

## Excited states and the metal-insulator transition in monovalent systems

R. N. Bhatt and Subir Sachdev

*AT&T Bell Laboratories, Murray Hill, New Jersey 07974*

(Received 12 June 1986)

We have used Gutzwiller's variational method to study the metal-insulator transition in a modified Hubbard model with low-lying excited states and one electron per site. Comparison is made with results obtained using conventional band screening arguments. Implications for many-valley semiconductors with and without central-cell effects are discussed.

Uncompensated  $n$ -doped semiconductors have been one of the important family of disordered systems in which the metal-insulator transition has been extensively studied.<sup>1</sup> While it is now established that the transition is continuous in a large variety of disordered systems, the nature of the conductivity variation in the metallic phase and, in fact, the conductivity onset exponent  $\nu[\sigma(n) \sim (n - n_c)^\nu]$  has been found to be rather different, varying from 0.5–0.7 for uncompensated doped semiconductors<sup>2,3</sup> to  $1.0 \pm 0.1$  for amorphous metal-semiconductor alloys,<sup>4</sup> magnetic semiconductors,<sup>5</sup> as well as heavily compensated doped semiconductors.<sup>6</sup> Furthermore, though the microscopic characterization of the electronic states of the doped semiconductor system is the most complete of the systems studied (see, e.g., Ref. 1), the nature of the states near the metal-insulator transition is not clear. Weak-localization corrections far on the metallic side are evaluated<sup>7</sup> assuming the full degeneracy of the conduction-band minima in many-valley semiconductors. On the other hand, the variation of the metal-insulator transition density  $n_c$  in phosphorus-doped silicon (Si:P) with uniaxial stress,<sup>2</sup> is explained only if one assumes a single-nondegenerate-band model,<sup>8</sup> rather than a degenerate band of effective-mass donor states.<sup>9</sup> In fact, the agreement using a single band derived from the (stress-dependent) isolated donor ground state in Si:P, neglecting all the higher states, turns out to be embarrassingly quantitative,<sup>10</sup> while the result for effective mass donors has the opposite sign.<sup>9</sup> This is despite the fact that the degeneracy splitting of the  $1s$  ground state at zero stress, due to the short-range control cell potential ( $\approx 12$  meV), is much less than the impurity bandwidth ( $\approx 45$  meV).

In this paper we address this issue by applying Gutzwiller's variational technique<sup>11</sup> to a modified Hubbard model<sup>12</sup> with one electron per site, each site having a nondegenerate (except for spin) ground state and a  $f$ -fold degenerate excited state. As shown by Brinkman and Rice,<sup>13</sup> the "half-filled band" Hubbard model (without excited states) in the Gutzwiller variational approximation yields a metal-insulator transition as the on-site repulsion  $U$  is increased, from a metal at  $U=0$  to an insulator with one electron per site (zero double occupancy) for  $U > U_c$ . Since Gutzwiller's variational technique basically keeps track of correlation effects only through the average number of double occupied sites (which is the variational parameter in the many-electron wave function), we would expect a similar metal-insulator transition in our model.

The parameters at the critical point, however, depend on the energy and degeneracy of the excited states. In addition, we find that the nature of the metallic phase can differ depending on the parameters in a way which is markedly different from what is expected using simple band theory and metallic screening arguments, as done originally by Mott<sup>14</sup> and elaborated by others.<sup>15</sup> Our results not only explain the success of Bhatt's single band calculation,<sup>8</sup> but also suggest that one should be rather careful when extracting exponents for the conductivity onset in uncompensated, doped many-valley semiconductors, from samples not too close to  $n_c$ .

Our model Hamiltonian can be written as

$$H = H_0 + H_1, \quad (1)$$

where

$$H_0 = \sum_{\substack{\langle ij \rangle \\ \sigma}} t_{ij} a_{i\sigma}^\dagger a_{j\sigma} + \sum_{\substack{\langle ij \rangle \\ \sigma, \lambda}} t_{ij} b_{i\sigma\lambda}^\dagger b_{j\sigma\lambda} + \epsilon_0 \sum_{i\sigma\lambda} b_{i\sigma\lambda}^\dagger b_{i\sigma\lambda} \quad (2)$$

is the noninteracting two-band Hamiltonian with creation and annihilation operators  $a^\dagger/a$  and  $b^\dagger/b$ , spin index  $\sigma = \pm \frac{1}{2}$ , and a further degeneracy (e.g., valley) index  $\lambda = 1$  to  $f$  in the case of the excited state, which is an energy  $\epsilon_0$  above the ground state.  $t_{ij} = t$  for nearest neighbors and 0 otherwise is the hopping matrix elements.  $H_1$  is a purely on-site electron-electron interaction, which we take to be the simple form

$$H_1 = \begin{cases} 0 & \text{for one electron on a site,} \\ U & \text{for two electrons on a site of any kind,} \\ \infty & \text{for more than two electrons on a site.} \end{cases} \quad (3)$$

Several nonessential features of the doped semiconductor system have been ignored. The hopping matrix element  $t$  is taken the same for both bands. Interband hopping elements are ignored. Such terms are expected to be unimportant as long as the  $t$  for both bands are not widely separated. Inclusion of such terms will greatly increase the complexity of the calculation. Also neglected are the small intrasite exchange terms.

Gutzwiller's variational scheme starts with the ground-state wave functions of the noninteracting Hamiltonian  $H_0$ , and transforming to a site representation, projects out terms which involve double occupancy in terms of a variational parameter. We follow the approach used by

Vollhardt.<sup>16</sup> Considering only the nonmagnetic case, and assuming all the  $f$  upper orbitals are equivalent, we may write

$$n_a + n_b = \frac{1}{2}, \quad (4)$$

where  $n_a/n_b$  are the number of  $a/b$  electrons with spin up or down. Assuming the number of doubly occupied  $a/b$  sites is  $d_a/d_b$  and the number of sites with one  $a$  and one  $b$  electron is  $g$ , giving weights  $\alpha$ ,  $\beta$ , and  $\gamma$  to these three possible cases of double occupancy, we find from the dominant term condition that

$$\alpha^2 = \frac{ed_a}{x_a^2}, \quad \beta^2 = \frac{ed_b}{x_b^2}, \quad \gamma^2 = \frac{eg}{x_a x_b}, \quad (5)$$

$$q_a = \frac{x_a(\sqrt{e} + \sqrt{d_a})^2 + 4f\sqrt{gx_b}(\sqrt{ex_a} + f\sqrt{gx_b} + \sqrt{d_a x_a})}{n_a(1 - n_a)}, \quad (8a)$$

$$q_b = \frac{x_b[\sqrt{e} + (2f - 1)\sqrt{d_b}]^2 + 4\sqrt{gx_a}[\sqrt{ex_b} + \sqrt{gx_a} + (2f - 1)\sqrt{d_b x_b}]}{n_b(1 - n_b)}. \quad (8b)$$

Analytic results could only be obtained for the case  $\epsilon_0 = 0$  when we can put  $d_a = d_b = g$  and  $n_b = fn_a$ . We now minimize Eq. (7) by varying  $d_a$ . For the case  $f = 1$  we get a metal-insulator transition at

$$U_c = -13.20 \left[ 2 \sum_{k < k_a} E_k^q \right].$$

This differs from the more complete calculation of Chao and Gutzwiller<sup>17</sup> which included interband hopping by about 1%. For  $f \rightarrow \infty$ ,

$$U_c = -11.66 \left[ (f + 1) \sum_{k < k_a} E_k^q \right],$$

indicating the insensitivity of  $U_c$  to the value of  $f$ .

Minimizing Eq. (7) with respect to  $d_a$ ,  $d_b$ , and  $g$ , we find three possible phases, depending on the parameters  $\epsilon_0/U$  and  $t/U$ : (1) the insulating phase with  $d_a = d_b = g = n_b = 0$  and  $n_a = \frac{1}{2}$ , (2) a one-band metal  $M_1$  with  $d_b = g = n_b = 0$  and  $n_a = \frac{1}{2}$ , but  $d_a \neq 0$ , and (3) a two-band metal  $M_2$  where all occupation numbers are nonzero.

Figure 1 summarizes the phase diagram in the  $\epsilon_0/U - t/U$  plane for the cases (a)  $f = 1$  and (b)  $f = 5$  (which would be relevant to Si:P, with five valley-orbit-split excited  $1s$  states).<sup>18</sup> Results for  $f \rightarrow \infty$  are not very different from the latter. As can be seen, for a given  $\epsilon_0/U$ , as  $t/U$  is increased (which corresponds to increasing the donor concentration in the experimental system), one has, for  $\epsilon_0/U$  not too small, a transition from an insulating phase to a one-band metallic phase, which is continuous in that the number of double occupied increases continuously from zero. It is only at a larger value of  $t/U$  that there is a (first-order) transition from a one-band to a multiband metal in this model. The noteworthy result is that one gets a transition to a multiband metal directly, only for abnormally small values of  $\epsilon_0/U < 0.04$ . This implies that the variation of the metal-insulator transition density for most donors with even moderately small central cell effects

where

$$\begin{aligned} x_a &\equiv n_a - d_a - 2fg, \\ x_b &\equiv n_b/f - (2f - 1)d_b - 2g, \\ e &\equiv 1 - 2n_a - 2n_b + d_a + f(2f - 1)d_b + 4fg. \end{aligned} \quad (6)$$

The total energy of the system is then found to be

$$\begin{aligned} E &= 2q_a \left[ \sum_{k < k_a} E_k^q \right] + 2fq_b \left[ \sum_{k < k_b} E_k^q \right] \\ &+ U(d_a + f(2f - 1)d_b + 4fg), \end{aligned} \quad (7)$$

where  $E_k^{a/b}$  are the band energies for the two tight-binding bands, and the Gutzwiller parameters for the Fermi surface discontinuity are given by

( $\epsilon_0/U < 0.5$ ) can be described in terms of a single nondegenerate ground-state band, and explains why such a calculation neglecting the excited-state bands was so successful in Si:P. Another interesting result is the relatively small variation (less than a factor of 2) in the ratio  $t/U$  at the transition for all values of  $f$  and  $\epsilon_0/U$ . The hopping integral is very sensitive to the donor separation ( $t \sim e^{-r/a}$ ); consequently, the small variation in  $t/U$  would explain the near universality of the Mott number ( $n_c^{1/3}a$ ) observed<sup>19</sup> (on a 10% level). This is in contrast to results obtained using simple metallic screening arguments,<sup>15</sup> following Mott's original argument.

These results will now be compared with the results of a

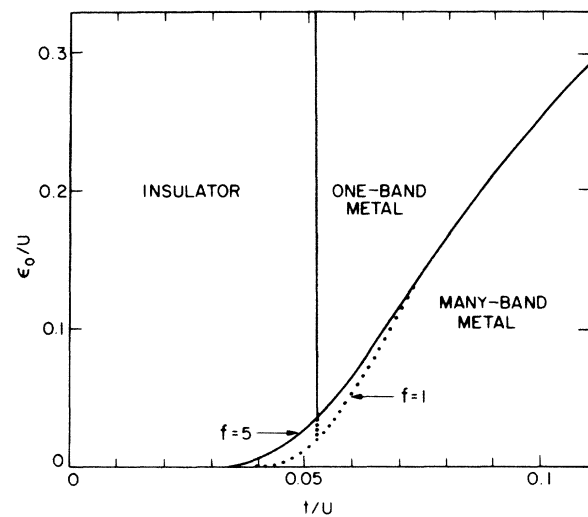


FIG. 1. Phase diagram of model Hamiltonian (1), using a Gutzwiller-type wave function. The parameter  $t$  is the overlap of the wave function of nearest neighbors,  $\epsilon_0$  is the splitting between the two states on a site,  $U$  is the Hubbard repulsion, and  $f$  is the degeneracy of the upper level.

simple band picture of the metal along with Mott's screening argument for the metal-insulator transition. We have constructed a phase diagram for a two-band model with parabolic (effective mass) bands. For a nondegenerate ground state and an  $f$ -fold degenerate excited state separated by  $\epsilon_0$ , we obtain a Thomas-Fermi screening wave vector:

$$K = 2 \left( \frac{k_F}{\pi a_B} \right)^{1/2} \left[ 1 + f \left( 1 - \frac{\delta}{k_F^2 a_B^2} \right)^{1/2} \theta(k_F^2 a_B^2 - \delta) \right]^{1/2}, \quad (9)$$

where  $k_F$  is the Fermi wave vector in the ground-state band,  $\theta$  is the Heaviside step function,  $a_B$  the effective-mass Bohr radius  $a_B = \epsilon \hbar^2 / m^* e^2$ , and  $\delta$  is  $\epsilon_0$  divided by the effective-mass Rydberg  $E_R = me^4 / 2\epsilon_2 \hbar^2$  ( $\epsilon$  is the host dielectric constant). The critical density at which the Thomas-Fermi potential

$$V_{TF} = - \frac{e^{-Kr}}{\epsilon r} \quad (10)$$

binds an electron is given by<sup>20</sup>

$$K = 1.19/a_B. \quad (11)$$

The critical value of  $x = k_F a_B$  is therefore obtained by substituting Eq. (11) in Eq. (9), and the number obtained from

$$n_c^{1/3} a_B = x [1 + f(1 - \delta/x^2)^{3/2} \theta(x^2 - \delta)]^{1/3}. \quad (12)$$

The resulting phase diagram that is obtained is shown in Fig. 2 for  $f=1$  and  $f=5$ . The vertical axis is comparable to Fig. 1, since  $U \approx 0.95$  Ry for hydrogenic systems,<sup>21</sup> while the correspondence to the horizontal axis in Fig. 1 is monotonic but not linear. Despite this, the phase diagram is clearly very different—the critical density is much more sensitive to the presence and degeneracy of the excited state, and starts to shift from the single band value as soon as the excited state comes within an energy of the order of the Rydberg, as one might have easily guessed. (The difference in the shift is in fact underplayed by comparing Figs. 1 and 2 directly, since in Fig. 1 the horizontal axis is exponential in the impurity separation, so even a small shift in  $n_c^{1/3}$  would register a big change. In Fig. 2, on the other hand, the scale is linear.) In addition, unlike Fig. 1, the transition in Fig. 2 is for the most part directly from the many-band metal to the insulator. These differences in the phase diagrams serve to illustrate the crucial role of electron correlations in determining both the metal-insulator transition density, as well as the nature of the metallic phase. This role is captured in the Gutzwiller technique, which is therefore able to explain both the success of the one-band calculation in Si:P as well as the observed constancy of the Mott number.

We expect that in real many-valley doped semiconduct-

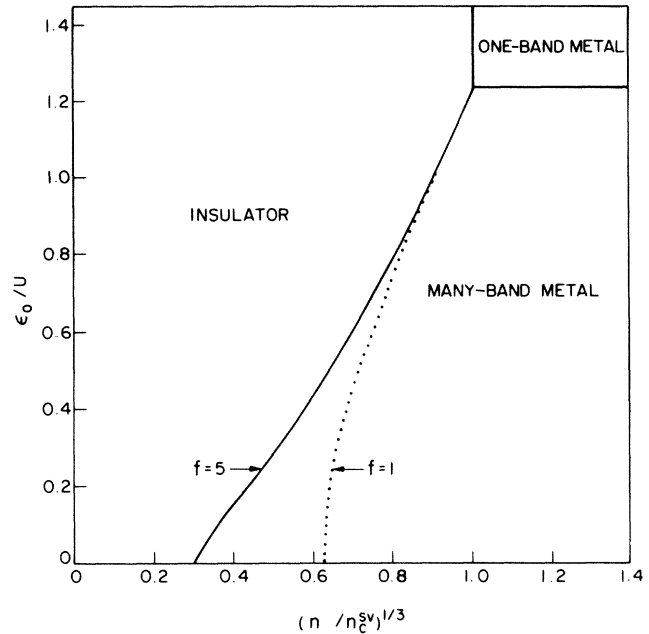


FIG. 2. Phase diagram of a doped many-valley semiconductor using the Mott screening argument. The constants  $\epsilon_0$  is the valley splitting.  $U$  and  $f$  have the same meaning as in Fig. 1. The density of electrons in  $n$  and  $n_c^{sv}$  is the critical density for the single-valley case.

ors the insulator multiband metal transition will not be suppressed quite as much as Fig. 1 suggests. This is in part because the (symmetric) ground-state wave function is sucked in and has therefore a larger  $U$  and smaller  $t$  than the excited states, and exchange terms would favor multiband occupancy also. In addition, the hybridization terms will tend to keep the bands apart; this could be particularly important if the bands cross,<sup>22</sup> or have nonparallel dispersion.

A final point of relevance is the single-many-band metal boundary in Fig. 1. In the real system there is disorder, which our model lacks; consequently, we would not really expect a true phase change. However, a crossover behavior would probably persist, and if the boundary is close to the metal-insulator transition, then this would render determinations of conductivity onset exponents inaccurate. From Fig. 1, using hydrogenic overlap integrals, we would conclude that one is relatively safe from this effect if determining results from samples with  $n < 2n_c$  in Si:P and Si:As. However, shifts due to effects contained in the previous paragraph would make the crossover closer, and it would therefore be prudent to use a criterion more stringent than the one given above.

<sup>1</sup>For a review, see R. F. Milligan, T. F. Rosenbaum, R. N. Bhatt, and G. A. Thomas, in *Electron-Electron Interactions in Disordered Systems*, edited by A. L. Efros and M. Pollak (North-Holland, Amsterdam, 1985), p. 231.

<sup>2</sup>M. A. Paalanen, T. F. Rosenbaum, G. A. Thomas, and R. N. Bhatt, *Phys. Rev. Lett.* **48**, 1284 (1982); T. F. Rosenbaum, K. Andres, K. A. Thomas, and R. N. Bhatt, *ibid.* **45**, 1723 (1980).

- <sup>3</sup>P. F. Newman and D. F. Holcomb, *Phys. Rev. Lett.* **51**, 2144 (1983); W. N. Schaffarman, T. G. Castner, J. S. Brooks, K. P. Martin, and M. J. Naughton, *ibid.* **56**, 980 (1986).
- <sup>4</sup>B. Dodson, W. L. McMillan, J. Mochel, and R. C. Dynes, *Phys. Rev. Lett.* **46**, 46 (1981); N. Nishida, M. Yamagnchi, T. Furubayashi, K. Morigaki, H. Ishimoto, and K. Ono, *Solid State Commun.* **44**, 305 (1982); G. Hertel, D. J. Bishop, E. G. Spencer, J. M. Rowell, and R. C. Dynes, *Phys. Rev. Lett.* **50**, 743 (1983).
- <sup>5</sup>S. von Molnar, A. Briggs, J. Floquet, and G. Remenyi, *Phys. Rev. Lett.* **51**, 706 (1983); F. Holtzberg, S. von Molnar, and J. Stankiewicz, *Bull. Am. Phys. Soc.* **31**, 679 (1986).
- <sup>6</sup>G. A. Thomas, Y. Ootuka, S. Katsumoto, S. Kobayashi, and W. Sasaki, *Phys. Rev. B* **25**, 4288 (1982).
- <sup>7</sup>T. F. Rosenbaum, K. Andres, G. A. Thomas, and P. A. Lee, *Phys. Rev. Lett.* **46**, 568 (1981); R. N. Bhatt and P. A. Lee, *Solid State Commun.* **48**, 755 (1983).
- <sup>8</sup>R. N. Bhatt, *Phys. Rev. B* **26**, 1082 (1982).
- <sup>9</sup>R. N. Bhatt, *Phys. Rev. B* **24**, 3630 (1981).
- <sup>10</sup>An unpublished analysis by R. N. Bhatt and M. A. Paalanen suggests an agreement of the full (nonlinear) variation with stress much better than that claimed in Refs. 2 and 8.
- <sup>11</sup>M. C. Gutzwiller, *Phys. Rev.* **134**, A927 (1964); **137**, A1726 (1965).
- <sup>12</sup>J. Hubbard, *Proc. R. Soc. London, Ser. A* **276**, 238 (1963).
- <sup>13</sup>W. F. Brinkman and T. M. Rice, *Phys. Rev. B* **2**, 4302 (1970).
- <sup>14</sup>N. F. Mott, *Proc. Phys. Soc. London* **62**, 416 (1949).
- <sup>15</sup>J. B. Krieger and M. Nightingale *Phys. Rev. B* **4**, 1266 (1971); F. Martino, G. Lindell, and K. F. Berggren, *ibid.* **8**, 6030 (1973); R. L. Greene, C. Aldrich, and K. K. Bajaj, *ibid.* **15**, 2217 (1977).
- <sup>16</sup>D. Vollhardt, *Rev. Mod. Phys.* **56**, 99 (1984).
- <sup>17</sup>K. A. Chao and M. C. Gutzwiller, *J. Appl. Phys.* **42**, 1420 (1971); K. A. Chao, *Phys. Rev. B* **4**, 4034 (1971).
- <sup>18</sup>We neglect the small splitting between the triplet and doublet excited states.
- <sup>19</sup>P. P. Edwards and M. J. Sienko, *Phys. Rev. B* **17**, 2575 (1978).
- <sup>20</sup>See, e.g., J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (Wiley, New York, 1952).
- <sup>21</sup>Since the  $H^-$  ion is bound only by 0.05 Ry,  $U = 1.0 - 0.05 = 0.95$  Ry.
- <sup>22</sup>C. M. Varma, W. Weber, and L. J. Randall, *Phys. Rev. B* **33**, 1105 (1986).