PHYSICAL REVIEW B VOLUME 32, NUMBER 6 15 SEPTEMBER 1985

Density-functional theory of the band gap

L. J. Sham
Department of Physics, University of California—San Diego, La Jolla, California 92093

M. Schlüter
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

(Received 15 April 1985)

There is a correction to the energy band gap obtained from the eigenvalues of the exact one-electron density-functional equation due to the discontinuity in the exchange-correlation potential in an insulator. This correction is expressed in terms of the electron self-energy. If the local-density approximation is used for the exchange-correlation potential, then an additional correction to the gap is needed. Both terms are evaluated for a two-plane-wave model with exchange only. Empirical evidence is also given for the discontinuity.

I. INTRODUCTION

The density-functional theory[1,2] gives a framework for the interaction effects in an inhomogeneous system which, in the simple local-density approximation (LDA), has proven to be quantitatively rather good for many ground-state properties.[3] The one-electron Schrödinger equation with the effective LDA potential derived from the density-functional theory[2] yields wave functions which form a quite accurate ground-state density. This is the basis of the good approximation for the ground-state properties. The eigenvalues from the same equation do not, however, yield accurate band gaps for semiconductors and insulators. Errors of the order or larger than 50% are not uncommon.[4] This is the famous "band-gap problem."

One approach is to say that there is no theoretical basis for interpreting the eigenenergies of the effective one-particle Schrödinger equation (even with the exact density-functional potential) as the one-electron excitation energies. Instead, one should construct the self-energy functional to obtain energy eigenvalues defined correctly in terms of the poles of the one-electron Green's function.[5] Much recent progress has been made in this direction.[6]

A second approach is to investigate whether density-functional theory can, in principle, predict band gaps, and whether improvements beyond the LDA would yield better band gaps. This approach does have a firm theoretical basis for the problem of fundamental band gaps even though it does not rigorously apply to the general band structure. In a semiconductor or insulator the band gap is rigorously defined[7] as the difference between the lowest conduction-band energy and the highest valence-band energy, which, in turn, are defined as the lowest energies to add and remove, respectively, an electron from the N-electron system. By emphasizing the N dependence of the density functional of energy, we can examine the change in total energy with N and arrive at an expression for the energy gap. The perturbative expression for the total energy[8] gives a basis for approximations.

One factor of an accurate determination of the energy gap is the correct exchange-correlation potential for the N-particle ground state. Whether improvement over the LDA will change the gap needs to be investigated. A more interesting finding of our investigation,[9] which is independently reached by Perdew and Levy,[10] is the existence of a discontinuity in N in the density-functional derivative of the exchange-correlation energy, i.e., in the potential. This discontinuity contributes to the band gap.

Our goal is to investigate the relative importance of these two contributions to the band gap, namely the improvement over the LDA potential and the discontinuity in the exchange-correlation potential. We have made some progress towards this goal by finding field-theoretic expressions for the exact exchange-correlation potential[8] and for the discontinuity, and by evaluating these expressions for a two-plane-wave model for an insulator in the exchange-only approximation. This simple model calculation is meant as an illustration for the unfamiliar concepts introduced here. The effect of correlation is studied in another paper.[11]

As a lead-in, the discontinuity in $\delta T_s/\delta n$, where T_s is the kinetic-energy functional, is examined in Sec. II. In Sec. III an expression for the band gap is derived within the density functional theory. The discontinuity Δ in the exchange-correlation potential V_{xc} is given in terms of the self-energy. Section IV describes the two-plane-wave model for a semiconductor. Section V gives some empirical evidence for the contribution of the discontinuity in V_{xc} to the energy gap in real materials. Section VI summarizes our findings.

II. DISCONTINUITY IN THE KINETIC-ENERGY FUNCTIONAL DERIVATIVE

As a prelude to the demonstration of the discontinuity in the functional derivative of the exchange-correlation energy, we study a similar discontinuity in the kinetic energy of the noninteracting system, $T_s[n]$. Throughout this paper we shall consider an extensive system, i.e., a system of N particles with $N \rightarrow \infty$ such that, in an equation, errors, which become negligible in the large-N limit, are understood. We also restrict our attention to spincompensated systems and neglect spin. Extension to inclusion of spin is straightforward.

Consider a noninteracting electron system with a monotonically increasing energy spectrum $\epsilon_1, \epsilon_2, \ldots$ with a gap between ϵ_N and ϵ_{N+1}. To be more precise, as $N \to \infty$ the energy difference between successive levels is $O(N^{-p})$, p being a positive fraction, except that $\epsilon_{N+1} - \epsilon_N = O(1)$. Then the N-electron ground state is an insulator, with an energy gap

$$\epsilon_g = \epsilon_{N+1} - \epsilon_N . \tag{2.1}$$

The ground state with M electrons will have a density distribution $n(r)$ given by

$$\frac{\delta T_{S,M}}{\delta n} + V = \mu_M , \tag{2.2}$$

where V is the potential and a subscript M is attached to the kinetic energy to emphasize the dependence on the total number of electrons as well as the density-functional dependence. The chemical potential μ_M is well defined, being

$$\mu_M = \epsilon_M + O(M^{-p}) \tag{2.3}$$

for all M except $M = N$. μ_M can be anywhere in the gap between ϵ_N and ϵ_{N+1}. Taking the difference of the two limits of Eq. (2.2) with M tending from above to $N+1$ and with M from below to N, we obtain

$$\epsilon_g = \frac{\delta T_s}{\delta n_+} - \frac{\delta T_s}{\delta n_-} , \tag{2.4}$$

where

$$\frac{\delta T_s}{\delta n_\pm} = \lim_{M \to N \pm 1} \frac{\delta T_{S,M}}{\delta n} . \tag{2.5}$$

This demonstrates the discontinuity in the functional derivative of the kinetic energy.

The discontinuity arises from the gap in the energy spectrum. In the Hohenberg-Kohn theory,[1] the energy functional is defined only for each given M. The variational equation (2.2) from their theory is strictly valid only for a given M. The derivative $\delta T_{S,M}/\delta n$ is then restricted to variations δn not changing M. To extrapolate the meaning of the Lagrange multiplier μ_M to be the energy of adding an electron to the system would involve the difference in functionals $T_{S,M+1}$ and $T_{S,M}$. The functional derivative is well defined if $T_{S,M}$ is a sufficiently smooth function of M. It is valid for conductors but not for insulators across a gap.

These subtleties come from the fact that the Hohenberg-Kohn theorem establishes a one-to-one mapping between the N-particle density $n_N(r)$, and not just the potential $V(r)$ but the combined set of $V(r)$ excluding an additive constant V_0 *plus* the total number of particles N. An energy functional of $n_N(r)$ is defined specifically for a subspace of $n_N(r)$ functions with N particles. As N changes we see an example above a functional not changing smoothly.

The discontinuity also shows up in the zero-temperature limit of the energy functional defined by Mermin's procedure.[12] For a given potential $V(r)$ (excluding V_0) and a chemical potential μ, there is an M which minimizes the free energy and a unique corresponding density $n_M(r)$.

Consider still the noninteracting case. The density is a functional of $V(r)$ and μ given by

$$n(r) = \sum_j \Theta(\mu - \epsilon_j) \, |\Psi_j|^2 , \tag{2.6}$$

where Ψ_j is the wave function associated with the energy ϵ_j. A change δn due to a small change in δV suffers a discontinuous jump when μ goes from just below ϵ_N to just above ϵ_{N+1}. Similarly, let us express the kinetic energy as

$$T_s[n] = \sum_j \Theta(\mu - \epsilon_j) \int dr \, \Psi_j^* (-\tfrac{1}{2}\nabla^2) \Psi_j . \tag{2.7}$$

It is a functional of n because Ψ_j, ϵ_j are functionals of (V, μ), which are, in turn, functionals of n. Explicit perturbation calculation of $\delta T_s/\delta n$ involves a term $\delta(\mu - \epsilon_j)$. The gap between ϵ_N and ϵ_{N+1} then yields a discontinuity in $\delta T_s/\delta n$ given by Eq. (2.4).

III. THE BAND GAP

A. Definition for an interacting system

The band structure in a semiconductor can be rigorously defined as the energies of one-particle excitations, which are, in turn, related to the difference between total energies of states differing by one electron.[7] In particular, the lowest conduction-band energy is given by

$$\epsilon_c = E_{N+1} - E_N , \tag{3.1}$$

where E_M is the total energy of the M-particle ground state, and the N-particle state corresponds to the insulating ground state. Similarly, the highest valence-band energy corresponds to

$$\epsilon_v = E_N - E_{N-1} . \tag{3.2}$$

The fundamental band gap is then

$$E_g = \epsilon_c - \epsilon_v , \tag{3.3}$$

and the zero-temperature-limit chemical potential at midgap is

$$\mu = (\epsilon_c + \epsilon_v)/2 . \tag{3.4}$$

B. Relation to the Kohn-Sham eigenenergies

For an M-particle state the density is determined by the wave functions of the Schrödinger equation[2]

$$[-\tfrac{1}{2}\nabla^2 + V(M)]\Psi_j(M) = \epsilon_j(M)\Psi_j(M) , \tag{3.5}$$

where $V(M)$ is the Kohn-Sham potential for M particles:

$$V(M) = V_{\text{ext}} + \int u n(M) + V_{\text{xc}}(M) , \tag{3.6}$$

where, for simplicity, the coordinate dependence of the integral is understood. The right-hand side of Eq. (3.6) is the sum of the external potential V_{ext}, the Coulomb potential through the interaction u due to the charge distribution $n(M)$ and the exchange-correlation potential

$$V_{\text{xc}}(M)=\frac{\delta E_{\text{xc},M}}{\delta n} . \tag{3.7}$$

E_{xc} is the exchange-correlation part of the total energy.[2] Similar to the metal case, where the highest "occupied" eigenenergy for the N-electron system $\epsilon_N(N)$ is the Fermi energy,[2] it is easy to show that

$$\epsilon_c=\epsilon_{N+1}(N+1) , \tag{3.8a}$$

$$\epsilon_v=\epsilon_N(N) . \tag{3.8b}$$

Define a density function

$$n_M(N)=\sum_{j=1}^{M} |\Psi_j(N)|^2 . \tag{3.9}$$

By Ref. 2, $n_N(N)$ is the density of the N-particle ground state. From the definition of the valence-band edge, Eq. (3.2),

$$\epsilon_v=E[n_N(N)]-E[n_{N-1}(N-1)] . \tag{3.10}$$

We wish to find terms of $O(1)$ on the right-hand side and can neglect the difference between $E[n_{N-1}(N-1)]$ and $E[n_{N-1}(N)]$, which is, by the variational theorem,[1] of second order in the difference in the $(N-1)$-particle density $n_{N-1}(N)-n_{N-1}(N-1)$, and thus is $O(N^{-1})$. Thus, Eq. (3.10) becomes

$$\begin{aligned}\epsilon_v&=E[n_N(N)]-E[n_{N-1}(N)]\\&=\int\frac{\delta E}{\delta n_{N-1}}|\Psi_N(N)|^2=\epsilon_N(N) ,\end{aligned} \tag{3.11}$$

by virtue of the variational equation (2.2), which, with V replaced by $V(N)$, is equivalent to Eq. (3.5).[2] The proof of Eq. (3.8a) is similar.

Thus, the energy gap is related to the eigenenergies of Eq. (3.5):

$$E_g=\epsilon_{N+1}(N+1)-\epsilon_N(N) . \tag{3.12}$$

Note that it differs from the naive definition of the band gap in terms only of the N-particle eigenenergies

$$\epsilon_g=\epsilon_{N+1}(N)-\epsilon_N(N) , \tag{3.13}$$

only in replacing $\epsilon_{N+1}(N)$ by $\epsilon_{N+1}(N+1)$. We write the deviation as

$$E_g=\epsilon_g+\Delta_{\text{xc}} , \tag{3.14}$$

with the subscript xc in anticipation that it is due to the exchange-correlation effects. The deviation is given by

$$\Delta_{\text{xc}}=\epsilon_{N+1}(N+1)-\epsilon_{N+1}(N) . \tag{3.15}$$

C. Band gap in terms of the discontinuity in V_{xc}

For an interacting system, the variational equation (2.2) is valid provided that V is replaced by the total effective potential $V(M)$ in Eq. (3.6). The difference of the two limits of M tending to $N+1$ from above and to N from below then gives

$$\epsilon_{N+1}(N+1)-\epsilon_N(N)=\frac{\delta T_S}{\delta n_+}-\frac{\delta T_S}{\delta n_-}+V_{\text{xc}}^{(+)}-V_{\text{xc}}^{(-)} , \tag{3.16}$$

where

$$V_{\text{xc}}^{(\pm)}=\delta E_{\text{xc}}/\delta n_\pm \tag{3.17}$$

are defined similarly to Eq. (2.5). Using Eqs. (2.4) and (3.14), we obtain

$$\Delta_{\text{xc}}=V_{\text{xc}}^{(+)}-V_{\text{xc}}^{(-)} . \tag{3.18}$$

The discontinuity in V_{xc} is a constant.

Alternatively, from the total-energy difference (3.1),

$$\epsilon_c=E[n_{N+1}(N)]-E[n_N(N)] , \tag{3.19}$$

where the variational theorem has again been invoked to replace the exact density $n_{N+1}(N+1)$ in E_{N+1} by $n_{N+1}(N)$. By the usual separation of the total energy,[2]

$$E[n]=T_s[n]+\int V_{\text{ext}}n+\tfrac{1}{2}\int\int nun+E_{\text{xc}}[n] , \tag{3.20}$$

where, for simplicity, the coordinate dependence in the integrals is understood, Eq. (3.19) becomes, to first order in the density change,

$$\begin{aligned}\epsilon_c=\;&\epsilon_{N+1}(N)+E_{\text{xc}}[n_{N+1}(N)]-E_{\text{xc}}[n_N(N)]\\&-\int V_{\text{xc}}(N)|\Psi_{N+1}(N)|^2 .\end{aligned} \tag{3.21}$$

In the evaluation of the total-energy difference, we have not implicitly assumed that E_{xc} has a smooth derivative across N. Thus, we take the limit

$$V_{\text{xc}}(N)=V_{\text{xc}}^{(-)} , \tag{3.22}$$

but

$$E_{\text{xc}}[n_{N+1}(N)]-E_{\text{xc}}[n_N(N)]=\int\frac{\delta E_{\text{xc}}}{\delta n_+}|\Psi_{N+1}(N)|^2 . \tag{3.23}$$

By the definitions (3.13) and (3.14) of ϵ_g and Δ_{xc}, we have

$$\Delta_{\text{xc}}=\int(V_{\text{xc}}^{(+)}-V_{\text{xc}}^{(-)})|\Psi_{N+1}(N)|^2 . \tag{3.24}$$

The results in Sec. II up to this point have been independently obtained by Perdew and Levy.[10]

D. The discontinuity in V_{xc}

In the preceding subsection it is established through the density-functional theory that the contribution to the band gap can be divided into two components: ϵ_g, which is the gap from the Kohn-Sham equation for the ground state, and Δ_{xc}, which is due to the discontinuity in the exchange-correlation potential. An improvement over the LDA would yield a better value for ϵ_g. It is not yet established if the LDA gives a good approximation to ϵ_g.

It is, however, clear that the LDA yields no discontinuity in V_{xc} and, thus, gives zero Δ_{xc}. It is therefore imperative to demonstrate that, beyond the LDA, such a discontinuity in V_{xc} does exist for an insulator. We shall do so by deriving an expression for the discontinuity from the perturbation expression of $E_{\text{xc}}[n]$ in the preceding paper.[8]

The root of the discontinuity lies in the existence of a gap in the one-electron energy spectrum. The discontinuity in V_{xc} arises in a way similar to $\delta T_s/\delta n$ from expression (2.7). We take the field-theoretic expression for E_{xc} in the preceding paper[8] and calculate the difference on the left-hand side of Eq. (3.23) to yield the $V_{\text{xc}}^{(+)}$ on the right-hand side. The unperturbed Green's function in Ref. 8, G_0, is given by the effective potential with the exchange-correlation component $V_{\text{xc}}^{(-)}$ in Eq. (3.22). For simplicity, denote the $(N+1)$th orbital of the equation for N particles by a subscript c. The change in G_0 corresponding to the density change of $|\Psi_{N+1}(N)|^2$ is

$$\delta G_0(r,r';\omega)=2\pi i\,\delta(\omega-\epsilon_c)\Psi_c(r)\Psi_c^*(r') \; . \tag{3.25}$$

From Ref. 8,

$$E_{\text{xc}}=i\,\text{Tr}[\ln(1-\Sigma G_0)+\Sigma G]+Y_{\text{xc}} \; , \tag{3.26}$$

where Y_{xc} is the sum of skeleton diagrams in G. The functional derivative of the first term is

$$i\,\text{Tr}[\Sigma\delta G-(\Sigma+\Sigma G\Sigma)\delta G_0] \; . \tag{3.27}$$

The functional derivative of the second term is

$$\delta Y_{\text{xc}}=-i\,\text{Tr}[\Sigma\delta G] \; . \tag{3.28}$$

Together, they give, for Eq. (3.23),

$$\Delta_{\text{xc}}=\int dr\int dr'\Psi_c^*(r)\widetilde{\Sigma}(r,r';\epsilon_c)\Psi_c(r') \; , \tag{3.29}$$

where the improper self-energy is given by the integral equation

$$\widetilde{\Sigma}=(\Sigma_{\text{xc}}-V_{\text{xc}}^{(-)})(1+G_0\widetilde{\Sigma}) \tag{3.30}$$

in terms of the exchange-correlation part of the self-energy.[8]

IV. THE TWO-PLANE-WAVE MODEL

A. The model

Unlike the metal, where the jellium serves as an excellent simple model to study the many-electron effects, the semiconductor lacks a simple counterpart. The Penn model,[13] a spherical two-wave model, which serves admirably for the purpose of a dielectric function, has some inconsistencies in its application to the exchange-energy calculation.[14] We therefore settle for a quasi-one-dimensional two-plane-wave model which can be solved exactly.

The basis set for the wave function consists of two plane waves, $\phi_{k\pm B}$, where

$$\phi_k(\mathbf{r})=L^{-1/2}e^{ikx}\zeta(\mathbf{R}) \; , \tag{4.1}$$

L being the length of the system along the x axis and $\zeta(\mathbf{R})$ the wave-function dependence in the y-z plane decaying in a distance of the order $1/B_\perp$. B is the Brillouin-zone boundary in the x direction so that $2B=Q$ is the reciprocal-lattice vector. The effective Hamiltonian is

$$\begin{bmatrix} \frac{1}{2}(k+B)^2 & V \\ V & \frac{1}{2}(k-B)^2 \end{bmatrix} , \tag{4.2}$$

where V is the strength of the Fourier component of wave vector Q of the one-electron potential. The energy dispersion of the two-band model is given, for $-B\le k\le B$, by

$$\left.\begin{matrix}\epsilon_{ck}\\ \epsilon_{vk}\end{matrix}\right\}=\tfrac{1}{2}(k^2+B^2)\pm[(kB)^2+V^2]^{1/2} \; . \tag{4.3}$$

The higher band is taken to be the conduction band and the lower the valence band. The associated wave functions are given by

$$\Psi_{vk}=\sum_{\sigma=\pm}C_\sigma(vk)\phi_{k+\sigma B} \; , \tag{4.4}$$

with coefficients for the conduction and valence band given by

$$\begin{bmatrix} C_+(ck) \\ C_-(ck) \end{bmatrix}=\begin{bmatrix} \cos\theta_k \\ -\sin\theta_k \end{bmatrix} , \qquad \begin{bmatrix} C_+(vk) \\ C_-(vk) \end{bmatrix}=\begin{bmatrix} \sin\theta_k \\ \cos\theta_k \end{bmatrix} , \tag{4.5}$$

and

$$-\theta_k=\tfrac{1}{2}\tan^{-1}(V/kB) \; . \tag{4.6}$$

The Coulomb interaction matrix element for a plane-wave scattered from k_1 to k_2 while the second scatters from k_3 to k_4 is

$$\int dr\int dr'\phi_{k_2}^*(\mathbf{r})\phi_{k_1}(\mathbf{r})u(\mathbf{r}-\mathbf{r}')\phi_{k_4}^*(\mathbf{r}')\phi_{k_3}(\mathbf{r}')=\frac{e^2}{L}\ln\left[\frac{(k_2-k_1)^2+B_\perp^2}{(k_2-k_1)^2}\right]\delta_{k_3-k_4,\,k_2-k_1}\equiv u(k_2-k_1)\delta_{k_3-k_4,\,k_2-k_1}/L \; . \tag{4.7}$$

B. The exchange potential and its discontinuity

With this simple model of the band structure and interaction it is a straightforward matter to construct the exchange energy $E_x[n]$ and demonstrate the discontinuity in V_x directly. We also verify that the formula (4.3) in Ref. 8 of V_x in terms of the exchange self-energy confirms the result.

Following the procedure of Ref. 8 we construct the single-particle band structure of the Kohm-Sham equation. Thus, the energy bands in Sec. IV A are the eigenenergies of Eq. (3.5), with the potential V in (4.2) corresponding to the Q component of the total effective potential (3.6). In the approximation with exchange only,

$$V=V_{\text{ext}}(Q)+u(Q)n(Q)+V_x(Q)\ . \tag{4.8}$$

The density distribution

$$n(x)=n(0)+2n(Q)\cos(\mathbf{Q}\cdot\mathbf{x}) \tag{4.9}$$

is determined by the two Fourier components $n(0)$ and $n(Q)$. If the system is filled to the Fermi level μ, the functional relation between $n(x)$ and $V(x)$ is given by the two functions of $n(0),n(Q)$ in terms of μ and V. Later, we shall determine $V_x(Q)$ and thus obtain a functional relation between the density $n(x)$ and the bare lattice potential $V_{\text{ext}}(x)$.

It is convenient to use the Fermi wave vector k_F in place of μ,

$$\mu=\tfrac{1}{2}(k_F^2+B^2)\pm[(k_FB)^2+V^2]^{1/2}\ , \tag{4.10}$$

where the plus sign corresponds to μ in the conduction band (n-type semiconductor) and the minus sign for the valence band (p-type semiconductor). Then,

$$n(0)=\frac{B_1^2}{2\pi^2}(B\pm k_F)\ , \tag{4.11a}$$

$$n(Q)=\frac{B_1^2V}{4\pi^2B}\ln\left[\frac{Bk_F+(B^2k_F^2+V^2)^{1/2}}{B^2+(B^4+V^2)^{1/2}}\right]\ . \tag{4.11b}$$

The $\pm$ signs in (4.11a) have the same meaning as above. $k_F=0$ corresponds to the insulating state.

From Eq. (4.1) of Ref. 8, we can express the exchange functional $E_x[n]$ in terms of k_F and V:

$$E_x[n]=-L\sum_{\alpha,\beta,\alpha',\beta'}\delta_{\alpha-\beta,\alpha'-\beta'}\left[\int_{-B}^{B}\frac{dk}{2\pi}\rho_{\alpha\beta}^{-}(k)\pm\int_{-k_F}^{k_F}\frac{dk}{2\pi}\rho_{\alpha\beta}^{\pm}(k)\right]\times\left[\int_{-B}^{B}\frac{dk'}{2\pi}\rho_{\beta'\alpha'}^{-}(k')\pm\int_{-k_F}^{k_F}\frac{dk'}{2\pi}\rho_{\beta'\alpha'}^{\pm}(k')\right]u(k-k'+(\beta-\beta')B)\ , \tag{4.12}$$

where the indices α, etc. run over ±1, and the one-particle density matrices are given by

$$\rho_{\alpha\beta}^{\pm}(k)=\tfrac{1}{2}\begin{bmatrix}1\pm\cos(2\theta_k) & \mp\sin(2\theta_k)\\ \mp\sin(2\theta_k) & 1\mp\cos(2\theta_k)\end{bmatrix}\ . \tag{4.13}$$

The exchange potential, as $\delta E_x/\delta n$, is given by the Fourier components

$$V_x(0)=\frac{1}{\Omega}\frac{\partial E_x}{\partial n(0)}\ ,\qquad V_x(Q)=\frac{1}{2\Omega}\frac{\partial E_x}{\partial n(Q)}\ , \tag{4.14}$$

where

$$\Omega=4\pi L/B_1^2\ . \tag{4.15}$$

Since, in (4.12), $E_x[n]$ is a function of k_F and V, we use the chain rule to obtain the exchange potential

$$V_x(G)=\frac{1}{2L}\frac{\partial E_x}{\partial V}\bigg/\frac{4\pi}{B_1^2}\frac{\partial n(Q)}{\partial V}\ , \tag{4.16}$$

$$V_x(0)=\left[\frac{1}{L}\frac{\partial E_x}{\partial k_F}\bigg/\frac{4\pi}{B_1^2}\frac{\partial n(0)}{\partial k_F}\right]-2V_x(Q)\frac{\partial n(Q)/\partial k_F}{\partial n(0)/\partial k_F}\ . \tag{4.17}$$

We take the semiconductor limits by letting $k_F\to0$ for μ given by Eq. (4.10) in the valence and conduction bands. We find that

$$V_x^{(+)}(Q)=V_x^{(-)}(Q) \tag{4.18}$$

and

$$\Delta_x=V_x^{(+)}(0)-V_x^{(-)}(0)=-\int_{-B}^{B}\frac{dk}{2\pi}u(k)\sin(2\theta_k)-2V_x^{(-)}(Q)\ . \tag{4.19}$$

Alternatively, the results for V_x and Δ_x can be obtained directly in terms of the self-energy Σ_x in the insulating ground state.

Since

$$\Sigma_x(r,r')=-\tfrac{1}{2}u(r-r')\rho(r,r') \tag{4.20}$$

in terms of the one-particle density matrix $\rho(r,r')$, the matrix elements of Σ_x are given by

$$\Sigma_{ss}(k)=\int dr\int dr'\phi_{k+sB}^{*}(r)\Sigma_x(r,r')\phi_{k+s'B}(r')=-L^{-1}\sum_{\kappa,\sigma}u(k-\kappa+(s-\sigma)B)C_\sigma(vk)C_{\sigma-s+s'}^{*}(vk)\ . \tag{4.21}$$

From Eq. (4.3) of Ref. 8, the exchange potential is given by

$$V_x^{(-)}(Q)X(Q,Q)=-\int_{-B}^{B}\frac{dk}{2\pi}\frac{\cos(2\theta_k)}{[(kB)^2+V^2]^{1/2}}\{\tfrac{1}{2}[\Sigma_{++}(k)-\Sigma_{--}(k)]\sin(2\theta_k)+\Sigma_{+-}(k)\cos(2\theta_k)\}\ , \tag{4.22}$$

where $\mathcal{X}(Q,Q)$ is the Fourier transform of the density response $\mathcal{X}(r,r')$, given by

$$\mathcal{X}(Q,Q)=\frac{1}{\pi B}\left[\frac{B^2}{(B^4+V^2)^{1/2}}+\ln\left[\frac{(B^4+V^2)^{1/2}-B^2}{V}\right]\right]. \tag{4.23}$$

The resulting expression for the exchange potential is the same as that from Eq. (4.16). When Σ_x is substituted in Eq. (3.29), we obtain Eq. (4.19).

C. The local-density approximation

Since for this model we have the exact exchange potential, it would be interesting to compare it with the LDA. However, since the interaction given by (4.7) is not exactly Coulombic, we need to produce the appropriate LDA for the exchange potential, which is given by the one-plane-wave approximation for Σ_x at the chemical potential, i.e.,

$$\mu_x(n)=-e^2[\ln(\tfrac{5}{4})+\tan^{-1}2](2n/\pi)^{1/3}\,. \tag{4.24}$$

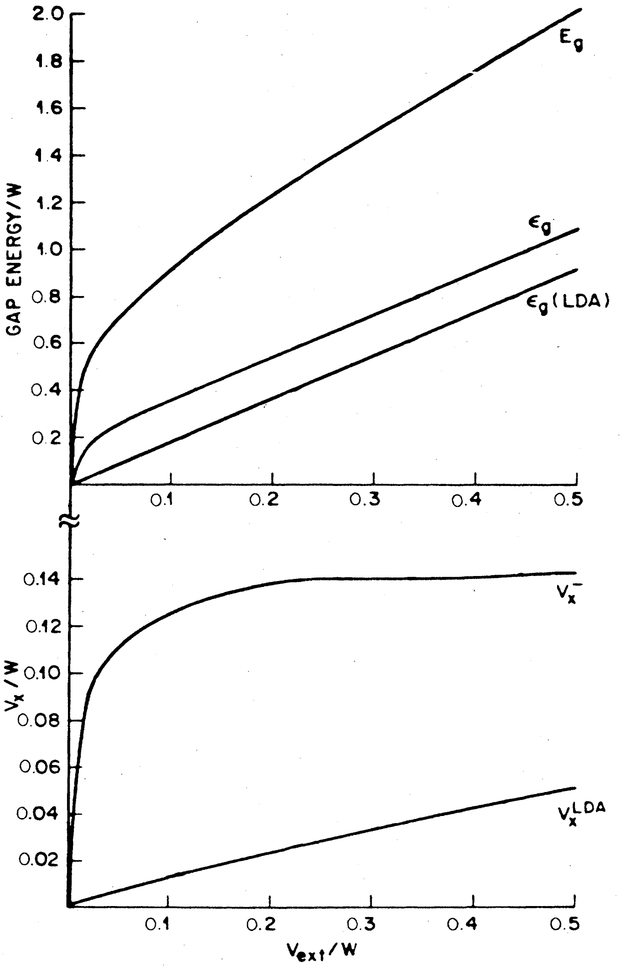

FIG. 1. Exact and LDA exchange potentials $V_x(Q)$ plotted vs the eternal potential V_{ext} in units of W, the valence-band width (bottom). Exact gap E_g, pseudogap ϵ_g (calculated from the ground-state equation), and the LDA gap ϵ_g^{LDA}, plotted for a range of external potential V_{ext} (top).

For simplicity, we have set $B_\perp=B$.

D. Numerical results

It is convenient to express energy in units of the valence-band width,

$$W=\tfrac{1}{2}B^2\,, \tag{4.25}$$

and Coulomb energy in r_s,

$$r_s=\left[\frac{3\pi}{16}\right]^{1/3}\frac{B}{W}\,. \tag{4.26}$$

The results for V_x, which scales with r_s, are plotted for $r_s=2$ and a range of external potential V_{ext} in Fig. 1 (bottom), and compared with the LDA as defined by (4.24). In this model the LDA is not a particularly good approximation to the exact V_x. In Fig. 1 (top), for a range of values of V_{ext}/W, we plot the exact gap E_g, the partial gap ϵ_g given by Eq. (3.13), and the LDA gap. The difference between ϵ_g and the LDA indicates the extent of improvement possible in the gap by improving the exchange potential. The contribution from the discontinuity in V_x, as indicated by the difference between E_g and ϵ_g, is clearly substantial, being of the order of ϵ_g itself. By comparison, the deviation of the LDA V_x from the exact V_x is small.

V. REAL SEMICONDUCTORS

The study of the exchange and correlation potential beyond the LDA and its discontinuity Δ_{xc} for real systems is underway. We do not yet know if the LDA is a good approximation to $V_{\text{xc}}^{(-)}$. However, if we assume that this is the case, as suggested by the success of computations of

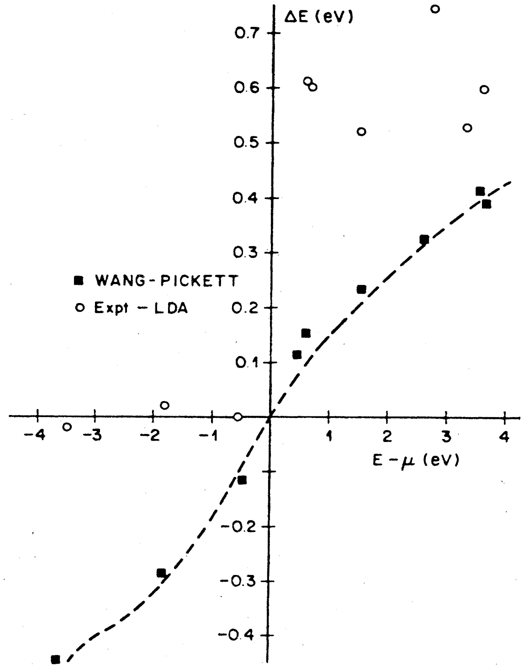

FIG. 2. Difference between measured and LDA band energies at symmetry points in silicon vs the measured energies. The points calculated by Wang and Pickett are taken from Ref. 16.

some of the ground-state properties, then the discrepancy between the measured gap and the LDA value is a measure of the discontinuity Δ_{xc}. In Fig. 2 we plot the difference in energy values between experiment and the LDA for various symmetry points in silicon[15] versus the measured energy $E-\mu$, placing μ at midgap and assuming that the top of the valence band is given by the LDA value. The steplike distribution of data points appears to be an indication that the low-lying conduction-band energies deviate from the LDA gap more or less by a constant, which might be interpreted as Δ_{xc}. The corresponding Wang-Pickett[16,6] values, shown for comparison, do not exhibit such a steplike feature. The values are from a version of their calculation where only the energy dependence but not the *nonlocality* of the self-energy is changed from the LDA. This suggests that in semiconductors nonlocality plays an essential role in determining the gap.

VI. SUMMARY

We have shown that the exchange-correlation potential jumps by a constant when an electron is added to an insulator. This discontinuity gives the difference between the true band gap and the one obtained from the ground-state one-particle equations. This partial contribution to the gap arises from the difference between the conduction-band valleys obtained from the ground-state self-consistent equations for the N and $(N+1)$-particle systems, which, in turn, comes from the difference in the exchange and correlation effects between the two systems.

The simple model for an insulator illustrates the exchange potential beyond the LDA and its discontinuity. It serves as a counterexample to arguments[17] purporting to show that an exact exchange-correlation potential in the self-consistent potential for the N-particle ground state suffices to yield the exact band gap, and that, by implication, the discontinuity in V_{xc} is zero. We hope that this paper will serve as a starting point for the study of both the exchange and correlation contributions to the energy gap of real materials.

ACKNOWLEDGMENTS

One of us (L.J.S.) acknowledges the support of the National Science Foundation under Grant No. DMR-80-18440 and of the Guggenheim Foundation.

---

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[3] For reviews, see, for example, M. Schlüter and L. J. Sham, Physics Today **35**(2), 30 (1982), and in *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundquist and N. H. March (Plenum, New York, 1983).

[4] D. R. Hamann, Phys. Rev. Lett. **42**, 662 (1979); M. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 5668 (1982).

[5] L. J. Sham and W. Kohn, Phys. Rev. **145**, 561 (1966).

[6] G. W. Bauer, Phys. Rev. B **27**, 5912 (1983); G. Strinati, H. J. Mattausch, and W. Hanke, *ibid.* **25**, 2867 (1982); W. E. Pickett and C. S. Wang, *ibid.* **30**, 4719 (1984); M. Hybertson and S. G. Louie (unpublished).

[7] L. J. Sham, Phys. Rev. **150**, 720 (1966).

[8] L. J. Sham, preceding paper, Phys. Rev. B **32**, 3876 (1985).

[9] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

[10] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[11] M. Lannoo, M. Schlüter, and L. J. Sham, following paper, Phys. Rev. B **32**, 3890 (1985).

[12] N. D. Mermin, Phys. Rev. **137**, A1441 (1965).

[13] D. R. Penn, Phys. Rev. **128**, 2093 (1962).

[14] F. Guinea and C. Tejedor, J. Phys. C **13**, 5515 (1980).

[15] H. Stöhr and M. Bross, Phys. Status Solidi B **90**, 497 (1978).

[16] C. S. Wang and W. E. Pickett, Phys. Rev. Lett. **51**, 600 (1983).

[17] A. R. Williams and U. von Barth, in *Theory of the Inhomogeneous Electron Gas*, Ref. 3, p. 189.