

## Density-functional theory of the metal-insulator transition

J. H. Rose

*Physics Department, University of Michigan, Ann Arbor, Michigan 48109*

H. B. Shore

*Physics Department, San Diego State University, San Diego, California 92182*

L. M. Sander

*Physics Department, University of Michigan, Ann Arbor, Michigan 48109*

(Received 20 March 1979)

The spin-density-functional theory is used to investigate the phase transitions of electrons in a lattice of fixed point charges (a model first studied by Mott). We find a first-order metal-insulator transition at  $r_s = 2.84$  and a second-order magnetic transition at  $r_s = 2.74$ . Results are presented for the magnetization and the spin susceptibility. Extrapolating our results for the case of donor impurities in silicon (neglecting disorder), we estimate that the first-order metal-insulator transition would not be observable above  $T_c \approx 50$  mK.

### I. INTRODUCTION

The metal-insulator transition is a subtle and interesting subject. Different qualitative phenomena can cause electrons to localize. For example, Anderson<sup>1</sup> has shown that noninteracting electrons will become localized in an environment which is in a certain sense sufficiently random. On the other hand, in a perfectly ordered solid, electrons should localize when their interaction energy becomes sufficiently larger than their kinetic energy. There are several phenomena which are thought to be associated with the metal-insulator transition including distortion of the lattice leading to a crystallographic change of phase.<sup>2</sup> Further, the ground state of the solid often becomes magnetically ordered either near or at the metal-insulator transition.<sup>2</sup>

A general solution embracing all of these features is beyond grasp at the present time. However, there is a simplified problem suggested by Mott<sup>3</sup> some time ago as an approach to the metal-insulator transition. Consider an array of hydrogen atoms at zero temperature fixed rigidly on a lattice. What is the electronic ground state of the system as a function of the separation of the atoms? Some features of Mott's problem are fairly clear. At very high density the system will be a paramagnetic metal, while at low densities it will be an insulator and probably antiferromagnetic. Between these two limits there must be a transition from metallic to insulating behavior, and also from paramagnetic to antiferromagnetic behavior. The statement of the problem rules out both the effects of randomness and crystallographic phase changes.

Beyond these rather general features, the na-

ture of the electronic ground state has not been satisfactorily elucidated. In particular the number of phase transitions and their nature has not been determined, although Mott has argued on quite general grounds that the metal-insulator transition must be first order. Further, Mott has suggested that the transition from metallic to insulating and from paramagnetic to spin-ordered behavior might occur at the same density (i.e., there would be only a single transition).<sup>2</sup> On the other hand, it has been known for some time that the most widely studied model of Mott's problem (the Hubbard model<sup>4</sup>) may exhibit two separate phase transitions in contradiction to Mott's suggestion.<sup>5</sup> However, the results of the Hubbard approach cannot be considered conclusive for the Mott problem, primarily because the Hubbard Hamiltonian treats the long-range electrostatic interaction as a short-range phenomenon. A proper treatment of the electrostatic force might well change the order and number of the phase changes.

We note that the fixed nuclear background of Mott's original problem reduces the problem to that of determining the ground-state properties of a collection of electrons moving in a fixed external potential; that is, the entire problem is determined by the properties of the inhomogeneous interacting electron gas. Such problems are very difficult and not yet subject to exact solutions. However, the spin-density-functional formalism<sup>6,7</sup> provides a practical method for approximately determining the ground state of many such systems. Basically, density-functional theory<sup>8,9</sup> is a Hartree-type method which also includes the effects of exchange and correlation in the self-consistent field. Good results are generally obtained both for the properties of isolated atoms

(including hydrogen) and for the properties of bulk metals.<sup>10,11</sup> Further density-functional theory has been used to study the Wigner transition in the low-density electron gas with reasonable results.<sup>12</sup> The method yields reliable answers not only for the ground-state energy but also for the ground-state spin densities.

Ghazali and Leroux-Hugon<sup>13</sup> have recently used the density-functional theory to make an estimate of the critical density of the metal-insulator transition for the Mott problem. They used a rather restricted approach which implicitly assumed that there was only a single transition in agreement with Mott's speculation. Further, their approach did not allow them to determine the order of the transition. In this paper we report a similar but more general calculation which does allow us to examine in detail the number and order of the phase transitions. We find in this work that there is a first-order metal-insulator phase transition whose critical density satisfies  $a_0 n_c^{1/3} = 0.22$ , where  $a_0$  is the Bohr radius. Our microscopic calculation confirms and illustrates Mott's qualitative arguments for the first-order nature of the transition. However, in agreement with the calculations based on the Hubbard Hamiltonian we find a separate second-order phase transition associated with the onset of spin ordering. It has a critical density about 10% greater than the metal-insulator transition. Our picture is this, as we go from high-density solid to the low-density solid: first, we have a paramagnetic metal, then a spin-ordered metal, and finally a spin-ordered insulator.

The work to be reported has three major restrictions which will be discussed in more detail in Sec. V. The first restriction derives from the fact that the true magnetically ordered ground state near the metal-insulator transition is unknown but probably antiferromagnetic. For ease of computation we have assumed here that the spin-ordered state has ferromagnetic symmetry. The second restriction is the use of the local approximation for exchange and correlation energy. Finally, we have ignored structural differences between different possible lattices and have used a Wigner-Seitz sphere in our calculations.

The structure of the paper is as follows. First we discuss some general features of the transition. Then we discuss the applications of the density-functional formalism to the Mott problem. The fourth section gives our results. The final section contains a discussion of our results and a summary.

## II. GENERAL FEATURES OF THE TRANSITIONS

We are primarily concerned with two features in this system: the conductivity  $\sigma$  and the mag-

netization  $M$ . We determine these quantities by minimizing an energy functional,  $E[n^\uparrow, n^\downarrow]$ , which is discussed in Sec. III. Minimizing  $E$  is equivalent to solving a fully self-consistent spin-dependent band-structure problem. As first pointed out by Slater,<sup>14</sup> the Coulomb repulsion that prevents two electrons from occupying the same site can be thought of in terms of nondegenerate spin-up and spin-down bands.

The band-structure approach gives a straightforward signal for the conductivity. The system is insulating when the lower band is completely filled and the upper band is completely empty. It is conducting when the two bands are each partially filled. Besides the metal-insulator transition, we find a second transition where a magnetic moment begins to form in the metallic phase. Here the order parameter is the magnetic moment in the unit cell, which is given directly in a spin band-structure calculation.

## III. CALCULATIONAL PROCEDURE

The total ground-state energy of a collection of  $N$  electrons can be written as a unique functional of the spin density.<sup>6,7</sup> Further, for the true ground-state spin density, the energy is a minimum. We write the energy as a functional of the spin-up and the spin-down densities  $n^\uparrow(r)$  and  $n^\downarrow(r)$ .

$$E[n^\uparrow, n^\downarrow] = T[n^\uparrow] + T[n^\downarrow] + \frac{1}{2} \int \frac{\rho(r)\rho(r')}{|\vec{r} - \vec{r}'|} d^3r d^3r' + \int V_{\text{ext}}(r)\rho(r)d^3r + E_{\text{xc}}[n^\uparrow, n^\downarrow]. \quad (1)$$

Here  $\rho(r)$  is the total electronic density, and  $T$  is the kinetic energy for a band of noninteracting electrons with density  $n^\uparrow$  or  $n^\downarrow$ . The third term on the right is the electrostatic energy, while the fourth term yields the potential energy of the electrons moving in the field of the protons. The final term  $E_{\text{xc}}$  is a functional of the densities  $n^\uparrow$  and  $n^\downarrow$  which represents the exchange and correlation energy. In solving a problem this is the only term which must be approximated since an exact expression is unknown. We use the local approximation, i.e.,

$$E_{\text{xc}} = \int d^3r \epsilon_{\text{xc}}(n^\uparrow(\vec{r}), n^\downarrow(\vec{r})). \quad (2)$$

Here  $\epsilon_{\text{xc}}$  is the exchange and correlation energy per unit volume for a uniform electron gas with spin densities  $n^\uparrow$  and  $n^\downarrow$ . We use the formulas of Gunnarson and Lundqvist<sup>7</sup> for the function  $\epsilon_{\text{xc}}$ .

The next step in solving our problem is to minimize the energy which leads to a spin-generalized version of the self-consistent equations given by Kohn and Sham<sup>6,7</sup> (we use atomic units throughout).

$$-\frac{1}{2} \nabla^2 \psi_k^\uparrow(\vec{r}) + V_{\text{eff}}^\uparrow(r; n(r)) \psi_k^\uparrow(\vec{r}) = \epsilon_k^\uparrow \psi_k^\uparrow(r), \quad (3)$$

$$n^\dagger(r) = \sum_k^{k_f} |\psi_k^\dagger(r)|^2, \quad (4)$$

and

$$V_{\text{eff}}^\dagger(\vec{r}) = -Z \sum_{\vec{R}_i} \frac{1}{|\vec{r} - \vec{R}_i|} + \int \frac{\rho(\vec{r}') d^3r'}{|\vec{r} - \vec{r}'|} + \frac{d}{dn^\dagger} \epsilon_{xc}(n^\dagger, n^\dagger). \quad (5)$$

Equivalent equations for the spin-down band are found by substituting  $\downarrow$  for  $\uparrow$  in all the equations above.  $\vec{R}_i$  denotes the lattice positions of the protons. Here  $\epsilon_k^\uparrow$ ,  $\psi_k^\uparrow$  are the eigenvalues and eigenvectors for the spin-up band; similar quantities  $\epsilon_k^\downarrow$ ,  $\psi_k^\downarrow$  are defined for the spin-down band. The symbol  $\vec{k}$  denotes the wave vector. The Fermi energies are given by  $\epsilon_{k_F}^\uparrow$  and  $\epsilon_{k_F}^\downarrow$ . These quantities are related by

$$N = \frac{1}{3\pi^2} (k_F^{\uparrow 3} + k_F^{\downarrow 3}). \quad (6)$$

Finally, the total energy may be evaluated using Eq. (1) and noting that

$$T^\dagger = \sum_k^{k_F} \epsilon_k^\dagger - \int n^\dagger(\vec{r}) V_{\text{eff}}^\dagger(\vec{r}) d^3r. \quad (7)$$

To solve Eqs. (3)–(5) one chooses an arbitrary trial potential  $v_{\text{eff}}$ . One then obtains via Eqs. (3) and (4) an estimate for the spin densities. These densities yield a new estimate for the effective potential via Eq. (5). This process is iterated to self-consistency.

Since the potential is periodic we need to solve only for  $\psi_k$  over the unit cell. Hence as we noted before the entire procedure is simply the determination of a self-consistent band structure. In order to reduce the numerical computation we follow a method suggested by Tong<sup>15</sup> and use a spherical approximation, replacing the true Wigner-Seitz cell by a Wigner-Seitz sphere about each proton. It is this approximation which restricts us to only paramagnetic and ferromagnetic ground states, since there is no simple spherical approximation for a unit cell containing more than one proton (e.g., the antiferromagnetic case).

In solving for the wave functions, we now apply periodic boundary conditions on the sphere. That is, we require

$$\psi_k(\vec{r}) = \psi_k(-\vec{r}) \exp(2ikr_s \cos\theta), \quad |\vec{r}| = r_s \quad (8)$$

and

$$\frac{\partial \psi_k(\vec{r})}{\partial r} = -\exp(2ikr_s \cos\theta) \frac{\partial \psi_k(-\vec{r})}{\partial r}, \quad |\vec{r}| = r_s. \quad (9)$$

Here  $r_s$  is the radius of the Wigner-Seitz sphere,  $k$  is the magnitude of the wave vector assumed

in the  $z$  direction, while  $\theta$  is the angle between  $\vec{k}$  and  $\vec{r}$ . The restriction  $|\vec{r}| = r_s$  ensures that both  $-\vec{r}$  and  $\vec{r}$  lie on the surface of the Wigner-Seitz sphere.

The wave functions  $\psi_k(\vec{r})$  are constructed in the following way. First, we solve the radial Schrödinger equation, using an effective potential which reflects the spherical symmetry of the unit cell at an energy  $E$ . We denote the solution of the radial wave equation as  $\phi_l(\vec{r}, E)$ . Assuming that  $\psi_k(\vec{r})$  can be expanded in terms of spherical harmonics,

$$\psi_k(\vec{r}) = \sum_l \eta_l c_l P_l(\cos\theta) \phi_l(\vec{r}, E). \quad (10)$$

Here  $\eta_l = 1$  if  $l$  is even, and  $\eta_l = i$  if  $l$  is odd. The expansion coefficients are real. If we substitute this expansion into the boundary condition and average  $r$  over the surface of the sphere, we obtain a secular equation (see Brooks<sup>16</sup>). For a given value of  $k$  the secular equation has solutions only for a discrete set of  $E$  (the energy bands).<sup>15,16</sup> Our procedure is thus to determine the  $E$  vs  $k$  relation for the lowest spin-up and spin-down band. We then choose the relative occupation of the two bands and calculate the density and the kinetic energy  $T_s$  from the wave functions.<sup>15</sup> Equation (1) leads to the total energy as a function of  $n^\dagger$  and  $n^\dagger$ . By studying this function one determines the location and nature of the phase transition and the magnetic susceptibility of the system.

#### IV. RESULTS

Our results can be qualitatively summarized as follows: At high densities there is a paramagnetic metal with the spin-up and the spin-down bands degenerate and equally occupied. As the density is decreased there is a second-order phase transition in which a magnetic moment forms spontaneously on each proton; i.e., one spin band lies lower in energy than the other and is occupied by more electrons. Finally, at even lower densities the system undergoes a first-order phase transition to a state in which the upper spin band is completely empty and the lower band is full, which we take to be equivalent to an insulating state.

To explicate the origin of these results, we show in Fig. 1 a plot of the total ground-state energy per unit cell at  $r_s \approx 2.84(4\pi r_s^3 = 1/n)$  as a function of the magnetic moment  $M$ . We fix the occupation of two spin bands (to constrain the net magnetization of the cell) and find the self-consistent solution subject to that constraint. The actual ground-state energy and magnetization for this value of  $r_s$  are then obtained from the absolute minimum

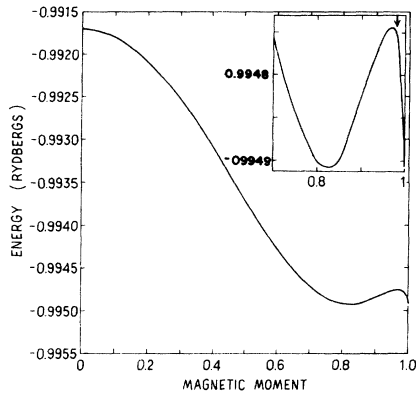


FIG. 1. The total energy at the first-order metal-insulator transition as a function of the magnetic moment in the unit cell at  $r_s = 2.84$ . The inset shows an enlargement of the double-well structure. The arrow indicates that value of  $M$  for which the bottom of the upper spin band crosses the top of the lower spin band.

energy in Fig. 1. We have chosen to present a plot for an  $r_s$  in the vicinity of the metal-insulator transition; in this density regime there are two approximately degenerate minima of the energy curve. For a slightly higher density (smaller  $r_s$ ) the minimum centered at  $M=0.82$  will be the preferred state, while at slightly lower density the completely spin-aligned solution will be the ground state. Hence, as  $r_s$  is varied through 2.84 there is a sharp transition in which  $M$  changes discontinuously from 0.82 to 1; that is, the ground state changes from that of a spin-polarized metal to an insulator.

The double-well structure in Fig. 1 is the signature of the first-order phase transition, and it arises in the following way. Still considering the case of  $r_s = 2.84$  we show the band occupation in Fig. 2(a). The upper band lies a small but finite energy  $\Delta E$  above the lower band if  $M=1$ , i.e.,

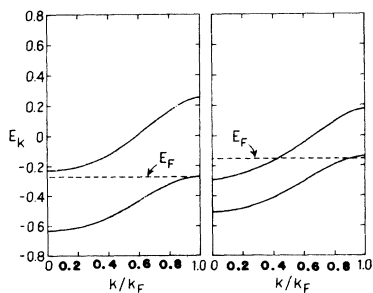


FIG. 2. The spin-split bands at the first-order phase transition. (a) shows the bands and the existence of an energy gap when the electrons are completely spin aligned ( $M=1$ ). (b) shows the overlapping bands which occur at the second energy minimum (see Fig. 1) for  $M=0.82$ .

for complete spin polarization of the system. Consider taking a few electrons out of the lower band and putting them in the upper band. We must pay energy  $\Delta E$  per electron as we transfer electrons, and this accounts for the increase in energy shown in Fig. 1 as we move away from the completely spin-aligned solution ( $M=1$ ). However, there is a second effect which causes this increase in energy to stop and leads to the formation of the second minimum at  $M=0.82$ . This effect results from the fact that the electrons and holes created by the transfer interact and reduce the energy gap between the two bands. In fact, this effect is sufficiently large so that, when we put enough electrons in the upper band, the bottom of the upper band crosses the Fermi surface of the lower band and energy is gained by transferring electrons to the upper band. This transfer continues until balanced by the loss of exchange and correlation energy. The result is the minimum at  $M=0.82$ . In Fig. 2(b) we show the band occupation at this minimum. The qualitative behavior of our model is described by Mott's original arguments for the first-order nature of the metal-insulator transition (basically his arguments concern the reduction of the band gap due to electron-hole interaction). We note that owing to the small size of the energy barrier separating the insulating state from the metallic state, the critical temperature for the first-order transition will be quite low, about 20 K.

The formation of the spin moment results in a second-order phase transition at  $r_{sc} = 2.74$ . In this regime there is no double-well structure and the energy minimum moves continuously out from  $M=0$ . The behavior of the magnetization is shown in Fig. 3. As expected from mean-field theory,  $M$  grows as the square root of the difference between the density and the critical density:  $M = 2.57 (r_s - r_{sc})^{1/2}$ . This behavior is obeyed to within a few percent up to the occurrence of the first-order transition. Similarly the differential magnetic susceptibility has a mean-field form:  $\chi = 4.59 / (r_{sc} - r_s)$  for  $r_s$  less than the critical value for  $r_{sc} = 2.74$ .

Several approximations were made in this study. We are attempting to investigate the nature of the phase transitions in a mean-field theory (since we require each cell to be equivalent and do not allow for fluctuations). As always fluctuations could change the picture drastically near the critical density. We have also disregarded effects of disorder and of lattice distortions, both of which could be important for transitions in real materials. Additionally we have required that the magnetic state have ferromagnetic symmetry. The true nature of the magnetic ground state in the

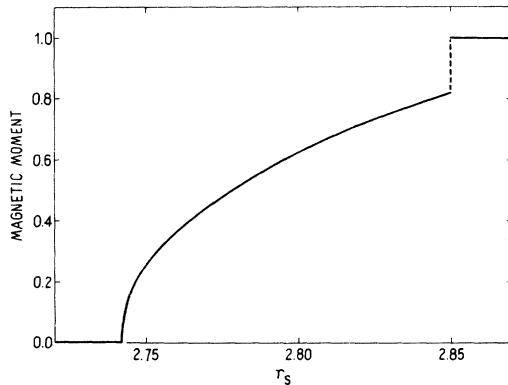


FIG. 3. The magnetic moment in the unit cell as a function of  $r_s$ . The behavior between  $r_s = 2.74$  and  $2.84$  is square-root-like. At  $r_s = 2.84$  there is a discontinuous change to the fully spin-aligned state.

neighborhood of the transitions is unknown. At much lower densities where tight binding is appropriate, the antiferromagnetic ground state is preferred. It is quite possible that the transition from a paramagnetic to an antiferromagnetic metal is preferred to the situation studied in this paper. In this case, rather than an energy splitting between the spin-up and spin-down bands, the onset of antiferromagnetic order is accompanied by a reduction in the size of the magnetic Brillouin zone. There would again be partially overlapping upper and lower bands, but the detailed picture would be somewhat different from that of the ferromagnetic case. We are currently extending our calculations to allow for these more complicated magnetic ground states. Preliminary studies suggest that the band-crossing effects which are the heart of the metal-insulator transition become much more complicated, and, as a consequence, the specific crystal geometry may become important in determining the nature of the transitions.

A final limiting feature of the calculation is the local-density approximation for the exchange and correlation. The structure of the transitions we studied depends on small differences in energy. A complete theory of the exchange-correlation energy might provide systematic changes in our results. Within the local approximation our results are insensitive to the particular choice of energy functional. For example, if we use the exchange-correlation energy of von Barth and Hedin<sup>6</sup> we find that the transitions are shifted to larger  $r_s$  by less than 3%. All other features of the calculation remain essentially the same.

The insensitivity of the results to the choice of exchange-correlation energy functional is easily understood, and gives hope that the results of a

more complete exchange-correlation energy functional will not change our conclusions. Basically the transitions depend on a trade-off of the interaction (exchange-correlation) energy with the kinetic energy. Over the range of the transitions the interaction energy is relatively constant, while the kinetic energy is changing rapidly. Hence the location of the transitions depends on getting the size of the interaction term correct and treating the kinetic energy as accurately as possible.

The calculation has an interesting qualitative feature which should be remarked upon. Consider a very low-density system; the bands are completely split and all the electrons occupy the lower spin-up band. An electron in the lower (ground-state) band sees the potential due to a static array of protons (electron-electron effects are very small at low density). On the other hand, consider an excited electron in the upper band. It should see essentially the potential of neutral H, since this state corresponds to two electrons on the same site. Qualitatively this effect occurs to some degree at all densities lower than the magnetic transition including situations where the two bands overlap. This effect is crucial in any proper understanding of the metal-insulator transition. It is represented phenomenologically in the Hubbard model by the quantity  $U$  which is just the extra energy required to occupy one of the sites with two electrons.

The self-consistent band-structure approach does mimic this effect qualitatively in terms of the total energy. The electrons in the lower band have a strong exchange term in the local spin-density approximation, while the electrons in the upper band have very little exchange. In the low-density limit the statistical approximation to the exchange is a self-interaction correction which cancels the self-electrostatic energy almost completely in a global sense. It is just these features that lead to the splitting of the spin bands in our model.

## V. DISCUSSION AND SUMMARY

We now compare our results with experiments. Edwards and Sienko<sup>17</sup> have reviewed studies of impurity atoms embedded in various host materials such as P in Si and Na in Ar. These experiments have the common property of a single loosely bound electron about the impurity. For this whole class of experiments there is a metal-insulator transition for  $n_c^{1/3} a_0^* = 0.26 \pm 0.05$ . Here  $a_0^*$  is a Bohr radius chosen appropriately to represent the localization of the electron about a single isolated impurity in the matrix; in Si, for example,  $a_0^*$  is the effective-mass radius corrected for central-

cell effects. This empirical relation for the critical density holds for systems with ratios of density up to  $10^{10}$ . Our determination of  $n_c^{1/3}a_0 = 0.22$  is in agreement with the empirical relation, whereas the estimate of the critical density by Ghazali and Leroux-Hugon<sup>11</sup> mentioned above gives a critical density of a factor of 2 smaller than the experimental one. We attribute the difference to the several approximations in their calculation which we did not make.

For the particular case of the transitions of donors or acceptors in semiconductors a further comment is in order. For definiteness consider donors in silicon where the energy scale is set by the donor Ry,  $\approx 30$  meV. The critical temperature for the first-order metal-insulator transition scales to approximately 40 mK. We expect that the maximum transition temperature for the se-

cond-order transition may be similarly small. In consequence one must work at much lower temperatures than is usual to observe these transitions, if, in fact, they are not washed out by other effects such as disorder.

In summary, we have examined the metal-insulator transition using a parameter-free model for the interactions. We find that there are two separate transitions for the Mott problem. The metal-insulator transition is of first order, while the formation of magnetic moments at the proton sites is of second order.

#### ACKNOWLEDGMENTS

We would like to acknowledge useful discussions with Professor Thomas Witten and Dr. Sudip Chakravarty. We would also like to thank the National Science Foundation for its support.

<sup>1</sup>P. W. Anderson, Phys. Rev. 109, 1492 (1958).

<sup>2</sup>N. F. Mott, *Metal Insulator Transitions*, (Taylor and Francis, London, 1974), particularly Chap. 4 and references contained therein.

<sup>3</sup>N. F. Mott, Proc. Phys. Soc. London, Sect. A 62, 416 (1949); Can. J. Phys. 34, 1356 (1956); Philos. Mag. 6, 287 (1964).

<sup>4</sup>J. Hubbard, Proc. R. Soc. London, Ser. A 277, 237 (1964); *ibid.* 281, 401 (1964).

<sup>5</sup>M. Cyrot, Philos. Mag. 25, 1031 (1972).

<sup>6</sup>U. von Barth and L. J. Hedin, J. Phys. C 5, 1629 (1972).

<sup>7</sup>O. Gunnarson and B. I. Lundqvist, Phys. Rev. B 13, 4274 (1976).

<sup>8</sup>P. Hohenberg and W. Kohn, Phys. Rev. 136, B864

(1964).

<sup>9</sup>W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).

<sup>10</sup>J. Callaway and C. S. Wang, Phys. Rev. B 7, 1096 (1973).

<sup>11</sup>J. F. Janak, Phys. Rev. B 16, 255 (1977); J. P. Perdew, *ibid.* 16, 1525 (1977).

<sup>12</sup>H. B. Shore, E. Zaremba, J. H. Rose, and L. Sander, Phys. Rev. B 18, 6506 (1978).

<sup>13</sup>A. Ghazali and P. Leroux-Hugon, Phys. Rev. Lett. 41, 1569 (1978).

<sup>14</sup>J. C. Slater, Phys. Rev. 82, 538 (1951).

<sup>15</sup>B. Y. Tong, Phys. Rev. B 6, 1189 (1972).

<sup>16</sup>H. Brooks, Nuovo Cimento, Suppl. 7, 165 (1958).

<sup>17</sup>P. Edwards and M. Sienko, Phys. Rev. B 17, 2575 (1978).