tilde{\epsilon}_{2\vec{k}\sigma}&=\epsilon_{2\vec{k}}+U_{12}(2-n)\,.\end{aligned} \tag{24}$$

Note that $\tilde{\epsilon}_{2\vec{k}\sigma}=w_{2\vec{k}}$ [cf. (21)]. Also $\bar{N}_{\nu i}$ is now the average with respect to the HF density matrix, of course.

III. ASYMPTOTIC BEHAVIOR AT LOW TEMPERATURE

To determine the low-T behavior of $\mathfrak{a}$, we first study n. By definition of the case we are considering, a semiconductor, n is exponentially small at low T. Consider the localized case first. From (19) and (12), it is clear that we then must have, at zero temperature, the condition $\mu<w_{2\min}$, the minimum value of $w_{2\vec{k}}$. Then we have

$$n=O\left[e^{-\beta(w_{2\min}-\mu)}\right];$$

actually one can see that

$$n\simeq AT^{3/2}e^{-\beta(w_{2\min}-\mu)} \tag{25}$$

if $w_{2\vec{k}}$ is parabolic about its minimum value, where

$$A=\tfrac{1}{2}\pi^{1/2}(\Omega/2\pi^2N)(2m_e/\hbar^2)^{3/2}\,,$$

Ω/N is the volume per site, and m_e is the effective mass. Now (14) gives

$$\mu=w_1+kT\ln x \tag{26}$$

and (20) shows that

$$x=[2(1-n)/n]e^{\beta U_{11}}(1+\delta)\,, \tag{27}$$

where $\delta=O(v^n)$ where $v=n$ or $e^{-\beta U_{11}}$, whichever is larger. Thus (26) becomes, using (21),

$$\mu=b_1+U_{12}n+kT\ln[2(1-n)/n]e^{\beta U_{11}}(1+\delta)\,. \tag{28}$$

It follows readily, using (25) and (21), that

$$\mu \underset{T\to 0}{\longrightarrow} \tfrac{1}{2}(b_1+\epsilon_{2\min}+2U_{12}+U_{11}) \equiv \mu_0 , \tag{29}$$

so that the exponential dominating n at low T is

$$w_{2\min}-\mu_0=\tfrac{1}{2}(\epsilon_{2\min}-b_1+2U_{12}-U_{11})\equiv\tfrac{1}{2}G_{\rm loc}. \tag{30}$$

Finally, with the help of (26) and (27), (22) becomes at low T

$$\mathcal{A}_{\rm loc}=2b_1+U_{11}-2kTn+O(T^{-1/2}n^2) . \tag{31}$$

In finding the order of magnitude of the correction in (31) we used the fact that

$$N_a^{-1}\sum_{\vec{k}\sigma}\bar{n}_{2\vec{k}}{}^2\simeq AT^{3/2}e^{-2\beta(w_{2\min}-\mu)}$$

under the same conditions as those used for (25).

We now turn to the extended case. With the help of the well-known fact that at low temperature

$$\mu\simeq\tfrac{1}{2}[(\tilde{\epsilon}_{1\vec{k}})_{\max}+(\tilde{\epsilon}_{2\vec{k}})_{\min}] , \tag{32}$$

we obtain again [cf. Eq. (31)]

$$\mathcal{A}_{\rm ext}\equiv(1/N_a)A\,(\rho_{\rm ext})\simeq 2b_1+U_{11}-2kTn , \tag{33}$$

where now at low T

$$n=O(e^{-\beta[(\tilde{\epsilon}_{2\vec{k}})_{\min}-\mu]})\equiv O(e^{-\beta G_{\rm ext}/2}) . \tag{34}$$

The effective gap at low temperature in this Bloch-wave HF approximation is, from (32) and (24) and (34),

$$G_{\rm ext}=(\epsilon_{2\vec{k}})_{\min}-(\epsilon_{1\vec{k}})_{\max}+2U_{12}-U_{11} . \tag{35}$$

Comparing with (30), we see that

$$G_{\rm loc}-G_{\rm ext}=(\epsilon_{1\vec{k}})_{\max}-b_1>0 , \tag{36}$$

the last inequality holding when the bandwidth Δ_v of band 1 is not zero, since b_1 is the average $N_a^{-1}\times\sum_{\vec{k}}\epsilon_{1\vec{k}}$. Hence $n_{\rm ext}>n_{\rm loc}$ for $T\neq 0$. So we have the required result

$$\mathcal{A}_{\rm ext}<\mathcal{A}_{\rm loc} \tag{37}$$

for $\Delta_v\neq 0$, U_{11} finite, and T sufficiently small.

This result is very reasonable on the general grounds that at low enough T, there are so few holes present that the probability of finding two on one site is negligible even in the extended picture, so that a large U_{11} does not hurt the extended picture seriously.

IV. POSSIBLE TRANSITION FROM EXTENDED TO LOCALIZED STATES

There remains the possibility that the localized picture becomes preferred over this extended picture at $T>$ some temperature T_0 in which case a change in the nature of the excitations with increasing T might be expected. In fact, it seems clear that T_0 not only should exist for large U_{11}/Δ_v but it should approach zero as $U_{11}/\Delta_v\to\infty$, keeping (35) positive.[8] Thus we should worry whether T_0 might be absurdly small whenever FK's assumption $U_{11}/\Delta_v\gg 1$ holds. To show that this is not the case, we shall calculate the free energies for both states, assuming parameter values that are consistent with FK's assumption. For simplicity we will consider low T, i.e., $kT\ll$ either bandwidth and the gap. We will also assume parabolic behavior at the pertinent band extrema.

Considering the extended case first, we put (24) in $\bar{n}_{\nu\vec{k}\sigma}=f(\tilde{\epsilon}_{\nu\vec{k}\sigma})$ and equate the number of electrons to the number of holes, obtaining for the chemical potential

$$\mu_{ext}=\mu^{\circ}_{\rm ext}-\tfrac{1}{4}U_{11}n_{\rm ext}-\tfrac{3}{4}kT\ln(m_e/m_h) . \tag{38}$$

Here, m_e and m_h are the electron and hole masses, respectively, and

$$\mu^{\circ}_{\rm ext}=\tfrac{1}{2}(\epsilon_{1\max}+\epsilon_{2\min}+2U_{12}+U_{11}) .$$

Also, then (25) (which holds for both the extended and the localized cases) becomes

$$n_{\rm ext}\simeq A(m_h/m_e)^{3/4}(kT)^{3/2}\times\exp\{-\tfrac{1}{2}\beta[G_{\rm ext}+\tfrac{1}{2}(U_{11}-4U_{12})n_{\rm ext}]\} . \tag{39}$$

Using (28) for $\mu_{\rm loc}$ in (25) we obtain

$$n_{\rm loc}\simeq[2A(kT)^{3/2}]^{1/2}\exp[-\tfrac{1}{2}\beta(G_{\rm loc}-2U_{12}n_{\rm loc})] , \tag{40}$$

$G_{\rm loc}$ being the zero-temperature gap defined by (30).

To get a very rough idea as to the orders of magnitude that may be involved, we assume the parameter values $U_{11}\simeq 1$ eV $>U_{12}$, $\Delta_v\simeq 0.04$ eV, $G_{\rm ext}\simeq 0.04$ eV, $m_e\simeq 10$ electron masses, $m_h/m_e\simeq 20$, and $\Omega/N\simeq 30$ Å^3. These are similar to values recently obtained phenomenologically[9] for Ti_2O_3; they are clearly consistent with FK's assumption that $U_{11}\gg\Delta_v$. Considering $kT/G_{\rm ext}=0.1$ ($T\simeq 50\,^{\circ}$K), and first neglecting the term in the exponential of (39) proportional to $n_{\rm ext}$, we find $n_{\rm ext}\approx 10^{-4}$; then $\frac{1}{4}\beta U_{11}n_{\rm ext}\approx 10^{-2}$, justifying the neglect of the term. We similarly estimated $n_{\rm loc}$, assuming in addition that in (36) $(\epsilon_{1\vec{k}})_{\max}-b_1\simeq\frac{1}{2}\Delta_v$. We found

$$n_{\rm ext}/n_{\rm loc}\simeq 3 ,$$

so that $\mathcal{A}_{\rm ext}<\mathcal{A}_{\rm loc}$ for these numbers.

Thus we see that the transition temperature would be $>50\,^{\circ}$K in this case, and, in fact, might be appreciably higher, T_0 possibly occurring in a range where physical transitions are observed ($\approx 500\,^{\circ}$K for Ti_2O_3).

V. SUMMARY AND DISCUSSION: INELASTIC MAGNETIC NEUTRON SCATTERING

In summary, Eq. (37) shows that no matter how large U_{11} and U_{12}, the FK picture, which uses localized valence-band states and extended conduction-band states, is a poorer approximation (in the variational sense) at sufficiently low T ($<T_0$) than the

usual HF picture with extended states for both bands. This result becomes quite plausible on considering the general physical grounds that at low enough T, there are so few holes present that the chances of finding two on one site is negligible even in the extended picture. It was shown that the temperature T_0 is not expected to be unattainably low even when $U_{11} \gg \Delta_v$, and, in fact, the possibility that T_0 might occur in experimentally interesting ranges was noted.

It should be pointed out that the behavior found[1] in the FK picture is reproduced qualitatively in the extended picture (which is the usual HF Bloch approximation). In particular it is possible[9] within this HF extended-state picture to describe the interesting type of semiconductor-to-metal transition discussed by FK[1] to roughly the same accuracy as in the FK picture. The type of behavior that would occur if there is a transition to a localized-hole picture needs further investigation.

Another possible state is one in which localized states occur in both bands (a single-determinantal description of excitons)—this is under consideration.

Finally, we ask the question, is it possible in principle to distinguish between the simple extended picture and the localized-hole extended-electron picture? To this end we calculated the spin-spin correlation function whose space-time Fourier transform gives essentially the neutron scattering cross section.[10] That calculation, summarized in the Appendix, leads to the following requirements on the incident and scattered neutron wave vectors $\vec{q}$ and $\vec{q}'$ (zero T is discussed here for simplicity). For both pictures we have energy conservation

$$\frac{\hbar^2}{2M}(q^2 - q'^2) = \begin{cases} \tilde{\epsilon}_{2\vec{k}_2} - \tilde{\epsilon}_{1\vec{k}_1} & \text{(extended picture)} \\ \tilde{\epsilon}_{2\vec{k}_2} - b_1 - U_{11} & \text{(localized picture)}, \end{cases} \tag{41}$$

where M is the neutron mass. For the extended case we find, in addition, the usual crystal momentum conservation

$$\vec{q} = \vec{q}' + \vec{k}_2 + \vec{k}_1 + \vec{K} \quad \text{(extended picture)}; \tag{42}$$

$\vec{K}$ is 2π times a reciprocal vector and $\vec{k}_1$ and $\vec{k}_2$ are the wave vectors of the hole and the electron created in the neutron scattering ($\vec{k}_1$ is minus the wave vector of the valence-band electron which is destroyed). But for the localized picture, (42) is *not* required (this is quite reasonable, resulting from the distribution in $\vec{k}$ of a localized hole).

It can be seen that the scattering in both cases is spatially diffuse [despite the diffraction condition (42), the involvement of *two* excitations, electron and hole, causes this diffuseness]. Nevertheless, the existence of (42) for the extended case and its nonexistence for the localized picture provides an interesting handle (at least in principle) on distinguishing these pictures. Consider the "absorption edge," i.e., the minimum neutron energy $E_{\min} = \hbar^2 q^{*2}/2M$ for which scattering can occur. For energies in this neighborhood we need only consider the neighborhood of the minimum of the conduction band and the maximum of the valence band, so we can put

$$\epsilon_{1\vec{k}_1} = -\hbar^2 \delta k_1^2/2m_h, \quad \epsilon_{2\vec{k}_2} = G + \hbar^2 \delta k_2^2/2m_e,$$

where $\delta\vec{k}_\nu$ are the deviations of the wave vectors from the respective extrema. Putting

$$k_c^2 = 2MG/\hbar^2, \tag{43}$$

(41) becomes

$$q^2 = q'^2 + k_c^2 + \alpha_1 \delta k_1^2 + \alpha_2 \delta k_2^2, \tag{44}$$

where $\alpha_1 = M/m_h$ and $\alpha_2 = M/m_e$; $m_h = \infty$ for localized holes. Thus from energy conservation alone E_{min} would occur for $q' = 0$ and $q^* = k_c$. Therefore

$$E_{\min} = G \quad \text{(localized holes)} \tag{45}$$

(G being the gap for the localized case, of course). It is important to note that for "small-gap" semiconductors with G a few tens of meV, Eq. (45) can be satisfied with neutrons in the thermal range. However (45) is not necessarily the case for extended holes, since then (42) requires

$$\vec{q}^* = \vec{k}_{10} + \vec{k}_{20} + \vec{K}, \tag{46}$$

where $\vec{k}_{\nu 0}$ are the locations of the band extrema, which cannot, in general, be satisfied. For example, suppose $\vec{k}_{10} = \vec{k}_{20} = 0$; then there would have to be a $\vec{K}$ with magnitude k_c. If it can be satisfied (in our example $K_1 = k_c$), it will be satisfied only for certain orientations of the crystal ($\vec{K}_1$ parallel to the incident neutron beam) so that the scattering can be turned off or on by rotating the crystal. Another way of looking at this is that the edge q^* will depend on crystal orientation, or, again, the scattered intensity will depend strongly on crystal orientation. This is in direct contrast to the situation in the localized case. If (46) cannot be satisfied with $q^* = k_c$, the threshold will occur at a higher energy than G in the extended case, but a similar situation with respect to crystal rotation will hold.

It is interesting to see that a striking physical difference between the localized- and extended-hole pictures is predicted at least in principle, particularly in light of the historically held notion that there is in principle no possibility of such a distinction for a "filled-band" semiconductor[11] (the error made historically is that only the minimum-energy determinant is considered; the distinction actually occurs in the excited states).

The question of whether this kind of scattering can be observed in practice is certainly of interest and is under study. As far as we know, inelastic magnetic neutron scattering in nonmagnetic semiconductors involving excitation of itinerant electrons has not been discussed previously.

ACKNOWLEDGMENTS

I would like to thank H. J. Zeiger for a valuable discussion of the FK paper and R. A. Bari, L. Van Zandt, and S. D. Mahanti for helpful remarks.

APPENDIX: SPIN-SPIN CORRELATION FUNCTION

The cross section for magnetic scattering of unpolarized neutrons by the spin magnetic moment of the electrons is proportional to[10]

$$S(\vec{\kappa},\omega)=\int_{-\infty}^{\infty} dt \int d\vec{r}\, e^{i\vec{\kappa}\cdot\vec{r}-\omega t}\gamma(\vec{r},t)\,, \tag{A1}$$

where $\vec{\kappa}=\vec{q}-\vec{q}'$ (see text), $-\hbar\omega$ is the neutron energy change, $d\vec{r}$ goes over the crystal volume Ω, and

$$\gamma(\vec{r},t)=(1/\Omega)\int d\vec{r}'\,\langle S_z(\vec{r}',0)S_z(\vec{r}'+\vec{r},t)\rangle \tag{A2}$$

is the spin-spin correlation function. The z component of the spin-density operator

$$\vec{S}(\vec{r})=\sum_i \vec{S}_{io}\,\delta(\vec{r}-\vec{r}_{io}) \tag{A3}$$

is the component perpendicular to $\vec{\kappa}$; for our models, which are isotropic, $\langle S_u(\vec{r},0)\,S_u(r',t)\rangle$ will be independent of the component u. The time dependence comes through

$$\vec{S}(\vec{r},t)=e^{iHt/\hbar}\,\vec{S}(\vec{r})e^{-iHt/\hbar}\,. \tag{A4}$$

We define our spin operators so that the eigenvalues of S_u are $\pm\frac{1}{2}$.

The sum in (A3) goes over all the electrons. In terms of creation and destruction operators $b_n^\dagger$ and b_m, (A3) is

$$\vec{S}(\vec{r})=\sum_{nm}(n\,|\,\vec{S}_{op}\delta(\vec{r}-\vec{r}_{op})\,|m)\,b_n^\dagger b_m,$$

so that for a two-band model the appropriate expression is

$$S_z(\vec{r})=\tfrac{1}{2}\sum_{\nu\mu ij\sigma} w_{\nu i}(\vec{r})^* w_{\mu j}(\vec{r})\sigma b_{\nu i\sigma}^\dagger b_{\mu j\sigma}\,, \tag{A5}$$

where $w_{\nu i}(\vec{r})$ is the Wannier function for band ν, site i, $\sigma=\pm1$. These give

$$\gamma_{\vec{\kappa}}(t)\equiv\int e^{i\vec{\kappa}\cdot\vec{r}}\gamma(\vec{r},t)d\vec{r}$$
$$=\tfrac{1}{2}\sum_{\substack{\nu\mu\nu'\mu'\\ ij\sigma i'j'\sigma'}}\sigma\sigma'\,\langle b_{\nu i\sigma}^\dagger b_{\mu j\sigma}b_{\nu' i'\sigma'}^\dagger(t)b_{\mu' j'\sigma'}(t)\rangle$$
$$\times e^{-i\vec{\kappa}\cdot(\vec{R}_i-\vec{R}_{i'})}\,\Omega^{-1}\,I_{\nu\mu}(-\vec{\kappa},\vec{R}_{ij})I_{\nu'\mu'}(\vec{\kappa},\vec{R}_{i'j'})\,, \tag{A6}$$

where $\vec{R}_i$ is the position of the ith site, $\vec{R}_{ij}=\vec{R}_i-\vec{R}_j$, and

$$I_{\nu\mu}(\vec{\kappa},\vec{R})=\int d\vec{r}\, e^{i\vec{\kappa}\cdot\vec{r}}w_\nu(\vec{r})^* w_\mu(\vec{r}+\vec{R})\,. \tag{A7}$$

We shall calculate the correlation functions corresponding to the extended and the localized states by using Hamiltonians

$$\tilde{H}_{ext}=\sum_{\nu\vec{k}\sigma}\tilde{\epsilon}_{\nu\vec{k}}n_{\nu\vec{k}\sigma} \tag{A8}$$

and [as in Eq. (5)]

$$\tilde{H}_{loc}=U_{11}\sum_i N_{1i\uparrow}N_{1i\downarrow}+\sum_i wN_{1i}+\sum_{\vec{k}\sigma}w_{2\vec{k}}n_{2\vec{k}\sigma}\,, \tag{A9}$$

respectively, both in the density matrix $\propto e^{-\beta\tilde{H}}$ needed for the average in (A6) and for the time dependence

$$s(t)=e^{i\tilde{H}t/\hbar}s\,e^{-i\tilde{H}t/\hbar}\,.$$

For the extended case the calculation is straightforward but lengthy. The Wannier function operators $b_{\nu i\sigma}$ are conveniently transformed to the Bloch-function operators $a_{\nu\vec{k}\sigma}$; use is also made of

$$a_{\nu\vec{k}\sigma}(t)=a_{\nu\vec{k}\sigma}\,e^{-i\tilde{\epsilon}_{\nu k}t/\hbar}\,,$$

well-known for the Hamiltonian (A8). We finally obtain the function (A1) for the extended case

$$S_{ext}(\vec{\kappa},\omega)=\frac{2\pi\hbar}{\Omega}\sum_{\vec{k}\vec{k}'\vec{K}}\sum_{\nu\mu}\delta_{\vec{k}'-\vec{k}-\vec{\kappa},\vec{K}}\,\delta(\tilde{\epsilon}_{\mu\vec{k}'}-\tilde{\epsilon}_{\nu\vec{k}}-\hbar\omega)$$
$$\times\langle n_{\nu\vec{k}\sigma}(1-n_{\mu\vec{k}'\sigma})\rangle\,|J_{\mu\nu}(\vec{k},\vec{\kappa})|^2\ \text{for}\ \omega\neq0, \tag{A10}$$

where

$$J_{\mu\nu}(\vec{k},\vec{\kappa})=\sum_m e^{-i\vec{k}\cdot\vec{R}m}\,I_{\mu\nu}(\vec{\kappa},\vec{R}_m); \tag{A11}$$

also it follows from (A8) that

$$\langle n_{\nu\vec{k}\sigma}(1-n_{\mu\vec{k}'\sigma})\rangle=f(\tilde{\epsilon}_{\nu\vec{k}\sigma})[1-f(\tilde{\epsilon}_{\mu\vec{k}'\sigma})]$$
$$\text{for}\ \nu\vec{k}\neq\mu\vec{k}'\,. \tag{A12}$$

For $\omega=0$, one must add a term to (A10) which arises from purely elastic scattering and which vanishes at $T=0$ (we omit it here for simplicity).

The calculation for the localized case is somewhat more complicated, particularly since in (A6) terms with $\nu=1$ have to be treated differently from those for $\nu=2$. Although the general calculation is not prohibitive, for simplicity we will restrict ourselves to zero temperature.

The Hamiltonian (A9) clearly implies

$$a_{2\vec{k}\sigma}(t)=a_{2\vec{k}\sigma}e^{-i\omega_{2\vec{k}}t/\hbar}\,. \tag{A13}$$

We also have

$$\dot{b}_{1i\sigma}=(i/\hbar)[\tilde{H}_{loc},\ b_{1i\sigma}]=-(i/\hbar)b_{1i\sigma}(w_1+U_{11}N_{1i-\sigma}),$$

so that

$$b_{1i\sigma}(t)=b_{1i\sigma}\,e^{-(i/\hbar)(w_1+U_{11}N_{1i-\sigma})t}. \tag{A14}$$

With these relations we can readily see that the zero-T average $\langle b_{\nu i\sigma}^\dagger\cdots\rangle$ in (A6) is zero unless $\nu=1$, $\mu=2$, $\nu'=2$, $\mu'=1$, or $\nu=\mu=\nu'=\mu'=1$. It

can be shown that the latter contributes nothing to $\gamma_{\vec{k}}(t)$, so we are left with the former, namely,

$$\langle b^{\dagger}_{1i\sigma}\, b_{2j\sigma} b^{\dagger}_{2i'\sigma'}(t) b_{1j'\sigma'}(t)\rangle$$

$$= \frac{1}{N}\sum_{\vec{k}\vec{k}'} e^{i\vec{k}\cdot\vec{R}_j - i\vec{k}'\cdot\vec{R}_{i'}} \langle b^{\dagger}_{1i\sigma} a_{2\vec{k}\sigma} a^{\dagger}_{2\vec{k}'\sigma'}(t) b_{1j'\sigma'}(t)\rangle$$

$$= \delta_{ij'}\delta_{\sigma\sigma'} \frac{1}{N}\sum_{\vec{k}} e^{-i\vec{k}\cdot(\vec{R}_{i'}-\vec{R}_j)} e^{i(w_{2\vec{k}}-w_1)t/\hbar} \times \langle 1-n_{2\vec{k}\sigma}\rangle \langle N_{1i\sigma} e^{-iU_{11}N_{1i},-\sigma t/\hbar}\rangle$$

$$\xrightarrow[T\to 0]{} \delta_{ij'}\delta_{\sigma\sigma'} \frac{1}{N}\sum_{\vec{k}} e^{-i\vec{k}\cdot(\vec{R}_{i'}-\vec{R}_j)} e^{i(w_{2\vec{k}}-w_1-U_{11})t/\hbar}$$

$$\longrightarrow \delta_{ij'}\delta_{\sigma\sigma'} \frac{1}{N}\sum_{\vec{k}} e^{-i\vec{k}\cdot(\vec{R}_{i'}-\vec{R}_j)} e^{i(\epsilon_{2\vec{k}}-b_1-U_{11})t/\hbar}. \tag{A15}$$

[In the last step we used (21).] Thus (A6) gives

$$\gamma_{\vec{\kappa}}(t) \xrightarrow[T\to 0]{} \frac{1}{N\Omega}\sum_{ij i'}\sum_{\vec{k}} e^{-i\vec{k}\cdot(\vec{R}_{i'}-\vec{R}_j)} e^{i(\epsilon_{2\vec{k}}-b_1-U_{11})t/\hbar} \times e^{-i\vec{\kappa}\cdot(\vec{R}_i-\vec{R}_{i'})} I_{12}(-\vec{\kappa}, \vec{R}_{ij}) I_{21}(\vec{\kappa}, \vec{R}_{i'i})$$

$$= \frac{1}{\Omega}\sum_{\vec{k}} e^{i(\epsilon_{2\vec{k}}-b_1-U_{11})t/\hbar} J_{12}(\vec{k}, -\vec{\kappa}) J_{21}(\vec{k}-\vec{\kappa}, \vec{\kappa}).$$

Using the general relation

$$J_{\mu\nu}(\vec{q}, \vec{k})^* = J_{\nu\mu}(\vec{k}+\vec{q}, -\vec{k}), \tag{A16}$$

we finally obtain

$$S_{\rm loc}(\vec{\kappa}, \omega) \xrightarrow[T\to 0]{} \frac{2\pi\hbar}{\Omega}\sum_{\vec{k}} \delta(\tilde{\epsilon}_{2\vec{k}} - b_1 - U_{11} - \hbar\omega) \times |J_{12}(\vec{k}, -\vec{\kappa})|^2. \tag{A17}$$

Note that if $\tilde{\epsilon}_{1\vec{k}} = b_1 + U_{11} \equiv \tilde{\epsilon}_1$, then (A10) reduces to (A17): Using (A16), (A10) then gives

$$S_{\rm ext}(\vec{\kappa}, \omega) \xrightarrow[T\to 0]{} \frac{2\pi\hbar}{\Omega}\sum_{\vec{k}\vec{k}'\vec{K}} \delta_{\vec{k}'-\vec{k}-\vec{\kappa},\vec{K}} \times \delta(\tilde{\epsilon}_{2\vec{k}'} - \tilde{\epsilon}_1 - \hbar\omega) \; |J_{12}(\vec{k}+\vec{\kappa}, -\vec{\kappa})|^2$$

$$= \frac{2\pi\hbar}{\Omega}\sum_{\vec{k}\vec{k}'\vec{K}} \delta_{\vec{k}'-\vec{k}-\vec{\kappa},\vec{K}}\, \delta(\tilde{\epsilon}_{2\vec{k}'} - \tilde{\epsilon}_1 - \hbar\omega) |J_{12}(\vec{k}', -\vec{\kappa})|^2$$

$$= \frac{2\pi\hbar}{\Omega}\sum_{\vec{k}'} \delta(\tilde{\epsilon}_{2\vec{k}'} - \tilde{\epsilon}_1 - \hbar\omega) |J_{12}(\vec{k}, -\vec{\kappa})|^2.$$

In other words, if the valence-band width is zero, the zero-T scattering is identical for the two pictures, which is quite reasonable.

Finally, we note that for narrow bands,

$$J_{12}(\vec{q}, \vec{\kappa}) \simeq I_{12}(\vec{\kappa}, 0) = \int d\vec{r}\, e^{i\vec{\kappa}\cdot\vec{r}}\, w_1(\vec{r})^* w_2(\vec{r}),$$

so the form factor is the Fourier transform of the "charge density" $w_1(\vec{r})^* w_2(\vec{r})$—this is reasonable since the process involved is the excitation of an electron from the valence to the conduction band.

---

*A brief account of this work has been presented by T. A. Kaplan, Bull. Am. Phys. Soc. 16, 425 (1971).

[1]L. M. Falicov and J. C. Kimball, Phys. Rev. Letters 22, 997 (1969).

[2]R. Ramirez, L. Falicov, and J. Kimball, Phys. Rev. B 2, 3383 (1970).

[3]T. A. Kaplan, Bull. Am. Phys. Soc. 13, 386 (1968).

[4]T. A. Kaplan and P. N. Argyres, Phys. Rev. B 1, 2457 (1970).

[5]A. Huber, in *Mathematical Methods in Solid State and Superfluid Theory*, edited by R. C. Clark and G. H. Derrick (Plenum, New York, 1968), Chap. 14.

[6]John M. Blatt, *Theory of Superconductivity* (Academic, New York, 1964).

[7]This refutes what appears to be the implication of A. Minesaki [Phys. Letters 32A, 235 (1970)] that FK calculated the entropy incorrectly (at least for the special J, J', and m values considered here).

[8]This can be indicated by comparing (39) and (40) below: as $U_{11}\to\infty$, $n_{\rm ext}\to 0$ whereas $n_{\rm loc}$ does not $\to 0$. Thus (31) and (33) suggest that for large enough U_{11}, $\mathcal{A}_{\rm loc} < \mathcal{A}_{\rm ext}$. (The limit $\Delta_v \to 0$ is not included since $kT \ll \Delta_v$ was assumed in these equations.)

[9]H. J. Zeiger, T. A. Kaplan, and P. M. Raccah, Phys. Rev. Letters 26, 1328 (1971).

[10]P. G. deGennes, in *Magnetism*, edited by G. Rado and H. Suhl (Academic, New York, 1963), Vol. III.

[11]For a recent expression of this view see H. Brooks, Rev. Mod. Phys. 40, 795 (1968).