

Density-Functional Theory of the Energy Gap

L. J. Sham

*Department of Physics, University of California San Diego, La Jolla, California 92093,^(a) and
Institute for Theoretical Physics, University of California Santa Barbara, California 93106*

and

M. Schlüter

Bell Laboratories, Murray Hill, New Jersey 07974

(Received 23 June 1983)

The energy-band gap of an insulator is obtained from the eigenvalues of the one-particle density-functional equation for the ground state and a finite correction due to the discontinuity of the functional derivative of the exchange and correlation energy. This correction is expressed in terms of the improper self-energy and the density-functional exchange-correlation potential. It is evaluated for a two-plane-wave model including exchange only.

PACS numbers: 71.45.Gm, 71.25.Cx

In the density-functional theory,¹ the ground-state energy of a many-electron system is treated as a functional of its density distribution $n(r)$. It may be separated² into the electrostatic potential energy, the kinetic energy $T_s[n]$ of a non-interacting electron system with the same n , and the remainder, $E_{xc}[n]$, termed the exchange-correlation energy. The variation principle yields a single-particle Schrödinger equation with the potential given by the sum of the electrostatic potential v_{es} due to the external and electronic charges and the exchange-correlation potential v_{xc} given by $\delta E_{xc}/\delta n$. The self-consistent solution of this so-called Kohn-Sham (KS) equation gives the density distribution and the ground-state energy. The local-density approximation² (LDA), where $E_{xc}[n]$ is given in each small neighborhood by the exchange and correlation energy of the homogeneous electron gas at the local density, makes the self-consistent computation no more difficult than the Hartree approximation and has proved to be of satisfactory accuracy for ground-state properties of atoms,³ molecules,³ metals,⁴ semiconductors,⁵ surfaces,⁶ and defects.⁷ However, when the eigenvalues of the KS equation are taken to be the single-particle excitation energies, the reliability varies. While it is reasonably satisfactory for metals,⁴ the errors in semiconductor band gaps^{5,8} can be as large as 50%.

Two approaches may be taken to rectify this situation. One has been forewarned⁹ that the eigenvalues of the KS equation are not justified to be the excitation energies (except for the chemical potential in a metal) and that the proper procedure is to obtain the one-particle excitation

from the self-energy operator constructed as a functional of the density. An approximate construction of the self-energy which takes into account the band gap has recently been given by Wang and Pickett.¹⁰ The alternative approach is to examine if a full knowledge of the exchange-correlation potential v_{xc} would lead to the correct energy gap.

Here we take the second approach and note that the fundamental gap for an N -particle insulator ground state ($N \rightarrow \infty$) is given by

$$E_g = (E_{N+1} - E_N) - (E_N - E_{N-1}), \quad (1)$$

where E_M is the ground-state energy of M particles. If ϵ_g denotes the energy gap obtained from the exact KS equation for the N -particle ground state, then consideration of the density functionals of the energies on the right-hand side of Eq. (1) leads to the correction for the gap,

$$E_g - \epsilon_g \equiv \Delta = v_{xc}^{(+)} - v_{xc}^{(-)}, \quad (2)$$

where

$$v_{xc}^{(\pm)} = \delta E_{xc} / \delta n_{\pm}, \quad (3)$$

with the change in density distribution δn_{\pm} (or δn_{\pm}) resulting in an increase (or nonincrease) in the total number of electrons. $v_{xc}^{(-)}$ is the appropriate potential for the N -particle ground state KS equation.

One implication of Eq. (2) is clear. Improvement of the exchange-correlation potential over LDA, such as gradient corrections¹¹ or modification of the exchange-correlation hole,³ might improve ϵ_g but yields zero for Δ . In the following, we shall establish the existence of the discontinuity of the functional derivatives of E_{xc} for an

insulator¹² and also provide a formula for Δ by the many-body perturbation theory. In this paper, we restrict our consideration to spin-compensated systems. There is a similar problem for the band splitting of ferromagnetic systems.¹³ Our work can be straightforwardly extended to include the spin density.

We shall also be only concerned with extensive systems with $N \rightarrow \infty$ so that it is understood that relative errors of $O(N^{-p})$ with $p > 0$ are neglected. We establish first the discontinuity of $\delta T_s / \delta n$ for a noninteracting system with an insulating ground state,

$$\delta T_s / \delta n_+ - \delta T_s / \delta n_- = \epsilon_g, \quad (4)$$

by taking the limits of the variational equation¹

$$\delta T_s / \delta n|_M + v = \mu_M, \quad (5)$$

for M particles with M tending to N from below and tending to $N + 1$ from above, and subtracting one limit from the other. While each term $\delta T_s / \delta n_{\pm}$ is position dependent, the difference is a position-independent functional of the density.

Alternatively, we note that the Hohenberg-Kohn theorem¹ establishes the functional relationship between the N -particle density $n_N(r)$ and one of the three sets of functional variables: (i) the external potential $v(r)$ (excluding the knowledge of an additive constant v_0) and the total number of particles N , (ii) $v(r)$ (excluding v_0) and the chemical potential μ , or (iii) $v(r)$ including v_0 with μ fixed. If we choose (iii), then the density of the noninteracting system is given, in terms of the wave function ψ_k and energy ϵ_k , by

$$n = \sum_k \theta(\mu - \epsilon_k) |\psi_k|^2, \quad (6)$$

with μ fixed anywhere between the conduction-band minimum ϵ_c and the valence-band maximum ϵ_v . If we wish an infinitesimal change of the density δn_+ with $\int dr \delta n_+(r) > 0$ to correspond to an infinitesimal change of the external potential δv , we must fix μ just below ϵ_c so that the original state contains N electrons while δv increases the

number of electrons. Similarly for δn_- with $\int dr \delta n_-(r) < 0$, μ must be fixed just above ϵ_v . Since

$$T_s[n] = \sum_k \theta(\mu - \epsilon_k) \int dr \psi_k^* (-\frac{1}{2} \nabla^2) \psi_k, \quad (7)$$

the discontinuity in $\delta T_s / \delta n$ comes from the $\delta(\mu - \epsilon_k)$ term, with the two values of μ .

The first proof of Eq. (4) may similarly be followed to deduce Eq. (2) by taking the limits of the version of the variational equation for the interacting system and by noting the property for the extensive system

$$E_N - E_{N-1} = \mu_N + O(N^{-p}). \quad (8)$$

This procedure shows that the true gap is equal to the sum of the discontinuities in $\delta T_s / \delta n$ and in v_{xc} but does not demonstrate that there is necessarily a nonzero discontinuity in v_{xc} . The second proof of Eq. (4) can be extended to establish the discontinuity of v_{xc} . This follows from an expression for $E_{xc}[n]$ obtained by the many-body perturbation theory in terms of the unperturbed Green's function G_0 which has the same sort of discontinuity in its functional derivative as T_s in Eq. (7) for exactly the same reason of having the chemical potential fixed at two different values for δn_{\pm} .

The unperturbed Green's function G_0 is given in terms of the self-consistent potential $v_{es} + v_{xc}$ ⁽⁻⁾ where the second term is given in Eq. (3). The perturbation can be treated by standard field theoretic techniques to obtain a coupling-constant integral for E_{xc} .¹⁴ One can further deduce *à la* Luttinger and Ward^{15,16}

$$E_{xc}[n] = i \text{Tr} \{ \ln(1 - \Sigma G_0) + \Sigma G \} + Y_{xc}', \quad (9)$$

where Tr stands for trace, Σ for the self-energy, and G for the full Green's function. Y_{xc}' is $-i$ times the sum of all exchange and higher-order energy diagrams in terms of the full Green's function and the Coulomb interaction only. Functional derivative of Eq. (9) yields the equation for the "exact" exchange-correlation potential v_{xc} ,

$$\int dr' v_{xc}(r') \int \frac{d\omega}{2\pi} G_0(r, r'; \omega) G(r', r; \omega) = \int dr_1 \int dr_2 \int \frac{d\omega}{2\pi} G_0(r, r_1; \omega) \Sigma_{xc}(r_1, r_2; \omega) G(r_2, r; \omega), \quad (10)$$

where Σ_{xc} is the exchange-correlation part of the self-energy, which, in contrast to v_{xc} , is nonlocal and energy dependent. Equation (10) also follows from the Dyson equation for G . From Eq. (10), v_{xc} in LDA is easily obtained. If only the exchange term is kept in Σ_{xc} and if G is replaced by G_0 in Eq. (10), we obtain the local potential in the Hartree-Fock approximation^{17,18} (see the numerical example below).

Let $n_M(N)$ be the density obtained by summing M orbitals in the KS equation of the interacting N -electron ground state. The energy gap from Eq. (1) is given in terms of energy functionals by

$$E_g = E[n_{N+1}(N+1)] - 2E[n_N(N)] + E[n_{N-1}(N-1)] = E[n_{N+1}(N)] - 2E[n_N(N)] + E[n_{N-1}(N)], \quad (11)$$

where the error incurred by the density replacements is guaranteed by the variational principle to be $O(N^{-1})$. The differences in total energies may now be evaluated to yield

$$\Delta = \int d\mathbf{r} \psi_c^* \{v_{xc}^{(+)} - v_{xc}^{(-)}\} \psi_c, \quad (12)$$

where ψ_c is the eigenfunction of the KS equation for the conduction-band minimum. Equation (12) is equivalent to Eq. (2) if we note that the term within the curly brackets is position independent.

To obtain the positive change in density δn_+ equal to $|\psi_c|^2$, we fix μ just below ϵ_c . The corresponding change in G_0 is

$$\delta G_0(\mathbf{r}, \mathbf{r}'; \omega) = 2\pi i \delta(\omega - \epsilon_c) \psi_c(\mathbf{r}) \psi_c^*(\mathbf{r}'). \quad (13)$$

Equation (9) may then be used to evaluate Eq. (12), resulting in

$$\Delta = \int d\mathbf{r} \int d\mathbf{r}' \psi_c^*(\mathbf{r}) \tilde{\Sigma}(\mathbf{r}, \mathbf{r}'; \epsilon_c) \psi_c(\mathbf{r}'), \quad (14)$$

where

$$\tilde{\Sigma} = \{\Sigma_{xc} - v_{xc}^{(-)}\} (1 + G_0 \tilde{\Sigma}). \quad (15)$$

This relation may alternatively be derived by considering the solution of the Dyson equation for G in terms of G_0 .

Estimates of Δ for model systems and for semiconductors are underway. If one assumes that LDA gives a good approximation for $v_{xc}^{(-)}$, then the gap correction would have to be of the same order of magnitude as ϵ_g . Although we do not have rigorous arguments which show the correction to be this large, there are three indications why it cannot be zero:

(1) The LDA for the self-energy⁹ can be used to

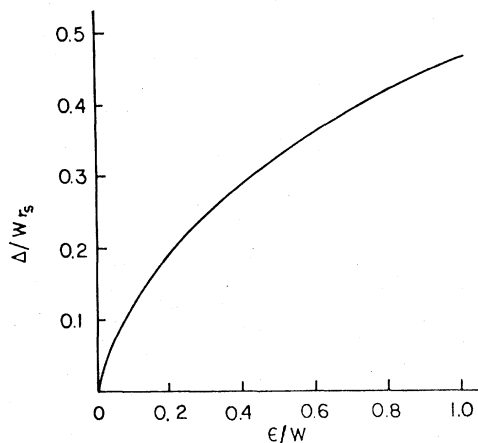


FIG. 1. The gap correction Δ vs the density functional gap ϵ for a two-plane-wave model in the Hartree-Fock approximation.

give, for slowly varying densities,

$$\Delta \simeq \epsilon_g [\langle m^*(n(\mathbf{r})) \rangle_c^{-1} - 1]. \quad (16)$$

This amounts only to a 5% correction for Si and is an underestimate for Δ because of the absence of a gap in the LDA for the self-energy.¹⁰

(2) Equation (12) relates Δ to the discontinuity of the fundamental derivative of E_{xc} . It is unlikely that this discontinuity will vanish when the functional derivative of $T_s[n]$ is discontinuous.

(3) We have calculated the gap correction for a two-plane-wave model including exchange only. The Brillouin zone is taken to be a cylinder of equal length and diameter. Each state is a mixture of two plane waves connected by the reciprocal-lattice vector in the direction parallel to the axis of the cylinder. The self-energy Σ is given by the lowest-order Hartree and exchange diagrams. The lowest-order approximation of Eq. (10), with G replaced by G_0 , is then solved to give the sole Fourier component of the exchange potential $v_x(G)$. Note that the LDA has been avoided. In Fig. 1, the gap correction Δ is plotted against the density functional gap ϵ_g . ϵ_g is in units of W and Δ is in units of $W r_s$, where W is the valence-band width and r_s is the conventional mean electron radius. The two-plane-wave exchange-only model illustrates how the density-functional equation underestimates the Hartree-Fock gap. The correction is of the order of the gap itself. Correlation contributions to the gap correction will alter the numerical values but not change the qualitative picture.

One of us (L.J.S.) wishes to thank Professor W. Kohn for the hospitality at the Institute of Theoretical Physics in Santa Barbara and for stimulating conversations. One of us (M.S.) acknowledges the hospitality of the Department of Physics, University of California San Diego, where part of this work was done. This work was supported in part by the National Science Foundation, under Grant No. DMR-08440-A02.

(a) Present address.

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

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

³O. Gunnarsson and R. O. Jones, Phys. Scr. **21**, 394 (1980).

⁴V. L. Moruzzi, J. F. Janak, and A. R. Williams,

Calculated Electronic Properties of Metals (Pergamon, Oxford, 1978).

⁵M. T. Yin and M. L. Cohen, Phys. Rev. B 26, 5668 (1982).

⁶N. D. Lang, Solid State Phys. 28, 225 (1973); J. A. Appelbaum and D. R. Hamann, Rev. Mod. Phys. 48, 479 (1976); K. C. Pandey, Phys. Rev. Lett. 49, 223 (1982).

⁷G. A. Baraff and M. Schlüter, Phys. Rev. B 28, 2296 (1983).

⁸D. R. Hamann, Phys. Rev. Lett. 42, 662 (1979).

⁹L. J. Sham and W. Kohn, Phys. Rev. 145, 561 (1966).

¹⁰C. S. Wang and W. E. Pickett, Phys. Rev. Lett. 51, 597 (1983).

¹¹D. C. Langreth and M. J. Mehl, Phys. Rev. Lett. 47, 446 (1981).

¹²J. P. Perdew, R. G. Parr, M. Levy, and J. L. Bal-

duz, Phys. Rev. Lett. 49, 1691 (1982), have constructed a density functional of energy for fractional number of electrons and found discontinuities in the functional derivatives of E_{xc} for *finite* systems. J. P. Perdew has informed us that the result can be extended to apply to bulk insulators: J. P. Perdew and Mel Levy, preceding Letter [Phys. Rev. Lett. 51, 1884 (1983)].

¹³L. Kleinman and K. Mednick, Phys. Rev. B 24, 6880 (1981).

¹⁴J. Harris and R. O. Jones, J. Phys. F 4, 1170 (1974).

¹⁵J. M. Luttinger and J. C. Ward, Phys. Rev. 118, 1417 (1960).

¹⁶L. J. Sham, to be published.

¹⁷R. T. Sharp and G. K. Horton, Phys. Rev. 90, 317 (1953).

¹⁸J. D. Talman and W. F. Shadwick, Phys. Rev. A 14, 36 (1976).