PHYSICAL REVIEW B VOLUME 30, NUMBER 8 15 OCTOBER 1984

Local-density approximation for dynamical correlation corrections to single-particle excitations in insulators

W. E. Pickett
*Condensed Matter Physics Branch, Naval Research Laboratory, Washington, D.C. 20375**
and Daresbury Laboratory, Science and Engineering Research Council, Daresbury, Warrington, WA4 4AD, England, United Kingdom

C. S. Wang
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
and Condensed Matter Physics Branch, Naval Research Laboratory, Washington, D. C. 20375

(Received 18 May 1984)

This paper presents a general approach to the calculation of single-particle excitations in insulating crystals, with application to silicon and diamond. The method includes the energy dependence of the self-energy which is evaluated in the Green's-function—screened-Coulomb-interaction ("GW") approximation of Hedin [Phys. Rev. **139**, A796 (1965)], which is applied in the local-density approximation of Sham and Kohn. This self-energy describes changes in exchange-correlation effects beyond those described by the ground-state exchange-correlation potential. The essential ingredient is the model of a homogeneous insulating electron gas, built on the ideas of Penn and of Levine and Louie. It is shown that this quasiparticle local-density-approximation (QPLDA) self-energy for an insulator with gap E_g is topologically distinct from its metallic counterpart due to a nonanalytic contribution to the self-energy proportional to $E_g(k-k_F)\ln|k-k_F|$. This gives a discontinuity in the energy derivative of the dynamic self-energy, which, in turn, leads to an increase in band gaps over the ground-state values. Both effects can be related directly to the gap between occupied and unoccupied states; charge inhomogeneities play only a minor role in this approach. Direct and indirect band gaps in both silicon and diamond are in much better agreement with experiment than are the results from the ground-state theory. The zone-boundary optical gaps and the valence-band width in Si appear to be predicted very accurately by the QPLDA, while a 0.24-eV discrepancy remains in the indirect band gap. In diamond the valence-band width and fundamental direct and indirect band gaps are within 5% of the experimental values. To provide a basis for an intuitive grasp of exchange-correlation corrections to excitation energies, the full nonlocal interaction for the model homogeneous system is presented and analyzed in some detail.

I. INTRODUCTION

Theoretical studies of the electronic structure of solids are based almost entirely on the Hartree-Fock (HF) and density-functional (DF) methods. However, both of these methods provide variational equations for ground-state properties (viz., energy), while most experimental probes sample the electronic properties through excited states. While the DF method[1] in the local-density approximation (LDA) is known[2] to give a reasonable description of excited single-particle ["quasiparticle" (QP)] states in many metals, the resulting energy bands typically underestimate gaps in semiconductors and insulators by 20—50%. The HF method, on the other hand, produces gaps which can be as much as a factor of 2 too wide.

In this paper we shall concentrate on taking steps to transcend the deficiencies of the DF-LDA approach. DF theory provides an effective one-electron potential $v_{\rm DF}(\vec{r};n)$ which describes the effects of exchange and correlation on *ground-state properties.* Hohenberg and Kohn[1] established a one-to-one correspondence between the external potential $v_{\rm ext}$ and the ground-state charge density n (within certain restrictions, which appear not to concern the present work). Thus *all properties* of the interacting system, which are functions of $v_{\rm ext}$ *a priori,* may just as well be considered functionals of n. The effective potential $v_{\rm DF}$ may depend on n in a very complex manner; however, it is a *local operator* in real space. The LDA is based on the supposition that this one-electron potential can be approximated locally as that of a homogeneous system at the same density:

$$v_{\rm DF}(\vec{r};n)\rightarrow\mu_{\rm ex}(n(\vec{r}))\;, \tag{1.1}$$

where $\mu_{\rm xc}(n)$ is the exchange-correlation (xc) potential of a homogeneous electron gas of density n. This approximation is formally justified if the density is slowly varying on the relevant length scale (i.e., Fermi wavelength or Thomas-Fermi screening length).

The breakdown of LDA in semiconductors has most often been ascribed to a violation of the "slowly varying density" criterion. From this point of view, the remedy is to "go beyond" LDA by constructing ground-state exchange-correlation potentials which depend nonlocally on the density. The weighted-density approximation[3] (WDA) represents one such approximate nonlocal-density representation of the ground-state exchange-correlation potential. In his application of the WDA to Si, Kerker[4] introduced a numerical simplification to make the calculation tractable. In the homogeneous limit his simplification increases the Kohn-Sham exchange ($\alpha=\frac{2}{3}$) to Slater exchange ($\alpha=1$) and increases the correlation potential correspondingly, an alteration which is known to widen band gaps in semiconductors. Therefore the increased band gaps found by Kerker are not representative of the full WDA, which itself reduces to the LDA [$\alpha(n)<1$] in the homogeneous limit. More recently, Langreth and Mehl[5] have constructed a nonlocal-density potential which interpolates between the correct small- and large-wavelength contributions to exchange and correlation. Although this potential appears to improve ground-state properties, it leads to only small increases[6] in the band gaps of semiconductors and insulators.

Two observations suggest that violation of the "slowly varying density" criterion may not be the paramount consideration. First, variations in density are no more rapid in covalent semiconductors than in transition-metal systems which are often described rather well by the LDA. Second, the recent determination within the LDA of several true ground-state properties[7] (cohesive energy, crystal structure, lattice constant, bulk modulus, and phonon spectra), all in good agreement with experiment, suggests that the breakdown is with DF theory itself—the application of a ground-state formalism to excited states—rather than with the LDA itself.

In this paper we investigate excited states directly via the construction of the Green's function $G(\vec{r},\vec{r}\,';E)$, which describes the propagation of an excited electron or hole at energy E from $\vec{r}\,'$ to $\vec{r}$. G is the solution to the Dyson equation

$$\left[-E-\frac{1}{2m}\nabla^2+v_C(\vec{r})\right]G(\vec{r},\vec{r}\,';E)+\int d\vec{r}\,''M(\vec{r},\vec{r}\,'';E)G(\vec{r}\,'',\vec{r}\,';E)=-\delta(\vec{r}-\vec{r}\,')\ , \quad (1.2)$$

where M is the mass operator. We separate the "classical" Coulomb potential v_C (arising from ionic and electronic Hartree contributions) from M and treat it exactly. The remaining mass operator arises from exchange and correlation among the electrons.

A formal solution to the Dyson equation is given by

$$G(\vec{r},\vec{r}\,';E)=\sum_n \frac{\chi_n(\vec{r},E)\chi_n^\dagger(\vec{r}\,',E)}{E-\epsilon_n(E)}\ , \quad (1.3)$$

where χ_n and $\chi_n^\dagger$ are the right ("column") and left ("row") eigenvectors, respectively,

$$\left[-\epsilon_n(E)-\frac{1}{2m}\nabla^2+v_C(\vec{r})\right]\chi_n(\vec{r},E)+\int d\vec{r}\,'M(\vec{r},\vec{r}\,';E)\chi_n(\vec{r}\,',E)=0\ , \quad (1.4)$$

$$\left[-\epsilon_n(E)-\frac{1}{2m}\nabla^2+v_C(\vec{r})\right]\chi_n^\dagger(\vec{r},E)+\int d\vec{r}\,'\chi_n^\dagger(\vec{r}\,',E)M(\vec{r}\,',\vec{r};E)=0\ , \quad (1.5)$$

corresponding to the eigenvalue $\epsilon_n(E)$, which is complex in general. In these Schrödinger-type equations (1.4) and (1.5), M assumes the role of a complex, non-Hermitian, energy-dependent, nonlocal potential operator.

The terminology "nonlocal" deserves comment and clarification. A *local* potential is a function of a single variable $\vec{r}$ alone and operates multiplicatively on the wave function. In general, any more general operator may be referred to as *nonlocal*. For example, energy-dependent pseudopotentials of the form $V_{\text{ps}}(\vec{r},E)$ are often called nonlocal. However, we will refer to this form as an *energy-dependent local* potential, and reserve the term *nonlocal* for those operators such as $M(\vec{r},\vec{r}\,';E)$ which depend on two real-space variables and therefore require a nontrivial real-space integral in Eqs. (1.4) and (1.5). (It need not be energy dependent.) Finally, in attempting to go beyond the LDA, *non-local-density* (NLD) potentials are constructed. This is distinct from the nonlocal operator defined above; for example, for the homogeneous electron gas [$n(\vec{r})=\text{const}$] M is still a nonlocal operator. Conversely, the non-local-density potential of Langreth and Mehl[5] is a local operator since it depends on $n(\vec{r})$ and its derivatives only at a single point in space.

Hedin[8] provided the first approximate evaluation of M for the electron gas in the Green's-function—screened-Coulomb-interaction ("GW") approximation, which is described in Sec. II. Several variants of the GW approximation have been applied to investigate the gaps in semiconductors. The "Coulomb hole plus screened exchange" (COHSEX) model of Hedin[8] has been applied by Brinkman and Goodman,[9] Lipari and Fowler,[10] Kane,[11] Brener,[12] and Bennett and Inkson.[13] The COHSEX model consists primarily of a static approximation for M (evaluated at the chemical potential) which neglects renormalization of the single-particle spectrum by electron-hole—pair excitations, but includes the effects of virtual excitation of plasmons. More recently, Strinati *et al.*[14] have utilized a localized-orbital representation to provide a more sophisticated evaluation of M in the dynamic GW approximation. Each of these calculations present M as a correction to the HF approximation, and, typically, a sizeable correction to the gap is found. Fulde and collaborators[15] have approached the problem from the viewpoint of correlated wave functions, both for the ground state and for excited states.

These studies must be contrasted with the situation when M (minus its value at the Fermi level) is applied in metals as a correction to the LDA. Reported eigenvalue changes[16] are usually less than 0.1 eV. Several considerations invite an adaptation of this historically metallic based approach to insulators. First, the magnitude of the correction in the fundamental gap is much smaller—1.8 eV (upward) for the LDA versus 7.5 eV (downward) for the smallest direct gap at Γ in diamond (see Sec. V), for example. The approximations which must be made (Secs. II–IV) will then produce smaller absolute errors, or, alternatively, more simplifying approximations may be made in attaining the same accuracy. Second, the initial LDA (or its NLDA derivative) calculation produces quite realistic ground-state properties; particularly important from our point of view is the ground-state charge density which is needed to evaluate M. Third, building a rigorous theory of excited states on a ground-state formalism allows a more direct investigation and understanding of the excitation process.

In this paper we develop a local-density approximation for the excited states in semiconductors and insulators. A brief report[17] of our results has already been published. The mass operator is evaluated in the GW approximation of Hedin discussed in Sec. II. Section III is a description of our point of departure, for which the electron gas is replaced by a *model semiconducting electron gas* whose dielectric response corresponds to a simplified analytic Penn model[18] introduced recently by Levine and Louie.[19] In Sec. IV we present calculations of M for homogeneous systems. The semiconductor mass operator is compared to the metallic analog studied in detail by Hedin, Lundqvist, and Lundqvist.[20] This leads to a discussion of the local-density condition[21] for quasiparticles, whereby it is shown that, within our model, (a) the gap typically increases the eigenvalue correction by an order of magnitude, and (b) the true gap will always be *larger* than the one obtained from ground-state DF calculations. It is also shown that, within the present theory, M for a semiconducting system differs *qualitatively* from its metallic counterpart, even for an extremely small gap. This nonanalytic change appears to be related to the discontinuity of the (exact) exchange-correlation potential with respect to electron number at that number which exactly fills the valence bands, discussed recently by Perdew and Levy[22] and by Sham and Schlüter.[23]

In Sec. V we apply our model to silicon and diamond. Compared to the ground-state eigenvalues, the present model represents a substantial improvement. For example, we find the direct band gaps in Si widen by 0.5 to 0.8 eV over the LDA value, resulting in close agreement with experiment. The remaining discrepancies and their possible origin are discussed. Section VI is devoted to a critical review.

II. M IN THE GW APPROXIMATION

Formally, M is given in terms of the renormalized Green's function G, the screened Coulomb interaction W, and the vertex function Γ, all of which must be determined simultaneously and self-consistently. Even for the relatively simple homogeneous electron gas this problem has never been solved. Hedin[8] suggested keeping only the first term in an expansion for M_h (h denotes homogeneous) in terms of fully renormalized electron and Coulomb-interaction propagators, G and W, respectively, thereby implicitly assuming $\Gamma \sim 1$. While Hedin has given plausibility arguments why this may be a reasonable approximation for M, and some perturbative investigations[24] have been made, its accuracy and range of validity remain unclear.

Hedin's GW approximation for the Fourier transform $\mathcal{M}_h$, then, is (with $\hbar = 1$)

$$\mathcal{M}_h(k,E) = \frac{i}{(2\pi)^4} \int_{-\infty}^{\infty} d\omega' \int d^3k'\, G(\vec{k}-\vec{k}', E-\omega') \times W(k',\omega') e^{i\omega' 0^+} , \tag{2.1}$$

$$W(k,\omega) = \frac{v(k)}{\epsilon(k,\omega)} = \frac{4\pi e^2}{\Omega k^2 \epsilon(k,\omega)} . \tag{2.2}$$

Here, v is the Fourier transform of the Coulomb potential, ϵ is the dielectric function, and Ω is the normalization volume. The Green's function itself is defined in terms of $\mathcal{M}_h$ by

$$G(\vec{k},E) = [E - E_{\vec{k}} - \mathcal{M}_h(k,E)]^{-1} . \tag{2.3}$$

As noted by Hedin, it is a very difficult numerical problem to include the full (k,ω) dependence of $\mathcal{M}_h$ in the Green's function appearing under the integral in Eq. (2.1), and the resulting behavior of $\mathcal{M}_h$ is not expected to be sensitive to fine details in the integrand. We therefore follow Hedin by approximating $\mathcal{M}_h$ in Eq. (2.3) by its value at the chemical potential $\mathcal{M}_h(k_F,\mu_h) = \mu_{xc}$, the exchange-correlation contribution to the chemical potential μ_h.

The expression to be evaluated is

$$\mathcal{M}_h(k,E) = \frac{i}{(2\pi)^4} \int_{-\infty}^{\infty} d\omega' \int d\vec{k}' \frac{v(k')}{\epsilon(k',\omega')} \frac{e^{-i\omega' 0^+}}{E - \omega' - E_{\vec{k}-\vec{k}'} + i\delta\, \mathrm{sgn}(|\vec{k}-\vec{k}'| - k_F)} . \tag{2.4}$$

This expression has been separated by Hedin into the static "Coulomb hole plus screened exchange" part $\mathcal{M}_h^C$, and the term involving dynamic screening $\mathcal{M}_h^d$:

$$\mathcal{M}_h^C(k) = \frac{e^2}{\pi} \left[\int_0^{\infty} dk' [\epsilon^{-1}(k',0) - 1] \int_0^{\infty} dk' \int_{-1}^{1} d\xi\, \epsilon^{-1}(k',0) \Theta(\mu - E_{\vec{k}-\vec{k}'}) \right] , \tag{2.5}$$

$$\mathcal{M}_h^d(k,E) = \frac{e^2}{\pi} \int_0^{\infty} dk' \int_{-1}^{1} d\xi \left\{ [\epsilon^{-1}(k', E - E_{\vec{k}-\vec{k}'}) - \epsilon^{-1}(k',0)][\Theta(E - E_{\vec{k}-\vec{k}'}) - \Theta(\mu - E_{k-k'})] - \frac{1}{2\pi} \int_{-\infty}^{\infty} dw [\epsilon^{-1}(k',iw) - \epsilon^{-1}(k',0)] \frac{1}{E - E_{\vec{k}-\vec{k}'} - iw} \right\} \tag{2.6}$$

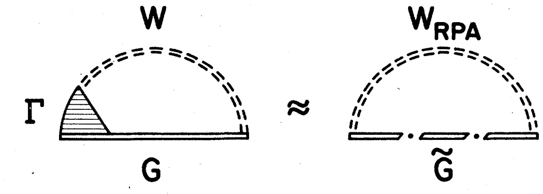

FIG. 1. Schematic representation of the GW approximation to the Coulomb self-energy.

$[\zeta=\vec{k}\cdot\vec{k}'/(|\vec{k}|\,|\vec{k}'|)$ and Θ is the unit step function]. The remaining integrals must be computed numerically. These expressions have been written in terms of a general unrenormalized, spherically symmetric spectrum $E_{\vec{k}}$, which will be discussed further in the next section. Denoting the corresponding Green's function by $\tilde{G}$, the relevant approximation for $\mathscr{M}_h$ is pictured schematically in Fig. 1.

III. SEMICONDUCTING HOMOGENEOUS ELECTRON GAS

To generalize the treatment of Hedin and Lundqvist[8,20] to a system with a gap between occupied and unoccupied states, both the dielectric function and the single-particle spectrum appearing in G under the integral (i.e., $\tilde{G}$) in Eq. (2.1) must be modified. To model semiconducting or insulating behavior in the dynamical screening, we use the dielectric function $\epsilon^{\rm LL}=\epsilon_1^{\rm LL}+i\epsilon_2^{\rm LL}$ of Levine and Louie[19] (LL), which is defined in terms of the random-phase-approximation (RPA) dielectric function $\epsilon^{\rm RPA}$ of Lindhard:[25]

$$\epsilon_2^{\rm LL}(k,\omega)=\begin{cases}0, & |\omega|<\lambda E_F\\ \epsilon_2^{\rm RPA}(k,\omega_-), & |\omega|>\lambda E_F\end{cases} \qquad (3.1)$$

where

$$\omega_-=(\omega^2-\lambda^2E_F^2)^{1/2}\,{\rm sgn}\omega\ ,$$

and $\epsilon_1^{\rm LL}(k,\omega)$ can be obtained analytically[19] by Kramers-Kronig transformation. This function satisfies the important f-sum rules on both ϵ and ϵ^{-1} and reduces to $\epsilon^{\rm RPA}$ for $\lambda\rightarrow 0$. The dielectric gap λE_F, which will be larger than the minimum direct gap, is determined from the experimental value of the static, long-wavelength dielectric constant ϵ_0,

$$\lambda E_F=\omega_p/(\epsilon_0-1)^{1/2},\quad \omega_p^2=4\pi ne^2/m\ . \qquad (3.2)$$

Figure 1 of Ref. 17 illustrates that $\epsilon^{\rm LL}$ provides a reasonable model of both the k and ω behavior of screening in Si ($r_s=2$, $\lambda=0.4$).

For the approximate single-particle spectrum $E_{\vec{k}}$ appearing in G, we introduce a gap E_g between occupied and unoccupied states in an otherwise free-particle spectrum,

$$E_{\vec{k}}=\frac{k^2}{2m}+\frac{E_g}{2}{\rm sgn}(|\vec{k}|-k_F)\ . \qquad (3.3)$$

This spectrum, which was used much earlier by Callaway,[26] is contrasted with the free-particle spectrum in Fig. 2. E_g is distinct (numerically) from λE_F, but is certainly the same order of magnitude. In the application of the self-energy to solids (Sec. V), we take the viewpoint that E_g is an average over the Brillouin zone of the direct gap between occupied and unoccupied states. Therefore in solid-state calculations E_g is determined self-consistently and λ is taken from ϵ_0 and thus no free parameters appear.

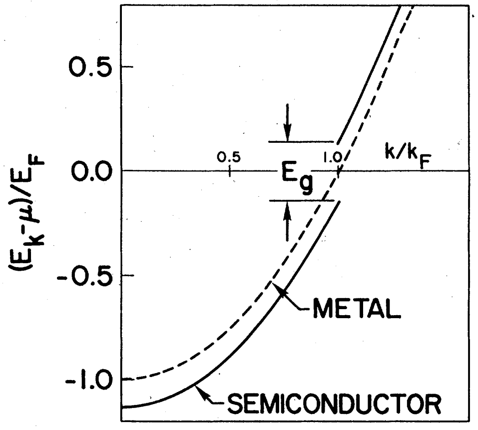

FIG. 2. Spectrum of the model homogeneous semiconductor, compared to its metallic counterpart.

IV. LOCAL-DENSITY APPROXIMATION

Sham and Kohn[21] considered the case of a system with a slowly varying density $n(r)$. Since $M(\vec{r},\vec{r}\,';E)$ is short range[21] in $|\vec{r}-\vec{r}\,'|$, it depends only on the density near $(\vec{r}+\vec{r}\,')/2$. This suggests that it can be approximated by the replacement

$$M(\vec{r},\vec{r}\,';E)\rightarrow M_h(\vec{r}-\vec{r}\,',E-\mu+\mu_h(\bar{n});\bar{n})\ , \qquad (4.1)$$

where $M_h(\vec{r}-\vec{r}\,',E;n)$ is the mass operator of the homogeneous electron gas with density n, and $\bar{n}=n[(\vec{r}+\vec{r}\,')/2]$. The local-density approximation is obtained by treating $X(r,E)$ in Eq. (1.4) as a superposition of locally plane-wave-like functions,

$$X(\vec{r},E)\approx e^{i\vec{k}(\vec{r},E)\cdot\vec{r}}\ . \qquad (4.2)$$

Inserting this expression into Eq. (1.4) and neglecting nonlocal terms (i.e., the $\vec{r}$ dependence of $\vec{k}$) gives the solution $\epsilon(E)=E$ if k satisfies the *local-density condition* $k=k_{\rm LD}$, with

$$-E+\frac{1}{2m}k_{\rm LD}^2+v_C(r)+\mathscr{M}_h(k_{\rm LD},E-\mu+\mu_h(n);n)=0\ , \qquad (4.3)$$

where μ and μ_h are the chemical potentials of the inhomogeneous and homogeneous systems, respectively. Using the Thomas-Fermi–type relation, valid for slowly varying density,

$$v_C(r)=\mu-\mu_h(n(r))\ , \qquad (4.4)$$

and the definition

$$\mu_h(n)=\frac{1}{2m}k_F^2(n)+\mu_{\mathrm{xc}}(n)$$
$$=\frac{1}{2m}k_F^2(n)+\mathcal{M}_h(k_F,\mu_h(n);n) \quad (4.5)$$

gives finally

$$(k_{\mathrm{LD}}^2-k_F^2)/2m=(E-\mu)-[\mathcal{M}_h(k_{\mathrm{LD}},E-\mu+\mu_h(n);n)$$
$$-\mathcal{M}_h(k_F,\mu_n(n);n)] \quad (4.6)$$

as the condition defining the local-density wave vector

$$k(\vec{r},E)=k_{\mathrm{LD}}(E,n(\vec{r}))\ .$$

When operating on the plane wave (4.2), the effect of the *nonlocal* operator $M_h(r-r',E)$ can be reproduced *exactly* by the *local* operator

$$\mathcal{M}(k_{\mathrm{LD}}(E,n(\vec{r})),E;n(\vec{r}))\delta(\vec{r}-\vec{r}\,')\ . \quad (4.7)$$

The *local-density approximation for the mass operator,* which we call the quasiparticle local-density approximation (QPLDA), consists of using this local operator when operating on the actual wave function at energy E. Since $M_h(\vec{r}-\vec{r}\,',E)$ is a short-range operator, we expect this approximation to be a reasonable one beyond the regime (slowly varying density) of its formal validity.

Since considerable sophistication has been applied to the calculation of the exchange-correlation potential

$$\mu_{\mathrm{xc}}=\mathcal{M}_h(k_F,\mu_h;n)$$

for the electron gas, it is useful to write

$$\mathcal{M}_h(k,E;n)=\mu_{\mathrm{xc}}(n)+[\mathcal{M}_h(k,E;n)-\mathcal{M}_h(k_F,\mu_h;n)]$$
$$\equiv\mu_{\mathrm{xc}}(n)+\Delta(k,E;n)\ , \quad (4.8)$$

where Δ will be called the self-energy. This approach has two advantages. First, the low-order GW approximation is used only for the small quantity Δ rather than for the large quantity $\mathcal{M}_h$, resulting in greater accuracy in the results. Second, the exchange-correlation potential μ_{xc} is already included in standard DF ground-state calculations, making it simple to append Δ to obtain a correction to the ground-state eigenvalues.

Hedin and Lundqvist[27] provided numerical results for $\mathrm{Re}\Delta_{\mathrm{LD}}$ in the QPLDA, where

$$\Delta_{\mathrm{LD}}(r_s,E)=\Delta(k_{\mathrm{LD}},E;r_s)\ ,$$

a decade ago. [In the following we revert to the electron Wigner-Seitz radius $r_s=(4\pi n/3)^{-1/3}$ rather than the density n.] The few applications of their results to real metals have given very small corrections to the ground-state bands; for Ni, for example, corrections were always smaller than 0.1 eV (Watson *et al.*, Ref. 16).

The origin of the small QPLDA corrections is evident from Fig. 3(a), where a contour plot of $\mathrm{Re}\Delta(k,E;r_s=2)$ is shown in a region of k,E including $k=k_F$, and $E=\mu_h$. For typical, *unrelated* values of k and E, Δ is of the order of several electron volts. However, along the LD line $k=k_{\mathrm{LD}}(E,r_s)$, Δ is of the order of a few $\times 10^{-2}$ eV. In Fig. 3(a), in fact, k_{LD} cannot be distinguished from the zero-contour line $k=k_0(E,n)$ satisfying

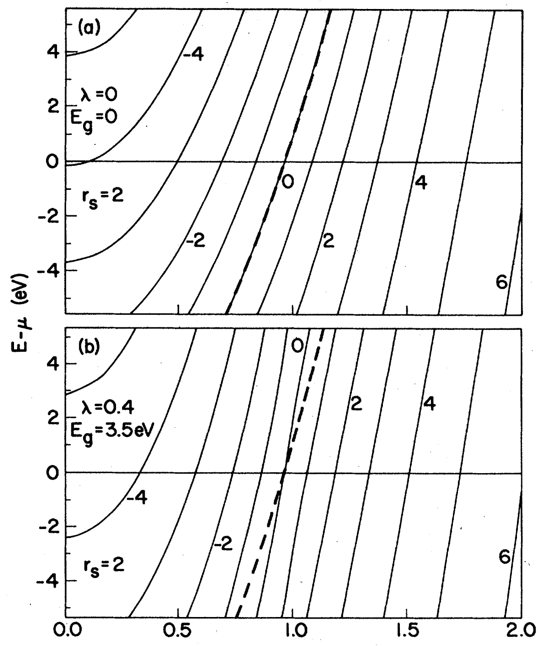

FIG. 3. Contour plot of the real part of the nonlocal self-energy $\Delta(k,E)$ (in eV) for $r_s=2$ for (a) the metallic case and (b) a semiconductor characteristic of Si. The dashed line traces the local-density condition.

$$\Delta(k_0,E;n)=0\ . \quad (4.9)$$

Both k_{LD} and k_0 must pass through k_F at $E=\mu$.

The situation is clarified upon comparison with the corresponding contour plot in Fig. 3(b) for a semiconductor with $\lambda=0.4$ and $E_g=3.5$ eV (parameters characteristic of silicon). The LD line, shown as the dotted line, is similar to that in the metal. However, the gap in the semiconductor has caused the zero-contour line k_0 to shift away from the LD line k_{LD}. The slope of k_0 in the (k,E) plane is given by

$$\left[\frac{dk_0}{dE}\right]^{-1}=-\left[\frac{\partial\Delta}{\partial k}\Big/\frac{\partial\Delta}{\partial E}\right]_{k=k_0}\ , \quad (4.10)$$

and the gap increases $\partial\Delta/\partial k$, while it decreases $|\partial\Delta/\partial E|$, each changing by $\sim 30\%$ for the gap characteristic of Si. As discussed in more detail elsewhere,[28] the increase in $d\Delta_{\mathrm{LD}}/dE$, given by

$$\frac{d\Delta_{\mathrm{LD}}}{dE}=\frac{d}{dE}\Delta(k_{\mathrm{LD}}(E),E)=\left[\frac{\partial\Delta}{\partial k}\frac{dk_{\mathrm{LD}}}{dE}+\frac{\partial\Delta}{\partial E}\right]_{k=k_{\mathrm{LD}}}\ , \quad (4.11)$$

results, in large part, from the decrease in the magnitude of $\partial\Delta/\partial E$, which is negative.

In addition to increasing the magnitude of Δ_{LD}, the gap introduces a qualitative distinction between insulating and metallic systems. Figure 4(a) reveals the (r_s,E) plane to

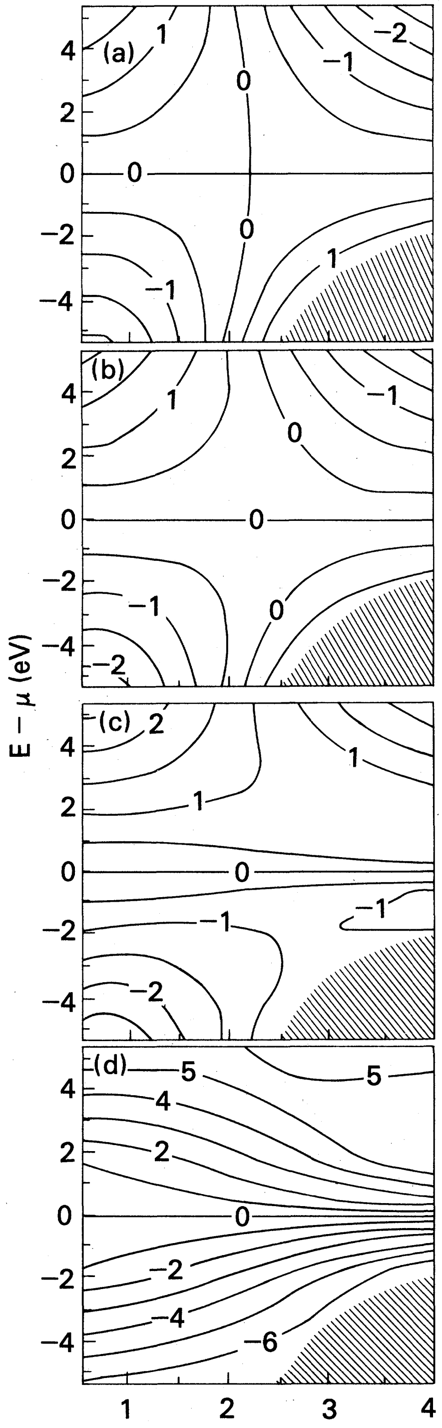

FIG. 4. Contour plots of the QPLDA self-energy $\Delta_{\rm LD}(r_s,E)$, in units of 0.1 eV, for (a) $\lambda=E_g=0$; (b) $\lambda=0.033$, $E_g=0.4$ eV; (c) $\lambda=0.1$, $E_g=1.2$ eV; (d) $\lambda=0.4$, $E_g=3.5$ eV. In the cross-hatched region there is no solution of the local-density condition for real $k_{\rm LD}$.

be divided into quadrants by $\Delta_{\rm LD}$, with $\Delta_{\rm LD}$ positive for $E>\mu, r_s<2$ and $E<\mu, r_s>2$, and $\Delta_{\rm LD}$ negative in the other two quadrants. When a gap is introduced, the *topology* of the contours of $\Delta_{\rm LD}$ is altered such that $\Delta_{\rm LD}$ has the same sign as $E-\mu$ for $|E-\mu|\lesssim E_g$. This is illustrated for a gap characteristic of Si in Fig. 4(d) and for gaps of one-third and one-tenth this size in Figs. 4(c) and 4(b) respectively. Since μ lies in the gap of an insulator, the *valence bands are lowered* and the *conduction bands are raised* by dynamic exchange-correlations corrections, regardless of the density which their wave functions sample. As a result the QPLDA will always lead to an increase in the gap over the LDA value.

To understand this behavior in more detail we differentiate $k_{\rm LD}$ in Eq. (4.6) to obtain

$$\left.\frac{dk_{\rm LD}}{dE}\right|_{E=\mu}=\frac{m}{k_F}\,\frac{1+\lambda_E}{1+\lambda_k}\,, \tag{4.12}$$

where

$$\lambda_E\equiv-\left.\frac{\partial\Delta}{\partial E}\right|_{k=k_F,\,E=\mu}\quad\text{and}\quad\lambda_k\equiv\frac{m}{k}\left.\frac{\partial\Delta}{\partial k}\right|_{k=k_F,\,E=\mu}.$$

Then

$$\left.\frac{d\Delta_{\rm LD}}{dE}\right|_{E=\mu}=\frac{\lambda_k-\lambda_E}{1+\lambda_k}\equiv\Delta'_{\rm LD}(\mu)\,. \tag{4.13}$$

The effect of the gap is to decrease λ_E while it increases λ_k (both are positive), resulting in a larger, or possibly less negative, energy derivative $\Delta'_{\rm LD}$. This does not, however, explain why $\Delta'_{\rm LD}$ is *always positive* for a nonvanishing gap, a property which is indicated by Fig. 4.

The nonanalyticity which guarantees $\Delta'_{\rm LD}>0$ in an insulator can be demonstrated in a simplified QPLDA. We consider the simplification $\tilde{\Delta}$ of Δ in which $\epsilon_{\rm LL}$ is replaced by the plasmon-pole dielectric function $\epsilon_{\rm pp}$, given by

$$\epsilon_{\rm pp}^{-1}(\omega)=1+\frac{\omega_p^2}{\omega^2-\omega_p^2}\,. \tag{4.14}$$

It is straightforward to show that $\tilde{\Delta}$ contains, to lowest order in the gap E_g in the underlying spectrum, a contribution $\delta\tilde{\Delta}$ given by

$$\delta\tilde{\Delta}\sim E_g(k-k_F)\ln|k-k_F|\,. \tag{4.15}$$

This leads to a divergent contribution to λ_k, $\delta\lambda_k\sim E_g\ln|k-k_F|$ with $k\to k_F$, when E_g is nonzero. As a result,

$$\Delta'_{\rm LD}(\mu)=\begin{cases}(\lambda_k-\lambda_E)/(1+\lambda_k), & E_g=0\\ 1, & E_g\neq0\,.\end{cases} \tag{4.16}$$

This behavior is just that shown in Fig. 4. It can be of either sign, depending on r_s, for $E_g=0$, but is positive for all r_s for $E_g\neq0$. This nonanalyticity in E_g is evidently related to the discontinuity[22,23] in $v_{\rm xc}$ for nonzero gap.

V. APPLICATION TO INSULATING CRYSTALS

We have applied this formalism to the QP bands in Si, the prototype semiconductor, and to diamond, a moderately-wide-gap insulator. In both cases we have replaced the ionic core with the ionic norm-conserving pseudopotentials generated according to Hamann *et al.*[29]

These pseudopotentials are nonlocal, being comprised to separate s, p, and d potentials. For Si the published form of Bachelet, Hamann, and Schlüter[29] was used. The diamond pseudopotential is discussed in Sec. V B. For Si (diamond), the cutoff in plane waves (PW's) for expansion of the Hamiltonian was taken at an energy of 11 Ry (38 Ry), resulting in matrices of approximately 170×170 (300×300). The charge density and screening potential were expanded in 609 PW's for Si and 1153 PW's for diamond. Self-consistency to a high degree of accuracy on the 10-special-point grid was achieved within 6–7 iterations.

For the calculation of the self-energy correction it is necessary to determine λ and E_g. For the latter we take the zone-average direct gap, equal to 3.5 eV in Si and 12.75 eV in diamond. This latter value was obtained by adding to the 10-special-$\vec{k}$-point average of the LDA calculation (10.22 eV) in anticipated 2.5-eV self-energy correction. We follow Levine and Louie[19] in using $\lambda=0.4$ for Si. For diamond, $\epsilon_0=5.7$ leads to $\lambda=0.5$.

For both Si and diamond we have carried out three calculations. The first and second consist of self-consistent LDA calculations using Hedin-Lundqvist and Levine-Louie exchange-correlation potentials, respectively. The third (and ultimate) results include QPLDA corrections to the Levine-Louie band structure, which we consider the best currently known approximation to the exact exchange-correlation potential for systems with a gap.

In this section we neglect $\mathrm{Im}\Delta_{\mathrm{LD}}$ (see Sec. VI A), in which case the spectral function becomes sharp. In this case $\mathrm{Re}\Delta_{\mathrm{LD}}$ can be included in existing band-structure codes with little modification, since the "Hamiltonian" remains Hermitian. The only modification arises from the energy dependence of Δ_{LD}. In the calculations for Si we have rediagonalized the energy-dependent QP Hamiltonian iteratively for each QP energy of interest, evaluating Δ_{LD} at the eigenvalue obtained from the previous diagonalization. Three iterations were sufficient to determine an excitation energy to better than 0.01-eV accuracy. For diamond this direct approach becomes very time consuming (three additional 300×300 diagonalizations for each eigenvalue sought), so the following efficient method was developed.

We first use the result (Ref. 17, and below) that most (75–95 %) of the self-energy correction arises from the constant ($\vec{G}=\vec{0}$) term in $\Delta_{\mathrm{LD}}(\vec{G},E)$, the Fourier transform of $\Delta_{\mathrm{LD}}(r_s(\vec{r}),E)$. Therefore we solve the scalar equation

$$\bar{\epsilon}=E+\Delta_{\mathrm{LD}}(\vec{G}=\vec{0},\bar{\epsilon}) \tag{5.1}$$

iteratively for the first approximation $\bar{\epsilon}$ to the QPLDA eigenvalue ϵ corresponding to the LDA eigenvalue E. This requires negligible time. Then, noting that

$$\langle i\,|\,(H_{\mathrm{LDA}}+\Delta_{\mathrm{LD}})\,|\,j\,\rangle=E_i\delta_{ij}+\langle i\,|\,\Delta_{\mathrm{LD}}\,|\,j\,\rangle \tag{5.2}$$

in terms of the eigenvectors $|\,i\,\rangle$ and eigenvalues E_i of the ground-state Hamiltonian H_{LDA}, the Dyson equation $\epsilon=E+\Delta_{\mathrm{LD}}(\epsilon)$ is solved iteratively from Eq. (5.2) beginning with the initial value $\bar{\epsilon}$ from Eq. (5.1). If all of the eigenstates are used in Eq. (5.2), this procedure leads to an exact solution. We have found, however, that only the lowest 20 (or less) eigenstates $|\,i\,\rangle$ are required in (5.2) due to the smallness of Δ_{LD} and to its short range in reciprocal vector space [only its $\vec{G}=(2\pi/a)(1,1,1)$ component is important]. Typically, only two iterations of (5.2) are necessary to determine QPLDA eigenvalues to better than 1 mRy. Using this procedure, six eigenvalues (see Table I) can be determined in less than the time required for a *single* additional diagonalization of the full 300×300 matrix.

A. Silicon

The numerical results for Si have been presented previously.[17] These results are presently graphically in Fig. 5. The (Hedin-Lundqvist[27]) LDA bands are superimposed on the QPLDA bands (self-energy added to Levine-Louie[19] exchange correlation), with chemical potentials (taken to be at gap center) aligned. The separation between valence and conduction bands is increased everywhere. The effect of the Levine-Louie exchange-correlation potential (results *not* shown in Fig. 5) is very small, except for the lower conduction bands, where it raises the L_3 and L_1 eigenvalues by about 0.10 eV and the X_1 eigenvalue (and thereby increases the indirect gap) by 0.15 eV. The self-energy correction is primarily an energy-dependent factor, behaving *roughly* as

$$\Delta E=\zeta(E-\mu),\quad |\,E-\mu\,|\leq 4\text{ eV}\,, \tag{5.3}$$

with $\zeta\simeq\frac{1}{8}$. The energy dependence of ΔE decreases for $|\,E-\mu\,|>4$ eV. Superimposed on this energy dependence is an additional density dependence which lowers the uppermost valence bands and raises the lowest conduction bands, in each case by roughly 0.1 eV. This behavior has been presented in Fig. 2 of Ref. 17.

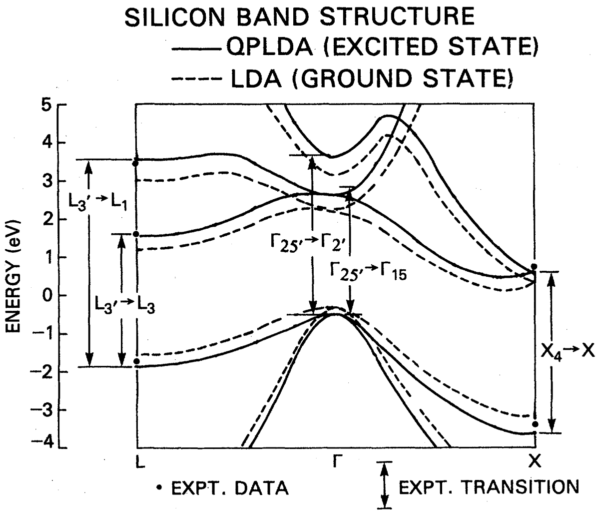

FIG. 5. QPLDA excited-state Si bands (solid lines) superimposed on the LDA ground-state bands (dashed lines) using the exchange-correlation potential of Hedin and Lundqvist. Dots and arrows indicate experimental eigenvalues and transitions, respectively, and were obtained from the literature as described in Ref. 17. The experimental transitions have been placed adjacent to the corresponding initial- and final-state bands.

TABLE I. Eigenvalues (in eV) at high-symmetry points and the indirect gap E_g^{ind} in diamond from different calculations. QPLDA eigenvalues in parentheses were obtained in the mean-density approximation (QPMDA). For this table the convention of setting the zero of energy at $E(\Gamma_{25}^{\prime v})$ has been used. The experimental values are discussed in the text. Notation: ZF (Zunger and Freeman, Ref. 34), GSA (Glotzel, Segall, and Andersen, Ref. 35), BGBS (Bachelet *et al.*, Ref. 37), YC (Yin and Cohen, Refs. 7 and 36), SMH (Strinati, Mattausch, and Hanke, Ref. 14), and Expt. (experiment). Exchange-correlation potentials: HL (Hedin and Lundqvist), W (Wigner), and LL (Levine and Louie). PW denotes plane wave, LO denotes local orbital, and LAPW denotes linearized augmented plane wave.

			Ground state							Excited state	
	ZF (HL)	GSA (HL)		BGBS (W)		YC (W)	Present		QPLDA	SMH	Expt.
			200 PW's	40 LO's	LAPW		HL	LL			
Γ_1^v	−20.44	−21.28	−21.45	−21.68	−21.36	−21.24	−21.38	−21.18	(−23.4)	−25.2	−24.0
$\Gamma_{25}^{\prime v}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Γ_{15}^c	6.33	5.55	5.40	5.59	5.66	5.51	5.51	5.63	7.36	7.4	7.3
$\Gamma_2^{\prime c}$	14.07	13.72	13.38	13.21	13.34	13.62	13.56	13.61	16.10	17.6	15.3
X_1^v	−12.17	−12.59	−12.65	−12.90	−12.61	−12.49	−12.67	−12.54	(−14.1)	−14.75	
X_4^v	−6.09	−6.18	−6.22	−6.43	−5.82	−6.23	−6.31	−6.21	−7.39	−7.9	
X_1^c	8.60	4.78	4.63	4.65	4.78	4.77	4.64	4.95	6.41	7.45	
X_4^c	16.69	17.01	16.73	16.87	17.30	16.45	16.81	16.80	19.52	21.35	
$L_2^{\prime v}$	−15.17	−15.43	−15.57	−15.79	−15.53	−15.36	−15.53	−15.41	(−17.2)	−19.7	
L_1^v	−12.18	−13.35	−13.35	−13.73	−13.27	−13.26	−13.47	−13.25	(−14.9)	−15.0	
$L_3^{\prime c}$	−2.82	−2.80	−2.81	−2.86	−2.44	−2.76	−2.81	−2.78	−3.09	−2.8	
L_1^c	10.51	9.37	8.92	8.90	8.90	9.11	8.97	9.10	11.26	12.6	
L_3^c	10.39	8.55	8.31	8.47	8.56	8.34	8.37	8.54	10.55	11.0	
E_g^{ind}	5.45	~4.5				4.15	4.05	4.31	5.74		5.47

In Fig. 5 experimentally inferred values of the excitations at X and L are shown as dots, *relative to the QPLDA valence-band maximum.* From this comparison discrepancies of the order of 0.2 eV remain. A table of eigenvalues has been published previously.[17] Also shown in Fig. 5 are experimentally determined transitions, which are simply juxtaposed with the corresponding region of the QPLDA band structure. These transition energies are less open to inaccuracies due to interpretation than are the corresponding initial- and final-state eigenvalues. The $X_4 \rightarrow X_1$, $L_3' \rightarrow L_3$, and $L_3' \rightarrow L_1$ zone-boundary transitions are reproduced extremely accurately by the QPLDA. The zone-center transition $\Gamma_{25'} \rightarrow \Gamma_{15}$ (the worst case) is underestimated by about 0.25 eV. The $\Gamma_{25'} \rightarrow \Gamma_{2'}$ transition also is reproduced by QPLDA. However, this transition involves the unusual antibonding $\Gamma_{2'}$ state, which is inordinately sensitive to computational details and which other *ab initio* LDA calculations[19,30] find to lie about 0.3 eV lower than our result. (Except for the $\Gamma_{2'}$ state, our bands are very close to previous calculations.[19,30]) Thus the perfect agreement here may be somewhat fortuitous. It is clear, however, that the QPLDA self-energy gives essential increases in transition energies which bring them into quantitative agreement with experiment. The calculated indirect gap of 0.93 eV, although much improved over the LDA value, still underestimates the experimental value by 0.24 eV. This discrepancy suggests some remaining density dependence, either in the exchange-correlation potential or in the self-energy, which has not been accounted for in the present theory. The $\Gamma_{25'}$ and X_1 states are known, from contour plots by Kerker[4] and by Zunger and Cohen,[30] to sample very different regions of density. To quantify this difference we have calculated the average value $\bar{r}_s$ of r_s, as well as the deviation δr_s from the average, for the high-symmetry states near the gap. For the $\Gamma_{25'}$ (X_1) state we find $\bar{r}_s = 1.75$ (2.30) and $\delta r_s = 0.32$ (0.65). Each of these values represents the extremum (minimum and maximum, respectively) for states within 5 eV of the gap, although the energies of these states differ by only about 1 eV. As a result an additional density dependence will affect the separation of these states more than for any other two states. It is for this reason that we assign the discrepancy in the indirect gap to some, as yet unknown, density dependence which may involve nonlocality in the self-energy or nonlocal-density corrections to the ground-state eigenvalues.

B. Diamond

We first review existing self-consistent calculations for diamond. The first self-consistent calculation carried out without making severe shape approximations to the charge density and potential were by Painter *et al.*[31] Using the $X\alpha$ scheme ($\alpha = 0.76$) they obtained an indirect gap of 5.4 eV and a bandwidth of ~21 eV. Euwema *et al.*[32] carried out a Hartree-Fock calculation, which was found to overestimate the fundamental gap by a factor of 2. Heaton and Lafon[33] performed a Slater exchange ("$\alpha = 1$") calculation, obtaining an indirect gap of 5.2 eV and a valence-band width of 21 eV. Zunger and Freeman[34] carried out a Hedin-Lundqvist LDA calculation, finding a valence-band width of 20.4 eV and an indirect gap of 5.5 eV. Their calculation shows a sharp upward dispersion of the lowest conduction band near the conduction-band minimum [near $k = (2\pi/a)(0.75,0,0)$] which is not shown by any other calculation. Using overlapping atomic and interstitial spheres within which the charge density and potential are spherically averaged, Glötzel *et al.*[35] obtain an indirect gap of ~4.5 eV using the Hedin-Lundqvist exchange-correlation potential. The eigenvalues of Zunger and Freeman[34] and of Glötzel *et al.*[35] are presented in Table I.

Of more interest for comparison here are the calculations of Yin and Cohen[7,36] and Bachelet *et al.*[37] Both of these studies used pseudopotentials generated according to the Hamann-Schlüter-Chiang "norm-conserving" prescription; however, each used pseudopotentials which were *generated independently* and represented differently than the Bachelet-Hamann-Schlüter potential. The differences in the resulting self-consistent eigenvalues (computed *exactly)* resulting from these "different" potentials are not known, but they are expected to be negligibly small. Bachelet *et al.*, using the potential shown in Fig. 6, have reported calculations using (a) 200 PW's and (b) 20 localized orbitals per atom, as well as a (full-core) linearized—augmented-plane-wave (LAPW) calculation using the method of Hamann,[38] finding differences in eigenvalues of 0.4 eV or less. The sufficiency of 20 localized orbitals was justified by Bachelet *et al.* upon comparison to their results using 200 PW's. The most highly converged[36] (with respect to basis set) plane-wave calculations are those of Yin and Cohen,[7] who used 350—400 PW's. All of these eigenvalues are presented in Table I.

Initially, we attempted to use the "pseudopotential that

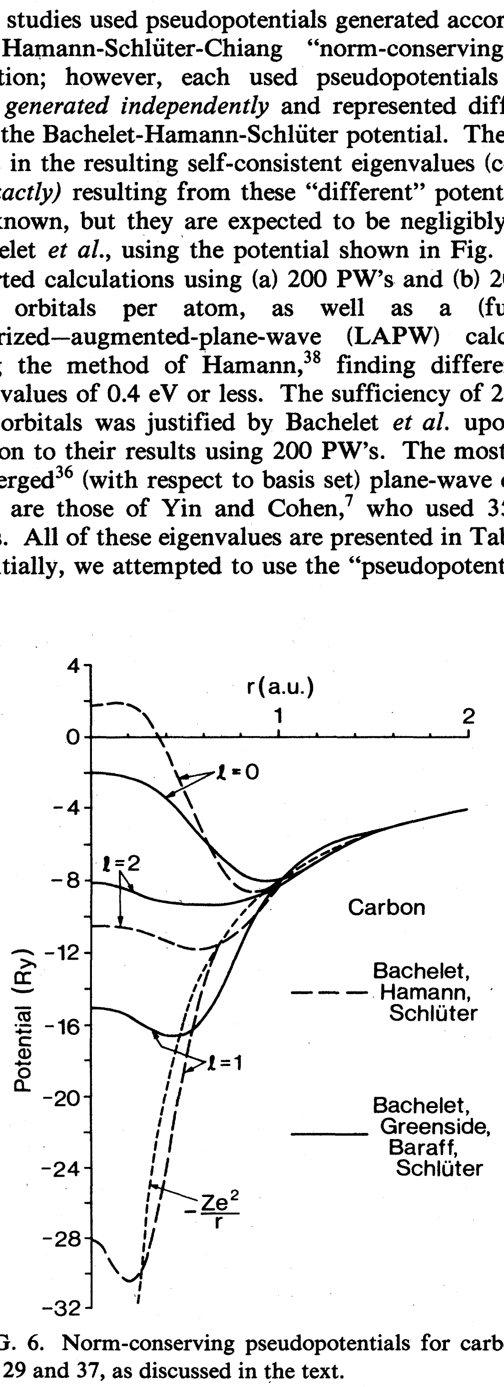

FIG. 6. Norm-conserving pseudopotentials for carbon from Refs. 29 and 37, as discussed in the text.

works" published by Bachelet, Hamann, and Schlüter[29] (BHS), shown as dashed lines in Fig. 6. However, using 300 PW's we obtained eigenvalues which differed from those of Yin and Cohen by as much as 0.5 eV. These discrepancies were due to lack of convergence, which could only be surmounted with an unacceptably large (perhaps 500–600 PW's) basis set. Thus this pseudopotential, chosen with a very conservative $l=1$ core radius, does not really "work" for a plane-wave basis.

In the calculations described below and in Table I we have used the carbon pseudopotential constants published by Bachelet, Greenside, Baraff, and Schlüter.[37] Using 300 PW's we obtain eigenvalues (Table I) similar to both those of Bachelet *et al.*[37] and of Yin and Cohen.[36] In particular, the eigenvalues involved in the smallest direct and indirect gaps are within 0.1 eV of those of Yin and Cohen.

For diamond the (Levine-Louie) LDA bands and the corresponding QPLDA bands are shown in Fig. 7. Here, the difference results *only* from the dynamic self-energy correction. The Levine-Louie correction alone increases the indirect gap by 0.26 eV. Before proceeding to a comparison of the QPLDA bands with experiment, we mention that the self-energy corrections again are primarily energy dependent. Equation (5.1) holds for diamond as well, but with $\zeta \simeq 0.3$.

The experimental indirect gap in diamond[39] is 5.47 eV. The QPLDA leads to the slight (5%) overestimate of 5.74 eV, compared to the large (Hedin-Lundqvist) LDA underestimate of 4.05 eV. The $\Gamma_{25'} \rightarrow \Gamma_{15}$ gap is about[38,39] 7.3 eV, and the QPLDA gives 7.36 eV (compared to the HL value of 5.51 eV). For most other transitions there remains considerable disagreement among the experimental interpretations. Generally, angle-resolved photoemission results have been considered as one of the most reliable standards for comparison, and Himpsel *et al.*[41] have reported a limited study for diamond. Their value of (6.0±0.2) eV for the $\Gamma_{25'} \rightarrow \Gamma_{15}$ transition disagrees considerably with all other data, and their quoted valence-band width of 21 eV is smaller than the 24-eV value found by McFeely *et al.*[42] They assign a 15.3-eV value to the $\Gamma_{25'} \rightarrow \Gamma_{2'}$ transition, for which the QPLDA value is 16.10 eV.

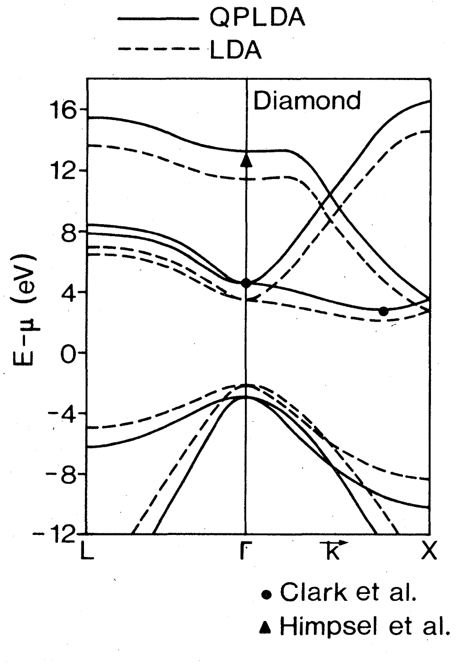

FIG. 7. QPLDA bands for diamond (solid lines) superimposed on the ground-state bands (dashed) lines using the Levine-Louie exchange-correlation potential. The experimental data, *placed relative to the QPLDA valence-band maximum,* are from Clark *et al.* (Ref. 38) and Himpsel *et al.* (Ref. 40).

VI. DISCUSSION

Up to this point we have defined the quasiparticle local-density theory and presented the principal results for silicon and diamond. In this section we provide a general critical discussion, aimed at understanding the QPLDA self-energy and its possible limitations, suggesting refinements which are possible within the local-density approximation and providing a more intuitive notion of the origin of the energy-dependent potential. We also compare our results with the recent work of Strinati *et al.*[14] and Horsch *et al.*[15] on exchange-correlation effects on excitations in insulators.

A. Negative energies and the valence-band width

The reader will not have failed to notice that the presentations of the QPLDA bands in silicon and diamond in Figs. 5 and 7 do *not* include the lower part of the valence bands. For these bands the hole excitation energy $E-\mu$ is sufficiently negative that the quasiparticle local-density condition, Eq. (4.6), has no solution *for positive* $k_{\rm LD}^2$ (i.e., for real $k_{\rm LD}$) for the low-density region in the diamond-lattice interstices. The solution to this situation is simple and was foreseen by Sham and Kohn:[21] simply analyti-

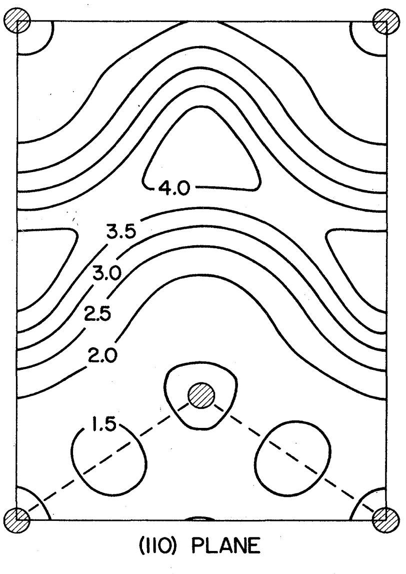

FIG. 8. Contour plot of r_s in the (110) plane of silicon. Numbers label the contour values and dashed lines connect the atom sites.

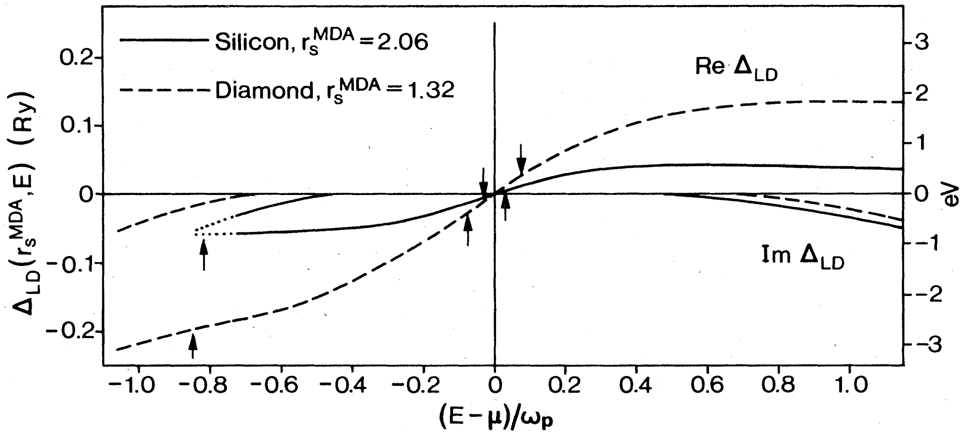

FIG. 9. Real and imaginary parts of the self-energy $\Delta_{\rm LD}$ of silicon and diamond in the mean-density approximation (MDA). Solid (dashed) lines refer to silicon (diamond), and the dotted lines indicate an extrapolation. Arrows indicate band edges (valence-band minimum and fundamental band gap).

cally continue the mass operator to complex momentum. We have not yet carried out this task.

The situation can be visualized using Fig. 8, which displays a contour plot of r_s, proportional to (density)$^{-1/3}$, for Si. In Si, r_s ranges from 1.4 in the bond center to 4.4 at the interstitial site of tetrahedral symmetry (the corresponding extrema in diamond are 0.9 and 2.1). The corresponding value of $k_F^2/2m$ ranges from 25 eV down to 2.5 eV. When $E-\mu$ drops too far below $k_F^2(r_s)/2m$, our present codes cannot assign a value to $\Delta_{\rm LD}(r_s,E)$. The strongly bonding states in the upper valence bands do not sample the interstitial region appreciably, which allows accurate self-energies to be obtained down to $E-\mu\approx-5$ eV in Si and down to ~-10 eV in diamond. We emphasize that this is merely a constraint imposed by the present computer codes rather than a fundamental limitation on the QPLDA.

The QPLDA self-energy is dominated by a position-independent contribution; the only appreciable spatial dependence arises from a much smaller part with $\vec{G}=(2\pi/a)(1,1,1)$ (which also tends to *increase* the indirect gap). This suggests the use of a quasiparticle mean-density approximation (QPMDA) to obtain band shifts in the lower valence bands. We have used the values $r_s^{\rm MDA}=2.06$ (1.32) for silicon (diamond) to obtain the shift in the lower-valence-band eigenvalues. The corresponding self-energies over the range $|E-\mu|\leq\omega_p$ are shown in Fig. 9.

For Si the QPMDA gives a downward shift of the Γ_1 eigenvalue of 0.8 eV, leading to a valence-band width W of 12.5 eV, in excellent agreement with the experimental data of Grobman and Eastman[43] [(12.4±0.6) eV] and of Ley *et al.*[44] [(12.6±0.6) eV]. For diamond the shift is 2.2 eV (see Table I), leading to $W=23.4$ eV. This is in good agreement with the photoemission results of McFeely *et al*;[42] who found $W=(24.2\pm1.0)$ eV. Himpsel *et al.*[41] have reported the value $W=21$ eV, which we believe to be less soundly based.

From Fig. 9 the QPMDA value of the half width of the Γ_1 state, given by $|{\rm Im}\Delta_{\rm LD}(E_{\Gamma_1})|$, is 0.7 and 0.25 eV for silicon and diamond, respectively. Unfortunately, there seems to be no experimental determination of these widths. The QPMDA gives a vanishing width for hole excitations below about 7.0 and 16.5 eV in silicon and diamond, respectively. In the QPLDA the variation in density sampled by these hole states will lead to a finite lifetime for substantially lower energy excitations, although the lifetime should remain long. Since the Γ_1 states have a very diffuse density (especially in silicon) the QPMDA result quoted above may be similar to the QPLDA value.

The ground-state charge density, usually determined from the Hohenberg-Kohn-Sham ground-state procedure, can be determined equally well from the single-particle Green's function by the relation

$$n(r)=-\frac{2}{\pi}\int_{-\infty}^{\mu}dE\,{\rm Im}G(r,r;E)\ . \tag{6.1}$$

In the QPMDA (and ignoring the imaginary part of the self-energy) the energies of the single-particle eigenfunctions are shifted, but the eigenfunction is left unchanged. Thus the QPMDA leads "trivially" to the same density as the ground-state density functional gives, upon which the QPLDA (and the QPMDA) is built. Of course, both prescriptions, if carried out exactly, must lead to the same density. It remains an intriguing question whether a fully-self-consistent QPLDA will lead to appreciable corrections to the LDA density.

B. Comparison with other work

Strinati, Mattausch, and Hanke[14] (SMH) have used a localized-orbital representation to evaluate the *crystalline* mass operator (minus the exchange term, which is included in their reference system) also in the GW approximation. M is then used to give the necessary corrections to a reference Hartree-Fock band structure. As noted in the Introduction, Hartree-Fock overestimates the direct gap in diamond by about 7.5 eV, i.e., a factor of 2. Thus the dynamic corrections necessary to reproduce the experimental gap are 4 times as large as in our approach (and of the opposite sign). Furthermore, the computational utility of their theory is most suitable for (and may be restricted to) systems which are easily described by localized orbitals, i.e., wide-gap solids. For diamond with a minimal basis set, numerical simplifications were made to make the calculations tractable. No corresponding numerical approximations are necessary in the QPLDA.

The approach of SMH does, however, have the merit that if the numerical procedures are carried to convergence, the resulting mass operator is the *appropriate one for the crystal* [subject to the GW approximation and to independent knowledge of what interactions to include in $\epsilon(q,\omega)$]. Therefore the approach is amenable to systematic improvement. The QPLDA, on the other hand, is an *ansatz* (just as the LDA is), and "higher-order" corrections may be difficult to obtain (see Sec. VI. D).

Some features of the two approaches are readily reconciled. For example, the inhomogeneity of the density is included by SMH as "crystalline local field effects" [matrix nature of $\epsilon(q,\omega)$], where the QPLDA accounts for it through a local-density-dependent self-energy. As corrections to the gaps of the reference system are of opposite sign, however, it is difficult to compare the "corrections" in the two approaches. A large part of the corrections to Hartree-Fock bands (reference system for SMH) are included in an average manner in the ground-state density-functional bands (reference system for the QPLDA).

Nevertheless it is easy to compare the results, the quasiparticle energies. Strinati, Mattausch, and Hanke obtain a direct gap at Γ of 7.4 eV and a valence-band width of 25.2 eV, in good agreement with experiment. They also find a Γ_1 half width of the order of 1 eV, considerably larger than our (QPMDA) estimate of 0.25 eV. It would be of great use to have an angle-resolved-photoemission and/or bremstrahlung isochromat spectrum study of diamond to assess the accuracy of higher QPLDA and SMH eigenvalues, listed in Table I, which differ by as much as 2.5 eV.

Horsch, Horsch, and Fulde[15] (HHF) have approached the quasiparticle problem using a correlated wave-function method. Their approach, also based on the Hartree-Fock reference system, included only interatomic correlations (which are expected to be dominant) within a minimal basis set of Gaussian orbitals. Horsch, Horsch, and Fulde obtained a direct gap of 7.4 eV and a valence-band width of 23.8 eV, again similar to experiment. Again, it can be said that comparison of the contributions to the gap with the QPLDA and with the theory of SMH is difficult. Horsch, Horsch, and Fulde conclude, however, that the major correlations in their theory are due to an electron-hole polarization cloud, in qualitative agreement with current implementations of the Green's-function approach by SMH and ourselves.

Sterne and Inkson[45] have applied a tight-binding representation to the GW mass operator to derive approximate exchange-correlation potential separately for the valence and conduction bands. Their approach appears to give approximately the correction necessary to account for experimental gaps.

The COHSEX approximation (see the Introduction), which essentially consists of using the static screening limit $\omega \to 0$ in $\epsilon(q,\omega)$, has been applied to diamond in varying degrees of numerical complexity.[9-12] Strinati, Mattausch, and Hanke[14] have summarized this approach as leading to approximately the correct gap, but leaving the Hartree-Fock bandwidth almost unchanged.

C. Interpreting the self-energy

Naturally the question arises of how one is to interpret the self-energy corrections in order to have, or to develop, an intuitive understanding which can be applied to other systems. Sometimes the question is posed from an anthropomorphic viewpoint: If I were in excited electron (or hole) moving through the crystal, how would I see the other electrons behaving? In this subsection we present calculations aimed at understanding the effect of the energy gap on exchange-correlation processes involved in excitations.

Levine and Louie[19] have shown that their model dielectric function, used here in the QPLDA self-energy, leads to a ground-state exchange-correlation functional which differs from its metallic counterpart by arising from a smaller exchange-correlation hole. To investigate the real-space manifestations of exchange-correlation effects, we turn to the nonlocal mass operator $M_h(\vec{r}-\vec{r}\,',E)$ of our homogeneous insulator, which is the Fourier transform of $\mathscr{M}_h(k,E)$. Hedin[8] has found, for the electron gas, that

$$M_h(r-r',E) \sim -\text{const} \times \frac{1}{|\vec{r}-\vec{r}\,'|}, \quad |\vec{r}-\vec{r}\,'| \to 0 , \tag{6.2}$$

and M_h quickly vanishes beyond $|\vec{r}-\vec{r}\,'|/r_s = 2$. When the gap is introduced, this gross behavior is unchanged, and a graph such as Hedin[8] presented, but including effects of the gap (the size of silicon's, say), would be *difficult to distinguish from the metallic case.* We therefore focus on *differences,* presented in terms of the dimensionless function

$$U(r/r_s E, r_s; \lambda, \nu) = 4\pi r^2 a_B M_h(r,E;r_s)/E_F , \tag{6.3}$$

where $\nu = E_g/E_F$. This representation builds in the scaling of real-space structure with r_s, and accounts for the volume factor $4\pi r^2$ which arises in the integral in the QP Schrödinger equation (1.2).

In Fig. 10(a) we plot the difference in U between insulating and metallic systems, evaluated at $E=\mu$. The insulator is represented by the constants $\lambda = 0.45$ and $\nu = 0.35$, which are midway between the corresponding values for silicon and diamond. This difference, arising from the gap and only weakly dependent on r_s, is oscillatory and is long ranged compared to $r^2 M_h$ itself. The real-space potential $M_h(r,\mu)$ is somewhat less attractive for $r < r_s$; however, it is considerably more attractive for $1 < r/r_s < 2.5$, especially for the higher densities.

In Fig. 10(b) we illustrate the dependence on excitation energy of M_h for an insulator by displaying the difference between the value of U at

$$E = \mu + \tfrac{1}{2}\nu E_F = \mu + \tfrac{1}{2} E_g$$

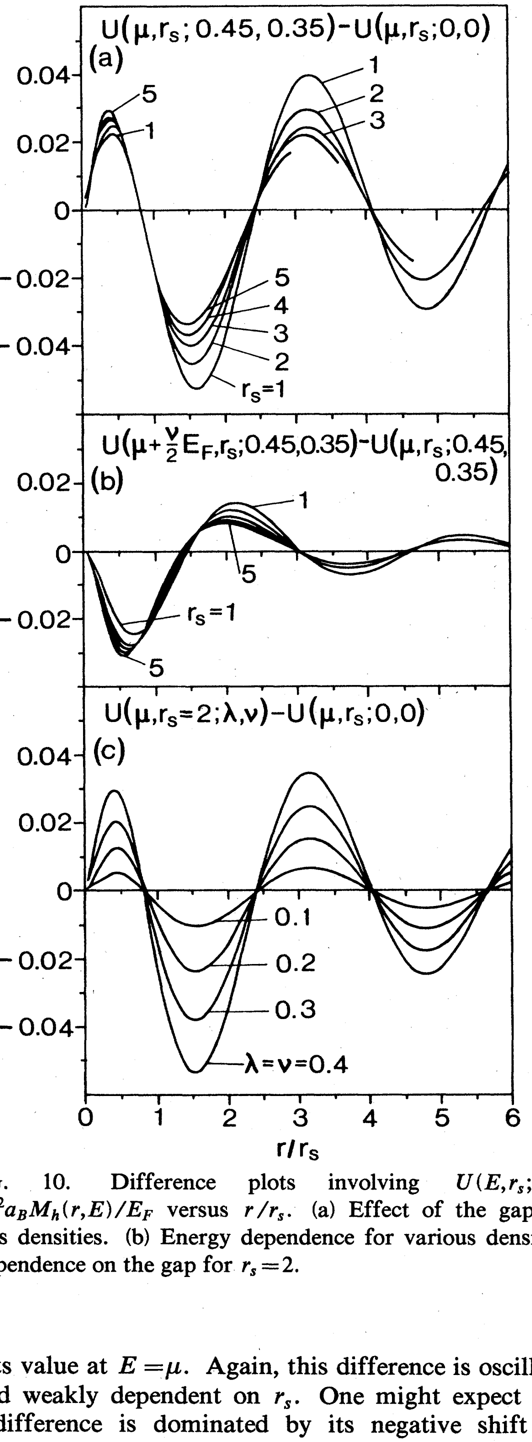

FIG. 10. Difference plots involving $U(E,r_s;\lambda,\nu) \equiv 4\pi r^2 a_B M_h(r,E)/E_F$ versus r/r_s. (a) Effect of the gap for various densities. (b) Energy dependence for various densities. (c) Dependence on the gap for $r_s=2$.

and its value at $E=\mu$. Again, this difference is oscillatory and weakly dependent on r_s. One might expect that this difference is dominated by its negative shift for $r/r_s < 1.5$, but once the QPLDA condition, Eq. (4.6), is applied, the energy shift in the eigenvalue is positive. The corresponding plot for $E=\mu-\frac{1}{2}E_g$ (not shown) is essentially the negative of Fig. 10(b), reflecting the near symmetry between electrons and holes as long as $|E-\mu|$ is not too near E_F.

Finally, in Fig. 10(c) we illustrate the effect of increasing gap size on the nonlocal potential at $E=\mu$ for $r_s=2$. This is a variant of Fig. 10(a) with gap size rather than density being varied, so the qualitative (long-range oscillating) behavior is similar. It is evident that, for gaps less than half the size of E_F, the correction is very closely *proportional to gap size.*

It must be kept in mind that the *differences* being discussed here are quite small (a few percent) compared to the mass operator itself, which is in many respects changed very slightly from its behavior in a metal. The value $\mathcal{M}_h(\mu,r_s)$ at $E=\mu$ gives the local-density ground-state potential, viz., Hedin-Lundqvist potential for metals or Levine-Louie potential for semiconductors. The fact that the differences in eigenvalues for these exchange-correlation potentials, ≤ 0.15 eV (≤ 0.2 eV) in silicon (diamond), confirms their overall similarity.

These observations allow some insight into the nonlocal behavior of the homogeneous insulating electron gas. In this paper we have applied this model self-energy in the QPLDA, which replaces $M(\vec{r}-\vec{r}\,',E)$ by $\mathcal{M}_{\rm LD}(E)\delta(\vec{r}-\vec{r}\,')$, and approximates $\mathcal{M}_{\rm LD}(E)$ by $\mu_{\rm xc}+\Delta_{\rm LD}(E)$. [We emphasize that the Hedin-Lundqvist and Levine-Louie exchange-correlation potentials $\mu_{\rm xc}$ are more sophisticated than the GW approximation to $\mathcal{M}_h$ gives. Hence the use of the GW approximation *only for the small difference* $\Delta_{\rm LD}(E)=\mathcal{M}_h(E)-\mathcal{M}_h(\mu)$ in the QPLDA.] It is *essential* that this local approximation be used in applications to inhomogeneous systems, since the vertex correction to $\mathcal{M}_h(k,E)$, neglected in our calculations, is important in the short-wavelength limit. This problem is avoided in the local-density approximation since $k_{\rm LD}$ does not become too large for the energies of interest. In a full nonlocal calculation there are substantial cancellations among large matrix elements of $\mathcal{M}_h(k,E)$ which require an accurate estimate of $\mathcal{M}_h(k,E)$ for the crystal to give a meaningful result. Our application of the full nonlocal self-energy [i.e., no QPLDA approximation, Eq. (4.7)] to silicon was found to *lower* the conduction bands at the zone boundary, tending to destroy the gap, even though the direct gaps are increased as in the QPLDA. One source of the strong nonlocality, viz., the k dependence of $\Delta(k,E)$, arises from the existence of a gap in the model insulator *only* at $k=k_F$, whereas crystalline insulating systems have a direct gap of $\sim E_g$ at all $\vec{k}$ points in the zone. In this respect the QPLDA point of view is consistent with the results of SMH,[14] who calculate the nonlocality of M in diamond and find it to be quite small.

The foregoing discussion leads to a novel viewpoint on the effect of exchange-correlation processes on single-particle excitations in insulators. The usual point of view has been described in some detail by Sterne and Inkson.[45] It is based on the Hartree-Fock reference system which is used as the point of departure for all the previous calculations discussed in the Introduction and in Sec. VI B. An excited particle, assumed for definiteness to be residing in the bond at the origin, polarizes other bonds. (This viewpoint is implemented *directly* in the correlated wave-function approach of HHF.[15]) This polarization is long ranged and arises from the Coulomb interaction which is incompletely screened in an insulator. This polarization cloud makes up the long-range correlation which compensates the long-range exchange interaction, thereby considerably reducing the size of the region which actually is disturbed in the presence of the excitation.

Although this viewpoint is usually restricted to insulators, when it is applied to a metal, as follows, the net effect is similar. Consider an excitation in a metal described initially in the Hartree-Fock approximation. Hedin[8] has noted that the exchange interaction is long ranged, but that when correlation effects are included, the resulting total interaction $M(r-r',E)$ is short ranged. Therefore in metals as well as in insulators correlation effects alone are long ranged (the infinitely better screening is compensated for by the infinitely more polarizable medium). Strinati, Mattausch, and Hanke[14] indeed have found the resulting total interaction to have negligible nonlocality, and have suggested that some local theory should be adequate. The present approach realizes such a theory. Instead of an impuritylike viewpoint, the resulting picture is one in which, as long as the excitation is bandlike, the electronic system is rearranged from its ground state in a manner which is describable in a local-density approximation.

D. Refinements in the QPLDA

Like Hedin and Lundqvist[20] in the metallic case, we have not calculated a self-consistent G for the homogeneous system to use in the GW expression for $\mathscr{M}_h$. In Fig. 11 we show the result of determining the *homogeneous* QP spectrum from $\mathscr{M}_h$, that is, the first step in determining G self-consistently. The results are shown both *including* the nonlocality (denoted QPNL),

$$\epsilon_k = E_k + \Delta(k,\epsilon_k)\ , \tag{6.4}$$

and within the QPLDA,

$$\epsilon_k = E_k + \Delta_{\rm LD}(\epsilon_k)\ , \tag{6.5}$$

where E_k is given by Eq. (3.3) and parameters appropriate to Si were chosen for illustration. As noted above, the QPLDA widens the gap (by about 0.3 eV in this case).

In an exact theory the full nonlocality of Eq. (6.4) should be taken into account. Figure 11 shows that the QPNL gap is *decreased* by about 15%, and states with $0.93 < k/k_F < 1.15$ are pulled nearer the gap while states outside this range are repelled from the gap. As these changes tend to cancel in the integrals for $\mathscr{M}_h$, we expect that the self-consistent QPNL spectrum will not differ significantly from the first iteration, shown in Fig. 11.

The other type of refinement would be to derive, or calculate, a better dielectric function. Although other models have been derived, for example, by Milchev,[46] we know of none which are superior to the *ad hoc* model of Levine and Louie.[19] Using a numerical dielectric function requires a three-dimensional numerical integration (as opposed to the two-dimensional integration we have implemented), and therefore severely increases the computational effort, although this approach may not be beyond question.

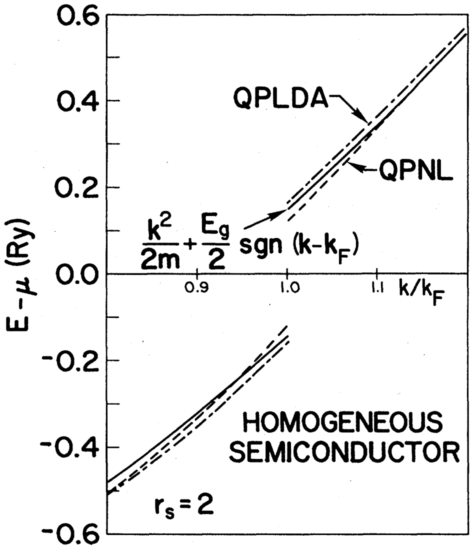

FIG. 11. Unrenormalized spectrum for the model homogeneous insulator (solid line) and the resulting spectrum for both the nonlocal (QPNL, dashed line) and local (QPLDA, short-dashed—long-dashed line) theories.

The dielectric function can also be improved by including the "local-field," or exchange-correlation, correction to screening, corresponding to the use of an *electron,* rather than a *test charge,* dielectric function. Strinati, Mattausch, and Hanke have found small but significant corrections to their gap from the "local field." Within the QPLDA approach these corrections are included with considerable sophistication in the ground-state bands, and as a result we expect these "local-field" exchange-correlation corrections to be even less than the factor of 4 smaller which would be expected from the overall magnitude of band shifts in the two approaches.

Note added in proof. P. A. Sterne (private communication) has found that the $\Gamma_{2'}$ state in silicon is lowered by $\sim$0.2 eV when 300 plane waves are used in the local-density calculation, while all other states are almost unaffected. As a result of this change the QPLDA overestimates the $\Gamma_{25'} \rightarrow \Gamma_{2'}$ transition by $\sim$0.2 eV. This result supersedes the perfect agreement with experiment for this transition which would be inferred from Fig. 5 and Ref. 17.

ACKNOWLEDGMENTS

At various stages of this work we have had useful discussions and/or correspondence with J. Callaway, M. L. Cohen, A. Glick, D. R. Hamann, V. Heine, S. G. Louie, A. H. MacDonald, J. P. Perdew, M. Schlüter, L. J. Sham, P. A. Sterne, K. Unger, and M. T. Yin. This work was supported in part by U.S. Office of Naval Research (under Contract No. N00014-79-WR-90028). One of us (C.S.W.) also wishes to acknowledge the support of the National Science Foundation (under Grant No. DMR-82-13768). Use of the computer facilities at the University of Maryland, and the Naval Research Laboratory Condensed Matter and Radiation Sciences Division VAX-11/780 computer, are gratefully acknowledged.

*Permanent address.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, *ibid.* **140**, A1133 (1965).

[2] The review of D. D. Koelling [Rep. Prog. Phys. **44**, 139 (1981)] contains an extensive list of references to band-structure calculations.

[3] O. Gunnarsson, M. Jonson, and B. I. Lundqvist, Phys. Rev. B **20**, 3136 (1979); J. A. Alonso and L. A. Girifalco, *ibid.* **17**, 3735 (1978).

[4] G. P. Kerker, Phys. Rev. B **24**, 3468 (1981).

[5] D. C. Langreth and M. J. Mehl, Phys. Rev. Lett. **47**, 446 (1981); Phys. Rev. B **28**, 1809 (1983).

[6] U. von Barth and R. Car (unpublished).

[7] See M. T. Yin and M. L. Cohen, Phys. Rev. Lett. **50**, 2006 (1983), and references therein.

[8] L. Hedin, Phys. Rev. **139**, A796 (1965); see also J. J. Quinn and R. A. Ferrell, *ibid.* **112**, 812 (1958).

[9] W. Brinkman and B. Goodman, Phys. Rev. **149**, 597 (1966).

[10] N. O. Lipari and W. B. Fowler, Phys. Rev. B **2**, 3354 (1970).

[11] E. O. Kane, Phys. Rev. B **5**, 1493 (1972).

[12] N. E. Brener, Phys. Rev. B **11**, 929, 1600 (1975); **12**, 1487 (1975).

[13] M. Bennett and J. C. Inkson, J. Phys. C **10**, 987 (1977); J. C. Inkson and M. Bennett, *ibid.* **11**, 2017 (1978).

[14] G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. Lett. **45**, 290 (1980); Phys. Rev. B **25**, 2867 (1982).

[15] S. Horsch, P. Horsch, and P. Fulde, Phys. Rev. B **28**, 5977 (1983); **29**, 1870 (1984); B. Kiel, G. Stollhoff, C. Weigel, P. Fulde, and H. Stoll, Z. Phys. B **46**, 1 (1982); B. Kiel, G. Stollhoff, P. Fulde, and H. Stoll, Phys. Lett. **87A**, 433 (1982).

[16] R. E. Watson, J. F. Herbst, L. Hodges, B. I. Lundqvist, and J. W. Wilkins, Phys. Rev. B **13**, 1463 (1976); L. Hedin, Ark. Fys. **30**, 231 (1966); F. Sacchetti, J. Phys. F **12**, 281 (1982); G. Arbman and U. von Barth, *ibid.* **5**, 1155 (1975); U. von Barth and L. Hedin, Il Nuovo Cimento **23B**, 1 (1974); M. Schreiber and H. Bross, J. Phys. F **13**, 1895 (1983).

[17] C. S. Wang and W. E. Pickett, Phys. Rev. Lett. **51**, 597 (1983).

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

[19] Z. H. Levine and S. G. Louie, Phys. Rev. B **25**, 6310 (1982).

[20] L. Hedin, Ref. 8; L. Hedin, B. I. Lundqvist, and S. Lundqvist, Solid State Commun. **5**, 237 (1967); B. I. Lundqvist, Phys. Kondens. Mater. **6**, 193, 206 (1967); **7**, 117 (1968); L. Hedin and S. Lundqvist, *Effects of Electron-Electron and Electron-Phonon Interactions on the One-Electron States of Solids,* in Vol. 23 of *Solid State Physics,* edited by F. Seitz and D. Turnbull (Academic, New York, 1969), p. 1.

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

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

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

[24] M. Bennett, J. Phys. C **11**, L515 (1978); M. Watabe and H. Yasuhara, Phys. Lett. **28A**, 329 (1968).

[25] J. Lindhard, Kgl. Dansk. Vidensk. Selsk. Mat. Fys. Medd. **28**, (1954).

[26] J. Callaway, Phys. Rev. **116**, 1368 (1959).

[27] L. Hedin and S. Lundqvist, J. Phys. C **4**, 2064 (1971).

[28] W. E. Pickett and C. S. Wang (unpublished).

[29] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979); G. Bachelet, M. Schlüter, and D. R. Hamann, Phys. Rev. B **26**, 4199 (1982).

[30] D. R. Hamann, Phys. Rev. Lett. **42**, 662 (1979); C. S. Wang and B. M. Klein, Phys. Rev. B **24**, 3393 (1981); A. Zunger and M. L. Cohen, *ibid.* **20**, 4082 (1979).

[31] G. S. Painter, D. E. Ellis, and A. R. Lubinsky, Phys. Rev. B **4**, 3610 (1971).

[32] R. N. Euwema, D. L. Wilhite, and G. T. Surratt, Phys. Rev. B **7**, 818 (1973).

[33] R. Heaton and E. Lafon, Phys. Rev. B **17**, 1958 (1978).

[34] A. Zunger and A. J. Freeman, Phys. Rev. B **15**, 5049 (1977).

[35] D. Glotzel, B. Segall, and O. K. Andersen, Solid State Commun. **36**, 403 (1980).

[36] M. T. Yin (private communication).

[37] G. B. Bachelet, H. S. Greenside, G. A. Baraff, and M. Schlüter, Phys. Rev. B **24**, 4736 (1981); **24**, 4745 (1981).

[38] D. R. Hamann, Phys. Rev. Lett. **42**, 662 (1979).

[39] C. D. Clark, P. J. Dean, and P. V. Harris, Proc. R. Soc. London, Ser. A **277**, 312 (1964).

[40] R. A. Roberts and W. C. Walker, Phys. Rev. **161**, 730 (1967).

[41] F. J. Himpsel, J. F. van der Veen, and D. E. Eastman, Phys. Rev. B **22**, 1967 (1980).

[42] F. R. McFeely, S. P. Kowalczyk, L. Ley, R. G. Cavell, R. A. Pollak, and D. A. Shirley, Phys. Rev. B **9**, 5268 (1974).

[43] W. D. Grobman and D. E. Eastman, Phys. Rev. Lett. **29**, 1508 (1972).

[44] L. Ley, S. Kowalczyk, R. Pollak, and D. A. Shirley, Phys. Rev. Lett. **29**, 1088 (1972).

[45] P. Sterne and J. Inkson, J. Phys. C **17**, 1497 (1984).

[46] A. Milchev, Phys. Status Solidi B **90**, 679 (1978).