Theory of temperature-induced Mott transitions

Ellen J. Yoffa*
Department of Physics and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

David Adler
Department of Electrical Engineering and Computer Science and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 30 November 1978)

The Hubbard model is extended to include long-range Coulomb interactions between electrons on different atomic sites. This results in a screening of the effective correlation energy by free carriers. Both the cases of an integral number of electrons per atom and a general electronic density are considered. The electronic free energies of both insulating and metallic states for finite bandwidths are calculated and compared. These results are used to generate complete phase diagrams as functions of temperature and bandwidth. A wide range of electronic behavior can be understood by use of the model. For certain materials, insulator-to-metal transitions are predicted as the temperature is increased. However, for somewhat larger bandwidths, a metallic ground state is present and two transitions are predicted: metal-to-insulator and, at a still higher temperature, insulator-to-metal. The model is used to analyze the anomalous transport properties of the $Ni_{1-x}Co_xS_2$ system.

I. INTRODUCTION

Metal-insulator transitions have been studied intensively for the past 50 years (for an excellent review, see Ref. 1), and yet remain one of the least understood and most controversial areas of solid-state physics. One of the reasons for the profound interest in this subject is the fact that such transitions often focus on the regions of validity of the one-electron approximation, the basis for much of the quantitative theory of solids. This approximation neglects the correlations between electrons in the material and often leads to the prediction that some of our most insulating materials should be metallic.[1,2] Another reason for the interest in the transitions themselves is the wide diversity of systems in which they have been observed; in particular, they are often found in transition-metal[2,3] and rare-earth[4] compounds, organic charge-transfer salts,[5] and disordered solids,[6] which all are systems where electronic correlations play an important role. Mott[7] was the first to suggest that such correlations could result in an insulating ground state, even when one-electron theory predicts metallic behavior. He also presented arguments that the transition between such an insulating state and the metallic state of ordinary band theory should be sharp; such metal-insulator transitions are now called *Mott transitions*.

Despite a great deal of experimental work on temperature-induced metal-insulator transitions in systems in which correlations are important, there has been a singular lack of quantitative theoretical investigation in this area. There have always been some major difficulties with models which explicitly include electronic correlations. In a series of papers, Hubbard[8-10] introduced a Hamiltonian which is exact in the two opposite limits of small correlations (one-electron limit) and narrow bandwidths (atomic limit). Thus this Hamiltonian could be used to analyze major aspects of both the insulating and metallic states of such systems, as well as the transition from one of these states to the other. The essence of Hubbard's approximation was to consider explicitly short-range intrasite correlations between electrons and to treat the remaining Coulomb interactions with a mean-field (Hartree-Fock) description. Several investigators have suggested mechanisms for semiconductor-to-metal transitions involving a temperature-induced collapse of the semiconducting gap.[11-15] In fact, Doniach[16] has shown that any system with an energy gap which varies as $E_G(f) = E_G(0) - U_{\text{eff}} f$, where U_{eff} is a positive effective interaction and f is the fraction of excited carriers, may exhibit a cooperative first-order transition.

In this paper, we develop a model which describes screening-induced metal-insulator transitions in Mott-Hubbard insulators. We first extend the Hubbard model to include long-range Coulomb interactions between electrons on different atomic sites. A detailed analysis then shows that excited carriers can screen the on-site Coulomb repulsion. In Sec. II, we derive an expression that describes this screening. We then generalize this derivation in Sec. III to include

systems in which free carriers are present. We find significant modifications of the screening behavior due to these free carriers. In the zero-bandwidth limit, we obtain two equations which simultaneously determine the value of the intrasite correlations and the density of excited carriers. We show in Sec. IV that the self-consistent solution of these equations predicts that the insulating state often cannot exist above a certain critical temperature. The remainder of this paper is devoted to an in-depth investigation of various aspects of the finite-bandwidth system. Section V deals with the screening-induced behavior of the insulating gap. In Sec. VI, we introduce the self-consistent metallic state and examine its properties. In order to describe the transition behavior, it is necessary to calculate and compare the free energies of the insulating and metallic states. We do this in Sec. VII and then generate complete phase diagrams as functions of temperature and bandwidth for these systems. In Sec. VIII, ideas presented in the previous sections are applied to $Ni_{1-x}Co_xS_2$, a system which exhibits many anomalous transport properties not explainable on the basis of conventional one-electron models.[17,18] However, we show that these anomalous properties can be understood in terms of a model in which excited carriers screen a Hubbard gap.

II. SCREENING OF THE CORRELATION ENERGY: HALF-FILLED-BAND CASE

The major feature of the Hubbard Hamiltonian[8] is its explicit treatment of the on-site interaction U between electrons in the same band. Interband and intersite processes do not appear explicitly and enter only through the Hartree-Fock field. For the systems we have chosen to consider, the correlated electrons are *simultaneously* the conduction electrons, so that these materials have no highly mobile conduction-electron gas to screen the intersite interactions as in the model of Falicov and Kimball.[11] Consequently, the off-site terms cannot be ignored. In the following, we develop a formalism for treating the effect of the most important class of intersite interactions on the effective intrasite Coulomb repulsion U.

The complete Coulomb-interaction matrix elements are given by

$$\langle ij\left|\frac{1}{r}\right|kl\rangle \equiv q^2\int d\vec{r}\,d\vec{r}\,'\,\phi^*(\vec{r}-\vec{R}_i)\phi^*(\vec{r}\,'-\vec{R}_j)\times\frac{1}{|\vec{r}-\vec{r}\,'|}\phi(\vec{r}-\vec{R}_k)\phi(\vec{r}\,'-\vec{R}_l)\,, \tag{1}$$

where the wave functions ϕ are Wannier orbitals centered on sites i, j, k, and l. In terms of these interactions, the Hubbard Hamiltonian can be written

$$H=\sum_{ij\sigma}T_{ij\sigma}c^\dagger_{i\sigma}c_{j\sigma}+\frac{1}{2}\sum_{\substack{ij\\kl}}\sum_{\sigma\sigma'}\langle ij\left|\frac{1}{r}\right|kl\rangle\, c^\dagger_{i\sigma}c^\dagger_{j\sigma'}c_{l\sigma'}c_{k\sigma}$$
$$-\sum_{\substack{ij\\kl}}\sum_{\sigma}\left[\nu_{jl,-\sigma}\langle ij\left|\frac{1}{r}\right|kl\rangle+\nu_{jl,\sigma}\left(\langle ij\left|\frac{1}{r}\right|kl\rangle-\langle ij\left|\frac{1}{r}\right|lk\rangle\right)\right]c^\dagger_{i\sigma}c_{k\sigma}\,, \tag{2}$$

where $T_{ij\sigma}$ is the transfer integral for a spin-σ electron from site j to site i, $c^\dagger_{i\sigma}$ ($c_{i\sigma}$) creates (annihilates) a spin-σ electron on site i, and

$$\nu_{ij,\sigma}\equiv N^{-1}\sum_{\vec{k}}\nu_{\vec{k}\sigma}\exp[-i\vec{k}\cdot(\vec{R}_i-\vec{R}_j)]\,,$$

where $\nu_{\vec{k}\sigma}$ is the average occupation number of the Bloch state $|\vec{k},\sigma\rangle$.

Because the overlap of localized wave functions centered on different lattice sites is so small, the direct Coulomb term

$$J_{ij}\equiv\langle ij\left|\frac{1}{r}\right|ij\rangle$$

is much larger than the other off-site integrals $\langle ij|1/r|kl\rangle$, all of which involve at least one such small overlap factor. In fact, i and j must differ by ten lattice constants before J_{ij} falls to the size of the next-largest nearest-neighbor term $\langle ii|1/r|ij\rangle$. Thus, for a three-dimensional lattice, there are 10^6 direct interactions that are stronger than the *nearest-neighbor* exchange term $\langle ij|1/r|ji\rangle$.

Suppose we include the entire set of direct Coulomb interactions, where i and j are now unrestricted. Kemeny[19] has considered these correlations in his general formulation of the Hubbard Hamiltonian, although he obtained solutions only for the case $J_{ij}(i\neq j)=0$. If the number of electrons N equals the number of lattice sites N_0 and the state is nonmagnetic ($\nu_{i\uparrow\sigma}=\frac{1}{2}$), then Eq. (2) reduces to

$$H=\sum_{ij}\sum_{\sigma}T_{ij\sigma}c^\dagger_{i\sigma}c_{j\sigma}+U\sum_i n_{i\sigma}n_{i\,-\sigma}$$
$$+\sum_{i>j}J_{ij}(n_i-1)(n_j-1)-\sum_{i>j}J_{ij}-\frac{NU}{2}\,, \tag{3}$$

where $n_{i\sigma} = c_{i\sigma}^{\dagger} c_{i\sigma}$.

Consider the term

$$\sum_{i>j} J_{ij}(n_i - 1)(n_j - 1) .$$

Using definition (1) and the fact that the charge (operator) on site i, Q_i, equals $-q(n_i - 1)$, we can rewrite the sum as

$$\sum_{i>j} \frac{Q_i Q_j}{|\vec{R}_i - \vec{R}_j|} ,$$

where we have assumed negligible overlap between wave functions centered on different atomic sites. Note that this term has the form of a long-range Coulomb interaction between sites having nonzero net charge. [Kaplan and Argyres[20] have considered these intersite terms and have shown rigorously (for $N = N_0$) that the charged sites occur only in excited states of the system.] Thus not only do the electrons *on* site i interact but, in addition, the charges Q_j are correlated with the charge on site i. Although the Q_j are fixed to the lattice points $\vec{R}_j$, the values Q_j may vary from -1 to $+1$. By influencing the charge density over regions surrounding $\vec{R}_i$, the charges Q_j can screen the on-site interaction U. The most important intersite interaction takes the form of a Coulomb interaction between charged sites. We shall not attempt to solve the Hamiltonian (3) exactly. Instead, we shall approximate it by incorporating the effects of the Coulomb terms into a dielectric screening function.

A static potential applied to a system of electrons leads to spatial variations in the local chemical potential ϵ_F,[21,22] and, consequently, in the local electron density n. Such behavior is described by the static dielectric screening function

$$\epsilon(\tilde{q}) = 1 + \frac{4\pi q^2}{\tilde{q}^2} \frac{dn}{d\epsilon_F} ,$$

which holds for small wave vector $\tilde{q}$.[23] This description is strictly valid for "weak coupling" only, i.e., in the limit where there are many charges within the screening volume. When Boltzmann statistics (and Debye-Huckel screening) apply, this requirement reduces to $kT \gg q^2/r$, i.e., the thermal kinetic energy must dominate the Coulomb interactions. For the cases examined here, the pseudoparticle bandwidths (and, hence, kinetic energies) are much smaller than, or comparable to, the on-site Coulomb repulsion energies. However, *even for localized electrons*, spatial variations in average site occupancies do lead to static charge fluctuations, and the intrasite repulsion is then reduced as a result of the interaction between a charged site and its altered neighborhood. Consequently, when the screening length exceeds the average interparticle distance, we can treat the charge redistribution as an effective screening mechanism.

In the atomic limit of the Hubbard model ($T_{ij\sigma} = 0$, $i \neq j$), the grand partition function corresponding to the Hamiltonian (2) is[20,24]

$$\begin{aligned} Z &= \left(\sum_s g_s \exp[-\beta(E_s - \epsilon_F N_s)] \right)^{N_0} \\ &= (1 + 2\exp[-\beta(T_0 - \epsilon_F)] \\ &\quad + \exp[-\beta(2T_0 + U - 2\epsilon_F)])^{N_0} , \end{aligned} \tag{4}$$

where $T_0 \equiv T_{ii}$, and g_s, E_s, and N_s are respectively the degeneracy, energy, and occupation number of single-site configuration "s"; the site may be empty, singly occupied, or doubly occupied. Since $N = (1/\beta)(\partial \ln Z / \partial \epsilon_F)$, we obtain

$$n = n_0 \left(\frac{2\exp[-\beta(T_0 - \epsilon_F)] + 2\exp[-\beta(2T_0 + U - 2\epsilon_F)]}{1 + 2\exp[-\beta(T_0 - \epsilon_F)] + \exp[-\beta(2T_0 + U - 2\epsilon_F)]} \right), \tag{5}$$

where n_0 is the site density and n is the electron density. Taking the derivative $dn/d\epsilon_F$, we find

$$\frac{dn}{d\epsilon_F} = 2\beta n_0 \left(\frac{\exp[-\beta(T_0 - \epsilon_F)] + 2\exp[-\beta(2T_0 + U - 2\epsilon_F)] + \exp[-\beta(3T_0 + U - 3\epsilon_F)]}{\{1 + 2\exp[-\beta(T_0 - \epsilon_F)] + \exp[-\beta(2T_0 + U - 2\epsilon_F)]\}^2} \right) . \tag{6}$$

For an average of one electron per site, $n = n_0$; Eq. (5) can then be solved to obtain $\epsilon_F = T_0 + \frac{1}{2}U$. Substitution of this result into Eq. (6) yields

$$\frac{dn}{d\epsilon_F} = \frac{\beta n_0}{1 + e^{\beta U/2}} . \tag{7}$$

However, we find from the grand partition function Eq. (4) that the density of doubly occupied sites n_d is

$$n_d = \frac{n_0}{2(1 + e^{\beta U/2})} . \tag{8}$$

Equations (7) and (8) then give

$$\frac{dn}{d\epsilon_F} = 2\beta n_d , \tag{9}$$

and Eq. (3) becomes $\epsilon(\tilde{q}, 0) = 1 + (4\pi q^2/\tilde{q}^2)(2\beta n_d)$. (10)

Thus the Coulomb repulsion q^2/r is screened to

$$V(r)=\frac{q^2}{r}\exp\{-[4\pi q^2\beta(2n_d)]^{1/2}r\}. \tag{11}$$

Hubbard introduces the bare intrasite Coulomb repulsion U_0 by defining $U_0\equiv\langle ii\,|\,1/r\,|\,ii\rangle$. By analogy with this, we approximate the screened interaction $\langle ii\,|\,V(r)/q^2\,|\,ii\rangle$ by

$$U_0\exp\{-[4\pi q^2\beta(2n_d)]^{1/2}d\},$$

where the average intersite distance

$$d=\int d\vec{r}\,d\vec{r}'\,|\vec{r}-\vec{r}'|\rho_i(\vec{r})\rho_i(\vec{r}')$$

is typically 0.5–3.0 Å, depending on the orbital configuration. Thus the on-site Coulomb repulsion can be written

$$U=U_0\exp\{-[4\pi q^2\beta(2n_d)d^2]^{1/2}\}. \tag{12}$$

Note that the inverse screening length ($q_s=[4\pi q^2\beta(2n_d)]^{1/2}$) increases with the number of doubly occupied sites. This follows from the fact that these sites (and the empty sites with density $n_e=n_d$) are the charge centers involved in the long-range Coulomb interactions, as discussed earlier. (Not surprisingly, Rice and Brinkman,[25] using Gutzwiller's[26,27] variational approach to study the Mott transition, find that the screening parameter in the strongly correlated metallic state also increases with n_d.) In the atomic limit, the density of doubly occupied sites is one-half the density n_u of electrons in the upper Hubbard pseudoparticle level.[22] The screened interaction is therefore

$$U=U_0\exp[-(4\pi q^2\beta n_u d^2)^{1/2}]. \tag{13}$$

By making this approximation to the exact Hamiltonian (3), we have generated an effective Hamiltonian which depends on temperature. When a careful free-energy variational procedure is used to derive an approximate Hamiltonian, fundamental thermodynamic and statistical relations are consistent.[28] Unfortunately, with the method described here, if the density operator $\rho\equiv Z^{-1}\exp[-\beta(H-\mu N)]$ and the thermodynamic potential $\Omega\equiv -kT\ln Z$, it is no longer true that the entropy $-k\,\mathrm{Tr}\rho\ln\rho=-\partial\Omega/\partial T$. In order to define the free energy unambiguously, we make the *ansatz* that $U(n_u)\equiv U(\bar{n}_u^i)$, where the $\bar{n}_u^i$ are determined by the self-consistency requirements

$$\left(\frac{\partial\Omega}{\partial n_u}\right)_{U,T}\Bigg|_{n_u=\bar{n}_u^i}=0$$

and

$$U=U_0\exp[-(4\pi q^2\beta\bar{n}_u^i d^2)^{1/2}].$$

With this condition, the inconsistency mentioned above is resolved and we can use the thermodynamic potential to determine, e.g., the equilibrium density of excited carriers $\bar{n}_u$: $\Omega(\bar{n}_u)=\min[\Omega(\bar{n}_u^i)]$.

Although we have derived Eq. (13) in the atomic limit, we wish to use it to study screening effects for finite bandwidths. For the case $U\gg\Delta$, but with $\Delta>0$, we demonstrate in the Appendix that $q_s=(4\pi q^2\beta n_u)^{1/2}d$ when the Boltzmann limit applies, just as the usual Debye-Huckel screening formula depends on the *total* number of conduction electrons, independent of the density of states in the band or the details of the level occupation. We shall then use Eq. (13) for the full range of model parameters in the insulating state. Equation (13) is an important result because it shows that the size of the Hubbard gap decreases as electrons are excited across it. Whenever the energy gap obeys a relationship of this form, we can expect a transition to occur. We shall examine the details of this transition in the sections that follow. But first, in Sec. III, we extend the results of the present section to those systems for which the number of electrons does not equal the number of sites ($n\neq n_0$). In that case, Eq. (13) must be modified in a significant manner.

III. SCREENING OF THE CORRELATION ENERGY: GENERAL CASE

In Sec. II we derived an expression governing the screening of the Hubbard gap for systems in which the number of electrons equals the number of lattice sites ($n=n_0$). However, it is often the case that these numbers are not equal, for example, in compounds that mix ionic charges (such as $Ni_{1-x}Li_xO$)[29,30] or in materials having incomplete charge transfer from cation to anion [such as tetrathiafulvalenium-tetracyanoquinodimethanide (TTF-TCNQ)].[5] Here we extend the results of Sec. II to include $n\neq n_0$.

Using the general expression Eq. (4) for the grand partition function Z in the atomic limit, we obtain

$$n^s=\frac{n_0 g_s\exp[-\beta(E_s-\epsilon_F N_s)]}{1+2\exp[-\beta(T_0-\epsilon_F)]+\exp[-\beta(2T_0+U-2\epsilon_F)]}, \tag{14}$$

where n^s is the density of lattice sites in state s. When we now relax the assumption that $n=n_0$, we modify the derivative Eq. (9) which becomes

$$\frac{dn}{d\epsilon_F}=\frac{\beta}{n_0}(n^+n^0+4n^+n^-+n^0n^-). \tag{15}$$

Here, n^+, n^0, and n^- are the densities of sites that are empty, singly occupied, and doubly occupied.

(Note that $n^- \equiv n_{d^-}$.) Clearly the superscript indicates the net charge of the site in the state s. Note that Eq. (15) is symmetric about $n = n_0$, so that extra carriers of either sign contribute equally to the screening.

Eliminating n^0 and n^+ from Eq. (15), we obtain

$$\frac{dn}{d\epsilon_F} = \beta[2n^- + n(1 - n/n_0)] = \beta[n_u + n(1 - n/n_0)]. \quad (16)$$

Consequently, the modified expression for the screened interaction is

$$U = U_0 \exp(-\{4\pi q^2 \beta[n_u + n(1 - n/n_0)]d^2\}^{1/2}). \quad (17)$$

The most important difference between Eq. (17) and Eq. (12) is that as $T \to 0$, $n_u \to 0$, whereas $n_u + n(1 - n/n_0) \to n(1 - n/n_0) \neq 0$. Of course, the assumptions of our derivation break down as $T \to 0$. However, the point we wish to make is that in the general case, even at low temperatures, screening can occur by means of extrinsic carriers and does not require intrinsic thermal excitation. Consequently, we expect that U will decrease as the value of n deviates from n_0. In fact, experimental measurements of the energy gap E_g of $Ni_{1-x}Co_xS_2$,[17,31] for example, show that E_g does decrease with x. (The bandwidth of $Ni_{1-x}Co_xS_2$ is expected to increase with added Co, so that the shrinking of the gap most likely results from a combination of both effects.)

We can thus conclude that extrinsic carriers can rearrange themselves in order to enhance the screening. In Sec. IV we shall return to the intrinsic case in order to study the transition in the limit of zero bandwidth.

IV. PROPERTIES OF THE INSULATING STATE IN THE NARROW-BAND LIMIT

In Secs. II and III we derived an expression describing the screening of the Hubbard energy in an intrinsic material and extended the result to systems having a general ratio of electrons to lattice sites. Although we have demonstrated that Eq. (13) is valid for finite bandwidths, we shall concentrate in this section on the case of $n = n_0$ in the narrow-band limit. The reason that we look at screening in this limit is that we can easily solve self-consistently for the number of carriers excited to the upper pesudoparticle band. We then obtain an equation determining the temperature at which the insulating state vanishes. Unfortunately, detailed solutions to the microscopic screening problem generally assume weak coupling between the electrons. Because, in the zero-bandwidth case, this assumption breaks down in the vicinity of the transition, we cannot accurately determine the transition temperature. However, the insight we gain from the analytic approach will help us understand the more detailed numerical calculations for finite bandwidths that follow.

U is determined by the equation

$$U = U_0 \exp[-(s\beta f)^{1/2}], \quad (18)$$

where $f = n_u/n_0$ is the fraction of electrons in the upper Hubbard band, and the screening parameter s is $s \equiv 4\pi n_0 q^2 d^2$; s depends only on n_0, the density of atoms, and d, the average on-site interelectron distance. Although n_0 will generally depend on temperature, the dependence is so weak that we can, with negligible error, treat s as a constant. For typical values of these parameters, $s \sim 0.01$–1 eV. But f is also given by

$$f = \int_{-\infty}^{\infty} \frac{g_u(E)\,dE}{1 + e^{\beta(E - U/2)}}. \quad (19)$$

Thus we have a pair of *coupled* equations determining U and f. As the temperature is raised, positive feedback occurs: as more carriers are excited across the gap, the Coulomb repulsion is more effectively screened [cf. Eq. (18)]. At the same time, as the gap shrinks, the fraction of electrons in the upper band grows [cf. Eq. (19)]. We must solve these equations self-consistently.

In the atomic limit, $g_u(E) = \delta(E - U)$, so that Eq. (19) reduces to

$$f = 1/(1 + e^{\beta U/2}). \quad (20)$$

The coupled equations, (18) and (20), can be solved graphically for f and U by plotting Eq. (18) and the inverse of Eq. (20):

$$U = (2/\beta)\ln(1/f - 1). \quad (21)$$

In this method, each value of T generates a unique pair of curves. However, because we would like to find $f(T)$ and $U(T)$, it is clearly best to simplify the problem analytically first.

As a result of our discussion in Sec. II, we must require that the average distance between screening electrons be much less than the screening length. Consequently, $q_s d$ will certainly be much less than unity and we can approximate Eq. (18) by

$$U = U_0[1 - (s\beta f)^{1/2}]. \quad (22)$$

As long as the bandwidth is very small, Eq. (22) will be valid for any value of βU. Nevertheless, suppose we first restrict ourselves to the Boltzmann limit, where

$$f = e^{-\beta U/2}. \quad (23)$$

This equation then becomes

$$f = \exp\{-(\beta U_0/2)[1 - (s\beta f)^{1/2}]\}. \quad (24)$$

Equations of this form were dealt with by Adler and Brooks.[32]

Defining $\alpha \equiv \exp(-\beta U_0/2)$ and $\gamma \equiv \exp[\frac{1}{2}(\beta^{3/2} U_0 s^{1/2})]$, Eq. (24) becomes

$$f = \alpha \gamma^{\sqrt{f}} . \tag{25}$$

With the additional definitions

$$x \equiv f/\alpha \tag{26}$$

and $\xi \equiv \gamma^{\sqrt{\alpha}}$, we obtain

$$x^{1/\sqrt{x}} = \xi , \tag{27}$$

or alternatively

$$x = \xi^{\left(\xi^{\left(\xi^{\left(\xi^{(\cdots)^{1/2}}\right)^{1/2}}\right)^{1/2}}\right)^{1/2}} . \tag{28}$$

In Fig. 1 we plot x (which is $x = fe^{\beta U_0/2}$) as a function of $x^{1/\sqrt{x}} = \xi$ (which depends only on temperature for fixed U_0 and s). The most important feature of the curve is the essential singularity that occurs in x (hence in f) at the temperature corresponding to $x^{1/\sqrt{x}} = \xi = e^{2/e}$. We can obtain $f(T)$ for *any* choice of U_0 and s from this single curve. For any temperature, we can first locate the appropriate point on the horizontal axis of Fig. 1 by using Eq. (27). Then, reading the corresponding value of x from the curve, we obtain f from Eq. (26). The value of U then follows easily from Eq. (22).

The dashed portions of the curve represent unphysical solutions, Equation (27) shows that the entire temperature range $0 \leq T \leq \infty$ corresponds to $1 \leq x^{1/\sqrt{x}} \leq e^{2/e}$; thus $x < 1$ does not correspond to real temperatures. The segment $x > e^2$ corresponds to a free-energy maximum.

Figures 2 and 3 show $f(T)$ and $U(T)$ for $U_0 = 1$ eV and several different screening strengths s; Figs. 4 and 5 show the same functions for $s = 0.625$ eV and several different values of U_0.

First, let us look at Fig. 2. There are several important features to consider. At low temperatures, the fraction of excited carriers $f(T)$ shows exponential activation. Furthermore, for all values of the screening parameter s, the functions $f(T)$ fall on a common curve. Both of these effects occur because there are not yet enough excited carriers for the screening to be appreciable. Each curve exhibits a singularity at a temperature above which we find no solution for $f(T)$. The singularity temperatures decrease as s is increased. The values of $f(T)$ at these temperatures range from $\sim 10^{-1}$ to 10^{-4}.

We can further understand this behavior by looking at Fig. 3. At low temperatures the gap is essentially unscreened for all values of s. It is interesting that, as s decreases, the transition not only occurs at a higher temperature but also at a smaller screened value of U.

Finally, from Figs. 4 and 5, we see that, for fixed s, the singularity temperature increases with U_0.

The function $f(T)$ has an essential singularity at the temperature β_s given by

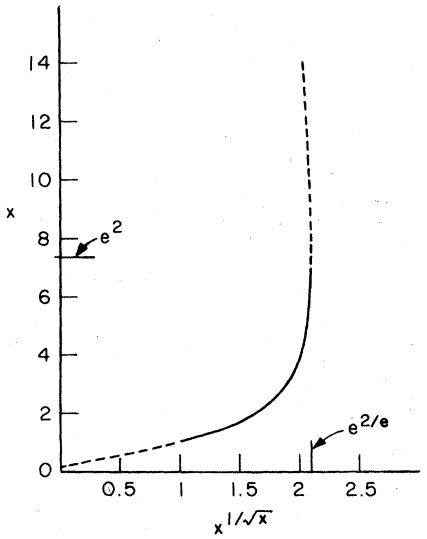

FIG. 1. Plot of x vs $x^{1/\sqrt{x}}$. This curve can be used to find the fraction of excited carriers at any temperature.

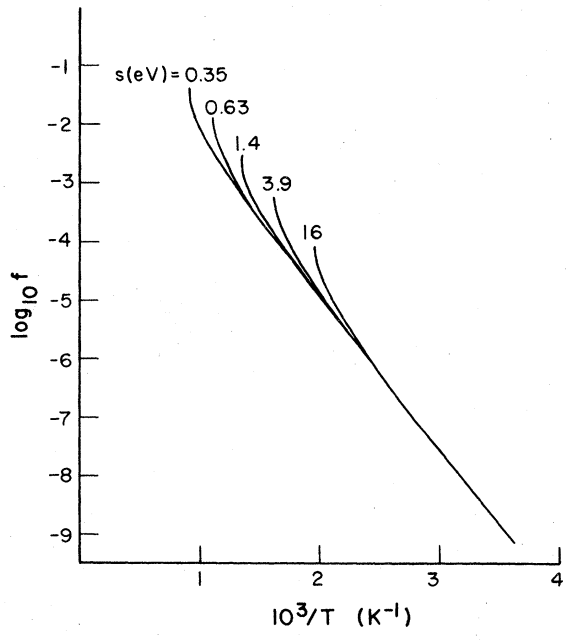

FIG. 2. Temperature dependence of the fraction of excited carriers for $U_0 = 1$ eV and several values of screening strength s.

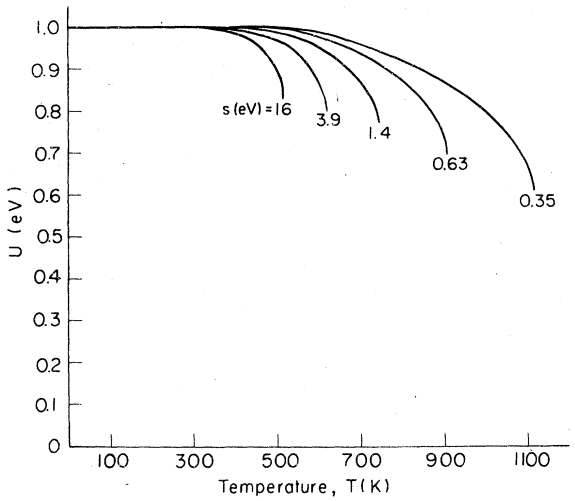

FIG. 3. Temperature dependence of the correlation energy U for $U_0=1$ eV and the same values of s as in Fig. 2.

$$\exp\{\tfrac{1}{2}[\beta_s^{3/2}U_0 s^{1/2}\exp(-\beta_s U_0/4)]\}=e^{2/e}. \quad (29)$$

The value of f at the singularity is

$$f_s=\exp(2-\beta_s U_0/2). \quad (30)$$

In Fig. 6 we plot the singularity temperature as a function of $\log_{10}s$ for various values of U_0. Note that for each U_0 there is a critical screening strength $s_{\min}$ less than which there is no singularity. From Eq. (13) we find that $s_{\min}=\frac{2}{27}eU_0$ and that $\beta_s(s_{\min})=6/U_0$.

The significance of the singularity in f is that a self-consistent insulating state [as described by Eqs. (22) and (23)] *cannot exist at higher tem-*

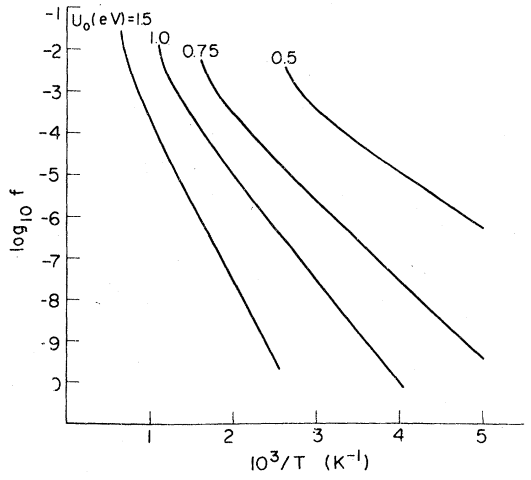

FIG. 4. Temperature dependence of the fraction of excited carriers for screening strength $s=0.625$ eV and several values of U_0.

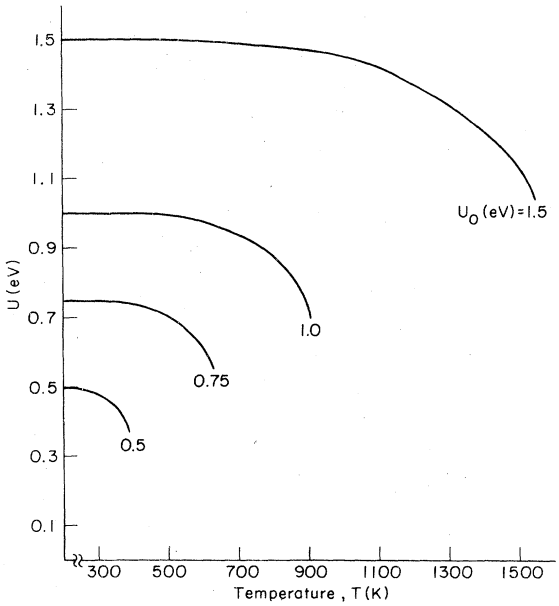

FIG. 5. Temperature dependence of the correlation energy U for screening strength $s=0.625$ eV and the same values of U_0 as in Fig. 4.

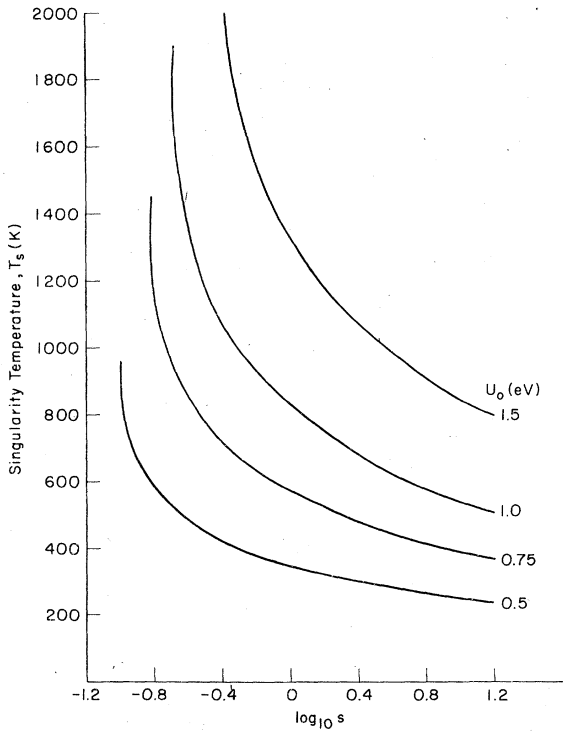

FIG. 6. Temperature of the singularity in the fraction of excited carriers as a function of the $\log_{10}$ of the screening strength.

peratures. At this temperature or below it, there must be a transition to another stable state. Although we have not yet described what the nature of this state might be, we do know that the final configuration will be the one with the lowest free energy.

Finally, we must check the results shown in Figs. 2–5 to verify that the assumptions made in our derivation of Eqs. (22) and (23) have not been violated. In all cases, the Boltzmann approximation is accurate to well within 5%. In deriving Eq. (22), we also assumed that the screening length is much greater than the average interparticle distance. This condition reduces to

$$s\beta f^{1/3} \ll \tfrac{1}{10}. \tag{31}$$

This is the requirement of weak coupling. The inequality (31) is most difficult to satisfy at temperatures near β_s, where the f's are largest. For the cases shown in the figures, (31) is satisfied at all temperatures sufficiently lower than the singularity point. However, at that point, $s\beta_s f_s^{1/3} \sim 1$. It is hoped that the stronger coupling at the highest temperatures will introduce only minor deviations from the predictions of the weak-coupling theories, as the detailed microscopic screening problem has been solved only in the weak-coupling limit.[23]

Thomas-Fermi screening theory, which is also confined to weak coupling, requires a much larger density of carriers than can exist in our systems, especially at temperatures well below the singularity temperatures. Nevertheless, we wish to emphasize that the type of singularity we are discussing is *not* limited to Debye-Huckel screening but will occur whenever the screening length decreases with the number of thermally excited carriers.

For a screened potential of the general form

$$V(r) = (q^2/r)\exp(-A\beta^a f^b r), \tag{32}$$

the function $f(T)$ has a singularity at the temperature satisfying the equation

$$\exp[\tfrac{1}{2}(AdU_0\beta_s^{(a+1)}e^{-\beta_s U_0 b/2})] = e^{1/be}. \tag{33}$$

At this temperature $f_s = \exp[(1/b) - (\beta_s U_0/2)]$. No singularity occurs when the screening strength is less than the minimum value

$$A_{\min} = \frac{1}{d}\left(\frac{beU_0}{2}\right)^a (a+1)^{-(a+1)}. \tag{34}$$

Thus we see that the type of singular behavior obtained from our previous calculation is *not* confined to one particular screening theory, but occurs whenever the interaction creating the insulating gap has the form given by Eq. (32).

Recall that the critical carrier density at the Mott transition[33] is $n_M = (0.25/a_0)^3$. It is interesting to note that the transition densities we have been discussing here are typically two to four orders of magnitude smaller than n_M. Consequently, whereas a Mott transition could occur in these materials only under applied pressure, the temperature-induced transitions under consideration here can take place at equilibrium lattice spacings and nondegenerate electron densities.

We can also solve Eqs. (18) and (20) using the graphical method described earlier. In this case, we do not need to assume that we are in the Boltzmann limit. For $U_0 = 1$ eV and various screening strengths, we obtain solutions for $f(T)$ and $U(T)$. In Figs. 7(a)–7(i), we show the behavior of $f(T)$ for increasing values of s, while the corresponding $U(T)$ curves are given in Figs. 8(a)–8(i). [These solutions $f(T)$ are $f(T) = \bar{n}_u^i/n_0$, where $\bar{n}_u^i$ are the densities discussed in Sec. II.]

As we expect from the behavior of our analytical solution, a singularity in U appears for s greater than a critical value, $s_{\min} \sim 0.95$ eV. (The assumptions made earlier gave $s_{\min} = 0.20$ eV.) It is interesting to observe the change in the character of the solutions as s passes through the critical value [cf. Figs. 7(e), 7(f), 8(e), and 8(f)]. Using this method, we find that f and U have solutions at *all* temperatures and that, without the approximations given by Eqs. (22) and (23), a third solution—large f, small U—appears. Most likely, the transition that takes place in the vicinity of the singularity is to a state of this sort. However, no matter how small U is, as long as it is greater than zero, there will always be an energy gap. Consequently, the system is never really "metallic." This is a problem which occurs only in the zero-bandwidth treatment. In the following sections, we shall present finite-bandwidth calculations.

V. PROPERTIES OF THE INSULATING STATE AT FINITE BANDWIDTHS

In the previous sections, we have shown that we can expect transitions to occur as a result of the screening-induced collapse of the insulating state in the atomic limit. Hubbard[10] has calculated the pseudoparticle density of states corresponding to an unperturbed density of the form

$$p(E) = \begin{cases} (8/\pi\Delta)\left[1 - \left(\frac{E}{\frac{1}{2}\Delta}\right)^2\right]^{1/2}, & |E| < \tfrac{1}{2}\Delta \\ 0, & |E| > \tfrac{1}{2}\Delta. \end{cases} \tag{35}$$

The solid curves in Fig. 9 show the resulting state density for several values of the ratio U/Δ. On the insulating side of the transition (which occurs at $U/\Delta = \frac{1}{2}\sqrt{3}$), we see that the band shape is very

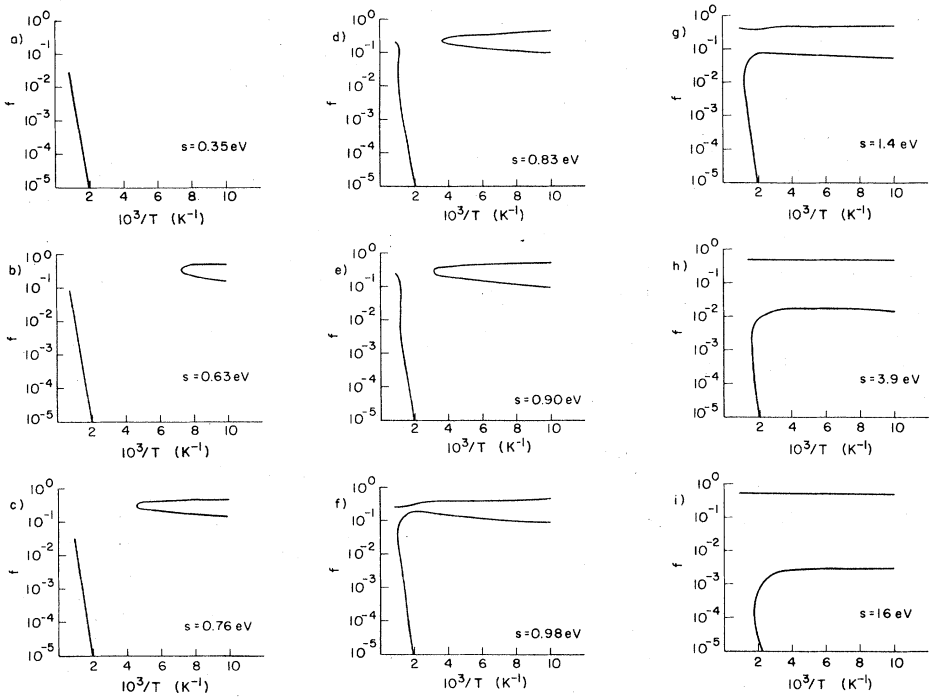

FIG. 7. Temperature dependence of fraction of excited carriers for $U_0=1$ eV and several values of s.

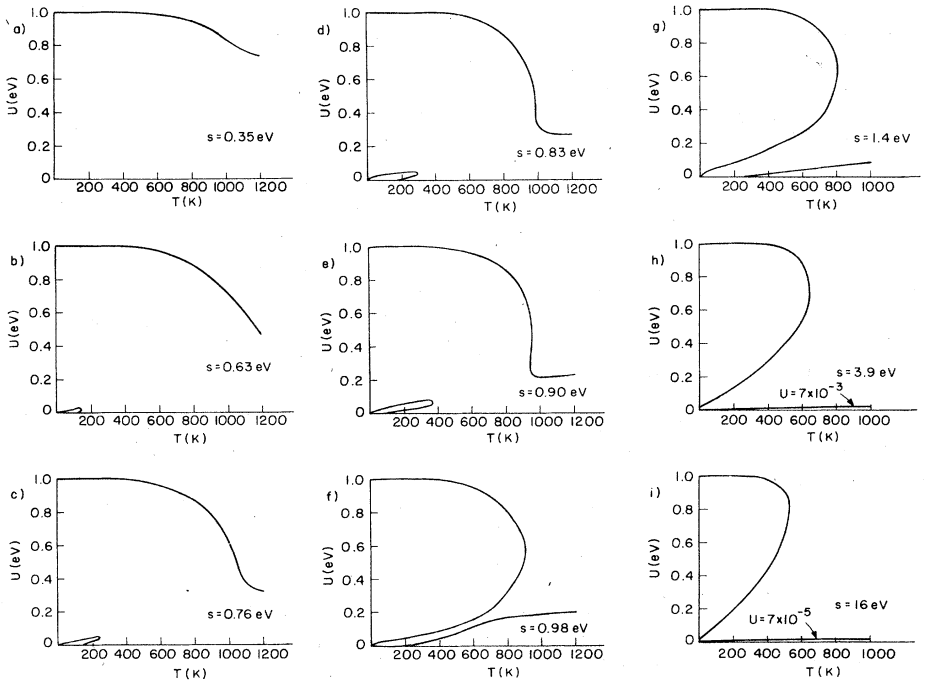

FIG. 8. Temperature dependences of the correlation energy for $U_0=1$ eV and the same values of s as in Fig. 7.

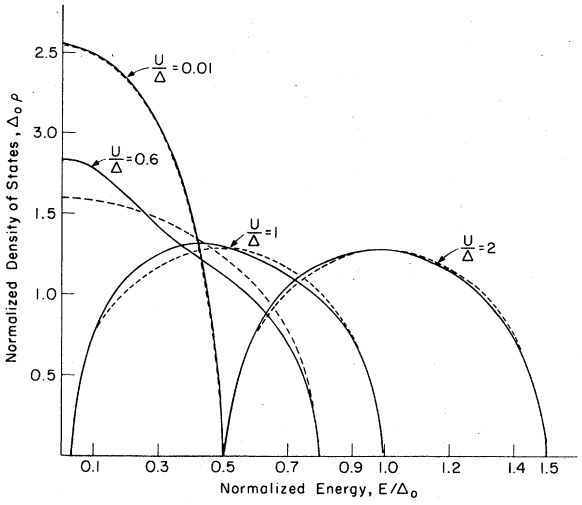

FIG. 9. Density of states for $E>0$ calculated using the method of Hubbard for Δ fixed at $\Delta_0=1$ and for several values of U/Δ on both sides of the transition. The dashed lines represent the approximations discussed in the text.

well represented by the density

$$p(E)=\begin{cases}\frac{4}{\pi\Delta}\left[1-\left(\frac{E\pm\frac{1}{2}U}{\frac{1}{2}\Delta}\right)^2\right]^{1/2}, & \left|E\pm\frac{U}{2}\right|\leq\tfrac{1}{2}\Delta\\ 0, & \left|E\pm\frac{U}{2}\right|>\tfrac{1}{2}\Delta,\end{cases} \tag{36}$$

which is indicated in the figure by the dashed line. Even for values of U/Δ as small as 1, Eq. (36) is a good approximation to the calculated curves. In any event, we shall be concerned with integrals over these densities of a slowly varying probability function so that the result will not be sensitive to the precise form of the pseudoparticle spectrum. Using Eq. (36), f becomes

$$f=\frac{4}{\pi\Delta}\int_{(U-\Delta)/2}^{(U+\Delta)/2}dE\,\frac{\left[1-\left(\frac{E-\frac{1}{2}U}{\frac{1}{2}\Delta}\right)^2\right]^{1/2}}{1+e^{\beta(E-U/2)}}, \tag{37}$$

so that f is a function of T, U, and Δ. At the same time, we can invert Eq. (18) to obtain

$$f=(1/s\beta)\ln^2(U/U_0)\ . \tag{38}$$

In this expression, f depends on T, U, U_0, and s. If we plot Eqs. (37) and (38) as functions of U, the point at which they intersect determines the self-consistent solution $\bar{f}(T, U, U_0, \Delta, s)$. Fortunately Eq. (37) does not involve s, so we can solve the pair of equations for various values of s by changing only Eq. (38); this involves shifting the entire curve Eq. (38) vertically by an amount equal to

$$\delta\log f=-\log(s/s_0), \tag{39}$$

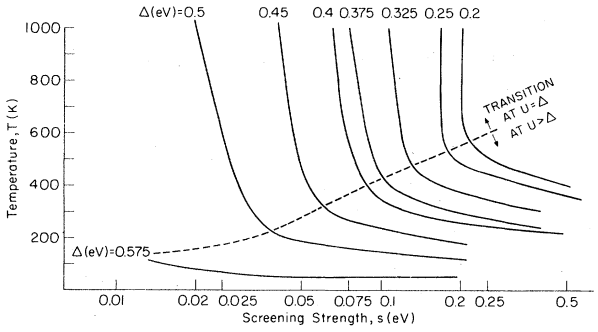

FIG. 10. Temperature above which the insulating state ceases to exist as a function of screening strength s for $U_0=0.625$ eV. Below (above) the dashed line the transition occurs at $U>(=)\,\Delta$.

where s_0 is the reference value used for the original plot.

An important feature of the pair of equations (37) and (38) is that, given T, U_0, and Δ, there is a critical value of s, s_{max} above which the equations have no solution. s_{max} is obtained from the curves using the formula $s_{max}=s_0 10^{V_{max}}$, where V_{max} is the maximum vertical separation $\delta\log f$ between the $\log f$ plots. For example, if $T=300$ K, $U_0=0.6$ eV, and $\Delta=0.25$ eV, we find $V_{max}=0.43$, so that $s_{max}=1.68$ eV. For s larger than s_{max}, there is no self-consistent insulating state at this temperature.

Looking at this result in another way, we can say that, for a screening strength s, the insulating state ceases to exist at $T=T_{max}$. At this temperature or below, the system will undergo a transition to an accessible, self-consistent metallic state, which has $U<\Delta$. We shall discuss this metallic state in Sec. VI.

In Fig. 10, we plot T_{max} as a function of s. When we are in the region above the dashed line in this figure, the maximum vertical separation V_{max} between the $\log f$ plots occurs at a value of U equal to Δ. As a result, T_{max} is the temperature at which the bands first touch. This does *not* imply that the dashed line marks a discontinuous change in the character of this transition, which is always abrupt. Below this line, screening by carriers in the upper Hubbard band causes the insulating gap to collapse from an initial value of U that is greater than Δ. Above the dashed line, the gap first shrinks continuously to zero. The density of states at the Fermi level rises very rapidly as U decreases through the transition value. Consequently, screening by the large number of carriers in the vicinity of ϵ_F facilitates the abrupt transition to the final metallic state.

The behavior of the gap in the insulator is shown in Fig. 11, where we plot U as a function of T

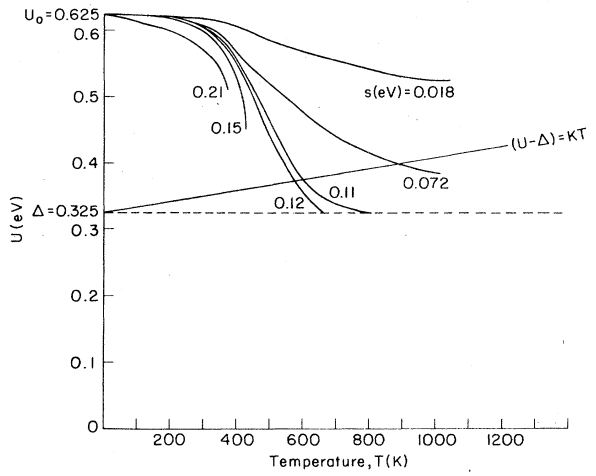

FIG. 11. Temperature dependence of the correlation energy in the insulator for $\Delta = 0.325$ eV and several values of s. The diagonal line indicates where the size of the gap equals kT.

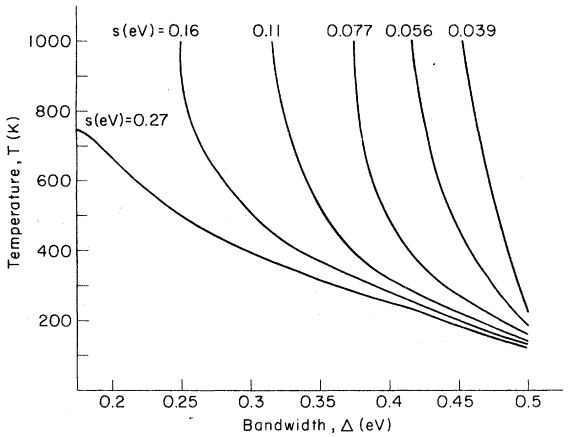

FIG. 12. Regions of temperature and bandwidth where insulating behavior can occur for $U_0 = 0.625$ eV and several values of s.

for various values of screening strength, with the parameters $U_0 = 0.625$ eV and $\Delta = 0.325$ eV. We see that three different types of behavior occur. For weak screening ($s < 0.11$ eV), the gap collapses slowly with increasing T. It approaches a finite value at high temperatures. For larger values of s ($0.11 < s < 0.13$ eV), the gap closes more rapidly, falling continuously to zero. When the screening is very strong ($s > 0.13$ eV), the gap collapses discontinuously. Typical values of the fraction of excited carriers just below the collapse of the gap range from 10^{-5} to 10^{-3}.

The diagonal line in the figure indicates where the size of the gap is equal to kT. When the gap shrinks below this value, the precise form which we have used for the screening is no longer strictly valid. Nevertheless, we can see that much of the interesting behavior, including the abrupt gap collapse, occurs well before this limit is reached. In addition, we expect that the error introduced by the use of our screening expression all the way to the point of overlap is not too great. In general, the amount of screening will be underestimated by such a procedure. This problem is offset by the advantage gained from the ability to treat the problem analytically.

Using Fig. 10, we can easily determine for each s the ranges of T and Δ for which the self-consistent insulating state can exist. The results are shown in Fig. 12. Below the curve for each value of s, insulating states are possible. Unless the screening is very strong, there is a minimum bandwidth corresponding to each value of s for which insulating behavior can occur at all temperatures. For $s \geq 0.26$ eV, a transition occurs even when $\Delta = 0$.

Although the ratio of U to Δ is the critical factor in determining the state of the system, it is important to remember that the variation of U/Δ occurs via a change in U with Δ held at a fixed value. The more familiar case is that treated by Hubbard,[10] in which U is a constant and Δ is varied. The choice of *particular* values for the ratio Δ/U and either U or Δ will, of course, be sufficient to determine most quantities of interest. Nevertheless, the *behavior* of these quantities (e.g., the density of states at the Fermi level and the position of the band edges), as a function of U/Δ, as well as the *physically realizable range* of U/Δ, depends very much on which of the two parameters is varied.

VI. CORRELATED METALLIC STATE

Before we can properly discuss the insulator-to-metal transition, we must also make the nature of the metallic state precise. In a manner similar to that used for the insulator, we can approximate the metallic density of states by

$$\rho(E) = \begin{cases} \dfrac{8}{\pi(\Delta+U)}\left[1-\left(\dfrac{E}{\frac{1}{2}(\Delta+U)}\right)^2\right]^{1/2}, & |E| \leq \tfrac{1}{2}(\Delta+U) \\ 0, & |E| > \tfrac{1}{2}(\Delta+U), \end{cases} \tag{40}$$

as shown in Fig. 9. The overall agreement between the approximations and the actual band shapes is very good, except for a small range of U/Δ just to the metallic side of the transition. As we shall see, this region of U/Δ is excluded for most of the abrupt transitions anyhow. It is important to understand that the value of the unscreened correlation energy U_0 *must be the same*

for both the insulator and the metal. These states are alternative solutions to the same screening problem with the physical parameters U_0, Δ, and s. The value of U must again be determined self-consistently. Because the bands have merged, the fraction of carriers in the upper Hubbard band is no longer a meaningful quantity. As mentioned in Sec. II, Rice and Brinkman,[25] using the variational method of Gutzwiller,[26,27] found that the screening in a strongly correlated metal at $T=0$ increases with the number of doubly occupied sites. Instead, we adopt an approach where the effect of correlations is incorporated into the metallic density of states using the approximation of Hubbard.[10] We find that the density of electrons in the vicinity of the Fermi level determines the extent to which U is screened.

To calculate an expression for U, we again use Eq. (3), where ϵ_F is determined by the equation

$$n=\int_{-(\Delta+U)/2}^{\epsilon_F}\frac{8n_0}{\pi(\Delta+U)}\left[1-\left(\frac{E}{\frac{1}{2}(\Delta+U)}\right)^2\right]^{1/2}dE\,, \tag{41}$$

provided $\Delta > kT$. For very small $\Delta \leq kT$, we must replace this equation by

$$n=\int_{-(\Delta+U)/2}^{(\Delta+U)/2}\frac{8n_0}{\pi(\Delta+U)}\times\left(\left[1-\left(\frac{E}{\frac{1}{2}(\Delta+U)}\right)^2\right]^{1/2}\Big/\{1+\exp[(E-\epsilon_F)/kT]\}\right)dE\,. \tag{42}$$

Integrating Eq. (41) and taking the derivative with respect to ϵ_F, we find

$$\frac{dn}{d\epsilon_F}=\frac{8n_0}{\pi(\Delta+U)}\left[1-\left(\frac{4\epsilon_F{}^2}{(\Delta+U)^2}\right)\right]^{1/2}, \tag{43}$$

which, for an average of one electron per site, is

$$\frac{dn}{d\epsilon_F}=\frac{8n_0}{\pi(\Delta+U)}\,. \tag{44}$$

Consequently,

$$\epsilon=1+\frac{32n_0q^2}{\tilde{q}^2(\Delta+U)}\,,$$

so that

$$U=U_0\exp\left[-\left(\frac{32n_0q^2d^2}{\Delta+U}\right)^{1/2}\right]$$

$$=U_0\exp\left[-1.6\left(\frac{s}{\Delta+U}\right)^{1/2}\right]. \tag{45}$$

When the effective bandwidth $\Delta+U$ decreases, the density of electrons near ϵ_F rises. This increases the strength of the screening. (Alternatively, we can say that as U decreases, the increase in the number of doubly occupied sites increases the screening strength.) This behavior is similar to that of Thomas-Fermi screening for which the screening parameter is proportional to the square root of the effective mass. We have the additional requirement that Δ is greater than U, i.e.,

$$\Delta > U_0\exp\left[-1.6\left(\frac{s}{\Delta+U}\right)^{1/2}\right]. \tag{46}$$

This condition is barely met for s equal to

$$s_{\min}=0.78\Delta\ln^2\frac{U_0}{\Delta}\,. \tag{47}$$

For values of s greater than $s_{\min}$, a self-consistent metallic state is possible. We can see from Eq. (45) that the screening strength decreases with increasing Δ. However, there is a concomitant decrease in the energy gap so that as the bandwidth increases, the value of $s_{\min}$ decreases. Because we treat the volume as a state variable, the screening parameter s, which depends only on the physical parameters n_0 and d, is fixed at essentially the same value for both the insulating and the metallic states.

Let us consider the range of parameters for which the insulating and metallic states can both exist. Recall from Fig. 12 that, for Δ less than a certain value, we find self-consistent solutions for the insulating state at all temperatures. At the same time Eq. (47) indicates that, for a given value of s, a minimum bandwidth is required to support a metallic state. Figure 13 shows that the critical value of Δ is the *same* in both cases. For the example in which $U_0=0.625$ eV and $s=0.16$ eV, this bandwidth is $\Delta=0.25$ eV.

As a result, the Δ-T plane is divided into three regions for each value of s: Below a critical bandwidth, metallic states cannot exist; above a cer-

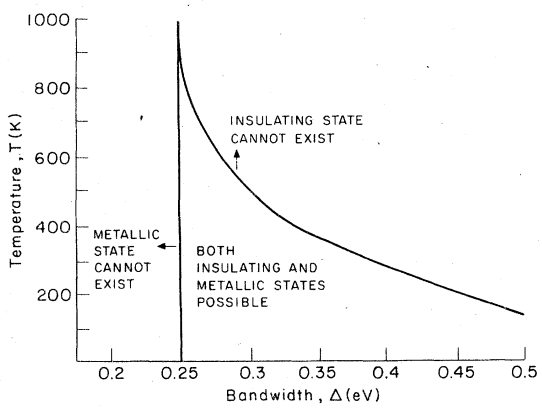

FIG. 13. Allowed range of temperature and bandwidth for the insulating and the metallic states for $U_0=0.625$ eV and $s=0.16$ eV.

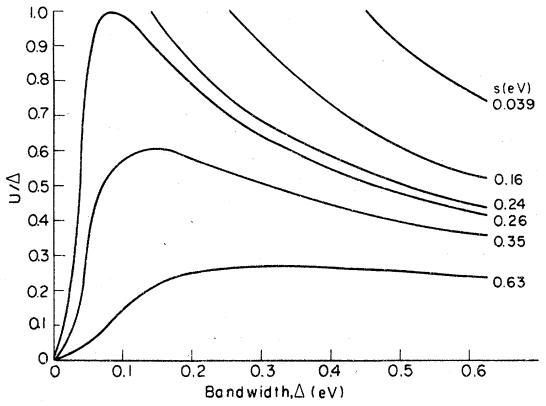

FIG. 14. Ratio U/Δ in the strongly correlated metal as a function of Δ for several screening strengths.

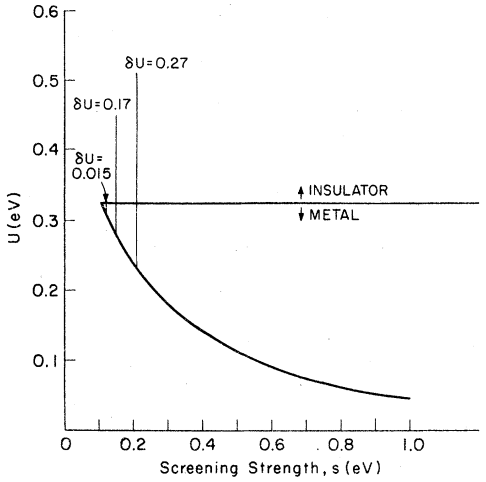

FIG. 15. Screened value of the correlation energy in the metal as a function of s, for $U_0 = 0.625$ eV and $\Delta = 0.325$ eV. The discontinuity in $U(\delta U)$ at the transition from the insulating to the metallic state is shown.

tain temperature (which decreases with increasing bandwidth), insulating states are not possible; in the remaining region of the plane, both states may exist. In order to determine which state actually occurs, we must calculate and compare their free energies. This is done in Sec. VII.

Let us first look in more detail at the metallic state. It is interesting to examine the values of U which satisfy Eq. (45). In Fig. 14 we plot the ratio U/Δ as a function of Δ. In the metallic state, U/Δ is less than one. We see from the figure that for $s \leq 0.26$ eV, metallic states occur when the bandwidth is greater than a certain minimum value. For large enough values of Δ, U/Δ eventually decreases, independent of the screening strength. When $s \geq 0.26$ eV, Eq. (45) has a metallic solution for all bandwidths. For these values of s, U decreases at small Δ as a consequence of the larger density of states at ϵ_F, which enhances the screening.

In Fig. 15, we plot U as a function of s for $U_0 = 0.625$ eV and $\Delta = 0.325$ eV. We can obtain directly from this curve the values of U that correspond to the final states of the transitions that were indicated in Fig. 11. The discontinuities in U at these transitions are shown. We should emphasize that even in the metallic state, the electrons are strongly correlated.

VII. PHASE DIAGRAMS

In the regime where kT is sufficiently larger than Δ^2/U that the system has no long-range magnetic order but is too low to excite a significant density of electrons, the free energy per electron in the insulating state can be written[34]

$$F_{\text{ins}} \approx -kT \ln 2 - \frac{4}{\sqrt{\pi}} \frac{\exp[-(U-\Delta/2)/2kT]}{\left(\frac{\Delta}{2kT}\right)^{3/2}} . \quad (48)$$

The electronic free energy of the metal can be calculated using the grand partition function and the density of states [Eq. (40)]. It can be expressed as:

$$F_{\text{met}} = \left(\frac{3\pi-4}{6\pi}\right)U - \frac{2\Delta}{3\pi} - \frac{8\pi(kT)^2}{3(\Delta+U)}$$

$$= (0.29U - 0.21\Delta) - \frac{8.4(kT)^2}{\Delta+U} . \quad (49)$$

We can easily interpret this result. The first term $(0.29U - 0.21\Delta)$ is the average energy which is obtained from the integral $\int \rho(E) E\, dE$. The second term is $(-TS)$, where the entropy S reflects the spin degree of freedom of the electrons within approximately $2kT$ of the Fermi energy. The product

$$k \ln 2 \times \rho(\epsilon_F) \times 4kT$$

$$= \frac{32k^2T}{\pi(\Delta+U)} \ln 2 \approx \frac{7k^2T}{(\Delta+U)} ,$$

which is very near our value for S, $S = 8.4k^2T/(\Delta+U)$. Chao and Berggren[35] have extended Gutzwiller's variational method[26,27] to calculate the free energy F_c of strongly correlated metallic electrons. They found that

$$F_c(T) = F(T)[1 - U/U_0(T)]^2 , \quad (50)$$

where $F(T)$ is the free energy per electron in the uncorrelated system and the critical correlation energy U_0 is $-8F(T)$, which is always non-negative. The function $F(T)$ is given by Eq. (49) with U set equal to zero. That is,

$$F(T) = -\frac{2\Delta}{3\pi} - \frac{8\pi}{3\Delta}(kT)^2 .$$

It then follows that

$$F_c = F\left(1+\frac{U}{8F}\right)^2$$

$$= F\left(1+\frac{U}{4F}+\frac{U^2}{64F^2}\right) = F+\frac{U}{4}+\frac{U^2}{64F} \ .$$

To first order in U/Δ, the correction to the free energy due to the correlations is just $\frac{1}{4}U$, which is precisely the increase we would expect from a Hartree-Fock treatment of the electronic interactions. The free energy is then

$$F_c = 0.25U - 0.21\Delta - 8.4(kT)^2/\Delta \ ,$$

which is very close to Eq. (49) for small values of U/Δ. Both our result and the calculations of Chao and Berggren[35] are unreliable in the immediate vicinity of the transition. However, as pointed out previously, we have chosen to use our method because we can derive *both* the insulating and metallic densities of states from one model that is continuous across the transition.

We can now combine the results we have developed so far in this paper in order to construct complete phase diagrams illustrating the effects of screening on Hubbard insulators. These are shown in Figs. 16–18 for $U_0 = 0.625$ eV and several different screening strengths s. The low-temperature metal-insulator boundaries are a consequence of the free-energy minimization discussed above and correspond qualitatively to those obtained by Kaplan and Bari[36] using the thermal single-determinant variational approximation. For several lattice geometries, Kaplan and Bari have also obtained transitions, which occur with increasing T, back to the delocalized state.[36] However, because U is taken to be constant in their calculations, there is no immediate correspondence between their high-T transitions and the screening-induced ones presented here. Figure 16 shows the regions of insulating and metallic behavior for the screening parameter $s = 0.16$ eV. For small values of Δ ($\Delta < 0.25$ eV), the gap decreases as the temperature is raised. However, because the gap [$\sim(U-\Delta)$] is so large, few carriers are excited across it. As a result, the strength of the screening is insufficient to cause a transition to the metallic state, and the system is insulating at all temperatures. For larger Δ ($0.25 < \Delta < 0.4$ eV), the gap closes more sharply with increasing T. At a critical temperature, which decreases as Δ increases, the screening becomes strong enough to cause the insulating state to vanish. The value of U undergoes an abrupt change at the transition. In general, the energy gap will be finite just before the transition. In the final state, U is determined self-consistently by taking into account the screening due to the metallic electrons in the vicinity of the Fermi level.

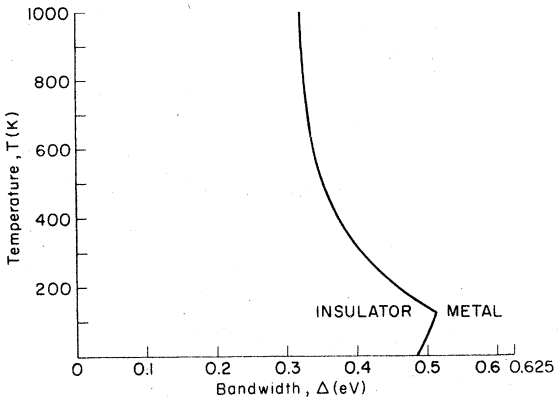

FIG. 17. Regions of insulating and metallic behavior as a function of temperature and bandwidth for $U_0 = 0.625$ eV and $s = 0.11$ eV.

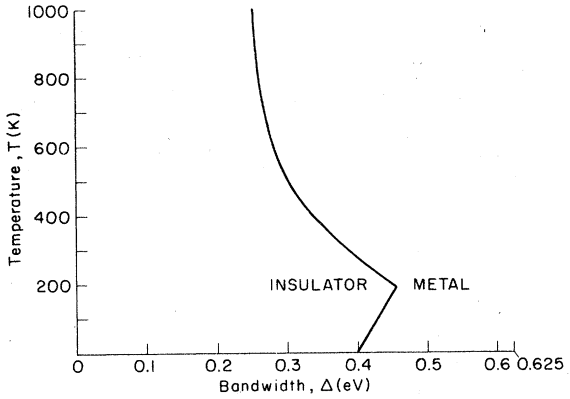

FIG. 16. Regions of insulating and metallic behavior as a function of temperature and bandwidth for $U_0 = 0.625$ eV and $s = 0.16$ eV.

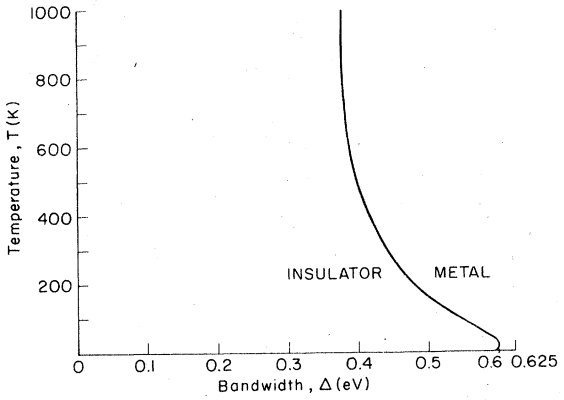

FIG. 18. Regions of insulating and metallic behavior as a function of temperature and bandwidth for $U_0 = 0.625$ eV and $s = 0.077$ eV.

For still larger bandwidths ($0.40 < \Delta < 0.45$ eV), the system is metallic at low T, insulating at intermediate T, and again metallic at high T. The low-temperature transition can be understood in terms of the electronic free energies [Eqs. (48) and (49)]. The low-temperature metallic state is a consequence of energy reduction due to electronic delocalization in the wide band. However, at higher temperatures, the large entropy in the insulator results in a metal-to-insulator transition. Finally, we have the high-temperature disappearance of the insulating state caused by screening. When $\Delta > 0.45$ eV, the system is metallic at all temperatures.

Figures 17 and 18 demonstrate the similar behavior that occurs for two other screening strengths $s = 0.11$ eV and $s = 0.077$ eV. We see that, as s is decreased, the range of insulating behavior expands, as expected. For most values of bandwidth, the gap decreases slowly with temperature but never vanishes. On the other hand, when s is increased much beyond its value in Fig. 16, i.e., $s = 0.16$ eV, an insulating gap cannot appear at any temperature and the system is always metallic.

In discussing these phase diagrams, we have traced the behavior of the system, as the temperature is raised, at a fixed value of Δ. We can equally well specify T and look at the effect of an increase in bandwidth (due, e.g., to a decrease in lattice constant or to an increase in the overlap of electronic wave functions brought about by pressure or compositional changes). It is clear from the figures that the system will eventually become metallic at any temperature if the bandwidth is increased to a sufficiently large value.

We have not extended our analysis to the temperature regime in which long-range magnetic order is present. The Hubbard insulator is known[37] to be antiferromagnetic below the Néel temperature $T_N \sim \Delta^2/U$. However, in our model, U is greater than Δ^2/U and thus the system remains insulating above T_N. In this case there is a wide range of temperatures for which we clearly have a nonmagnetic insulator. This nonordered insulating phase is the one considered in the present calculations.

As we discussed previously, it is often the case that there are excess electrons or holes present at low temperatures. These carriers, which do not require thermal activation, then can result in metallic behavior which suppresses the low-temperature insulating phase ordinarily occurring at small bandwidths. Such is the case for the system $Ni_{1-x}Co_xS_2$, which we analyze in detail in Sec. VIII.

Another subject which must be addressed is the order of the transitions shown in the phase diagrams. The low-temperature transition, driven by the spin entropy of the electrons in the insulator, is unambiguously of first order. The free energies of the two states intersect with different slopes at the transition. The gap closes abruptly, and the number of carriers increases discontinuously.

We reemphasize that we have calculated *electronic* free energies only, with the assumption that the energies of the lattice and the electron-lattice interaction do not vary much with the electronic state. Consequently, the plots of free energy versus volume which we have calculated should be used only to compare the relative energies of electronic states, and provide no information about the equilibrium volume of the system.

The order of the higher-temperature insulator to metal transition is more difficult to determine. In many ways this problem is similar to that of the pressure-induced Mott transition, which has been studied for nearly three decades but for which the order of the transition is still a subject of controversy.[38-41] The difficulty is due partially to the fact that the precise form which we have calculated for the screening parameter breaks down in the immediate vicinity of the transition. A more exact treatment would necessitate extending the general screening theory to the limit of strong coupling, a problem which has not yet been solved.[23] Nevertheless, it is very likely that the transition is of first order and is accompanied by an abrupt collapse of the insulating energy gap. In fact, it could well be the case that the actual transition occurs at a temperature slightly below that at which the self-consistent insulating solution ceases to exist. In either event, we can say unequivocably that the number of carriers will increase dramatically at the transition. This change should be reflected in the transport behavior of these systems.

In Sec. VIII, we discuss the application of these ideas to $Ni_{1-x}Co_xS_2$, for which transitions occur as a function of both temperature and cobalt concentration.[42] Although our screening analysis dealt only with nondegenerate s states, it is reasonable to expect that the qualitative behavior we have described carries over to the case of degenerate d bands,[9,43] especially since much of the degeneracy is lifted by the crystal field.

VIII. APPLICATION TO $Ni_{1-x}Co_xS_2$

Previous studies of the $Ni_{1-x}Co_xS_2$ system suggests that NiS_2 is a Mott insulator.[42,44,45] In collaboration with Mabatah, Eklund, and Dresselhaus, we have shown in an earlier paper that fea-

tures of the transitions observed in this system can be quantitatively explained by a model which involves the collapse of the semiconducting gap with increasing temperature.[17] Here, we briefly review the pertinent experimental results. We suggest that the reason for the gap collapse can now be understood on the basis of the screening model presented in this paper.

The experimental phase diagram obtained from the data of Mabatah and co-workers at MIT[17,18] is shown in Fig. 19. For temperatures above the solid line, the data suggest that conduction by a large number of electrons is taking place in *one* collapsed band. In fact, Kautz *et al.*[31] extrapolated their optical absorption data on nominally pure NiS_2 and found that the gap should close in the vicinity of 720 K. Unfortunately, the material decomposes at ~675 K, so there is no direct experimental test of this extrapolation. In the region between the solid and curved dashed lines, an insulating gap exists and the number of carriers is activated. Finally, there is a third, low-temperature region below the curved dashed line in which conduction occurs within the lower band. The addition of Co (as well as of impurities and nonstoichiometry) introduces holes into the valence band which dominate the conduction until the temperature is high enough for there to be an appreciable number of excited carriers.

In addition, we find that the mobility activation energy decreases with increasing Co and disappears at a Co concentration of ~13%. (This is indicated by the vertical dashed line.) Note, therefore, that the band collapse, which occurs with increasing temperature, is *not* the same as the insulator-to-metal transition caused by adding Co. This latter transition is due to a change in the mobility mechanism as a result of the increase in bandwidth.

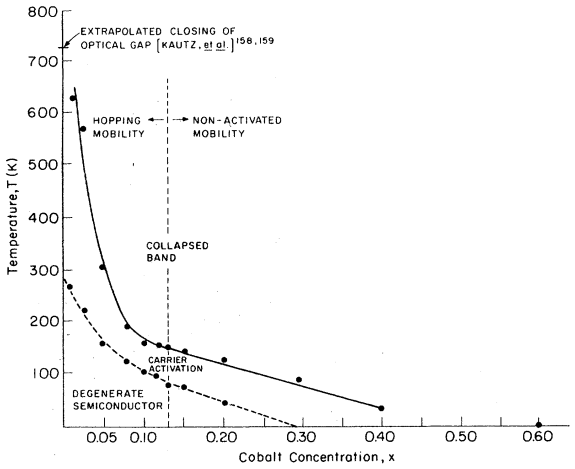

FIG. 19. Experimental diagram indicating regions of transport behavior for the system $Ni_{1-x}Co_xS_2$.

We suggest that the screening mechanism we have discussed is responsible for the collapse of the gap. First, it is likely that NiS_2 is an insulator by virtue of the interelectronic correlations. Studies have shown that $Ni_{1-x}Co_xS_2$ is an antiferromagnetic insulator at low temperatures.[46-49] The Néel temperature T_N ranges from ~40 K at $x=0$ to ~150 K at $x=0.1$ and then decreases for larger x.[45,49] An important feature of this material is that it remains insulating for $T > T_N$. Consequently, the insulating gap does not originate solely from the antiferromagnetic ordering. This is precisely the situation to which our model should apply.

There is direct evidence from the optical-absorption measurements that the semiconducting gap decreases as the temperature is raised. In addition, both conductivity and thermoelectric-power measurements suggest that the number of carriers is no longer activated at high T, and that conduction takes place in just one band.

The mobility mechanism is apparently unaffected by the band-collapse transition, which manifests itself instead as a change in carrier concentration. This is consistent with the assumptions of our model.

Finally, the model predicts that charged impurities or excess electrons or holes should lower the temperature at which the energy gap collapses. The introduction of Co into the lattice creates such screening charges and increases the bandwidth as well. We see from Fig. 19 that the transition temperature does decrease with increasing Co concentration.

It is important to note that the transition in this case is not characterized by sharp discontinuities in the transport properties. We can therefore infer that the bandwidths for these systems lie near, but below, the minimum width required for a first-order transition. This is just what we would expect for narrow-band materials. Nevertheless, the strength of the screening of the correlation energy is sufficient to close the gap, which is $E_g \simeq U - \Delta$. This collapse of the correlation-induced gap is driven by the screening interaction and occurs at a critical temperature, as shown in Fig. 19.

ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation under NSF-MRL Grant No. DMR-76-80895.

APPENDIX

In this appendix, we demonstrate, for the particular case $U \gg \Delta$, but with $\Delta > 0$, that Eq. (13) is a general result that holds for correlated systems when the Boltzmann limit applies.

For a general two-band system,

$$n = n_l + n_u = \int_{-\infty}^{\infty} dE \frac{g_l(E) + g_u(E)}{1 + e^{\beta(E-\epsilon_F)}}, \tag{A1}$$

where n_l and n_u are the electron densities, and g_l and g_u are the densities of states (per unit volume) in the lower and upper pseudoparticle bands. Taking the derivative of n with respect to ϵ_F, we obtain

$$\frac{dn}{d\epsilon_F} = \int_{-\infty}^{\infty} dE \frac{\beta[g_l(E) + g_u(E)]}{(1 + e^{\beta(E-\epsilon_F)})(1 + e^{-\beta(E-\epsilon_F)})} + \frac{dn}{d\epsilon_F} \int_{-\infty}^{\infty} dE \frac{(\partial g_l(E)/\partial n) + (\partial g_u(E)/\partial n)}{1 + e^{\beta(E-\epsilon_F)}} . \tag{A2}$$

Rearranging terms, this expression becomes

$$\frac{dn}{d\epsilon_F} = \int_{-\infty}^{\infty} dE \frac{\beta[g_l(E) + g_u(E)]}{(1 + e^{\beta(E-\epsilon_F)})(1 + e^{-\beta(E-\epsilon_F)})} \Bigg/ \left(1 - \int_{-\infty}^{\infty} dE \frac{(\partial g_l(E)/\partial n) + (\partial g_u(E)/\partial n)}{1 + e^{\beta(E-\epsilon_F)}}\right). \tag{A3}$$

We examine a system which has a density of states for uncorrelated electrons given by

$$g(E) = (2n_0/\Delta)\Theta(E - T_0 + \tfrac{1}{2}\Delta)\Theta(T_0 + \tfrac{1}{2}\Delta - E).$$

Hubbard[8] considered this case and found

$$g_l(E) = (2n_0/\Delta)\Theta(E - E_{-1,-1})\Theta(E_{-1,1} - E), \tag{A4a}$$

$$g_u(E) = (2n_0/\Delta)\Theta(E - E_{1,-1})\Theta(E_{1,1} - E), \tag{A4b}$$

where

$$E_{\alpha,\beta} = T_0 + \frac{U}{2} + \frac{\beta\Delta}{4} + \alpha\left[\left(\frac{U}{2} - \frac{\beta\Delta}{4}\right)^2 + \frac{\beta\Delta U n}{4n_0}\right]^{1/2}. \tag{A5}$$

In order to illustrate the range of validity of Eq. (13) analytically, we use Hubbard's first solution to the Hamiltonian (3) (with $J \equiv 0$).[8] This solution is invalid near the overlap transition $\Delta \simeq U$ and below the Néel temperature $T_N \sim \Delta^2/U$ so that, in addition to the Boltzmann requirement $kT \ll (U - \Delta/2)$, we have the restrictions $U \gg \Delta$ and $T > \Delta^2/U$. However, we expect that a solution within the Boltzmann limit, but additionally valid for $U \simeq \Delta$ and $kT \leq \Delta^2/U$, would also satisfy Eq. (13). Before setting $n = n_0$, we must find the derivatives $\partial g/\partial n$. We obtain

$$\frac{\partial g_l(E)}{\partial n} = -\frac{U}{4}\left(\frac{U^2}{4} + \frac{\Delta^2}{16}\right)^{-1/2}[\delta(E - E_{-1,-1}) + \delta(E - E_{-1,1})] \tag{A6a}$$

and

$$\frac{\partial g_u(E)}{\partial n} = \frac{U}{4}\left(\frac{U^2}{4} + \frac{\Delta^2}{16}\right)^{-1/2}[\delta(E - E_{1,-1}) + \delta(E - E_{1,1})]. \tag{A6b}$$

The form of these derivatives $\partial g(E)/\partial n$ is precisely what we would expect: as we fill rectangular bands, we add states symmetrically to both edges of the upper band and subtract them in the same fashion from the lower one. We can now let $n = n_0$. Solving Eq. (A1) for ϵ_F, we find $\epsilon_F = T_0 + \frac{1}{2}U$, just as in the atomic limit. Substituting this result and Eq. (A6) into Eq. (A3) and taking the Boltzmann limit $e^{\beta(U-\Delta/2)/2} \gg 1$, we obtain

$$\frac{\partial n}{\partial \epsilon_F} = \frac{8n_0}{\Delta} \exp[-\beta(\tfrac{1}{4}U^2 + \tfrac{1}{16}\Delta^2)^{1/2}] \sinh\tfrac{1}{4}\beta\Delta / [1 + \tfrac{1}{2}U(\tfrac{1}{4}U^2 + \tfrac{1}{16}\Delta^2)^{-1/2}] \simeq \beta n_0 e^{-\beta U/2} \frac{\sinh\frac{1}{4}\beta\Delta}{\frac{1}{4}\beta\Delta}. \tag{A7}$$

In a similar manner, we can evaluate n_u in the Boltzmann limit. We find

$$n_u = n_0 e^{-\beta U/2} \frac{\sinh\frac{1}{4}\beta\Delta}{\frac{1}{4}\beta\Delta}. \tag{A8}$$

Therefore $dn/d\epsilon_F = \beta n_u$, so that the screened interaction is again given by Eq. (13). Thus we have shown that this result is not only a zero-bandwidth approximation, but it also holds for finite bandwidths whenever the Boltzmann limit is valid. As does the conventional Debye-Huckel screening formula, the screened value of U depends on the *total* number of conduction electrons, independent of the density of states in the band or the details of the level occupation.

*Present address: IBM T. J. Watson Research Center, Yorktown Heights, N.Y., 10598.

[1] N. F. Mott, *Metal-Insulator Transitions* (Taylor and Francis, London, 1974).

[2] D. Adler, Solid State Phys. 21, 1 (1978).

[3] J. A. Wilson and A. D. Yoffe, Adv. Phys. 18, 193 (1969).

[4] M. R. Oliver, J. O. Dimmock, A. L. McWhorter, and

T. B. Reed, Phys. Rev. B 5, 1078 (1972); S. D. Bader, N. E. Phillips, and D. B. McWhan, *ibid.* 7, 4686 (1973).
[5]A. J. Epstein, S. Etemad, A. F. Garito, and A. J. Heeger, Phys. Rev. B 5, 952 (1972); R. H. Friend, M. Miljak, and D. Jerome, Phys. Rev. Lett. 40, 1048 (1978); J. B. Torrance and B. D. Silverman, Phys. Rev. B 15, 778 (1977).
[6]N. F. Mott, Philos. Mag. 17, 1259 (1968).
[7]N. F. Mott, Philos. Mag. 6, 287 (1961).
[8]J. Hubbard, Proc. R. Soc. A 276, 238 (1963).
[9]J. Hubbard, Proc. R. Soc. A 277, 237 (1964).
[10]J. Hubbard, Proc. R. Soc. A 281, 401 (1964).
[11]L. M. Falicov and J. C. Kimball, Phys. Rev. Lett. 22, 997 (1969).
[12]R. Ramirez, L. M. Falicov, and J. C. Kimball, Phys. Rev. B 2, 3383 (1970).
[13]H. Frohlich, in *Quantum Theory of Atoms, Molecules, and the Solid State*, edited by P. O. Löwdin (Academic, New York, 1966), p. 465.
[14]G. J. Hyland, Philos. Mag. 20, 837 (1969).
[15]Z. Zinamon and N. F. Mott, Philos. Mag. 21, 881 (1970).
[16]S. Doniach, Adv. Phys. 18, 819 (1969).
[17]A. K. Mabatah, E. J. Yoffa, P. C. Eklund, M. S. Dresselhaus, and D. Adler, Phys. Rev. Lett. 39, 494 (1977).
[18]E. J. Yoffa, A. K. Mabatah, P. C. Eklund, D. Adler, and M. S. Dresselhaus, in *Proceedings of the 14th International Conference on the Physics of Semiconductors*, edited by B. L. H. Wilson (Institute of Physics, Bristol, 1979), p. 473.
[19]G. Kemeny, Ann. Phys. 32, 69 (1965).
[20]T. A. Kaplan and P. N. Argyres, Phys. Rev. B 1, 2457 (1970).
[21]For a general correlated system, the chemical potential μ and Fermi energy ϵ_F are not equal (see Ref. 22). However, for the zero-bandwidth case discussed here, $\epsilon_F = \mu$.
[22]E. J. Yoffa, W. A. Rodrigues, Jr., and D. Adler, Phys. Rev. B 19, 1203 (1979).
[23]See, e.g., W. A. Harrison, *Solid State Theory* (McGraw-Hill, New York, 1970).
[24]E. J. Yoffa and D. Adler, Phys. Rev. B 12, 2260 (1975).
[25]T. M. Rice and W. F. Brinkman, in *Critical Phenomena in Alloys, Magnets, and Superconductors*, edited by R. E. Mills, E. Ascher, and R. I. Jaffee (McGraw-Hill, New York, 1971), p. 593.
[26]M. C. Gutzwiller, Phys. Rev. Lett. 10, 159 (1963).
[27]M. C. Gutzwiller, Phys. Rev. A 137, 1726 (1965).
[28]P. N. Argyres, T. A. Kaplan, and N. P. Silva, Phys. Rev. A 9, 1716 (1974).
[29]I. G. Austin and N. F. Mott, Adv. Phys. 18, 41 (1969).
[30]A. J. Bosman and H. J. VanDaal, Adv. Phys. 19, 1 (1970); D. Adler and J. Feinleib, Phys. Rev. B 2, 3112 (1970).
[31]R. L. Kautz, M. S. Dresselhaus, D. Adler, and A. Linz, Phys. Rev. B 6, 2078 (1972); R. L. Kautz, M. S. Dresselhaus, D. Adler, and A. Linz, in *Proceedings of the 11th International Conference on the Physics of Semiconductors*, edited by M. Miasek (Polish Scientific Publishers, Warsaw, 1972), p. 807.
[32]D. Adler and H. Brooks, Phys. Rev. 155, 826 (1967).
[33]N. F. Mott, Rev. Mod. Phys. 40, 677 (1968).
[34]E. J. Yoffa, Ph.D. thesis, MIT, 1978 (unpublished).
[35]K. A. Chao and K.-F. Berggren, Phys. Rev. B 15, 1656 (1977).
[36]T. A. Kaplan and R. A. Bari, in *Proceedings of the 10th International Conference on the Physics of Semiconductors*, edited by S. P. Keller, J. C. Hensel, and F. Stern (National Technical Information Service, 1970), p. 301.
[37]D. I. Khomskii, Fiz. Met. Metalloved. 29, 31 (1970).
[38]N. F. Mott, Proc. Phys. Soc. London A 62, 416 (1949).
[39]N. F. Mott, Philos. Mag. 6, 287 (1961).
[40]W. Kohn, Phys. Rev. Lett. 19, 789 (1967)..
[41]B. I. Halperin and T. M. Rice, Rev. Mod. Phys. 40, 755 (1968).
[42]J. A Wilson and G. D. Pitt, Philos. Mag. 23, 1297 (1971).
[43]J. Hubbard, Proc. R. Soc. A 296, 82 (1966).
[44]H. S. Jarrett, W. H. Cloud, R. J. Bouchard, S. R. Butler, C. G. Frederick, and J. L. Gillson, Phys. Rev. Lett. 21, 617 (1968).
[45]S. Ogawa, S. Waki, and T. Teranishi, Int. J. Magn. 5, 349 (1974).
[46]J. M. Hastings and L. M. Corliss, IBM J. Res. Dev. 14, 227 (1970).
[47]F. Gautier, G. Krill, M. F. Lapierre, and C. Robert, Solid State Commun. 11, 1201 (1972).
[48]T. Miyadai, K. Takizawa, H. Nagata, H. Ito, S. Miyahara, and K. Hirakawa, J. Phys. Soc. Jpn. 38, 115 (1975).
[49]J. M. D. Coey, Physica B 91, 59 (1977).