Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies

Mark S. Hybertsen and Steven G. Louie

Department of Physics, University of California, Berkeley, California 94720

(Received 31 March 1986)

We present a first-principles theory of the quasiparticle energies in semiconductors and insulators described in terms of the electron self-energy operator. The full dielectric matrix is used to evaluate the self-energy operator in the GW approximation: the first term in an expansion of the self-energy operator in terms of the dynamically screened Coulomb interaction (W) and the dressed Green's function (G). Quasiparticle energies are calculated for the homopolar materials diamond, Si, and Ge as well as for the ionic compound LiCl. The results are in excellent agreement with available experimental data. In particular, the indirect band gap is calculated as 5.5, 1.29, and 0.75 eV as compared with experimental gaps of 5.48, 1.17, and 0.744 eV for diamond, Si, and Ge, respectively. The Ge results include relativistic effects. The calculated direct gap for LiCl is within 5% of experiment. Viewed as a correction to the density-functional eigenvalues calculated with the local-density approximation, the present results show a correction dominated by a large jump at the gap. It is found that because of the charge inhomogeneity, the full dielectric screening matrix must be included, i.e., local-field effects are essential. The dynamical effects are also found to be crucial. The required dielectric matrices are obtained within the density-functional approach for the static case and extended to finite frequency with use of a generalized plasmon-pole model based on sum rules. The model reproduces the ω and ω^{-1} moments of the exact many-body response function. The qualitative features of the electron self-energy operator are discussed. Using the static Coulomb-hole—screened-exchange approximation for illustration, the role of local fields in the self-energy operator are explained. The role of dynamical renormalization is illustrated. The same qualitative features are observed in both the homopolar and ionic materials.

I. INTRODUCTION

Development of a *predictive* theory of quasiparticle energies and optical properties of semiconductors and insulators has been of long-standing difficulty. Because of the long-range Coulomb interaction between the electrons, an adequate treatment of the dynamical correlations in the motion of the electrons is a formidable many-body problem. This has been the central problem in first-principles treatment of the quasiparticle energies in real solids.

The density-functional formalism[1] provides a concise way to incorporate exchange and correlation effects into the calculation of the ground-state energy of the interacting-electron system in an external potential. Once the energy functional is known or approximated, the solution of the many-body problem is reduced to self-consistent solution of a set of effective one-particle equations.[2] This method has the advantage of being a practical computational scheme when coupled with the local-density approximation (LDA) for the exchange-correlation part of the functional. The success of that approach is well documented for many of the ground-state properties of a wide range of materials.[3] Unfortunately, the one-particle eigenvalues in the theory have no formal justification as quasiparticle energies although, in practice, these eigenvalues have been used to discuss the spectra of solids. For the case of simple metals, the overall results are reasonable,[3] although the results do not agree with experiment for the zone-edge gaps in such materials as Al (Ref. 4) or with recent measurements of the band-

width of Na.[5] Discrepancies in the case of semiconductors and insulators are much more dramatic. The band gap in the local-density-functional eigenvalues (Kohn-Sham gap ϵ_g) is typically 30—50 % less than the band gap observed in the optical spectrum.[3] For the case of Ge, the LDA conduction and valence bands in fact overlap when relativistic corrections are included.[6] The discrepancies for the details of the band dispersions are typically smaller, but are also material dependent. For example, the bandwidth of Ge is well represented by the LDA eigenvalues,[7] but for diamond, it is underestimated as compared to x-ray spectra.[8]

As briefly pointed out here and amply supported in the literature, this straightforward computational scheme for including exchange-correlation effects in solid-state calculations is inadequate for quasiparticle energies and optical properties. Furthermore, calculations with functionals that go beyond the LDA suggest that this is not a defect of the LDA *per se*.[7] In fact, Sham and Schluter[9] and Perdew and Levy[10] have shown formally that there is an explicit correction to the Kohn-Sham gap, ϵ_g, even when the exact exchange-correlation functional is used. The true minimum gap in the spectrum is given by $E_g = \epsilon_g + \Delta$ where Δ is related to a discontinuity in a functional derivative of the exchange-correlation energy with respect to density for systems with a gap. Calculations with a two-plane-wave model suggest that the correction Δ is substantial.[11]

A rigorous formulation for the quasiparticle properties is a Green's-function approach. Application of Green's-function techniques to the electron correlation problem

has been extensively reviewed by Hedin and Lundqvist.[12] Within the context of the one-particle Green's function, quasiparticles are associated with the peaks in the spectral function. If the peak is sufficiently sharp, a well-defined quasiparticle energy can be obtained. For the general case of an inhomogeneous system, the quasiparticle energies and wave functions are obtained by solving[12]

$$(T + V_{ext} + V_H)\psi_{nk}(\mathbf{r}) + \int d\mathbf{r}' \, \Sigma(\mathbf{r},\mathbf{r}';E_{nk})\psi_{nk}(\mathbf{r}')$$
$$= E_{nk}\psi_{nk}(\mathbf{r}) , \quad (1)$$

where T is the kinetic energy operator, V_{ext} is the external potential due to the ions, V_H is the average Coulomb (Hartree) potential due to the electrons, and Σ is the electron self-energy operator. Σ contains the effect of exchange and correlation among the electrons. It is in general a nonlocal, energy-dependent, non-Hermitian operator. The procedure for finding the quasiparticle energies requires evaluating Σ and then solving Eq. (1). Since Σ is non-Hermitian, the eigenvalues in Eq. (1), E_{nk}, are in general complex; the imaginary part gives the lifetime of the quasiparticle.

For the quasiparticle problem, the central difficulty is an adequate approximation for the self-energy operator, Σ. We use the formulation of Hedin[13] based on an expansion of the electron self-energy in a perturbation series in the fully screened (as opposed to the bare) Coulomb interaction. This formulation is quite general and has been basic to much subsequent development of theories of quasiparticle energies in semiconductors and insulators. In the present work, the first term in the expansion is used; i.e., the GW approximation for Σ. This requires the dressed Green's function, G, and the dynamical screened interaction, $W = \epsilon^{-1}v$. Our approach is to make the best possible approximation for G and W separately, calculate Σ, and then obtain the quasiparticle energies by solving Eq. (1) using the ab initio pseudopotential for the electron-ion interaction.[14] The crucial input required is the dielectric matrix which describes the dynamical screening in the solid. We obtain the static dielectric matrices from first principles within the density-functional approach using the LDA. The frequency dependence is then obtained with a generalized plasmon-pole (GPP) model. We emphasize that the full dielectric matrix is evaluated so that the local-field effects are included.

The resulting quasiparticle band structure has been obtained for the homopolar semiconductors diamond, Si and Ge as well as the ionic insulator LiCl.[15,16] (Lifetime effects are not considered here.) The calculated band gaps and band dispersions agree very well with available optical, photoemission, and inverse photoemission data. Table I illustrates the results of the present theory for the minimum gap in each material as compared to experiment.[17,18] The corresponding gaps in the LDA eigenvalues are shown for reference. The minimum gap opens up dramatically as compared to the LDA eigenvalues. The results of the present theory can be presented in the form of a correction to the LDA eigenvalues. In Fig. 1, the difference between the quasiparticles energies calculated here and the LDA eigenvalues [in the random-phase approximation (RPA)][19] are plotted as a function of the
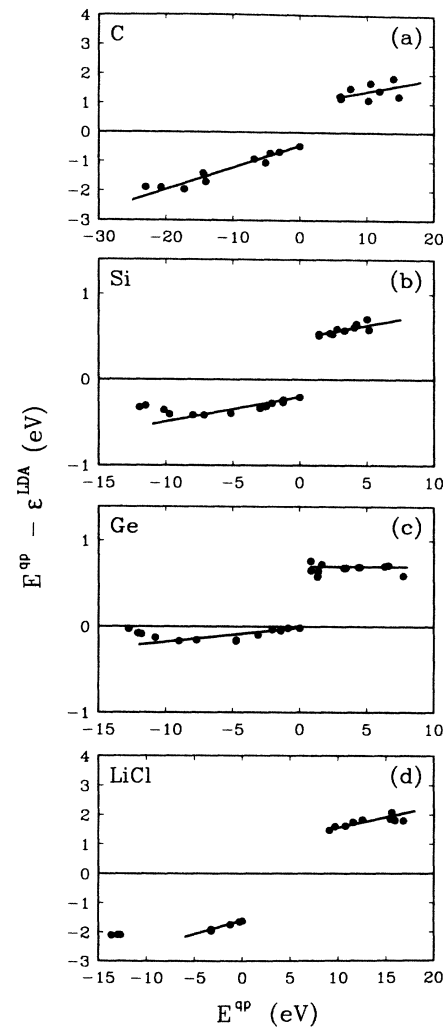

FIG. 1. The difference between the quasiparticle energy, E^{qp}, and the eigenvalue in the LDA, ϵ^{LDA}, is plotted against the quasiparticle energy for states at several high symmetry points in the Brillouin zone. The straight lines are drawn as a guide to the eye. Data for four materials is displayed: (a) diamond, (b) Si, (c) Ge, and (d) LiCl.

TABLE I. Results of the present work for the fundamental gap E_g of diamond, Si, Ge, and LiCl are compared to experiment and the gap in the LDA eigenvalues. The results for Ge include relativistic effects. All energies in eV.

	LDA	Present theory	Expt.
Diamond	3.9	5.6	5.48[a]
Silicon	0.52	1.29	1.17[a]
Germanium	<0	0.75	0.744[a]
LiCl	6.0	9.1	9.4[b]

[a]Reference 17.
[b]Reference 18.

quasiparticle energy. The required correction to the LDA eigenvalues is dominated by a jump at the gap. Away from the gap region the correction varies fairly smoothly with energy. Also, we find that the quasiparticle wave function $\psi_{n\mathbf{k}}$ is essentially unchanged from the LDA one-particle wave function $\phi_{n\mathbf{k}}$. These trends follow through the four materials studied ranging from wide to narrow gap as well as from covalent to ionic bonding.

The remainder of the paper is organized as follows. In Sec. II the Green's-function method is briefly reviewed and our approach discussed in detail. The dielectric matrix and the GPP model are described in Sec. III. We describe the details of the calculation of Σ and the solution of Eq. (1) in Sec. IV. In Sec. V, we discuss the qualitative features of the screened Coulomb interaction and the self-energy operator. The static Coulomb-hole—screened-exchange (COHSEX) approximation for Σ introduced by Hedin[13] provides a useful means for understanding the crucial role of local-field effects as well as the intrinsic nonlocality of the self-energy operator in real space. The role of dynamical renormalization is described. The trends in the calculation are discussed and a detailed comparison to the LDA potential is given. Section VI gives the results for diamond, Si, Ge, and LiCl as compared to available experimental data. In Sec. VII, previous work on quasiparticle energies in semiconductors and insulators is briefly discussed and the results of the present study are compared to those of four recent calculations. Finally, the paper concludes with a brief summary in Sec. VIII. Some of the details of the theory are described in the Appendixes. A reader interested only in the qualitative features of the present theory and the results might wish to go directly to Sec. V.

II. SUMMARY OF GREEN'S-FUNCTION APPROACH

In this section we summarize the relevant results from the Green's-function approach. The details are discussed in the review of Hedin and Lundqvist[12] and the original paper by Hedin.[13] Here we focus on the physical aspects related to evaluation of the self-energy operator in the GW approximation.

We start by introducing the spectral weight function $A(\mathbf{r},\mathbf{r}';E)$. The single-particle Green's function is related to A in the usual way

$$G(\mathbf{r},\mathbf{r}';E) = \int_C dE' \frac{A(\mathbf{r},\mathbf{r}';E')}{E-E'} . \tag{2}$$

The contour C runs infinitesimally above the real axis for $E' < \mu$ and below the real axis for $E' > \mu$ where μ is the chemical potential. The spin dependence is not shown explicitly. The spectral function is just proportional to the imaginary part of the Green's function:

$$A(\mathbf{r},\mathbf{r}';E) = \frac{1}{\pi} |\,\mathrm{Im}G(\mathbf{r},\mathbf{r}';E)\,| . \tag{3}$$

If the matrix element of A with respect to a one-particle state is sharply peaked as a function of energy, then it is meaningful to speak of a particlelike excitation of the many electron system. This corresponds to a pole in the Green's function at a complex energy. Finding this pole

is equivalent to solving the quasiparticle equation (1). The position of the peak in A is the quasiparticle energy (real part of $E_{n\mathbf{k}}$) and the width corresponds to the lifetime of the quasiparticle (imaginary part of $E_{n\mathbf{k}}$). Since Σ is energy dependent, Eq. (1) in general yields a spectrum that depends on the energy argument in Σ: $E_{n\mathbf{k}}(E)$. The pole in the Green's function requires the energy in Σ be the quasiparticle energy. This is just the energy self-consistency in Eq. (1). For sufficiently sharp peaks (long quasiparticle lifetime) we can write this as

$$E_{n\mathbf{k}}^{\mathrm{qp}} = \mathrm{Re}E_{n\mathbf{k}}(E_{n\mathbf{k}}^{\mathrm{qp}}) , \tag{4}$$

where we denote the quasiparticle energy by E^{qp}. The imaginary part may be neglected in that case.

The primary task is to evaluate the electron self-energy operator required in Eq. (1). The self-energy operator should in principle be obtained together with G in a self-consistent procedure. They are related to each other as well as the screened Coulomb interaction W and a vertex function Γ by a set of four integral equations. For completeness, these are

$$W(1,2) = v(1,2) + \int d(3\,4)v(1,3)P(3,4)W(4,2) , \tag{5a}$$

$$P(1,2) = -i \int d(3\,4)G(1,3)G(4,1^+)\Gamma(3,4;2) , \tag{5b}$$

$$\Sigma(1,2) = i \int d(3\,4)G(1,3)\Gamma(3,2;4)W(4,1^+) , \tag{5c}$$

$$\Gamma(1,2,;3) = \delta(1,2)\delta(1,3)$$
$$+ \int d(4\,5\,6\,7)\frac{\delta\Sigma(1,2)}{\delta G(4,5)}$$
$$\times G(4,6)G(7,5)\Gamma(6,7;3) . \tag{5d}$$

Here we adopt the compressed notation $1 \equiv (\mathbf{r},\sigma,t)$, etc. Also 1^+ means that $t \to t+\delta$ where delta is a positive infinitesimal. The bare Coulomb interaction is denoted $v(1,2)$ and P denotes the irreducible polarizability from which the screened interaction is derived as usual in Eq. (5a). These can be derived by the functional-derivative technique as outlined, for example, in Ref. 13.

The approach taken by Hedin[13] essentially generates a perturbation series in the screened interaction W. He starts with the vertex function to zeroth order in W:

$$\Gamma(1,2;3) = \delta(1,2)\delta(1,3) . \tag{6}$$

Then the polarizability is given by

$$P(1,2) = -iG(1,2^+)G(2,1) . \tag{7}$$

This corresponds to the random-phase approximation (RPA) for the dielectric matrix. To first order in W, the electron self-energy is given by

$$\Sigma(1,2) = iG(1,2)W(1^+,2) . \tag{8}$$

This is the GW approximation. One can iterate again to obtain Σ to second order in W, etc. We stop at first order

and proceed to calculate Σ. As in any perturbation theory, there is concern that important terms may be neglected. Strictly, thought of as an expansion in terms of the screened interaction, the expansion parameter here is approximately proportional to $r_s/(1+r_s)$ for the electron gas.[12] Thus there are not the usual convergence difficulties encountered in the perturbation expansion in the bare Coulomb interaction for r_s in the physically interesting range, e.g., $r_s = 2$ for Si.

It is convenient to make a division between the core and valence electrons in a solid. The Green's function is separated according to $G = G_c + G_v$ where G_c includes only the core states and G_v includes the balance of the states. Similarly, the polarizability is separated into a core and valence part: $P = P_c + P_v$. Then the electron self-energy operator breaks into three terms:[12]

$$\Sigma = iG_c W + iG_v W_v P_c W_v + iG_v W_v \ . \tag{9}$$

The first term is a core-valence exchange contribution. As noted by Phillips,[20] the screening is ineffective in this case so this term is essentially the bare core-valence exchange interaction. The second term is a screened polarization potential due to the cores. The final term is the self-energy of the valence electrons. The first two terms are generally small and will be discussed further in Sec. IV B. Here we focus on the final term which only involves the valence electrons. It will be implicitly understood that Σ refers to the valence-valence part only from this point forward.

After Fourier transformation to energy, the GW approximation for Σ is

$$\Sigma(\mathbf{r},\mathbf{r}';E) = i \int \frac{dE'}{2\pi} e^{-i\delta E'} G(\mathbf{r},\mathbf{r}';E-E')W(\mathbf{r},\mathbf{r}';E') \ , \tag{10}$$

where $\delta = 0^+$. The required frequency integrals can be carried out formally by making use of the spectral representation of the Green's function, Eq. (2), and an analogous decomposition of the screened interaction:

$$W(\mathbf{r},\mathbf{r}';E) = v(\mathbf{r},\mathbf{r}') + \int_0^\infty dE' \frac{2E'B(\mathbf{r},\mathbf{r}';E)}{E^2 - (E'-i\eta)^2} \ , \tag{11}$$

where $\eta = 0^+$. The bare Coulomb interaction is $v(\mathbf{r},\mathbf{r}') = e^2/|\mathbf{r}-\mathbf{r}'|$. $B(\mathbf{r},\mathbf{r}';E)$ is the spectral function for the screened interaction and is related to the imaginary part of W. The real part of Σ breaks into two terms:

$$\text{Re}\Sigma = \Sigma_{\text{SEX}} + \Sigma_{\text{COH}} \tag{12}$$

where the screened exchange part Σ_{SEX} arises from the poles in the Green's function and the Coulomb-hole part Σ_{COH} from the poles in the screened interaction. These are written explicitly for the real part:

$$\Sigma_{\text{SEX}}(\mathbf{r},\mathbf{r}';E) = - \int_{-\infty}^{\mu} dE' \, A(\mathbf{r},\mathbf{r}';E')$$

$$\times \text{Re}W(\mathbf{r},\mathbf{r}';E-E') \ , \tag{13a}$$

$$\Sigma_{\text{COH}}(\mathbf{r},\mathbf{r}';E) = - \int_{-\infty}^{\infty} dE' \, A(\mathbf{r},\mathbf{r}';E')$$

$$\times \text{P} \int_0^\infty dE'' \frac{B(\mathbf{r},\mathbf{r}';E'')}{E-E'-E''} \ . \tag{13b}$$

The principal part of the E'' integration is to be taken in the Coulomb-hole term. In the present paper, we will only be explicitly concerned with the real part of the self-energy operator.

In order to make further progress, approximations for G and W are required. The screened interaction is discussed in the next section. The Green's function is calculated iteratively with the simplest initial approximation being to replace the self-energy operator by an energy-independent effective potential:

$$\Sigma(\mathbf{r},\mathbf{r}';E) \rightarrow V_{\text{eff}}(\mathbf{r},\mathbf{r}') \ . \tag{14}$$

There are several reasonable choices that could be made, including the exchange-correlation potential from the density-functional theory. The essential requirement turns out to be that the resulting spectrum $\varepsilon_{n\mathbf{k}}$ and wave functions $\phi_{n\mathbf{k}}$ be reasonable. With this approximation, the expression for the Green's function simplifies because the resulting spectrum is real and independent of E. The explicit expression for the Green's function is

$$G(\mathbf{r},\mathbf{r}';E) = \sum_{n,\mathbf{k}} \frac{\phi_{n\mathbf{k}}(\mathbf{r})\phi_{n\mathbf{k}}^*(\mathbf{r}')}{E - \varepsilon_{n\mathbf{k}} - i\delta_{n\mathbf{k}}} \ , \tag{15}$$

where $\delta_{n\mathbf{k}} = 0^+$ for $\varepsilon_{n\mathbf{k}} < \mu$ and $\delta_{n\mathbf{k}} = 0^-$ for $\varepsilon_{n\mathbf{k}} > \mu$. The corresponding spectral function is simply

$$A(\mathbf{r},\mathbf{r}';E) = \sum_{n,\mathbf{k}} \phi_{n\mathbf{k}}(\mathbf{r})\phi_{n\mathbf{k}}^*(\mathbf{r}')\delta(E - \varepsilon_{n\mathbf{k}}) \ . \tag{16}$$

Then the real part of the self-energy operator is given by

$$\Sigma_{\text{SEX}}(\mathbf{r},\mathbf{r}';E) = - \sum_{n,\mathbf{k}}^{\text{occ}} \phi_{n\mathbf{k}}(\mathbf{r})\phi_{n\mathbf{k}}^*(\mathbf{r}')W(\mathbf{r},\mathbf{r}';E-\varepsilon_{n\mathbf{k}}) \ , \tag{17a}$$

$$\Sigma_{\text{COH}}(\mathbf{r},\mathbf{r}';E) = \sum_{n,\mathbf{k}} \phi_{n\mathbf{k}}(\mathbf{r})\phi_{n\mathbf{k}}^*(\mathbf{r}')\text{P} \int_0^\infty dE' \frac{B(\mathbf{r},\mathbf{r}';E')}{E-\varepsilon_{n\mathbf{k}}-E'} \ . \tag{17b}$$

In this form, the interpretation of the first term as a dynamically screened-exchange interaction is clear. Also, when taking matrix elements of Σ in Eq. (17), the electron only couples to the states of the same spin in the sum.

The advantage of formulating the self-energy operator in terms of the spectral function is that one see how to systematically improve the calculation of G and thus Σ. A better approximation to the spectral function is to include the shift and width of the peaks due to the self-energy operator. This yields

$$A(\mathbf{r},\mathbf{r}';E) = \sum_{n,\mathbf{k}} \phi_{n\mathbf{k}}(\mathbf{r})\phi_{n\mathbf{k}}^*(\mathbf{r}') \frac{\Gamma_{n\mathbf{k}}/\pi}{\Gamma_{n\mathbf{k}}^2 + (E - E_{n\mathbf{k}})^2} \ . \tag{18}$$

Here, the new spectrum $E_{n\mathbf{k}}$ is just the real part of the solution of Eq. (1) and $\Gamma_{n\mathbf{k}}$ is the imaginary part. We will show below that the corresponding wave function is insensitive to the choice of V_{eff} as opposed to the full Σ. If

it is further assumed that Γ_{nk} is sufficiently small, then the new spectral function in Eq. (18) reduces to that in Eq. (16) with the important exception that the spectrum is improved to include the effect of the full self-energy operator. This process is repeated until a self-consistent spectrum is obtained. The only important approximation in the spectral function then is concentration of all of the spectral weight in the quasiparticle peaks which are approximated by δ functions.

In the present calculation, we approximate the spectral function initially by the wave functions and spectrum from a density-functional calculation using the LDA for the exchange-correlation potential. Thus V_{eff} has the further advantage of being local. Then the self-energy operator is evaluated by the means to be described below and the quasiparticle energies are calculated from Eq. (1). This new spectrum is then used to construct an improved self-energy operator as just described. For the materials studied, we find that the final spectrum exhibits small, but not negligible, changes due to this self-consistent incorporation of the quasiparticle spectrum in the evaluation of the self-energy operator. A third iteration is not necessary.

The static COHSEX approximation to Σ of Hedin[13] is simpler to apply and interpret. As such, it is a useful reference point for understanding the important physical ingredients in the self-energy operator, particularly the role of local fields. It can be obtained from Eq. (17) by assuming that $E - \varepsilon_{nk} \rightarrow 0$. The result is

$$\Sigma_{\text{SEX}}(\mathbf{r},\mathbf{r}') = - \sum_{nk}^{\text{occ}} \phi_{nk}(\mathbf{r})\phi_{nk}^{*}(\mathbf{r}')W(\mathbf{r},\mathbf{r}';E=0) \ , \qquad (19a)$$

$$\Sigma_{\text{COH}}(\mathbf{r},\mathbf{r}') = \tfrac{1}{2}\delta(\mathbf{r}-\mathbf{r}')[W(\mathbf{r},\mathbf{r}';E=0)-v(\mathbf{r},\mathbf{r}')] \ . \qquad (19b)$$

The sum over all states in the Coulomb-hole term gives the δ function using the completeness relation. The screened-exchange term goes over to a static screened-exchange interaction and the Coulomb-hole term simplifies to a local interaction. In fact the physical interpretation of the Coulomb-hole part is clear in this approximation: it is the interaction of the quasiparticle with the induced potential due to the rearrangement of the electrons around the quasiparticle. The quantitative validity of the COHSEX approximation depends on whether the exchange frequency $E^{\text{qp}} - \varepsilon_{nk}$ is small compared to the characteristic frequency in the screened interaction, essentially the plasma frequency. Comparison to the results of the full calculation shows that the COHSEX approximation consistently overestimates the magnitude of the electron self-energy. We also note that the COH term in the full self-energy operator in Eq. (17b) is not a local potential; the sum over states is modulated by a frequency-dependent part.

III. EVALUATION OF THE DYNAMICALLY SCREENED INTERACTION

Unlike the case of simple metals which can be modeled by an electron gas, screening in an insulator is qualitatively different and more complex. The calculation therefore requires the full dielectric matrix $\epsilon^{-1}(\mathbf{r},\mathbf{r}';\omega)$ which depends separately on $\mathbf{r}$ and $\mathbf{r}'$. As will be discussed in de-

tail in Sec. V A, this then includes the important local fields in the screening. The local fields are the variations in screening due to the charge inhomogeneity. This is quite important in a semiconductor or insulator. The screened interaction is related to the dielectric matrix by

$$W(1,2) = \int d(3)\epsilon^{-1}(1,3)v(3,2) \ . \qquad (20)$$

Here we require the time-ordered dielectric matrix. This is simply related to the usual linear-response (causal) dielectric matrix: for positive frequencies, they are the same; for negative frequencies, the imaginary part of the causal dielectric matrix changes sign. Equation (20) can be Fourier transformed to the energy and wave-vector basis using the following convention for the latter

$$W(\mathbf{r},\mathbf{r}';E) = \sum_{\mathbf{q},\mathbf{G},\mathbf{G}'} e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} W_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)e^{-i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} \ . \qquad (21)$$

Here $\mathbf{G}$ is a reciprocal lattice vector and $\mathbf{q}$ is a wave vector in the first Brillouin zone. With this

$$W_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q},\omega)v(\mathbf{q}+\mathbf{G}') \ . \qquad (22)$$

The Fourier transform of the bare Coulomb interaction has the usual form $v(q) = 4\pi e^{2}/\Omega q^{2}$ with Ω being the crystal volume. In this form, the $\mathbf{G} \neq \mathbf{G}'$ components yield the local fields. Finally, for the choice of vertex function in Eq. (6), the dielectric matrix is related to the irreducible polarizability P by

$$\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \delta_{\mathbf{G},\mathbf{G}'} - v(\mathbf{q}+\mathbf{G})P_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) \ . \qquad (23)$$

This is simply the usual RPA when Eq. (7) is used for P.

There are two problems associated with calculating the full dynamical many-body response function, even in the RPA. The first problem is the establishment of an adequate approximation for the Green's function required as input to Eq. (7). Generally, empirical band structures have been used but such an approximation for G does not carry any a priori justification. The second problem is the technical difficulty of evaluating $\epsilon(\omega)$ numerically. This requires a careful evaluation of the required Brillouin-zone summations and is very time consuming. The approach adopted here is to obtain the static dielectric matrices from first principles using the density-functional approach. This is described in subsection A. These results are then extended to finite frequency using a generalization of the plasmon-pole model fixing the parameters with sum rules. The extension to finite frequency is discussed in subsection B.

A. Ab initio static dielectric matrices

Evaluation of the longitudinal dielectric matrix requires a ground-state expectation value of a pair of density-fluctuation operators, precisely the density-density correlation function. For the static case, there is no ambiguity as to whether this is a ground-state property. As such, it is obtainable from the density-functional approach. There has been some discussion of evaluation of ground-state expectation values using the density-functional approach in the literature.[21] In this paper, we formulate the problem in terms of the self-consistent response of the crystalline

charge density to an added periodic perturbation.

The ground-state calculation is performed using the density-functional-theory approach yielding the density, ρ. Then the external potential is perturbed by an infinitesimal δV_{ext} and the ground-state calculation is repeated with the new external potential yielding $\rho + \delta \rho_{ind}$. The polarizability is defined by just $\delta \rho_{ind} = P \delta V_{ext}$, adopting a matrix notation. P, in principle, includes all exchange-correlation effects from the density-functional approach. This then involves the comparison of two ground-state calculations. A procedure for obtaining the polarizability matrix within the linear-response regime can be carried through numerically along these lines.[22]

The RPA corresponds to using the independent particle polarizability, P^0, given generally by Eq. (7), for ϵ in Eq. (23). The relationship to P derived above is obtained from the definition of P^0 via

$$\delta \rho_{ind} = P^0(\delta V_{ext} + \delta V_{ind}) \ .$$

Interpreting the effective potential in density-functional theory as a self-consistent field, one obtains

$$P = [1 - P^0(K_{xc} + v)]^{-1} P^0 \ . \tag{24}$$

Here $K_{xc} = \delta V_{xc}/\delta \rho$ gives the change in the exchange-correlation potential in the density-functional theory due to the change in the density. Thus, within the self-consistent field interpretation, the polarizability obtained from the ground-state calculations leads unambiguously back to the irreducible polarizability P^0 and hence the RPA for ϵ^{-1}, given a form for E_{xc}. A direct approach for obtaining the RPA response function has also been recently proposed.[23]

Using the density-functional approach, it is straightforward to go beyond the RPA and include exchange-correlation effects in the response function, at least within the LDA. In the context of the Green's-function approach, this is equivalent to an approximate evaluation of the vertex function.[24] We have done calculations for Si within the COHSEX including exchange-correlation effects in the dielectric matrix within the LDA. The resulting relative quasiparticle energies differ from those within the RPA by less than 0.1 eV. This suggests that the vertex corrections may not be too important in the evaluation of the quasiparticle energies in insulators.

In the present calculation, we use the dielectric matrix within the RPA, i.e., Eq. (7) for the irreducible polarizability. Rather than doing the direct calculation suggested above, the equivalent Adler-Wiser[25,26] formulation is used:

$$P^0_{GG'}(\mathbf{q}, \omega=0) = \sum_{n,n'\mathbf{k}} \langle n, \mathbf{k} | e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} | n', \mathbf{k}+\mathbf{q} \rangle \langle n', \mathbf{k}+\mathbf{q} | e^{i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}} | n, \mathbf{k} \rangle \frac{f(\varepsilon_{n',\mathbf{k}+\mathbf{q}}) - f(\varepsilon_{n,\mathbf{k}})}{\varepsilon_{n',\mathbf{k}+\mathbf{q}} - \varepsilon_{nk}} \ . \tag{25}$$

The fermion occupation factors are denoted by $f(\varepsilon)$. Then the static RPA dielectric response matrix is obtained from Eq. (23) using $P = P^0$.

The irreducible polarizability is obtained from the spectrum and wave functions of a self-consistent ab initio pseudopotential[14,27] density-functional calculation within the LDA (Ref. 28) using Eq. (25). For the case of Ge, scalar relativistic effects are included.[29] The static dielectric matrices are well converged with respect to all numerical cutoffs. We have also compared our calculated dielectric matrices to results obtained using direct methods proposed recently.[22] The agreement is excellent.[30] More details of the calculation of the static dielectric matrices will be published at a future date.[30] For the present calculation, dielectric matrices of approximate sizes 140×140 are used for Si and 200×200 for diamond, Ge, and LiCl. We find this to be sufficient to represent the local-field effects in the present problem.

B. Generalized plasmon-pole model

To extend the dielectric matrix to finite frequencies, we propose a generalized plasmon-pole (GPP) model. Various forms of plasmon-pole models have been used in the evaluation of the self-energy.[11,31,32] In particular, Lundqvist[33] used that approach for evaluating the quasiparticle energies in the electron gas case. He found the results differed very little from those obtained using the full Lindhard response function. In order to extend this ap-

proach to the full dielectric matrix, we observe that realistic calculations of the response function show that $Im\epsilon_{GG'}(\mathbf{q},\omega)$ is generally a peaked function in ω. Figure 2 displays the real and imaginary parts of $\epsilon_{G,G'}(\mathbf{q},\omega)$ for Si from a previous calculation using an empirical pseudopotential band structure.[34] Both the cases where $\mathbf{G} = \mathbf{G}'$ and the off-diagonal case exhibit a peak. For cases where there is not a single well-defined peak, the amplitude tends to be small and $Im\epsilon(\omega)$ fluctuates in sign.

Based on this observation we generalize the plasmon-pole model to the full inverse dielectric matrix. For each set of momentum components $(\mathbf{q}, \mathbf{G}, \mathbf{G}')$, $Im\epsilon^{-1}$ is taken to be

$$Im\epsilon^{-1}_{GG'}(\mathbf{q},\omega) = A_{GG'}(\mathbf{q})\{ \delta[\omega - \widetilde{\omega}_{GG'}(\mathbf{q})]$$
$$- \delta[\omega + \widetilde{\omega}_{GG'}(\mathbf{q})]\} \ . \tag{26}$$

We work with the causal dielectric matrix here for convenience. The corresponding real part is given by

$$Re\epsilon^{-1}_{GG'}(\mathbf{q},\omega) = 1 + \frac{\Omega^2_{GG'}(\mathbf{q})}{\omega^2 - \widetilde{\omega}^2_{GG'}(\mathbf{q})} \ . \tag{27}$$

The effective bare plasma frequency $\Omega_{GG'}(\mathbf{q})$ introduced here is defined by Eq. (31) below. Thus the full ω-dependent dielectric matrix can then be obtained once the matrices $A_{GG'}(\mathbf{q})$ and $\widetilde{\omega}_{GG'}(\mathbf{q})$ are determined.

In order to evaluate A and $\widetilde{\omega}$ we employ exact sum rules and the static dielectric matrices described in Sec.
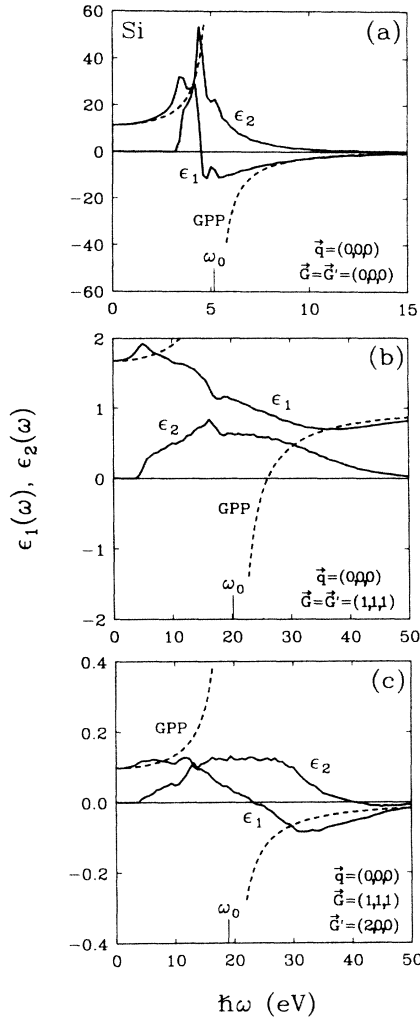
FIG. 2. Real and imaginary parts of the numerically calculated elements of the dielectric matrix based on the empirical pseudopotential results of Ref. 34 are plotted as a function of frequency for Si. In addition, the real part of the GPP model discussed in the text is plotted.

III A. The Kramer-Kronig relation provides one constraint

$$\mathrm{Re}\epsilon_{GG'}^{-1}(\mathbf{q}, \omega=0)=\delta_{GG'}+\frac{2}{\pi}P\int_0^\infty d\omega\frac{1}{\omega}\mathrm{Im}\epsilon_{GG'}^{-1}(\mathbf{q},\omega).$$

(28)

The second constraint is provided by a generalized f-sum rule relating the imaginary part of the many-body dielectric matrix to the plasma frequency and the crystalline charge density:

$$\int_0^\infty d\omega\,\omega\,\mathrm{Im}\,\epsilon_{GG'}^{-1}(\mathbf{q},\omega)$$

$$=-\frac{\pi}{2}\omega_p^2\frac{(\mathbf{q}+\mathbf{G})\cdot(\mathbf{q}+\mathbf{G}')}{|\mathbf{q}+\mathbf{G}|^2}\frac{\rho(\mathbf{G}-\mathbf{G}')}{\rho(0)}.$$

(29)

Equation (29) is analogous to the Johnson sum rules.[35]

The result established independently here[15] (see Appendix A) and discussed extensively by Taut[36] applies generally to the exact many-body linear-response function $\epsilon^{-1}(\omega)$.

Using the exact results Eqs. (28) and (29), the parameters $\widetilde{\omega}$ and A can be obtained for each set of momentum components $\mathbf{q},\mathbf{G},\mathbf{G}'$. The mode frequency is

$$\widetilde{\omega}_{GG'}^2(\mathbf{q})=\frac{\Omega_{GG'}^2(\mathbf{q})}{\delta_{GG'}-\epsilon_{GG'}^{-1}(\mathbf{q},\omega=0)}.$$

(30)

Here we define an effective bare plasma frequency based on the right-hand side of the sum rules Eq. (29):

$$\Omega_{GG'}^2(\mathbf{q})=\omega_p^2\frac{(\mathbf{q}+\mathbf{G})\cdot(\mathbf{q}+\mathbf{G}')}{|\mathbf{q}+\mathbf{G}|^2}\frac{\rho(\mathbf{G}-\mathbf{G}')}{\rho(0)}.$$

(31)

The amplitude A can similarly be determined:

$$A_{GG'}(\mathbf{q})=-\frac{\pi}{2}\frac{\Omega_{GG'}(\mathbf{q})}{\widetilde{\omega}_{GG'}(\mathbf{q})}.$$

(32)

There are no adjustable parameters. The model moreover reproduces the ω and ω^{-1} moments of the exact response function by construction. The present model as applied to the diagonal elements of the response function is similar to that used by Overhauser for the electron gas,[31] although derived from a different point of view.

Since the evaluation of the self-energy operator generally involves a sum over frequencies in the screened Coulomb interaction, the fine details of the frequency dependence of the dielectric function should not be important. Thus for the present purposes, the GPP model is valid if it represents the average features of $\mathrm{Re}\epsilon^{-1}(\omega)$ for all the important elements of the dielectric matrix. As one measure of the reliability of the GPP model, we compare it to the numerically calculated dielectric matrix from Ref. 34 for the elements depicted in Fig. 2. When the plasmon-pole model for $\epsilon^{-1}(\omega)$ is inverted to yield $\epsilon(\omega)$, the result is that $\mathrm{Im}\epsilon(\omega)$ has a δ-function peak at an average energy gap ω_0. This is related to the mode frequencies by

$$\widetilde{\omega}_{GG'}^2(\mathbf{q})=\Omega_{GG'}^2(\mathbf{q})+\omega_{0,GG'}^2(\mathbf{q}).$$

Then the same procedure described above can be used to generate a model for all the elements of the $\epsilon_{GG'}(\mathbf{q},\omega)$ matrix. The result is plotted for each case in Fig. 2. The frequency ω_0 is indicated. The model reproduces the numerically calculated $\mathrm{Re}\epsilon(\omega)$ very well except in the region of the resonance. The numerical calculation interpolates between the two branches of the model in the region where $\mathrm{Im}\epsilon(\omega)$ is large. This is true for $\mathbf{q}\rightarrow0$ as one might expect since the plasmon is a well-defined excitation. As seen in Fig. 2, it is also a reasonable approximation for larger values of $\mathbf{q}+\mathbf{G}$ and off-diagonal elements where there is no sharp excitation.

In the self-energy calculation the important contributions generally involve frequencies small compared to $\widetilde{\omega}$. Therefore a more useful comparison is for $\mathrm{Re}\epsilon^{-1}(\omega)$ in that range. In Fig. 3, the numerically calculated $\mathrm{Im}\epsilon^{-1}$ for Si based on the empirical pseudopotential[37] is shown for $\mathbf{q}=\mathbf{G}=\mathbf{G}'=0$ and $\mathbf{q}=(\pi/a,0,0)$, $\mathbf{G}=\mathbf{G}'=0$. The corresponding plasmon-pole model is also shown. The
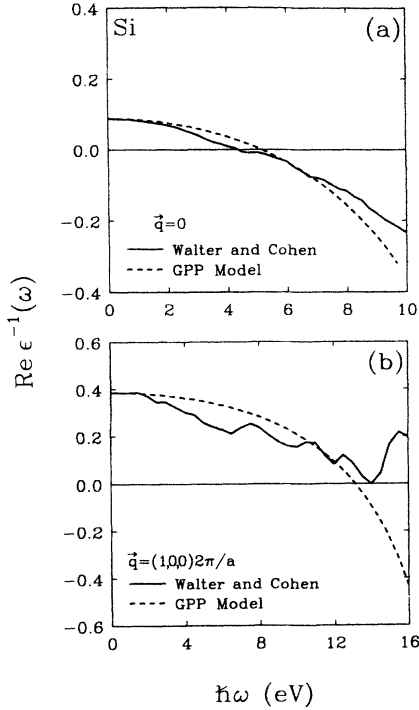
FIG. 3. The real part of ϵ^{-1} derived from the numerical results in Ref. 37 is plotted as a function of frequency in comparison to the GPP model results as described in the text.

comparison is quite favorable, even for q well away from zero where the imaginary part is broad as in the examples shown in Fig. 2.

In summary, we note that the GPP model essentially assumes the spectral function for the screened interaction to be a single narrow peak in frequency for each set of

momentum components. This is reasonable in the context of the self-energy calculation as discussed above. There are no adjustable parameters in the GPP model and it compares favorably to numerical results for the frequency dependence of the response function in the important frequency range.

IV. EVALUATION OF THE QUASIPARTICLE ENERGIES

The technical details of the calculation are described in this section. The approach for solving the quasiparticle equation is discussed in subsection A. Subsection B gives the details of the pseudopotential used for the electron-ion interaction. Details of numerical convergence are presented in subsection C.

A. Solution of the quasiparticle equation

In order to solve Eq. (1), the quasiparticle wave function is expanded in a basis set consisting of the self-consistent wave functions used in the construction of the Green's function as described in Sec. II:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{n'} \alpha_{n,n'}(\mathbf{k})\phi_{n'\mathbf{k}}(\mathbf{r}) \ . \tag{33}$$

The wave functions $\phi_{n\mathbf{k}}$ are derived from solving the effective one-particle equations with $V_{\text{eff}} = V_{\text{xc}}$ evaluated in the LDA. Thus Eq. (1) is transformed to a matrix eigenvalue equation with matrix elements of the self-energy operator required between the LDA states. Expanding the screened interaction in plane waves and using the form for the dielectric function given by the GPP model described in Sec. III B, the form of the matrix elements is given by

$$\langle n\mathbf{k} \mid \Sigma_{\text{SEX}}(\mathbf{r},\mathbf{r}';E) \mid n'\mathbf{k} \rangle = -\sum_{n_1}^{\text{occ}} \sum_{\mathbf{q},\mathbf{G},\mathbf{G}'} \langle n\mathbf{k} \mid e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} \mid n_1,\mathbf{k}-\mathbf{q} \rangle \langle n_1,\mathbf{k}-\mathbf{q} \mid e^{-i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} \mid n'\mathbf{k} \rangle$$

$$\times \left[1 + \frac{\Omega_{\mathbf{G}\mathbf{G}'}^2(\mathbf{q})}{(E - \varepsilon_{n_1\mathbf{k}-\mathbf{q}})^2 - \widetilde{\omega}_{\mathbf{G}\mathbf{G}'}^2(\mathbf{q})} \right] v(\mathbf{q}+\mathbf{G}') \ , \tag{34a}$$

$$\langle n\mathbf{k} \mid \Sigma_{\text{COH}}(\mathbf{r},\mathbf{r}';E) \mid n'\mathbf{k} \rangle = \sum_{n_1} \sum_{\mathbf{q},\mathbf{G},\mathbf{G}'} \langle n\mathbf{k} \mid e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} \mid n_1,\mathbf{k}-\mathbf{q} \rangle \langle n_1,\mathbf{k}-\mathbf{q} \mid e^{-i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} \mid n'\mathbf{k} \rangle$$

$$\times \frac{1}{2} \frac{\Omega_{\mathbf{G}\mathbf{G}'}^2(\mathbf{q})}{\widetilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})[E - \varepsilon_{n_1\mathbf{k}-\mathbf{q}} - \widetilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})]} v(\mathbf{q}+\mathbf{G}') \ . \tag{34b}$$

The states $\mid n\mathbf{k} \rangle$ are the $\phi_{n\mathbf{k}}$. In the sum over n_1, only the states of the same spin as $(n,\mathbf{k})$, $(n',\mathbf{k})$ contribute. Further details of the evaluation of the matrix elements are given in Appendix B.

In principle, a secular determinant in n,n' is required. In practice, this is unnecessary. We find that $\alpha_{nn'} \approx \delta_{nn'}$. As described in Appendix B, one can also obtain $\Sigma_{\mathbf{G}\mathbf{G}'}(\mathbf{k};E)$. Then Eq. (1) can be diagonalized directly in a plane-wave basis. This was done for several states at sym-

metry points in the Brillouin zone. We find that the resulting quasiparticle wave function has better than 99.9% overlap with the corresponding one-particle wave function from the density-functional band-structure calculation. That the quasiparticle wave functions should be so close to the LDA wave functions is a somewhat surprising result. The approximation of using only the diagonal element $\langle n\mathbf{k} \mid \Sigma \mid n\mathbf{k} \rangle$ in evaluating the quasiparticle energies leads to errors ranging from less than 0.01 eV for Si to

less than 0.05 eV for LiCl.

This allows the further simplification of only calculating the diagonal matrix elements of the self-energy operator. Equation (1) can be reduced to

$$E_{n\mathbf{k}}^{qp} = \epsilon_{n\mathbf{k}}^{LDA} - \langle n\mathbf{k} | V_{xc}^{LDA} | n\mathbf{k} \rangle + \langle n\mathbf{k} | \Sigma(E_{n\mathbf{k}}^{qp}) | n\mathbf{k} \rangle .$$

(35)

The self-energy operator must still be evaluated at the quasiparticle energy. This is done by expanding the matrix element of the self-energy operator to first order in the energy around $\epsilon_{n\mathbf{k}}$. Then the quasiparticle energy is obtained explicitly:

$$E_{n\mathbf{k}}^{qp} = E_{n\mathbf{k}}^0 + \frac{\Delta\Sigma_{n\mathbf{k}}(\epsilon_{n\mathbf{k}})/\Delta E}{1 - \Delta\Sigma_{n\mathbf{k}}(\epsilon_{n\mathbf{k}})/\Delta E}(E_{n\mathbf{k}}^0 - \epsilon_{n\mathbf{k}}) ,$$

(36)

where

$$E_{n\mathbf{k}}^0 = \epsilon_{n\mathbf{k}}^{LDA} - \langle n\mathbf{k} | V_{xc}^{LDA} | n\mathbf{k} \rangle + \langle n\mathbf{k} | \Sigma(\epsilon_{n\mathbf{k}}^{LDA}) | n\mathbf{k} \rangle .$$

(37)

A compact notation for the matrix element of the self-energy operator is used. The energy derivative is evaluated through a finite difference with ΔE typically 1 eV. The results are quite insensitive to a reasonable choice of ΔE. From Fig. 1 it is seen that E^{qp} does not differ by more than approximately 2 eV from ϵ for the materials considered here with LiCl exhibiting the largest changes. The difference between the first and second iteration quasiparticle energies is much smaller when the spectrum is updated and Σ recalculated as described in Sec. II.

B. The electron-ion interaction

We treat the electron-ion interaction using the *ab initio* pseudopotential technique.[14,27] The reference configuration was chosen to be singly ionized and the Ceperely-Alder form of the LDA was used.[28] For the case of Ge, relativistic effects were included in the generation of the pseudopotential.[29] Thus scalar relativistic effects are carried through the entire Ge calculation through the pseudopotential. The spin-orbit part is included at the end in first-order perturbation theory.[38] In each case, transferability of the potentials to atomic configurations nearby in energy was found to be of order 1 mRy or better, with only the Li potential being slightly worse (5 mRy).

An important point to note here is that the valence-core exchange-correlation term frozen into the pseudopotential is taken within the LDA. Thus we have taken

$$\Sigma_{X\text{ core-val}} + \Sigma_{\text{core-pol}} \rightarrow V_{xc\text{ core-val}} ,$$

(38)

where the core-valence exchange and core-polarization contributions to Σ were discussed in Sec. II. These terms are generally small as shown explicitly in the case of atomic Na.[12] They have also been estimated for solid Al.[39] These results suggest that the core-valence terms give a net contribution to the quasiparticle energy of approximately 1 eV relative to the bottom of the valence band in absolute terms. However, here we only need to estimate the much smaller error introduced by the *replace-*

ment of the proper core-valence terms by the LDA contribution as indicated in Eq. (38).

Among the four materials considered here (diamond, Si, Ge, and LiCl), we expect Ge to have the largest errors because of its relatively large and soft core. We have made estimates of the core-valence interaction for atomic Ge in the configuration $4s^1 4p^3$. The contribution of the core-valence part of the exchange-correlation potential is taken to be

$$\langle i | V_{xc\text{ core-val}} | i \rangle = \langle i | V_{xc}[\rho_{tot}] | i \rangle - \langle i | V_{xc}[\rho_{val}] | i \rangle ,$$

(39)

where i is the $4s$ or $4p$ atomic wave function. The bare valence-core exchange can be derived from an atomic Hartree-Fock calculation.[40] The bare core-polarization contribution can be estimated from the Rydberg states of the Ge^{3+} ion; it is given by the difference between the Hartree-Fock energy parameter and the experimental term value, corrected for relativistic effects. Our estimates of these quantities are listed in Table II for the $4s$ and $4p$ electrons. The final column gives the difference between the two sides of Eq. (38). The net difference is quite small. It suggests that the approximation in Eq. (38) could introduce, at most, approximately 0.3 eV shift between states of entirely p character and those of entirely s character.

In practice, the atomic estimates reported in Table II are hard to apply to the covalent crystal where the eigenstates are bonding and antibonding combinations of atomic orbitals. Further, the core-polarization term requires double screening in the solid. Also, most of the eigenstates are of mixed angular momentum character around the atomic sites. There is one exception to this. In the homopolar semiconductors, the conduction-band state with Γ_2 symmetry is of almost pure s character and further has all of its weight localized on the ion cores. This state will maximally probe the approximation in Eq. (38). It turns out that this state is systematically too low in energy in our calculation throughout the series of homopolar materials. The valence-core contribution from LDA is too deep for the s states relative to the p states. Therefore, proper account of the core-valence terms would tend to increase the energy for the s states relative to the p states at the valence-band edge. As will be shown in Sec. VI, this will bring the theoretical results into even better agreement with experiment.

TABLE II. The core-valence contribution to the atomic eigenvalues of atomic Ge in the configuration $4s^1 4p^3$. The exchange-correlation potential in LDA is compared to the bare-exchange plus core-polarization contribution. Energies are given in Ry.

Ge	LDA	Exchange	Core Polarization	Δ
$4s$	−0.322	−0.221	−0.077	−0.024
$4p$	−0.162	−0.108	−0.057	0.003

C. Numerical details

The self-consistent density-functional calculation is carried out in momentum space.[41] The wave functions are expanded in plane waves with an energy cutoff: 50 Ry for diamond, 17 Ry for Si, 20 Ry for Ge, and 25 Ry for LiCl. These were chosen to ensure complete convergence of the corresponding spectrum including the Γ'_{2c} state to within a few hundreths of an eV. Similar cutoffs were employed in the wave functions and spectrum used to compute the dielectric matrices. Summations over the Brillouin zone were carried out using the special points scheme.[42]

In the evaluation of the matrix elements of the self-energy operator in Eq. (34), another sum over the Brillouin zone is required. For this case, a different grid was used. The unit cell in reciprocal space was uniformly divided such that the grid included the point $q = 0$. These k points were then reduced to the irreducible wedge of the Brillouin zone. For the diamond and rock-salt structures, this generates sets including 3, 8, 16, etc. points in the irreducible part of the Brillouin zone. The point $q = 0$ requires special handling when $G' = 0$ in Eq. (34) because of the singularity in the bare Coulomb interaction. This is treated in the manner described by Phillips and Kleinman[43] as discussed in Appendix B. We have found that the quasiparticle energies are well converged for the set of 8 q points in the irreducible wedge of the Brillouin zone.[44] There are several other cutoffs required in the evaluation of the matrix elements of the self-energy in Eq. (34). The sum over umklapp scattering G, G' is cut off by a G_{max} such that $|q + G| \leq G_{max}$. For the parts involving the dielectric matrix, the sizes of the dielectric matrices were chosen to ensure convergence and are representative of the cutoffs used. It turned out that the COHSEX calculations were more sensitive to this cutoff than the final dynamical calculation. The sum over bands in the Coulomb-hole term in principle ranges over all states. In practice, the plane-wave matrix elements damp the contribution of the higher states. We find, however, that approximately 64 bands are required to achieve full numerical convergence.

In summary, we estimate that the results for the band gaps and band dispersions reported in this paper are reliable to within approximately 0.1 eV for the homopolar materials and somewhat worse for the gaps in LiCl. The absolute magnitude of the matrix elements of the self-energy operator, $\langle n\mathbf{k} | \Sigma | n\mathbf{k} \rangle$, are converged to within a few tenths of an eV. At that level, the numerical results for the band structure are representative of the fundamental approximations made in the theory and are within the experimental uncertainties of most of the available optical and photoemission data.

V. QUALITATIVE FEATURES
OF THE ELECTRON SELF-ENERGY OPERATOR

Two main features of the calculation to be sorted out are the role of local-field effects in the screening and the effect of dynamical renormalization. Both are crucial for obtaining the quasiparticle energies. For the purpose of analyzing these effects, the COHSEX approximation pro-

vides a useful intermediate step. The role of local fields is brought out clearly through the simpler COHSEX approximation to the self-energy operator. The nonlocality in real space is also illustrated. Then the role of dynamical effects can be traced, in part, through the differences between the COHSEX approximation and the full GW expression for the self-energy. As a separate issue, the degree to which the exchange-correlation potential, V_{xc}, in the LDA reproduces the effect of the self-energy operator is also quite interesting.

In Table III, the results for the fundamental gap are given for the materials studied here using four different approximations in comparison to experiment.[17,18] The four approximations are: the LDA potential, the COHSEX approximations without local fields, the COHSEX approximation including local fields in the screening and, finally, the full GW approximation evaluated within the GPP model. The trend for all the materials is clearly the same. The LDA potential gives band gaps that are too small as compared to experiment by 30% or more. For the case of Ge, the indirect gap shown here is still nonzero, but the direct gap has closed leading to semimetallic behavior. The COHSEX approximation with homogeneous screening (diagonal elements of the dielectric matrix only) generally gives a gap in better agreement with experiment, although not for Si. Inclusion of the full dielectric matrix in the COHSEX approximation dramatically opens up the gap. Finally, the full GW expression yields the gap in excellent agreement with experiment; taking account of the dynamical effects reduces the gap.

Figure 4 gives a different perspective on the same sequence of approximations. In this case, we focus on Si and track the magnitude of the matrix element of the self-energy operator for the valence-band edge and conduction-band edge (actually the X_{1c} state). The LDA potential is evaluated using the RPA for the correlation contribution.[19] The values given for the GW approximation are for the self-energy operator evaluated at the appropriate quasiparticle energy. To interpret this figure, note that the difference

$$\langle \text{CBM} | \Sigma | \text{CBM} \rangle - \langle \text{VBM} | \Sigma | \text{VBM} \rangle$$

TABLE III. Trends are shown for the fundamental gap E_g of diamond, Si, Ge, and LiCl for several approximations to the electron self-energy operator as described in the text in comparison to experiment. For Ge the indirect gap is shown including relativistic effects.

	LDA	COHSEX no LF	COHSEX LF	GW	Expt.
Diamond	3.9	5.1	6.6	5.6	5.48[a]
Si	0.52	0.50	1.70	1.29	1.17[a]
Ge	0.07	0.33	1.09	0.75	0.744[a]
LiCl	6.0	8.2	10.4	9.1	9.4[b]

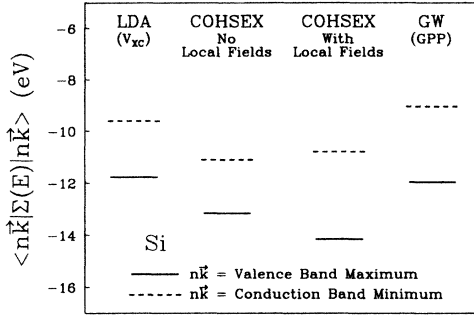[a]Reference 17.
[b]Reference 18.

FIG. 4. The matrix elements of the electron self-energy operator are compared for four different approximations discussed in the text for the case of Si. The matrix elements for the states at the valence-band edge and conduction-band edge are shown.

gives the exchange-correlation contribution to the band-gap energy. Thus, for example, a deepening of the valence-band self-energy relative to the conduction band leads to a larger band gap. Compared to the LDA potential, the COHSEX approximation gives larger self-energies, although the relative change for Si is small. Inclusion of local-field effects further deepens the valence electron self-energy while reducing the conduction-band self-energy slightly. This is the effect of local fields on the band gap. The net effect is a large increase in the band gap. The GW approximation yields self-energies comparable to the LDA potential with the valence electron self-energy being similar and the conduction electron self-energy less deep. This yields a gap that is larger than that found with the LDA potential. It is also clear from Fig. 4 that the COHSEX approximation systematically overestimates the magnitude of the self-energy operator by approximately 20%.

To supplement the data in Fig. 4, the matrix elements of the self-energy operator are broken down into the SEX and COH contributions in Table IV for the valence-band-edge and conduction-band-edge (X_{1c}) states in Si. For the COHSEX approximation, the cases without and with lo-

TABLE IV. The SEX and COH contributions to the matrix elements of the self-energy operator for the states at the valence- and conduction-band edges in Si are shown separately in three different cases: the static COHSEX approximation with no local fields (no LF), the COHSEX approximation including local fields (LF), and the full dynamical calculation including local fields.

Si	COHSEX		GW
	no LF	LF	
Σ_{SEX}			
Γ'_{25v}	−4.44	−3.85	−3.56
X_{1c}	−2.37	−2.08	−1.65
Σ_{COH}			
Γ'_{25v}	−8.72	−10.30	−8.41
X_{1c}	−8.72	−8.70	−7.40

cal fields are shown separately. The results of the dynamical calculation are also shown. In the COHSEX approximation, it is clear from Table IV that the largest effect of local fields is on the Coulomb-hole term. The Coulomb-hole term is primarily responsible for the dramatic increase in the band gap observed when local fields are included in Table III and Fig. 4. It is also seen from Table IV that the contribution of dynamical effects changes most the Coulomb-hole term. However, we emphasize that the screened-exchange and Coulomb-hole terms must be treated with the same degree of approximation to obtain quantitatively reliable quasiparticle energies.

The foregoing discussion is an overview of the qualitative features of the present calculation. In the remainder of this section, we will take up in detail the role of local fields and nonlocality using the COHSEX approximation in subsection A, the role of dynamical effects in subsection B, and finally an analysis of the LDA potential in comparison to the self-energy operator in subsection C.

A. Local fields and nonlocality in the self-energy operator

The local fields in the screening arise from the microscopic response of the inhomogeneous charge density in the crystal to a perturbation. The local fields are described by the off-diagonal elements of the dielectric matrix. It is intuitively clear that this is important for discussing the response to a perturbation in a semiconductor or an insulator. There is significant accumulation of charge in the covalent bonds or around the anion in ionic compounds. In a simple metal this effect is much less important. It is thus expected that many of the properties of a simple metal can be represented by a uniform charge density, the jellium model. With the full translation invariance in the jellium model, all two-point functions depend only on the *difference* between the two coordinates: $|\mathbf{r}-\mathbf{r}'|$. For example, the potential around an added point charge will not depend on its location, $\mathbf{r}'$. Furthermore, the screening potential will have spherical symmetry around $\mathbf{r}'$. However, for the inhomogeneous case of semiconductors or insulators, this is not an adequate approximation. In the example of the added point charge, the induced potential is a full two-point function, $V_{scr}(\mathbf{r},\mathbf{r}')$. The details of the screening potential depend dramatically on the location of the added charge, $\mathbf{r}'$, and in general the screening potential is anisotropic. First we will illustrate this for the case of Si responding to an extra point charge of one electron and then discuss the effects of local fields on the self-energy operator. Finally, the nonlocality of the self-energy operator will be discussed.

The expression for the electrostatic screening potential around an extra electron is straightforward to derive in terms of the dielectric matrix:

$$V_{scr}(\mathbf{r},\mathbf{r}') = \sum_{\mathbf{q},\mathbf{G},\mathbf{G}'} e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}[\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q},\omega=0)-\delta_{\mathbf{G}\mathbf{G}'}]$$
$$\times \frac{4\pi e^2}{\Omega|\mathbf{q}+\mathbf{G}'|^2}e^{-i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} . \tag{40}$$

For simplicity of discussion, only the static case is considered. The usual Coulomb interaction is present as the

perturbation from the point charge at $\mathbf{r}'$. In the case where the dielectric matrix is zero for $\mathbf{G} \neq \mathbf{G}'$ (local fields are negligible) this expression reduces to being a function of $\mathbf{r} - \mathbf{r}'$.

The screening potential around a point electron has been computed for Si in the $(1\bar{1}0)$ plane for the added electron at several points using Eq. (40). The static dielectric matrices computed in the RPA as described in Sec. III A are used.[45] The screening potential is shown in Fig. 5 for the added electron, $\mathbf{r}'$, at the bonding (a) and antibonding (c) sites. Negative contours are dashed and the figure reflects the fact that the added electron repels the other electrons, leading to an attractive screening potential.

For the case of the center of the bond, Fig. 5(a), the screening potential is close to spherical in the bond region but shows some elongation reflecting the underlying crystalline charge density in the bond. In Fig. 5(b), the contribution to the screening potential from local fields *only* is displayed for the case of an added electron at the center of the bond. This is the contribution from the terms in Eq. (40) with $\mathbf{G} \neq \mathbf{G}'$. It is clear from Figs. 5(a) and 5(b) that the local fields contribute more than one-third of the screening potential in the region near the center of the bond. This is also the contribution responsible for the anisotropy evident in Fig. 5(a). For the case of the added electron at the antibonding site, Fig. 5(c), the induced charge is both anisotropic and centered away from $\mathbf{r}'$ because the charge density is highly anisotropic around this

FIG. 5. In (a) and (c) the screening potential in response to a single electron at $\mathbf{r}'$ (indicated by the $+$) is displayed in the $(1\bar{1}0)$ plane of Si in units of Ry. (b) and (d) give the contribution from local fields *only*. The bond chain is indicated schematically.

site. The screening potential is correspondingly distorted towards the region of bond charge. Figure 5(d) shows the local-field contribution in this case. It is the local-field effect which shifts the screening potential, building it up near the bonds *and* reducing it in the interstitial region of low-charge density. The local fields reflect the difference in polarizability between the bond region and the interstitial region. The result displayed in Fig. 5 is precisely what one expects intuitively: because of the higher charge density in the bond region, an added electron can produce a deeper screening hole while in the interstitial region the low crystalline charge density leads to a shallower screening hole. Also, the short-range nature of the local-field effects are evident in Fig. 5. More than a bond length away from $\mathbf{r}'$, the screening potential is close to spherical and the local-field contribution is small.

We now examine the effect of local fields on the self-energy operator using the simpler COHSEX approximation of Eq. (19). As noted in the discussion of Table IV, the Coulomb-hole term is primarily responsible for the dramatic increase in the band gap observed when local fields are included. This was illustrated in Table III and Fig. 4. In the COHSEX approximation, the Coulomb-hole term can be written in terms of a local potential:

$$\Sigma_{\mathrm{COH}} = \delta(\mathbf{r} - \mathbf{r}') V_{\mathrm{COH}}(\mathbf{r}) ,$$

$$V_{\mathrm{COH}}(\mathbf{r}) = \tfrac{1}{2} V_{\mathrm{scr}}(\mathbf{r}, \mathbf{r}) , \qquad (41)$$

where V_{scr} is given by Eq. (40). The potential is the induced potential due to an electron at $\mathbf{r}$ evaluated at the site of the electron, $\mathbf{r}$. The factor of a half relates to the adiabatic buildup of the screening charge in evaluating the energy gained by building the screening hole around the added electron. It is clear that if V_{scr} depends only on $\mathbf{r} - \mathbf{r}'$, then $V_{\mathrm{COH}}(\mathbf{r})$ will be *uniform*. Therefore, if local fields are neglected in the COHSEX approximation, the Coulomb-hole term makes *no* contribution to dispersion within a band or the band gap. Inclusion of local fields leads to a potential plotted for the case of Si in the $(1\bar{1}0)$ plane in Fig. 6. In accordance with the discussion of V_{scr} above, V_{COH} is deeper in the bond chain and shallower in the interstitial region. This leads to a substantial contribution to the gap since the valence- and conduction-band states are concentrated in different places in the unit cell.[46] In particular, the valence-band-edge states are localized in the bond region. From Fig. 6 it is clear that local fields deepen the matrix elements of Σ_{COH} for the valence-band edge as observed in Table IV. This is the dominant contribution of the local fields seen in Table III and Fig. 4.

The screened-exchange term is nonlocal and, in the COHSEX approximation, can be factored in real space into the exchange charge

$$\rho_x(\mathbf{r}, \mathbf{r}') = \sum_{n,\mathbf{k}}^{\mathrm{occ}} \phi_{n\mathbf{k}}(\mathbf{r}) \phi_{n\mathbf{k}}^*(\mathbf{r}') , \qquad (42)$$

and the static screened Coulomb interaction. Accounting for the fact that a given electron only couples to the states of the same spin in the sum over states in Eq. (19) or Eq.
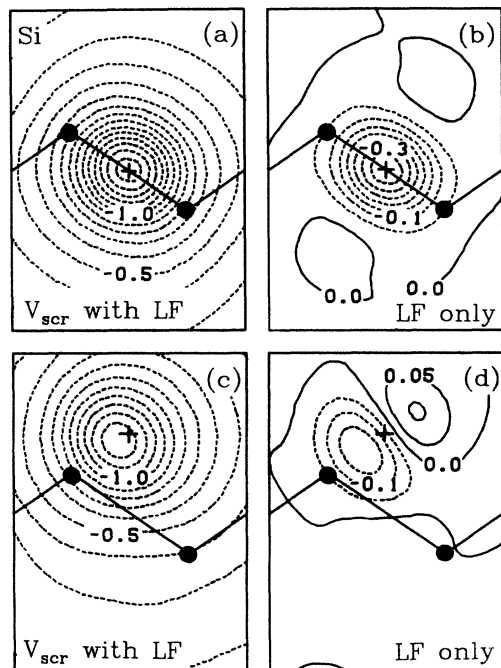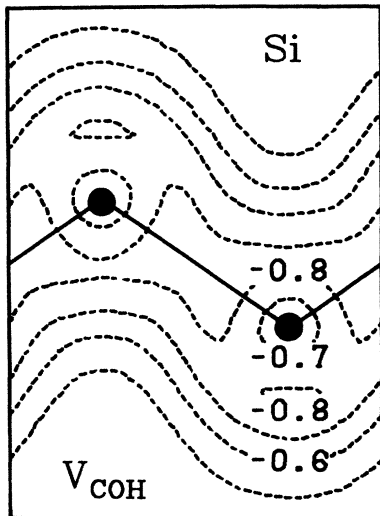
FIG. 6. The local potential that gives the Coulomb-hole part of the electron self-energy in the COHSEX approximation is plotted in the ($1\bar{1}0$) plane of Si. The units are Ry.

(42), the screened-exchange operator can be written as

$$\Sigma_{SEX}(\mathbf{r},\mathbf{r}') = -\tfrac{1}{2}\rho_x(\mathbf{r},\mathbf{r}')\left[V_{scr}(\mathbf{r},\mathbf{r}') + \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \right]. \qquad (43)$$

The effect of the local fields on V_{scr} are confined to the region near $\mathbf{r}'$. But in this region, the screened-exchange operator is dominated by the bare part of the interaction. Therefore the effect of local fields on Σ_{SEX} is relatively small as seen in Table IV. Generally speaking, the local-field contribution to the gap from Σ_{SEX} is less than 25% of the contribution from Σ_{COH}, but of *opposite* sign. Local fields must be included in both terms for quantitatively accurate results.

In the balance of this section, we examine aspects of the nonlocality of the self-energy operator using Σ_{SEX} in the COHSEX approximation for illustration. The exchange charge for Si in the ($1\bar{1}0$) plane has been calculated for $\mathbf{r}'$ at several points.[47] Then using the screening potential discussed above, the screened-exchange operator is calculated as in Eq. (43). These are displayed in Fig. 7 for $\mathbf{r}'$ fixed at the center of the bond, 7(a) and 7(b), and at the antibonding site, 7(c) and 7(d). Because of the singularity in the screened-exchange operator at $\mathbf{r}=\mathbf{r}'$, it has been multiplied by $|\mathbf{r}-\mathbf{r}'|$ in the figure. This yields a finite value at $\mathbf{r}=\mathbf{r}'$ and amplifies the lobes somewhat at large distance. Also, a logarithmic contour interval has been adopted for the plot of the screened-exchange operator.

The exchange charge in Figs. 7(a) and 7(c) reflects the buildup of charge in bonds and displays significant anisotropy. The value at $\mathbf{r}=\mathbf{r}'$ is precisely $\rho(\mathbf{r})$. Both the magnitude and shape as a function of $\mathbf{r}$ depend dramatically on $\mathbf{r}'$. In the jellium model, ρ_x is spherically symmetric with nodes at radii such that rk_F is a zero of the spherical Bessel function j_1. For an average density appropriate for Si, this places the first node at a radius of approximately 5

a.u. Since the bond length is approximately 4.4 a.u., one can see that the average distance of the first nodal surface indicated in Figs. 7(a) and 7(c) from $\mathbf{r}'$ is of that magnitude.

The degree of nonlocality in the self-energy operator is measured primarily by its range and secondarily by the detailed anisotropy. These properties are essentially given by the placement of the first nodal surface. In the COHSEX approximation, the nodal structure of Σ_{SEX} is taken precisely from the exchange charge. In particular, for an electron in the bond center, the nodal surface passes through the neighboring bonds. Therefore, in a local orbital picture the interaction with neighboring bonds should be relatively small. For the electron at the antibonding site, nodal surface passes through the second closest bonds so that the most significant interaction is with the neighboring bonds. The screened-exchange operator in a semiconductor is in principle long range due to the incomplete screening. However, far from $\mathbf{r}'$, the operator is reduced by a factor of ϵ_0 from the bare-exchange operator. Even the first lobe after the nodal surface is much reduced. From Figs. 7(b) and 7(d) one can see that the practical range is much shorter, at most a few bond lengths.

One interesting question regarding the nonlocality is its effect on the matrix elements of the self-energy operator. Referring to Table IV, the matrix elements of the
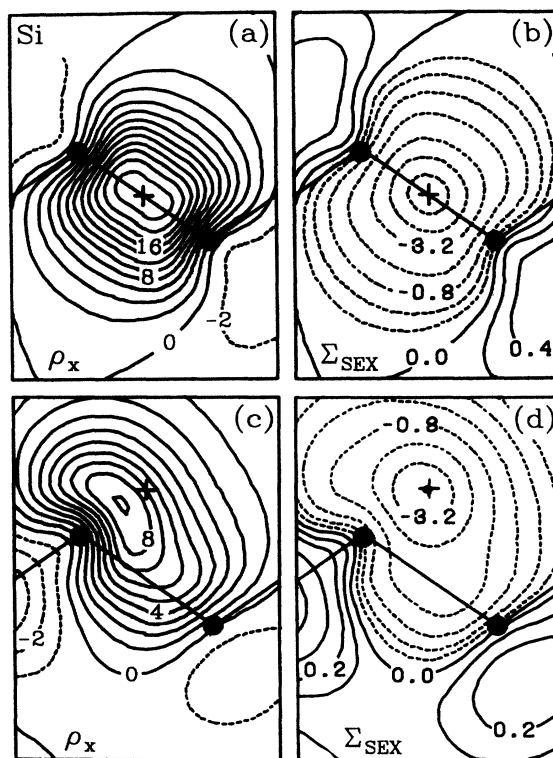

FIG. 7. In (a) and (c) the exchange charge is plotted in the ($1\bar{1}0$) plane of Si with $\mathbf{r}'$ fixed (indicated by the $+$). The units are electrons/cell. In (b) and (d) the screened-exchange part of the electron self-energy operator is plotted in the combination $|\mathbf{r}-\mathbf{r}'|\,\Sigma_{SEX}(\mathbf{r},\mathbf{r}')$ in the COHSEX approximation. The units are a.u. Ry/cell and the contours increase in powers of 2.

screened-exchange operator decrease in magnitude upon crossing the gap from occupied to empty states. This is even more dramatic for the bare-exchange operator where the relevant values are -12.54 eV (Γ'_{25v}) and -5.28 eV (X_{1c}). Unlike a local potential, the matrix elements of a nonlocal operator can be sensitive to the phase of the wave function. Therefore the matrix elements of the screened-exchange operator are sensitive to the extra nodes present in the conduction-band wave function. However, matrix elements of a local potential can also jump in magnitude when crossing the gap because the bonding states are concentrated in different parts of the unit cell than the anti-bonding states as emphasized above in connection with the Coulomb-hole term. Figure 7 suggests that this mechanism may also be relevant for the screened-exchange operator. The matrix elements of the screened-exchange operator with respect to the valence-band edge states will essentially average over the operator shown in Fig. 7(b) in the bond region. However, the conduction-band states at Γ are spatially located in the region around the antibonding site. That matrix element will essentially average over the operator shown in Fig. 7(d). From Fig. 7, these two cases display rather different magnitudes which can then account for the differences in the value of the matrix elements. Thus for an inhomogeneous charge density, both nonlocality and charge inhomogeneity seem to contribute to this effect.

B. Dynamical effects in the self-energy operator

In general terms, dynamical effects correspond to the energy dependence of the self-energy operator. For comparison, we summarize the general features expected for $\Sigma(E)$ here. A complete discussion is given in Ref. 12. Since $\Sigma(E)$ satisfies a dispersion relation, the energy dependence of the real part is related to the imaginary part of Σ being nonzero. In the jellium model, the dispersion of $\Sigma(k,E)$ displays resonances centered at $\mu \pm \omega_P$.[12] For hole states ($k < k_F$), the resonance at $\mu - \omega_P$ is stronger than the resonance at $\mu + \omega_P$ with the opposite behavior for electron states ($k > k_F$). On general grounds, one expects similar behavior in semiconductor systems, although the details of the imaginary part of the matrix elements of $\Sigma(E)$ will be richer. For energies near the quasiparticle energy, this leads to a general behavior of the real part of matrix elements of $\Sigma(E)$. The curves will have negative slope near the quasiparticle energy. They will be concave upwards for hole states and concave downwards for electron states. The nonzero slope at the quasiparticle energy is related to the renormalization constant

$$Z_{n\mathbf{k}} = \left[1 - \frac{\partial \Sigma_{n\mathbf{k}}(E)}{\partial E} \bigg|_{E=E_{n\mathbf{k}}^{qp}} \right]^{-1}. \tag{44}$$

The matrix element of Σ is indicated in a compressed notation. The renormalization constant gives the weight of the quasiparticle peak in the spectral function. Physically, the bare electron is renormalized: the quasiparticle includes the cloud of excitations associated with screening in the solid, electron-hole pairs and plasmons. A value near unity is indicative of a small associated cloud of vir-

tual collective excitations; the quasiparticle is a well-defined particlelike excitation.

These general ideas give the framework for studying the role of dynamical renormalization in the present theory. For each of the materials studied here, we plot matrix elements of Σ as a function of energy near the gap region. These are given in Fig. 8 for a few states near the gap. The corresponding quasiparticle energies are indicated by ticks on the horizontal axis. These plots summarize concisely many of the important results here. First, these give the magnitude of the self-energy for each material in the GW approximation using the GPP model. Second, the dispersion in energy conforms to the general features described above. Third, the curves for electron states are
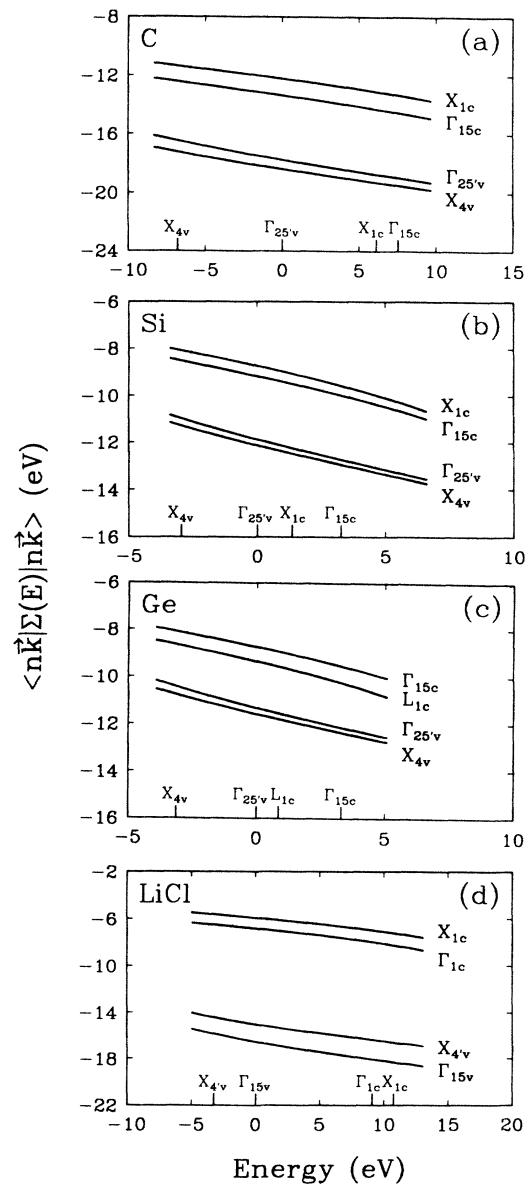

FIG. 8. Matrix elements of the electron self-energy operator evaluated in the GW approximation are drawn as a function of energy for selected states near the band gap. Results are displayed for (a) diamond, (b) Si, (c) Ge, and (d) LiCl.

close together as are the curves for the hole states. However, there is a substantial gap between the electron states and the hole states.

To analyze the contribution of the self-energy operator to the gap energy, consider breaking $\Sigma_{n\mathbf{k}}(E)$ into two terms: an energy-independent part $\Sigma_{n\mathbf{k}}(E_{\text{VBM}})$ and the remainder which contains the energy dependence. As an example, we consider the indirect gap in Si. The first term corresponds to drawing a vertical line at the valence-band-edge energy (Γ_{25v}') in Fig. 8. The intersection with each curve in Fig. 8 is the contribution of the first term to the self-energy operator. In particular, the $\Sigma_{n\mathbf{k}}(E)$ curves for the Γ_{25v}' and X_{1c} states are well separated. This first energy-independent term gives a large positive contribution to the gap energy. This contribution is due to real-space effects along the lines discussed in Sec. V A. The real-space contribution appears to be a large effect, but the energy dependence of the self-energy cannot be neglected: the slope of the curves is approximately -0.2. The self-energy operator must be evaluated at the quasiparticle energy, $\Sigma_{n\mathbf{k}}(E_{n\mathbf{k}})$. Therefore the energy-dependent term that remains accounts for moving from $E = E_{\Gamma_{25v}'}$ to $E = E_{X_{1c}}$ along the curve $\Sigma_{X_{1c}}(E)$. This gives a substantial negative contribution to the gap energy.

The suggested separation of real space and dynamical contributions to the gap energy gives us some insight into the trends shown in Tables III and IV and Fig. 4. In all the materials studied, inclusion of dynamical effects in comparison to the static COHSEX approximation reduces the gap energy. We emphasize that the COHSEX approximation is different in technical detail from evaluating the self-energy operator at a fixed energy such as the valence-band edge as suggested above. However, this analysis does give a picture of how introduction of dynamical renormalization leads to a reduced gap.

The final feature of Fig. 8 that we emphasize is that the renormalization constant Z introduced in Eq. (44) is roughly the same for the electron and hole states. Furthermore, Z is similar for all four materials studied. The actual values of Z for the valence- and conduction-band-edge states for the four materials studied are summarized in Table V. This similarity throughout is somewhat surprising in that semiconductors and insulators generally do *not* exhibit electron-hole symmetry. The similarity for the four materials may reflect the fact that the average electron density is similar for all four ($r_s \approx 2$) although the chemical aspects of the bonding vary considerably as to degree of metallicity and ionicity. With $Z \approx 0.8$ dynamical renormalization is not negligible; account of dynamical effects in the present theory is crucial for quantitative-

TABLE V. The renormalization constants Z for the hole state at the top of the valence-band model (VBM) and the electron state near the bottom of the conduction-band model (CBM) are shown for diamond, Si, Ge, and LiCl.

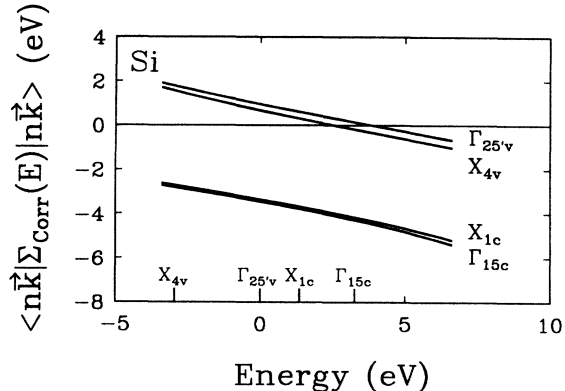|	Diamond	Si	Ge	LiCl
Z_{VBM}	0.86	0.78	0.79	0.83
Z_{CBM}	0.86	0.80	0.80	0.87

FIG. 9. For the case of Si, the matrix elements of the correlation part of the self-energy operator are displayed as a function of energy for several states near the gap region.

ly correct results. However, that Z is close to unity suggests that using the lowest-order term in a perturbation series, the GW approximation, is a reasonable approximation.

It is often instructive to separate out the correlation contribution to the self-energy; i.e., the bare exchange part of the operator is subtracted out. In Fig. 9, the correlation part of $\Sigma(E)$ is plotted as a function of E for Si. The trends for the other materials are similar. The results correspond to qualitative expectations. The correlation energy is positive for hole states and negative for electron states. The electron-hole *asymmetry* is evident as the magnitude of the correlation energy is substantially larger for the electron states. One interesting feature of Fig. 9 is that for states near the gap, the curves for the hole states are very close together. This is also true for the electron states. This suggests that the correlation energy near the gap region can be modeled by a constant shift plus an energy-dependent term for each case:

$$E_{e,\text{corr}}(E) \approx E_{e,\text{corr}}(E_{\text{CBM}}) + \alpha(E - E_{\text{CBM}}) \tag{45}$$

with a similar expression for the hole states. This is similar to the results of the simplest treatment of the polaron problem where the correlation energy consists of a self-energy shift plus a mass correction.[48]

As a final point regarding dynamical effects, general reasons for the overestimation of the self-energy operator in the COHSEX approximation can be given. This is illustrated in Fig. 4 and also can be seen to predominantly arise from the COH term in the self-energy operator from Table IV. An alternative view of dynamical effects is to note that the screened interaction of the added electron (or hole) with the other electrons in the crystal depends on the energy of the added electron. Loosely speaking, the degree of screening depends on how fast the electron is moving. This leads to analyzing the role of dynamical effects in terms of the details of the dynamical screening process. This can be done directly, but we suggest a more general argument here. Consider the case of a quasielectron corresponding to an added electron in the crystal. The COHSEX approximation as derived in Sec. II assumes

that the added electron interacts with the other electrons on a time scale long compared to the response time of the crystalline charge density. Essentially, it is assumed that the screening charge responds instantaneously to the motion of the added electron. The other electrons optimally screen the added electron leading to the lowest energy. However, in reality, the screening charge does not respond instantaneously so that the added electron is not optimally screened. The energy associated with the screening is not as low as that of the static case. Now the COH term in the self-energy operator is just the interaction of the added electron with the screening charge and so this argument suggests a reason for the overestimation of the magnitude of the self-energy operator in the COHSEX approximation. This picture corresponds to the common statement that the quasiparticle drags some of the screening cloud along with it.

C. Comparison to the LDA potential

Although our calculation does not proceed as a correction to the eigenvalues from the LDA potential, it is still interesting to see the character of such a correction. This is particularly true in light of recent interest in such corrections for the minimum gap.[9,10] We can examine *a posteriori* the form such a correction seems to have.

In Fig. 1, the difference between the quasiparticle energy computed using the GW approximation and the eigenvalue from the LDA potential in RPA (Ref. 19) is plotted versus the quasiparticle energy for several high-symmetry points in the Brillouin zone. The straight lines are drawn as a guide to the eye. As described in the Introduction, the main feature of these plots for all four materials (diamond, Si, Ge, and LiCl) is the large jump at the gap region. The correction near the gap region dominates. The other general feature is that the correction is a relatively smooth function (close to linear) of the quasiparticle energy away from the band edges. The scatter is small compared to the jump. The magnitude of the energy dependence varies with material, with diamond showing the largest energy dependence. Also, the conduction electrons show a different energy dependence, in general, than the valence electrons. Further, the distribution of the correction between the conduction electron self-energy and the valence electron self-energy is material dependent. Finally, the correction turns upwards deep in the valence band leading to relatively small corrections to the overall valence-bandwidth. This last effect may be related to the approximations used in evaluating the self-energy operator. In particular, one certainly expects the GPP model to work best for states near the gap region and deteriorate somewhat when the characteristic exchange frequencies required approach the magnitude of the plasma energy, as is the case for the states at the bottom of the valence band (or *s* band in LiCl).

There are general arguments which suggest that the energy of the highest occupied state should be given correctly by the exact density-functional eigenvalue for that state.[49] This implies that the correction for the valence-band-edge states should be zero (at $E^{qp}=0$ in Fig. 1). The

degree to which this is satisfied by the LDA potential gives some information on the reliability of the LDA for ground-state properties. The degree of lineup at the valence-band maximum also depends on the convergence of the absolute magnitude of the matrix elements of the self-energy operator (as opposed to the differences required for the gaps and band dispersions). From Fig. 1, we see that this is reasonably satisfied for the homopolar materials. There is a trend showing a smaller correction as a function of increasing metallicity (from diamond to Ge). For the case of LiCl, the correction at the valence-band edge is substantial suggesting that the LDA is not as good for this highly ionic (and inhomogeneous) material.

Figure 1 also suggests that the assumption usually made in applying the self-interaction correction to the LDA to solids is not correct.[50] The correction is not due to a lowering of the valence electron energy by removal of the spurious self-interaction energy. Rather, most of the correction is for the conduction-band electrons in the homopolar materials. For LiCl, the correction is evenly divided between valence and conduction bands.

The result shown in Fig. 1 together with the previous discussion of the quasiparticle wave functions give indirect evidence supporting the proposal of Sham and Schluter.[9] Through the occupied valence bands, the matrix element of the LDA potential tracks the matrix element of the self-energy operator fairly closely. But upon crossing the gap, there is a substantial decrease in the absolute magnitude of the matrix elements of the self-energy operator that is not reproduced by the LDA potential. This appears to be in accord with the picture proposed by Sham and Schluter. The added electron in the first excited state of the conduction band moves in a potential that differs from the potential seen in the N-electron system by a constant shift. This constant is traced to the discontinuity in a functional derivative of the exchange-correlation energy upon filling the system upto and across the gap. The near perfect overlap of the one-particle wave functions from the LDA potential with our calculated quasiparticle wave functions also supports this picture. Away from the gap region, the theory of Sham and Schluter does not apply and indeed Fig. 1 exhibits cases where the correction is dependent on energy away from band edges (diamond in particular).

VI. COMPARISON OF QUASIPARTICLE ENERGIES TO EXPERIMENT

We preface detailed discussion of the comparison of the present theory to experiment by noting that the quasiparticle energies calculated here strictly refer to experiments where an electron is added or removed from the solid. This corresponds, in principle, to inverse photoemission and photoemission. Optical experiments where both an electron and a hole are created do not strictly probe the quasiparticle energies. In principle, the optical response function can exhibit many-body effects beyond the quasiparticle interpretation, e.g., the effect of the electron-hole interaction. This has been shown to have a marked effect

on the amplitude of the features of the absorption spectrum of diamond[51] and Si.[52] However, a recent model calculation[53] suggests that the critical points in the spectrum are not shifted. The important point is that most of the data on critical points in the quasiparticle band structure have been unravelled from optical experiments or relatively low-energy photoemission experiments where such effects may manifest themselves. Values so derived may be shifted from the strict quasiparticle energies calculated here. We proceed on the assumption that such effects are relatively small, if present. The calculated gaps and direct transitions presented here will be taken as the difference between quasiparticle energies in keeping with a quasiparticle interpretation of the optical response.

The calculated results for diamond are compared to experimental results[17,54−56] in Table VI. The calculated results differ slightly from those reported earlier[15] because of improvements resulting from the second iteration in the calculation with the improved spectrum as described in Sec. II. The most reliable experimental results are the optical data for the fundamental gap (indirect at approximately 0.75Δ) and direct gaps at Γ. The theory agrees extremely well with the absorption edge[17] and the direct gaps derived from reflectance.[56] The calculated bandwidth is somewhat (5%) too narrow compared to the XPS measurement.[54] The angle-resolved ultraviolet photoemission spectroscopy (UPS) measurement of the bandwidth by Himpsel et al.[55] is significantly narrower than the x-ray photoemission spectroscopy (XPS) measurement and the theoretical value. This trend is also seen for the critical points measured at the L point in the Brillouin zone. Himpsel et al. deduce the conduction-band energy for the L'_{2c} state from the valence-band energies together with the photon energy.[55] The fact that the theoretical value is too small is also consistent with the compression of the valence bands in the UPS data. Based on constant initial-state spectroscopy, they deduce the energy of the Γ'_{2c} state.[55] The theoretical value is consistent with the experiment, although somewhat small. Roberts and Walker measured the reflectivity of diamond and obtained

ϵ_2.[56] The spectrum is dominated by a broad peak around 12 eV. Painter et al.[57] analyzed the reflectivity data in terms of an $X\alpha$ band structure and deduced that direct transitions near points X and L contribute to that broad peak. The current theory is in reasonable agreement with those values.

The results for Si are summarized in Table VII in comparison to experiment.[17,58−63] The calculated results are slightly different than those reported previously[15] because we make a second iteration to include an improved spectrum in the calculation of Σ. The fundamental gap (indirect at approximately 0.8Δ) is about 0.1 eV larger than experiment. The prominent critical points in the optical spectrum are well represented by the calculation. Here we make a comparison to features in the optical spectrum that are generally believed to be associated with direct transitions near high-symmetry points in the Brillouin zone. Based on the wavelength modulated reflectivity measurements of Zucca and Shen,[63] the E_1 feature is associated with a direct transition at L ($L'_{3v} \rightarrow L_{1c}$). The E'_1 transition is also assigned to L ($L'_{3v} \rightarrow L_{3c}$). The calculated second gap at the zone center (Γ'_{2c}) is a little too small in Si as noted for diamond. The second conduction-band edge in Si is at L. There are both absorption measurements of the second indirect gap[61] and angle-resolved inverse photoemission measurements of the empty states at L.[62] The two experiments nominally differ by 0.3 eV for the L_{1c}. The theory gives a value intermediate between

TABLE VI. Comparison of results (in eV) from the present calculation of the quasiparticle energies to experiment for diamond.

	Present theory	Expt.
Diamond		
E_g	5.6	5.48[a]
$\Gamma_{1v} \rightarrow \Gamma'_{25v}$	23.0	24.2±1,[b] 21±1[c]
$\Gamma'_{25v} \rightarrow \Gamma_{15c}$	7.5	7.3[d]
$\Gamma'_{25v} \rightarrow \Gamma'_{2c}$	14.8	15.3±0.5[c]
$L'_{2v} \rightarrow \Gamma'_{25v}$	17.3	15.2±0.3[c]
$L_{1v} \rightarrow \Gamma'_{25v}$	14.4	12.8±0.3[c]
$\Gamma'_{25v} \rightarrow L'_{2c}$	17.9	20 ±1.5[c]

[a]Reference 17.
[b]Reference 54.
[c]Reference 55.
[d]Reference 56.

TABLE VII. Comparison of results (in eV) from the present calculation of the quasiparticle energies to experiment for Si.

	Present theory	Expt.
Silicon		
E_g	1.29	1.17
$\Gamma_{1v} \rightarrow \Gamma'_{25v}$	12.04	12.5 ±0.6
$\Gamma'_{25v} \rightarrow \Gamma_{15c}$	3.35	3.4[h]
$\Gamma'_{25v} \rightarrow \Gamma'_{2c}$	4.08	4.2
$X_{4v} \rightarrow \Gamma'_{25v}$	2.99	2.9[b] 3.3 ±0.2[c]
$\Gamma'_{25v} \rightarrow X_{1c}$	1.44	1.3[d]
$L'_{2v} \rightarrow \Gamma'_{25v}$	9.79	9.3 ±0.4
$L_{1v} \rightarrow \Gamma'_{25v}$	7.18	6.7 ±0.2
$L'_{3v} \rightarrow \Gamma'_{25v}$	1.27	1.2 ±0.2,1.5[e]
$\Gamma'_{25v} \rightarrow L_{1c}$	2.27	2.1,[f] 2.4±0.15[g]
$\Gamma'_{25v} \rightarrow L_{3c}$	4.24	4.15±0.1[g]
$L'_{3v} \rightarrow L_{1c}$	3.54	3.45[h]
$L'_{3v} \rightarrow L_{3c}$	5.51	5.50[h]

[a]Reference 17 except where noted.
[b]Reference 58.
[c]Reference 59.
[d]Estimated from the indirect gap and longitudinal mass.
[e]Reference 60.
[f]Reference 61.
[g]Reference 62.
[h]Reference 63.

the experiments and does not give new insight into whether the electron-hole interaction is responsible for the difference between the experiments. The theory agrees with the inverse photoemission data for the L_{3c}. Turning to the XPS results for the occupied states, the overall agreement with experiment is quite good. The bandwidth is a few percent too narrow, but the L'_{2v} an L_{1v} critical points are a little too deep in energy. This reflects the bowing discussed in connection with Fig. 1. For both the X_{4v} and L'_{3v} the newer angle-resolved UPS data[59,60] place these critical points deeper than the older XPS data.[58] In both cases, the theoretical results fall between the two.

The calculated results for Ge are compared to experiment[17,59,64,65] in Table VIII. The calculation includes scalar relativistic effects at all stages. The effects of spin-orbit coupling are included in first order using the vector part of the pseudopotential which yields spin-orbit splittings in the crystal in good agreement with experiment.[38] The theoretically calculated indirect gap is in excellent agreement with experiment. Where direct transitions at symmetry points are associated with features in the electroreflectance spectrum, the present theory is compared to the data of Aspnes.[64] The direct gap to the Γ_{7c} is approximately 0.2 eV too small. This is discussed below. The transitions to the Γ_{6c} and Γ_{8c} states are very well reproduced by the theory. Also the spin-orbit splittings at Γ and at L agree very well with the electroreflectance data (not shown for L). The other conduction-band energies

were derived from direct transition analysis of UPS data where the spin-orbit splitting is not resolved. At both X and L the agreement with theory is good. There is also recent inverse photoemission data for Ge giving results for the conduction-band critical points at L.[65] The results for the $L_{6c}, L_{4,5c}$ complex, in contrast to Si, are somewhat smaller than that deduced from direct transition analysis. The data for L_{6c} agrees well with both the direct-transition data and the present theory. The theoretical results for the occupied states are in good agreement with the XPS data as well as the more recent angular-resolved UPS measurements reported by Wachs et al.[59] This is displayed clearly in Fig. 10 where the data taken for Ge(111) and Ge(001) surfaces from Ref. 59 are shown in comparison to the quasiparticle band structure. The zero of energy is fixed at the top of the valence bands. It does appear that the dispersion along the Δ direction observed in the experiment is consistently somewhat larger than that predicted by the quasiparticle bands. However, the overall agreement is excellent. As seen in Fig. 1, the corrections to the valence-band dispersions in Ge are small but Fig. 10 shows that they improve the agreement with experiment.

The calculated position of the s state in the conduction band of Γ'_{2c} symmetry seems to deviate from experiment consistently through the homopolar materials. Although the magnitude of the deviation is small, this is most obvious for Ge (Γ_{7c}) where every other point agrees well with experiment. As pointed out in Sec. IV B, this state has most of its weight centered on the atomic sites in the diamond lattice and thus maximally probes the approximate treatment given here of the valence-core part of the elec-

TABLE VIII. Comparison of results (in eV) from the present calculation of the quasiparticle energies to experiment for Ge. Calculated results include relativistic effects.

	Present theory	Expt.[a]
Germanium		
E_g	0.75	0.744
$\Gamma_{6v} \rightarrow \Gamma_{8v}$	12.86	12.6, 12.9±0.2[b]
$\Gamma_{7v} \rightarrow \Gamma_{8v}$	0.30	0.297[c]
$\Gamma_{8v} \rightarrow \Gamma_{7c}$	0.71	0.89[c]
$\Gamma_{8v} \rightarrow \Gamma_{6c}$	3.04	3.006[c]
$\Gamma_{8v} \rightarrow \Gamma_{8c}$	3.26	3.206[c]
$X_{5v} \rightarrow \Gamma_{8v}$	9.13	9.3±0.2[b]
$X_{5v} \rightarrow \Gamma_{8v}$	3.22	3.15±0.2, 3.5±0.2[b]
$\Gamma_{8v} \rightarrow X_{5c}$	1.23	1.3±0.2
$L_{6v} \rightarrow \Gamma_{8v}$	10.89	10.6±0.5
$L_{6v} \rightarrow \Gamma_{8v}$	7.82	7.7±0.2
$L_{6v} \rightarrow \Gamma_{8v}$	1.61	1.4±0.3
$L_{4,5v} \rightarrow \Gamma_{8v}$	1.43	
$\Gamma_{8v} \rightarrow L_{6c}$	4.33	4.3±0.2, 4.2±0.1[d]
$\Gamma_{8v} \rightarrow L_{4,5c}$	4.43	
$\Gamma_{8v} \rightarrow L_{6c}$	7.61	7.8±0.6, 7.8±0.1[d]

[a]Reference 17 except where noted.
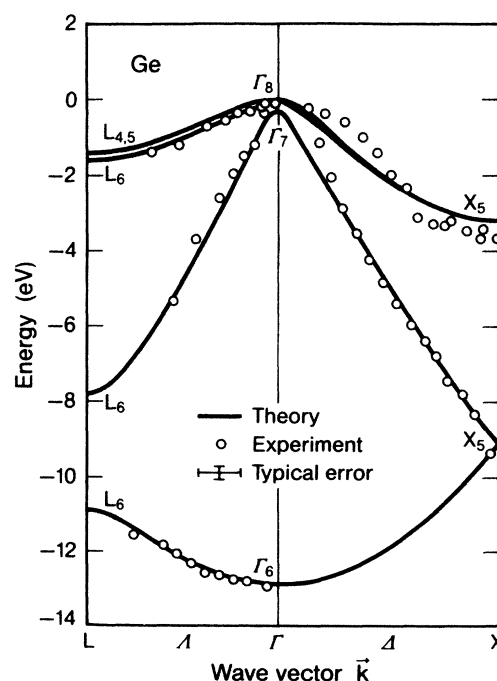[b]Reference 59.
[c]Reference 64.
[d]Reference 65.

FIG. 10. The theoretical valence bands along the symmetry directions Λ and Δ are drawn for Ge in comparison to results from recent angular-resolved photoemission measurements from Ref. 59. The theory includes relativistic effects.

TABLE IX. The results (in eV) of the present theory for LiCl are compared to experiment for the gap E_g, Cl $3p$ bandwidth W_{3p} and the separation between the Cl $3s$ and $3p$ bands $E_{3p} - E_{3s}$.

LiCl	Present theory	Expt.
E_g	9.1	9.4[a]
W_{3p}	3.8	4.0±0.2[b]
$E_{3p} - E_{3s}$	11.6	11.6±0.5[c], 11.0±0.6[d]

[a]Reference 18.
[b]Reference 66.
[c]Reference 67.
[d]Reference 68.

tron self-energy operator as described in Eq. (38). The atomic calculations summarized in Table II suggest the correction produced by proper treatment of the valence-core interaction would raise the energy of the s-like Γ'_{2c} (Γ_{7c}) state compared to the top of the valence band which is p-like. That is, the correction has the required sign. It is difficult to estimate the exact magnitude of the shift, but Table II gives corrections of the right order of magnitude.

We have previously discussed the available data for LiCl.[16] A comparison of theory to experiment[18,66-68] is summarized in Table IX for completeness. The theory gives the direct gap as 9.1 eV as compared to 9.4 eV deduced from reflectivity measurements.[18] The theoretical value is slightly larger than that given before in Ref. 16 because an improved procedure for including an updated spectrum in the calculation of the self-energy is used. The width and structure of the Cl $3p$ bands previously reported are unaffected. For reference, Table X gives a summary of the quasiparticle energies at symmetry points from the present theory in comparison to the eigenvalues

TABLE X. The results of the present theory are compared to the results from the electronic-polaron approach and the LDA eigenvalues for the quasiparticle energies at three symmetry points in the Brillouin zone. All energies are in eV and are referred to the energy of the Γ_{15v} state as zero.

LiCl	LDA	Electronic polaron[a]	Present theory
Γ_{15v}	0.0	0.0	0.0
Γ_{1c}	6.0	9.7	9.1
Γ'_{25c}	11.8	17.9	15.6
X'_{4v}	−3.0	−3.7	−3.3
X'_{5v}	−1.1	−1.0	−1.3
X_{1c}	7.5	10.2	10.7
X_{3c}	8.2	11.3	11.6
L'_{2v}	−2.9	−3.3	−3.2
L'_{3v}	−0.2	−0.5	0.3
L_{1c}	6.4	9.7	9.7
L_{3c}	9.03	14.7	12.5

[a]Reference 69.

in the LDA and the electron polaron approach implemented by Kunz,[69] to be discussed further in the next section.

VII. COMPARISON TO PREVIOUS WORK

Application of the Green's-function approach to the present problem of quasiparticle energies in semiconductors and insulators has a long, albeit sparse, history. Phillips proposed a generalized Koopman's theorem based on Hubbard's expression for the exchange-correlation energy.[20] The dominant term appeared to derive from a dynamically screened-exchange operator. Phillips and Kleinman made estimates of the matrix elements of the operator for Si.[43] Several calculations have been based on the COHSEX approximation to the self-energy operator of Hedin:[13] Brinkman and Goodman[70] and later Kane[71] for Si; Brener[72] for LiF and diamond; Lipari and Fowler[73] for Ar; and Lipari and Kunz[74] for NaCl. These calculations have an important deficiency in common: they all neglect the local fields in the screening. As illustrated by our results in Table III, neglect of local fields *and* dynamical effects can cancel to yield reasonable results fortuitously in some cases, although consistent, quantitatively accurate results require both.

Another line of development has been the "electronic-polaron" model introduced by Toyozawa[75] for application to wide gap insulators. This model envisions the quasiparticles interacting with an "exciton." The electron-hole pairs are replaced by a flat band of localized electron-hole pairs, essentially polarizable atoms, referred to as excitons in the theory. It is the coupling of the electrons and holes to these excitons which leads to the self-energy shift, the correlation correction, in the quasiparticle band structure. The electronic polaron can be seen as one particular manipulation of the more general second-order expressions for the correlation contribution as discussed by Pantelides, Mickish, and Kunz.[76] Their general picture involves interaction of the quasiparticle with virtual electron-hole pairs. The electronic-polaron scheme posits that the important excitation in a wide gap insulator is the exciton. We note that the present approach using the GPP model is quite similar to Overhauser's plasmon model[31] where the important excitations are taken to be plasmonlike with the electron-hole pair excitations collapsed to a single effective mode.

The electronic-polaron approach has been developed extensively by Kunz[77] for application to wide gap insulators. Generally, the model used for the electron-exciton coupling constant requires inclusion of a momentum space cutoff. In the recent application to the alkali-halide crystals,[69] Kunz further simplifies the resulting self-energy for the conduction electrons leaving the correlation contribution as **k** independent. Relaxation corrections are included based on molecular estimates. The results of Kunz[69] for LiCl are compared to the present work in Table X. Although the direct gap in both the present theory and the work of Kunz differs from experiment by about the same amount, the other quasiparticle energies deviate significantly.

There are three other recent calculations of quasiparticle energies in semiconductors. We summarize the results

of the LDA eigenvalues, the work of Horsch, Horsch, and Fulde,[78] the time-dependent screened Hartree-Fock (TDSHF) method,[79,80] the quasiparticle local-density approximation (QPLDA),[81] and the present theory in comparison to experiment for a few features in the spectra of Si and diamond (Table XI). For the fundamental gap and bandwidth, all the many-body theories give results in reasonable accord with experiment. However, it is evident that the results differ substantially in detail, e.g., for the direct transition at the zone edge in diamond. This suggests that more experimental data, particularly for diamond, is essential for sorting out the various theories. We now briefly compare the other theories to the present work.

The QPLDA of Wang and Pickett[81] is a modification of the original proposal of Sham and Kohn[82] for extending the density-functional theory to excited-state properties. The self-energy operator is treated as an energy-dependent functional of the local density thus extending the self-energy operator for the electron gas as a function of the density to the inhomogeneous crystal. In adapting this approach for the case of semiconductors and insulators, Wang and Pickett introduce two gap parameters to describe the self-energy operator of an insulating electron gas. The self-energy operator is local but energy dependent in the QPLDA. The results of this QPLDA approach clearly give systematic improvement over the LDA potential. It has the further advantage of being directly formulated as a correction to the LDA exchange-correlation potential for excitation energies. However, the detailed results show some systematic deficiencies. The gaps with respect to the valence-band edge in Si, e.g., the direct gap at Γ and the indirect gap, are too small. The better agreement for the zone-edge transitions can be traced to their correction being symmetric in energy around the chemical potential (midgap). With reference to Fig. 1, the QPLDA does not seem to provide the sharp jump at the gap region required to correct the LDA eigen-

values with respect to all the experimental data. These suggest that the local-density dependence does not adequately approximate the local-field contributions to the screening and the anisotropy introduced by consideration of the full crystalline Green's function.

The work of Strinati, Mattausch, and Hanke[79] has the same goal as the present theory and employs a similar formal structure. The essential conceptual difference between the present theory and TDSHF theory is in the formulation of the screened interaction W. They argue based on Ward identities that W should include the sum of the ladder bubble diagrams in the polarization propagator. Physically, this includes the continuum exciton effect in the screening of the Coulomb interaction: electron-hole pairs virtually excited interact via a screened interaction. For the calculation of the self-energy operator, it is not clear at present that this leads to systematic improvement over the use of the RPA. One might, in fact, expect this to be a small effect since the calculation of the self-energy operator in the GW approximation averages over the frequency structure in the screened interaction. Strinati, Mattausch, and Hanke have implemented the TDSHF approach using a minimal-basis-set tight-binding approach.[79] Therefore, there are only eight bands available in the self-energy calculation. As noted in Sec. IV C, we find approximately 60 bands are required to converge the magnitude of the Coulomb-hole part of the self-energy operator as well as the gap energies. It is unlikely that other aspects of the local representation employed in the TDSHF calculation will compensate for this. The numerical results of the TDSHF differ substantially in detail from the present theory for diamond. Although the direct and indirect gaps are similar, direct transitions at the zone edges differ by as much as 3 eV from the results of the present theory. Based on the limited results available for Si treated in the TDSHF,[80] the gaps are systematically overestimated for Si.

The variational approach of Horsch, Horsch, and

TABLE XI. Three recent calculations of quasiparticle energies in diamond and Si are compared to the results of the present theory, LDA eigenvalues, and experiment.

	LDA	Horsch *et al.*[a]	TDSHF[b]	QPLDA[c]	Present	Expt.[d]
Diamond						
E_g	3.9		5.7	5.7	5.6	5.48
$\Gamma_{1v} \rightarrow \Gamma'_{25v}$	21.6	23.8	25.2	(23.4)	23.0	24.2±1
$\Gamma'_{25v} \rightarrow \Gamma_{15c}$	5.5	7.4	7.4	7.4	7.5	7.3
$X_{4v} \rightarrow X_{1c}$	10.8	15.6	15.4	13.8	12.9	12.5
Silicon						
E_g	0.52		1.7	0.93	1.29	1.17
$\Gamma_{1v} \rightarrow \Gamma'_{25v}$	11.93		12.9		12.04	12.5±0.6
$\Gamma'_{25v} \rightarrow \Gamma_{15c}$	2.57		3.5	3.07	3.35	3.4
$L'_{3v} \rightarrow L_{1c}$	2.73			3.38	3.53	3.54
$L'_{3v} \rightarrow L_{3c}$	4.58			5.37	5.50	5.51

[a]Reference 78.
[b]References 79 and 80.
[c]Reference 81.
[d]Reference 17.

Fulde[78] is quite different conceptually than the others. A cluster of 25 bonds is treated using a variational approach to include correlation in the many-body wave function. For the quasiparticle problem, this system together with an added electron or hole is studied. Because of the long-range correlations for this case, the cluster must be embedded in a continuum with results for the finite cluster extrapolated to the limit of full crystal. The correlation energy has three contributions in this picture: a long-range polarization of the surrounding bonds, a correction to the ground-state correlation energy due to the presence of the quasiparticle, and a local relaxation of the electrons in the bonds. The last contribution is estimated from molecular calculations using quantum chemistry techniques. Since it is relatively large, of an order of eV, calculations with a more extended basis set are required to include relaxation effects in this approach and render the results more numerically precise. Extension of the method to other materials would facilitate systematic evaluation of it.

VIII. CONCLUSION

The present theory for the quasiparticle energy in semiconductors and insulators is based on evaluation of the electron self-energy operator within the GW approximation. We have shown that this approach, implemented in a numerically reliable fashion, yields quasiparticle energies in excellent agreement with available experimental data for the homopolar materials diamond, Si, and Ge as well as the ionic compound LiCl. In the evaluation of the self-energy operator, the local fields in the screening were shown to be crucial for an adequate description of the screened Coulomb interaction. The effects of dynamical screening are included using a generalized plasmon-pole model and were also found to be essential for quantitative results. The dynamical renormalization has been extensively discussed. The calculated modest degree of renormalization suggests the reasonableness of stopping with

the GW term in the perturbation series for the self-energy operator. The present results have been analyzed in comparison to the LDA potential and the COHSEX approximation to the self-energy operator. In particular, the correction required to the LDA eigenvalues is found to be dominated by a large jump at the gap region. This is consistent with the remarkable result that the quasiparticle wave function has near perfect overlap with the wave function found using the LDA potential. Finally, we emphasize that the present theory has been shown to be applicable to systems ranging from relatively small gap semiconductors (Ge) to wide gap ionic compounds (LiCl) with the same qualitative features extending across the entire range.

ACKNOWLEDGMENTS

APPENDIX A: GENERALIZED f-SUM RULE

In this appendix, we outline the derivation of the generalized sum rule used in Sec. III B, Eq. (29). The derivation generalizes the usual approach for the electron gas as outlined in, for example, Ref. 83.
Starting from the double commutator $[[H,\hat{\rho}_{q+G}],\hat{\rho}_{q+G'}]$ where $\hat{\rho}_{q+G}$ is the density-fluctuation operator, one can straightforwardly obtain

$$\sum_s (E_s - E_0)\frac{1}{2}\{\langle 0|\hat{\rho}_{q+G}|s\rangle\langle s|\hat{\rho}^\dagger_{q+G'}|0\rangle + \langle 0|\hat{\rho}^\dagger_{q+G'}|s\rangle\langle s|\hat{\rho}_{q+G}|0\rangle\} = (q+G)\cdot(q+G')\rho_{G-G'} \tag{A1}$$

which generalizes the usual f-sum rule. The states $|s\rangle$ are the exact excited states of the many-body system. $\rho_{G-G'}$ is the Fourier component of the electron density in the crystal. One can also show that

$$\int_0^\infty d\omega\, \omega\, \mathrm{Im}\epsilon_{GG'}^{-1}(q,\omega) = -\frac{\pi}{2}v(q+G)\sum_s (E_s - E_0)\frac{1}{2}\{\langle 0|\hat{\rho}_{q+G}|s\rangle\langle s|\hat{\rho}^\dagger_{q+G'}|0\rangle + \langle 0|\hat{\rho}^\dagger_{q+G'}|s\rangle\langle s|\hat{\rho}_{q+G}|0\rangle\}.$$

$$\tag{A2}$$

This follows directly from the definition of the linear-response dielectric function in terms of the ground-state matrix element of the commutator of the density operators (e.g., Ref. 13). From Eqs. (A1) and (A2), we obtain the set of sum rules in Eq. (29). Note that the geometric factors are normalized differently in Eq. (29) than in the Johnson result.[35] This is because the latter applies to the symmetric dielectric function describing the response to electric fields while the response function for a potential is

applicable in the present calculation. Otherwise the result can also be obtained from Johnson's result by relating the high-frequency limit of $\epsilon(\omega)$ to that of $\epsilon^{-1}(\omega)$. This yields precisely the sign change on the right-hand side of Eq. (29) compared to the Johnson sum rules.

The Johnson result applies to $\epsilon(\omega)$ in the RPA. As shown here (and discussed more generally by Taut[36]), the sum rules Eq. (29) for $\epsilon^{-1}(\omega)$ are quite general for the interacting many-body system. They are true for the exact

response function. In practice, they are used here with a crystalline charge density derived from a nonlocal pseudopotential calculation. In principle, this requires correction terms on the right-hand side of Eq. (29) because of the nonlocal terms in the pseudopotential. However, in practice, the pseudo charge density gives form factors so close to the experimental x-ray form factors[84] that those corrections should be small.

APPENDIX B: DETAILS OF EVALUATION OF THE SELF-ENERGY OPERATOR

In this appendix, we briefly discuss the evaluation of the matrix elements of the self-energy operator as given in Eq. (34). The expressions required for direct evaluation of the self-energy operator in a plane-wave basis are given.

Consider evaluation of the diagonal matrix elements of the self-energy operator in Eq. (34) for k along some symmetry direction in the Brillouin zone. The self-energy operator has the symmetry of the crystal (acting on r and r' together) so that degenerate states $|n\mathbf{k}\rangle$ will yield the same result in the matrix elements. If one averages over the states in a degenerate complex, then the operations in the little group of $\mathbf{k}$ can be used to reduce the summation

over $\mathbf{q}$ to an appropriate irreducible portion of the Brillouin zone. For $\mathbf{k}=0$, the full reduction to the usual irreducible wedge is achieved. This is a distinct computational advantage.

In Sec. IV A, results obtained by solving Eq. (1) in a plane-wave basis are discussed. The quasiparticle equation then has the form

$$\sum_{\mathbf{G}'}[H^0_{\mathbf{G}\mathbf{G}'}(\mathbf{k})+\Omega\Sigma_{\mathbf{G}\mathbf{G}'}(\mathbf{k};E_{n\mathbf{k}})]\psi_{n\mathbf{k}}(\mathbf{G}')=E_{n\mathbf{k}}\psi_{n\mathbf{k}}(\mathbf{G}) ,$$

(B1)

where the quasiparticle wave function has been expanded in plane waves

$$\psi_{n\mathbf{k}}(\mathbf{r})=\frac{1}{\Omega^{1/2}}\sum_{\mathbf{G}}\psi_{n\mathbf{k}}(\mathbf{G})e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} .$$

(B2)

The single-particle part of Eq. (1), H^0, includes the kinetic energy operator, external potential (ionic pseudopotential in this calculation), and Hartree potential. Formulation of the single-particle part for the nonlocal pseudopotential used here is given in Ref. 41. The formulation for the self-energy operator in the plane-wave basis is straightforward and the result is

$$\Sigma^{\text{SEX}}_{\mathbf{G}\mathbf{G}'}(\mathbf{k};E)=-\frac{1}{\Omega}\sum_{n_1}^{\text{occ}}\sum_{\mathbf{q},\mathbf{G}_1,\mathbf{G}_2}\psi_{n_1,\mathbf{k}-\mathbf{q}}(\mathbf{G}-\mathbf{G}_1)\psi^*_{n_1,\mathbf{k}-\mathbf{q}}(\mathbf{G}'-\mathbf{G}_2)\left[1+\frac{\Omega^2_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})}{(E-\varepsilon_{n_1,\mathbf{k}-\mathbf{q}})^2-\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})}\right]v(\mathbf{q}+\mathbf{G}_2) ,$$

(B3a)

$$\Sigma^{\text{COH}}_{\mathbf{G}\mathbf{G}'}(\mathbf{k};E)=\frac{1}{\Omega}\sum_{n_1}\sum_{\mathbf{q},\mathbf{G}_1,\mathbf{G}_2}\psi_{n_1,\mathbf{k}-\mathbf{q}}(\mathbf{G}-\mathbf{G}_1)\psi^*_{n_1,\mathbf{k}-\mathbf{q}}(\mathbf{G}'-\mathbf{G}_2)\frac{1}{2}\frac{\Omega^2_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})}{\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})[E-\varepsilon_{n_1,\mathbf{k}-\mathbf{q}}-\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})]}v(\mathbf{q}+\mathbf{G}_2) .$$

(B3b)

Since only ReΣ is considered, the secular equation in Eq. (B1) can be diagonalized by standard techniques.

Although evaluation of Eq. (B3) is as straightforward as evaluation of Eq. (34), the latter approach is numerically more efficient for the case of $n=n'$ in Eq. (34). It is easier to exploit symmetry. More importantly, the plane-wave matrix elements required in Eq. (34) need only be evaluated once for each $\mathbf{G}$, stored, and then retrieved for both $\mathbf{G}$ and $\mathbf{G}'$ when needed in the final summation in Eq. (34). Thus, the calculation in Eq. (34) scales roughly as N^3_G while evaluation of $\Sigma_{\mathbf{G}\mathbf{G}'}(\mathbf{k};E)$ scales roughly as N^4_G.

As noted in Sec. IV C, the singularity in the Coulomb interaction for $\mathbf{q}+\mathbf{G}'=0$ is handled following Ref. 43. Briefly, we require a Brillouin-zone summation of the form

$$F=\sum_{\mathbf{q}}\frac{f(\mathbf{q})}{\Omega q^2} .$$

(B4)

$f(\mathbf{q})$ is some function without singularities. Then assuming f to be smooth around $q=0$, the region of the Brillouin zone represented by $q=0$ is taken to be spherical and the singularity in the integrand is integrated out analytically. Thus we get the result for a finite uniform sampling including $q=0$:

$$F=\frac{f(q=0)}{2\pi^2}q_{sz}+\frac{1}{N}\sum_{\mathbf{q}_n\,(\neq 0)}\frac{f(\mathbf{q}_n)}{\Omega_c q^2_n} ,$$

(B5)

where there are N points in the sampling, Ω_c is the volume of the unit cell, and $q_{sz}=(6\pi^2/N\Omega_c)^{1/3}$ is the radius of a subzone equal to $1/N$ of the volume of the Brillouin zone. We note that the dielectric matrix for $\mathbf{q}\rightarrow 0$ also has a nonanalytic part.[85] However, when treated in the manner just described, the lowest-order contribution from that part goes as q^3_{sz}. These contributions can be made negligible by requiring $q_{sz}<1$. They are dropped by symmetrizing $\lim_{\mathbf{q}\rightarrow 0}\bar{\epsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega=0)$ with respect to the direction in which $\mathbf{q}\rightarrow 0$.

There are two aspects to the GPP model which should be noted. First, there is not always a real solution to Eq. (30) for $\tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$. This corresponds to the cases where Im$\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ does not have a single well-defined peak. Since Im$\epsilon^{-1}(\omega)$ typically oscillates for these momentum components and has small amplitude, we neglect these momentum component contributions in the evaluation of Eq. (34). These components affect the quasiparticle energies by approximately 0.1 eV or less in the extreme case of the COHSEX approximation and should contribute much less in the full dynamical calculation because of the sum over bands. The second point is that for energies near the

bottom of the valence bands, the unbroadened resonances in the screened-exchange term can be encountered. These contributions are not included since a full integration over the Brillouin zone would sample the positive and negative portions of the resonance approximately equally. In fact, this problem only occurs for some off-diagonal elements of the dielectric matrix where the resonance energy $\tilde{\omega}$ is small. These terms make a negligible contribution to the quasiparticle energies near the bottom of the valence band.

[1]P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
[2]W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
[3]Theory of the Inhomogeneous Electron Gas, edited by S. Lundqvist and N. H. March (Plenum, New York, 1983) and references therein.
[4]E. W. Plummer, Surf. Sci. 152&153, 162 (1985).
[5]E. Jensen and E. W. Plummer, Phys. Rev. Lett. 55, 1912 (1985).
[6]G. B. Bachelet and N. E. Christensen, Phys. Rev. B 31, 879 (1985).
[7]M. S. Hybertsen and S. G. Louie, Solid State Commun. 51, 451 (1984); Phys. Rev. B 30, 5777 (1984).
[8]A. Zunger and A. J. Freeman, Phys. Rev. B 15, 5049 (1977).
[9]L. J. Sham and M. Schluter, Phys. Rev. Lett. 51, 1888 (1983); Phys. Rev. B 32, 3883 (1985).
[10]J. P. Perdew and M. Levy, Phys. Rev. Lett. 51, 1884 (1983).
[11]M. Lannoo, M. Schlüter, and L. J. Sham, Phys. Rev. B 32, 3890 (1985).
[12]L. Hedin and S. Lundqvist, Solid State Phys. 23, 1 (1969).
[13]L. Hedin, Phys. Rev. 139, A796 (1965).
[14]D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. 43, 1494 (1979).
[15]Preliminary results for diamond and Si have been given in M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. 55, 1418 (1985).
[16]The results for LiCl have been briefly described by M. S. Hybertsen and S. G. Louie, Phys. Rev. B 32, 7005 (1985).
[17]Zahlenwerte und Funktionen aus Naturwissenschaften und Technik, in Vol. III of Landolt-Bornstein (Springer, New York, 1982), pt. 17a.
[18]G. Baldini and B. Bosacchi, Phys. Status Solidi 38, 325 (1970).
[19]The importance of consistently comparing to the RPA correlation energies in the LDA was pointed out to us by M. Schlüter (private communication). The GW approximation for the self-energy operator is equivalent to the RPA for the correlation energy. We use the parametrization of U. von Barth and L. Hedin, J. Phys. C 5, 1629 (1972). Their results are almost identical to the numerical results of Hedin for the correlation energy of the electron gas in the RPA given in Ref. 13.
[20]J. C. Phillips, Phys. Rev. 123, 420 (1961).
[21]G. E. W. Bauer, Phys. Rev. B 27, 5912 (1983).
[22]K. Kunc and E. Tosatii, Phys. Rev. B 29, 7045 (1984).
[23]R. Resta and A. Fleszar, Phys. Rev. B 31, 5305 (1985).
[24]One regards the LDA potential V_{xc} as an initial approximation to the self-energy operator Σ and then calculation of the vertex function is straightforward within the LDA using the general expressions for Γ from Ref. 13. The result is that the form of the new self-energy operator is still GW but with the screened interaction W including exchange-correlation effects explicitly for the case of an electron as test particle probing the screening potential.
[25]S. L. Adler, Phys. Rev. 126, 413 (1962).
[26]N. Wiser, Phys. Rev. 129, 62 (1963).
[27]The algorithm used is from G. P. Kerker, J. Phys. C 13, L189 (1980).
[28]D. M. Ceperley and B. I. Alder, Phys. Rev. Lett. 45, 566 (1980) as parametrized in J. P. Perdew and A. Zunger, Phys. Rev. B 23, 5048 (1981).
[29]L. Kleinmann, Phys. Rev. B 21, 2630 (1980); G. B. Bachelet and M. Schluter, ibid. 25, 2103 (1982). We use the algorithm from Ref. 27.
[30]M. S. Hybertsen and S. G. Louie (unpublished).
[31]A. W. Overhauser, Phys. Rev. B 3, 1888 (1971).
[32]M. Bennett and J. C. Inkson, J. Phys. C 10, 987 (1977); J. C. Inkson and M. Bennett, ibid. 11, 1017 (1978).
[33]B. I. Lundvist, Phys. Kondens. Mater. 6, 206 (1967).
[34]S. G. Louie and M. L. Cohen (unpublished).
[35]D. L. Johnson, Phys. Rev. B 9, 4475 (1974).
[36]M. Taut, J. Phys. C 18, 2677 (1985).
[37]J. P. Walter and M. L. Cohen, Phys. Rev. B 5, 3101 (1972).
[38]M. S. Hybertsen and S. G. Louie, Phys. Rev. B 34, 2920 (1986).
[39]G. Arbman and U. von Barth, J. Phys. F 5, 1155 (1975).
[40]C. Froese Fischer, The Hartree-Fock Method for Atoms (Wiley, New York, 1977).
[41]J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C 12, 4409 (1979).
[42]D. J. Chadi and M. L. Cohen, Phys. Rev. B 8, 5747 (1973). For diamond, a 28-k-point set is used. For Si, Ge, and LiCl, a 10-k-point set is used.
[43]J. C. Phillips and L. Kleinman, Phys. Rev. 128, 2098 (1962).
[44]The actual sets used were 16 q points for Si and 8 q points for diamond, Ge, and LiCl.
[45]In producing the plots for Fig. 5, the 8-q-point set described in Sec. IV C was used in the Brillouin-zone summation and approximately 140 G vectors in the sum over G depending on q. This ensures convergence of the local features of the screening potential but less accurately reproduces the features on the scale of a few bond lengths.
[46]M. S. Hybertsen and S. G. Louie, in Proceedings of the Seventeenth International Conference on the Physics of Semiconductors, San Francisco, 1984, edited by J. D. Chadi and W. A. Harrison (Springer, New York, 1985), p. 1001.
[47]The 8-k-point set has been used in the sum over the Brillouin zone.
[48]C. Kittel, Quantum Theory of Solids (Wiley, New York, 1963), pp. 137–141.
[49]C.-O. Almbladh and U. von Barth, Phys. Rev. B 31, 3231 (1985).
[50]R. A. Heaton, J. G. Harrison, and C. C. Lin, Solid State Commun. 41, 827 (1982); Phys. Rev. B 28, 5992 (1983).
[51]W. Hanke and L. J. Sham, Phys. Rev. Lett. 33, 582 (1974); Phys. Rev. B 12, 4501 (1975).

[52]W. Hanke and L. J. Sham, Phys. Rev. Lett. **43**, 387 (1979); Phys. Rev. B **21**, 4656 (1980).

[53]M. del Castillo—Mussot and L. J. Sham, Phys. Rev. B **31**, 2092 (1985).

[54]F. R. McFeely *et al.*, Phys. Rev. B **9**, 5268 (1974).

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

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

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

[58]W. E. Spicer and R. C. Eden, *Proceedings of the Ninth International Conference on the Physics of Semiconductors, Moscow, 1968* (Nauka, Leningrad, USSR, 1968), Vol. 1, p. 61.

[59]A. L. Wachs *et al.*, Phys. Rev. B **32**, 2326 (1985).

[60]F. J. Himpsel, P. Heimann, and D. E. Eastman, Phys. Rev. B **24**, 2003 (1981).

[61]R. Hulthen and N. G. Nilsson, Solid State Commun. **18**, 1341 (1976).

[62]D. Straub, L. Ley, and F. J. Himpsel, Phys. Rev. Lett. **54**, 142 (1985).

[63]R. R. L. Zucca and Y. R. Shen, Phys. Rev. B **1**, 2668 (1970).

[64]D. E. Aspnes, Phys. Rev. B **12**, 2797 (1975).

[65]D. Straub, L. Ley, and F. J. Himpsel, Phys. Rev. B **33**, 2607 (1986).

[66]R. T. Poole, J. G. Jenkin, R. C. G. Leckey, and J. Liesegang, Chem. Phys. Lett. **22**, 101 (1973).

[67]T. Ohta, S. Kinoshita, and H. Kuroda, J. Electron Spectrosc. Relat. Phenom. **12**, 169 (1977).

[68]L. I. Johnsson and S. B. M. Hagstrom, Phys. Scr. **14**, 55 (1976).

[69]A. B. Kunz, Phys. Rev. B **26**, 2056 (1982).

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

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

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

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

[74]N. O. Lipari and A. B. Kunz, Phys. Rev. B **3**, 491 (1971).

[75]Y. Toyozawa, Prog. Theor. Phys. (Kyoto) **12**, 421 (1954).

[76]S. T. Pantelides, D. J. Mickish, and A. B. Kunz, Phys. Rev. B **10**, 2602 (1974).

[77]A. B. Kunz, Phys. Rev. B **6**, 606 (1972).

[78]S. Horsch, P. Horsch, and P. Fulde, Phys. Rev. B **29**, 1870 (1984).

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

[80]W. Hanke, Th. Golzer, and H. J. Mattausch, Solid State Commun. **51**, 23 (1984).

[81]C. S. Wang and W. E. Pickett, Phys. Rev. Lett. **51**, 597 (1983); Phys. Rev. B **30**, 4719 (1984).

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

[83]D. Pines, *Elementary Excitations in Solids* (Benjamin, New York, 1963).

[84]M. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 5668 (1982).

[85]R. M. Pick, M. H. Cohen, and R. M. Martin, Phys. Rev. B **1**, 910 (1970).