Self-energy operators and exchange-correlation potentials in semiconductors

R. W. Godby
AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974
and Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, United Kingdom

M. Schlüter
AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974

L. J. Sham
Department of Physics, University of California at San Diego, La Jolla, California 92093
(Received 12 November 1987)

We show how the density-functional theory (DFT) exchange-correlation potential $V_{xc}(\mathbf{r})$ of a semiconductor is calculated from the self-energy operator $\Sigma(\mathbf{r},\mathbf{r}',\omega)$, and how Σ is obtained using the one-particle Green's function and the screened Coulomb interaction (the *GW* approximation). We discuss the nature of V_{xc} and the self-energy in real space, and investigate features and trends found in Si, GaAs, AlAs, and diamond. In each case the calculated quasiparticle band structure is in good agreement with experiment, while the DFT band structure is surprisingly similar to that with the common local-density approximation (LDA); in particular, about 80% of the severe LDA band-gap error is shown to be inherent in DFT calculations, being accounted for by the discontinuity Δ in V_{xc} upon addition of an electron. The relationship of the calculated V_{xc} to the LDA and its extensions is also examined.

I. INTRODUCTION AND FOUNDATIONS

In previous papers[1-3] we have given brief accounts of some results of calculations of exchange-correlation potentials and quasiparticle energies for silicon, gallium arsenide, aluminum arsenide, and diamond from the self-energies $\Sigma(\mathbf{r},\mathbf{r}',\omega)$, which are calculated in the so-called *GW* approximation using the one-particle Green's function and the screened Coulomb interaction.[4] The purpose of this paper is to give full details of these calculations and their results, and to present a more comprehensive discussion.

A. Density-functional theory and structural properties of solids

Much progress has been made in recent years in calculating structural properties of solids from first principles.[5] Equilibrium crystal structures and lattice parameters, dissolution and defect energies, phonon spectra, cluster energies, chemisorption bond strengths, and surface reconstructions are good example. Structural properties of the electronic ground state, since the Born-Oppenheimer approximation permits one to calculate the potential energy of the solid as a function of the configuration of its nuclei by finding the ground-state total electronic energy with the nuclei frozen in each configuration. This progress has been made possible by density-functional theory, (DFT)[6,7] in which the massive problem of calculating the ground state of the true system of 10^{23} interacting electrons is rigorously transformed into that of finding the ground state of a much simpler system of noninteracting electrons moving in a local effective potential $V_{\rm eff}(\mathbf{r})$. The exact ground-state electron density of the real system is then given by the electron density of the fictitious noninteracting system, and the exact ground-state total energy by a functional of that electron density.

The total ground-state electronic energy is written as

$$E_{\rm tot}[n]=T_s[n]+E_{\rm es}[n]+E_{\rm xc}[n] \tag{1}$$

where T_s is the kinetic energy of the fictitious noninteracting electrons, $E_{\rm es}$ the classical Coulomb electrostatic potential energy of the electron gas and the nuclei, and $E_{\rm xc}$ the remainder—the so-called exchange-correlation energy. The basic theorem of DFT states that apart from the classical external electrostatic potential energy $\int n(\mathbf{r})V_{\rm ext}(\mathbf{r})d\mathbf{r}$, $E_{\rm tot}[n]$ is a *universal* functional of the electron density $n(\mathbf{r})$. The corresponding effective one-electron Schrödinger equation, constructed by Kohn and Sham[7] to avoid needing to know the functional $T_s[n]$, is[8]

$$[-\tfrac{1}{2}\nabla^2+V_{\rm ext}(\mathbf{r})+V_H(\mathbf{r})+V_{\rm xc}(\mathbf{r})-E_{i\rm DFT}]\psi_{i,\rm DFT}(\mathbf{r})=0\ , \tag{2}$$

where V_H is the Hartree potential $\int n(\mathbf{r}')d\mathbf{r}'/|\mathbf{r}-\mathbf{r}'|$ and $V_{\rm xc}$ the exchange correlation potential $\delta E_{\rm xc}[n]/\delta n(\mathbf{r})$. The electron density is given by

$$n(\mathbf{r})=\sum_{\rm occ}|\psi_{i,\rm DFT}(\mathbf{r})|^2\ . \tag{3}$$

The only complications are that the effective potential in the fictitious system (i) is itself a functional of the electron density, so that the equations must be solved self-consistently, and (ii) must, in practice, be approximated, as the functional $E_{\rm xc}[n]$ is not known exactly for most densities. The first of these complications is trivial to overcome to any desired degree of accuracy; the second is not, but, in practice, very satisfactory results have been

obtained by using the well-known local-density approximation (LDA):

$$E_{xc} \approx E_{xc}^{LDA} \equiv \int n(\mathbf{r}) \varepsilon_{xc}^{hom}(n(\mathbf{r})) d\mathbf{r} \quad (4)$$

and, correspondingly

$$\begin{aligned} V_{xc}(\mathbf{r}) &\approx V_{xc}^{LDA}(n(\mathbf{r})) \\ &\equiv \mu_{xc}^{hom}(n(\mathbf{r})) \\ &= \varepsilon_{xc}^{hom}(n(\mathbf{r})) + n(\mathbf{r}) \frac{d\varepsilon_x^{hom}(n(\mathbf{r}))}{dn} , \end{aligned} \quad (5)$$

where $\varepsilon_{xc}^{hom}(n)$ is the exchange-correlation energy per electron of a *homogeneous* electron gas of density n, which *is* known accurately from numerical many-body theory.[9] Indeed, this works so well that two of the most important unsolved problems of the fundamentals of density-functional theory are (i) why does the LDA work so well, and (ii) when and why does it break down? As we shall see, this work will allow us to find the true V_{xc} of an extended system quite independently of any LDA, so that the physical basis for their similarities and difference may be studied.

B. Electronic excitations

The electronic and optical properties of a solid are quite distinct from its structural properties. In the latter, as described in the preceding section, one is keeping the electrons in their ground state and asking what the change is in the total energy of the system when one moves the nuclei adiabatically. In one-electron properties, one is keeping the nuclei stationary (except for the possibility of allowing some coupling to the phonons) and asking with what possible energies one can add an electron to the system (or take one from it). In the language of many-body theory, these energies are the energies of the *quasiparticles* of the system; that is, the energies that appear in a Schrödinger-like equation containing the nonlocal, energy-dependent *self-energy* in place of the exchange-correlation potential. In the case of an infinite, periodic solid, the electron's momentum is well defined, so that one can inquire about the quasiparticle energies as a function of the momentum, $\mathbf{k}$ These functions $E(\mathbf{k})$ are generally known as the *band structure* of the solid.

Optical measurements involve excitations of the system without adding or subtracting electrons. The simplest linear-optical process involves two quasiparticles, one electron and one hole, which attract each other, forming an exciton. The inferred continuum edge of the optical spectrum is taken to be the difference between quasiparticle energies in the conduction and valence bands. With this in mind we may talk about quasiparticle and excitation energies interchangeably.

Calculating the quasiparticle energies is by no means so refined a procedure as calculating the total energy. Formally, the energies are the solutions E of the equation[4]

$$\left[-\tfrac{1}{2}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) - E\right]\psi(\mathbf{r}) + \int \Sigma(\mathbf{r},\mathbf{r}',E)\psi(\mathbf{r}')d\mathbf{r}' = 0 , \quad (6)$$

where V_H is defined above, and $\Sigma(\mathbf{r},\mathbf{r}',E)$ is the *self-energy,* a nonlocal, energy-dependent operator. The complexity of Σ has been the main reason for the difficulty of realistic first-principles calculations of quasiparticle energies.

There has, however, been some exploration of the validity of various approximations to the self-energy for semiconductors, insulators, and metals. Earlier work includes that of Strinati, Mattausch, and Hanke,[10] who used the GW approximation[4] starting from a Hartree-Fock calculation; Pickett and Wang,[11,12] who implemented the local-density approximation of Sham and Kohn[13] for the self-energy operator; Horsch, Horsch, and Fulde,[14] who have taken a local-orbital-correlation approach; and Stern and Inkson,[15] who performed an analytic calculation of the GW self-energy within an extreme-tight-binding model. The best results have been obtained using the GW approximation starting with LDA wave functions and energies: Hybertsen and Louie[16–19] obtained results for silicon, germanium, lithium chloride, diamond, and a germanium surface using a plasmon-pole approximation for the frequency dependence of W, and the present authors[1–3] calculated the quasiparticle energies of silicon, diamond, gallium arsenide, and aluminum arsenide using a formally similar approach, but without making any assumptions about the frequency dependence of W. von der Linden and Horsch[20] have performed GW calculations using empirical-pseudopotential wave functions and energies to calculate the corrections to the LDA Kohn-Sham eigenvalues. Recently Gygi and Baldereschi[21] have completed LDA GW calculations with results similar to those reported here and in Refs. 16–19.

C. The connections between quasiparticle energies and DF theory

There is a formal resemblance of Eq. (6), which gives the quasiparticle energies, to the Schrödinger equation (2) for the effective-one-electron eigenvalues of density-functional theory: each is a Schrödinger equation for fictitious, noninteracting electrons moving in an effective potential (although in one case this "potential"—the self energy—is nonlocal and energy dependent). This has suggested that there might be a connection between the quasiparticle energies E_i and the DFT eigenvalues $E_{i,\mathrm{DFT}}$.

This speculation can be put on a firmer footing by noting that the quasiparticle energies within an Auger threshold of the gap can, from their definition, be written in terms of changes in the total ground-state energy of the system when electrons are added or subtracted, and DFT gives total ground-state energies exactly. Exploiting this relationship has led to the result[13] that, in a large system at zero temperature, the chemical potential μ (the Fermi energy, or, for metallic systems, the minimum energy to add an electron to the system; equal also to both the electron affinity and the ionization potential), is given correctly by the highest occupied DFT eigenvalue:

$$\mu^{(N)} \equiv \frac{\partial E_{tot}^{(N)}}{\partial N} = E_{\mathrm{DFT},N}^{(N)} , \quad (7)$$

where (N) denotes the system containing N particles. No

further exact relationship is known for energies away from E_F.

In the case of semiconductors, however, there is an *energy gap,* and the electron affinity and ionization potential are *not* equal, as the energy change in adding an electron to the system is not the same as in removing one. Nevertheless, the relationship (7) may be applied to the continuous changes $N-1\rightarrow N$ and $N\rightarrow N+1$, giving

$$E_g \equiv \mu^{(N+1)} - \mu^{(N)} = E^{(N+1)}_{N+1,\mathrm{DFT}} - E^{(N)}_{N,\mathrm{DFT}} , \tag{8}$$

as shown by Sham and Schlüter[22] and Perdew and Levy.[23] Both also showed that the difference between the highest occupied and lowest unoccupied DFT eigenvalues in the N-electron system (which is what is usually called the DFT band gap) is *not* the true quasiparticle band gap E_g, but differs from it by Δ, the discontinuity in the exchange-correlation potential when an electron is added to the system:

$$\Delta \equiv E_g - (E^{(N)}_{N+1,\mathrm{DFT}} - E^{(N)}_{N,\mathrm{DFT}}) = V^{(N+1)}_{\mathrm{xc}}(\mathbf{r}) - V^{(N)}_{\mathrm{xc}}(\mathbf{r}) \tag{9}$$

(independent of $\mathbf{r}$), where $V^{(N)}_{\mathrm{xc}}$ is the exact DFT exchange-correlation potential for the N-electron system. This is because the exchange-correlation potential is the only part of the DFT one-electron potential that can be a nonanalytic functional of the electron density—which changes only infinitesimally in a large system when an extra electron is added—and therefore the only means by which the DFT eigenvalues can be altered (by an amount of order 1) when an electron is added (Fig. 1). Moreover, V_{xc} can only change by an r-independent constant, Δ, because only this ensures that there is no effect (of relative order 1) on the electron density.

Δ has been shown[24] to be a significant fraction of the energy gap for a one-dimensional model semiconductor within a two-plane-wave basis set, but its size for a real semiconductor was not known. If Δ is small, exact DFT will give accurate band structures, and attention should be paid to going beyond the LDA in density-functional theory; if Δ is large, no attempt to add nonlocal-density corrections to the LDA will improve the calculated DFT band structures, and one must either go outside DFT, or take the discontinuity in V_{xc} into account explicitly as a correction to the DFT band structure when quasiparticle energies are needed. Thus Δ lies at the heart of the relationship between the quasiparticles and DFT for semiconductors and insulators. In a periodic solid, where the momentum $\mathbf{k}$ is a good quantum number for both the DFT and quasiparticle wave functions, the rigorous extension $\Delta(\mathbf{k})$ is possible to all the minimum direct band gaps throughout the Brillouin zone with energies below the Auger threshold.

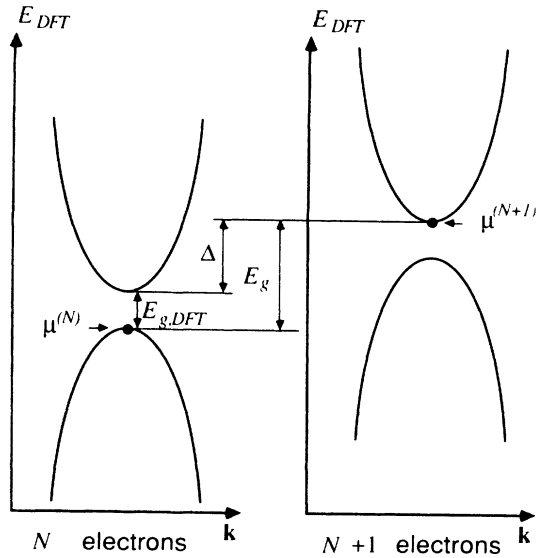

FIG. 1. Illustration of the significance of Δ, the discontinuity in V_{xc}. The exact DFT Kohn-Sham one-electron energies are shown in the form of a band structure for the N- and $(N+1)$-particle systems. The two differ in a uniform increase of the eigenvalues by Δ, as explained in the text. The quasiparticle gap E_g is the difference between the two eigenvalues indicated: $E_g = E^{(N+1)}_{N+1,\mathrm{DFT}} - E^{(N)}_{N,\mathrm{DFT}}$. It is evident that $E_g = E_g^{\mathrm{DFT}} + \Delta$.

D. Experimental facts

Owing to the general absence of true calculations of quasiparticle energies, experimental measurements of these energies (for example, in photoemission, optical absorption, and electronic transport) are usually compared with DFT band structures. It is found that there is often fairly good qualitative agreement. The worst quantitative discrepancies shown by careful calculations using the local-density approximation for exchange and correlation, and eliminating any other source of error, are that the *energy gaps in semiconductors and insulators are always much too small.* For example, the minimum band gap of silicon is 50% too small,[25] and that of germanium 100% too small[26]—the LDA causes the valence and conduction bands to overlap. After it was shown (see above) that the energy gap is not given correctly by even the exact DFT band structure for the N-particle ground state, it was realized that the belief that this error occurred wholly because of the use of the LDA in the DFT calculation was incorrect. However, the relative contributions of these two sources of error was unknown. In addition, the widths of individual bands, especially away from E_F, can be in error by up to about 50%.

E. The role of this work

In this paper we attempt to answer the following questions: (i) Do the discrepancies between the DFT band structures and the experimental quasiparticle band structures (especially the band gaps) arise from the use of the local-density approximation, or are they inherent even in exact DFT, or a combination of the two? What is the form of $\Delta(\mathbf{k})$? (ii) What is the nature and physical origin of the self-energy operator in real space, and which of its features are important for the quasiparticle properties? (iii) How does the true DFT exchange-correlation potential differ from the LDA approximation of it? Can the physical origins of the difference be used to suggest better approximations than the LDA?

The paper is organized as follows. In Sec. II we describe the calculation of the dielectric matrix ϵ, the

screened Coulomb interaction W, the self-energy Σ, and the exchange-correlation potential V_{xc}. Then in Secs. III and IV we give and discuss our detailed results. In Sec. III we concentrate on the self-energy operators and the quasiparticle properties calculated from them. In Sec. IV we examine V_{xc} in real space and the DFT one-electron energies, comparing the latter with the local-density approximation for V_{xc} and the quasiparticle energies, and determining the values of Δ. We also compare V_{xc} with previous extensions to the LDA.

II. THE CALCULATION

A. The dielectric function and screened Coulomb interaction

We begin with a well-converged LDA calculation for the material at its experimental lattice constant, using a plane-wave basis set and nonlocal pseudopotentials.[27] The whole calculation is performed within this plane-wave basis. (The different stages in the calculation, starting from the LDA band structure, are summarized as a flow diagram in Fig. 2.) The plane-wave energy cutoff was chosen to be 12.5 Ry for Si, GaAs, and AlAs, and 50 Ry for diamond. Eigenvalues are generally converged to 0.1 eV, with the exception of the Γ_2' conduction band, which converges very slowly as the number of plane waves is increased. Here the convergence error is less than 0.3 eV. Next, the LDA wave functions and energies are used to form the random-phase approximation (RPA) dielectric function, $\epsilon(\mathbf{r},\mathbf{r}',\omega)$. In each case about 169 reciprocal-lattice vectors and 65 bands were retained. Because of the periodicity of the crystal, ϵ (and all nonlocal quantities considered here) can be written in the form

$$\epsilon(\mathbf{r},\mathbf{r}',\omega)=\frac{1}{8\pi^3\Omega}\int d\mathbf{k}\sum_{\mathbf{G},\mathbf{G}'}\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{k},\omega)\times e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}e^{-i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}'}, \quad (10)$$

where the unit-cell volume is Ω, $\mathbf{G}$ and $\mathbf{G}'$ are reciprocal-lattice vectors, and $\mathbf{k}$ is restricted to the first Brillouin zone. We calculate the LDA band structure and the self-energy at two and six special $\mathbf{k}$ points[28] in the wedge, and the dielectric function at the six and ten $\mathbf{k}$ points, respectively, which are the difference vectors (not themselves special points: for example $\mathbf{k}=\mathbf{0}$ is included). We first calculate ϵ at the six $\mathbf{k}$ points, and then interpolate to obtain ϵ at the ten $\mathbf{k}$ points). Additionally, $\epsilon(\mathbf{k}=\mathbf{0})$ requires the LDA band structure at ten rather than two $\mathbf{k}$ points to be well converged. The final W used in the calculation of Σ is the composite of these calculations (see below).

The inclusion of the off-diagonal terms in the dielectric matrix $\epsilon_{\mathbf{G}\mathbf{G}'}$ corresponds to taking account of the so-called *local-field effects,* since a purely diagonal dielectric matrix would imply that ϵ depended only on $|\mathbf{r}-\mathbf{r}'|$, thereby ignoring the fact that the unit cell is not homogeneous but can respond to a perturbation by setting up locally varying fields. The local-field terms are crucial (though, as we shall see later, mainly in their effect on the strength of the screening hole rather than its shape) and contribute directly to the differing strengths of the self-energy operator at different points in the unit cell and therefore to the band-gap correction

The RPA expression for ϵ is[29] $\epsilon=1-V\chi$, where V is the Coulomb interaction $1/|\mathbf{r}-\mathbf{r}'|$ and χ the response function,

$$\chi(\mathbf{r},\mathbf{r}',\omega)=2\sum_{ij}(n_i-n_j)\frac{\psi_i(\mathbf{r})\psi_j^*(\mathbf{r})\psi_i^*(\mathbf{r}')\psi_j(\mathbf{r}')}{E_i-E_j-\omega-i\delta}, \quad (11)$$

where n_i, ψ_i, and E_i are the occupations (0,1), wave functions, and energies of the one-electron states (in this case the LDA Kohn-Sham states), and δ is an infinitesimal. Written in reciprocal space as described above, this becomes

$$\chi_{\mathbf{G}\mathbf{G}'}(\mathbf{k},\omega)=\frac{2}{M}\sum_{i,j}(n_i-n_j)\frac{\langle j|e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}|i\rangle\langle i|e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}'}|j\rangle}{E_i-E_j-\omega-i\delta} \quad (12)$$

(where $M\Omega$ is the crystal volume), which may be evaluated conveniently in the plane-wave basis set. Only states $|i\rangle$ and $|j\rangle$ differing in $\mathbf{k}$ vector by $\mathbf{k}+\mathbf{G}_0$ (where $\mathbf{G}_0$ is any reciprocal-lattice vector) contribute, and each of the two terms in the numerator is calculated most easily using a fast Fourier transform (FFT) by evaluating $\psi_i(\mathbf{r})\psi_j^*(\mathbf{r})$ on the real-space grid.

For $\mathbf{k}=\mathbf{0}$ (which is one point in the mesh used) the first (second) term in the numerator vanishes when $\mathbf{G}=\mathbf{0}$ ($\mathbf{G}'=\mathbf{0}$), but it is important to calculate the limiting behavior as $\mathbf{k}\to\mathbf{0}$ as χ will later be combined with the Coulomb interaction $V(\mathbf{k})=4\pi/k^2$. We follow Ref. 30 and use the identity

$$\lim_{\mathbf{k}\to\mathbf{0}}\langle\psi_{\mathbf{k}_1},n|e^{-i\mathbf{k}\cdot\mathbf{r}}|\psi_{\mathbf{k}_1+\mathbf{k}},n'\rangle = i\mathbf{k}\cdot\langle\psi_{\mathbf{k}_1},n|\nabla|\psi_{\mathbf{k}_1},n'\rangle/(E_{\mathbf{k}_1,n}-E_{\mathbf{k}_1,n'}) \quad (13)$$

(where n and n' are the band indices), which may be evaluated easily when ψ is expressed in a plane-wave basis set. We omit the small correction to this expression arising from the nonlocality of the pseudopotential, which we expect to be insignificant when ϵ is combined into Σ.

Despite the simplifications of periodicity, the evaluation of ϵ is still very demanding computationally, and only in the last few years have calculations of $\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ been published,[31] and then only for a very limited range of $\mathbf{G}$, $\mathbf{G}'$, $\mathbf{q}$, or ω. We achieve a considerable reduction in computation while maintaining full accuracy by calculating ϵ for imaginary, rather than real, ω, using the analytic continuation of the RPA expression to imaginary frequencies, so that there is no need to take elaborate precautions for handling the poles of ϵ that lie just off the real ω axis. This is possible because the calculation of Σ involves an integration over frequencies, and we choose

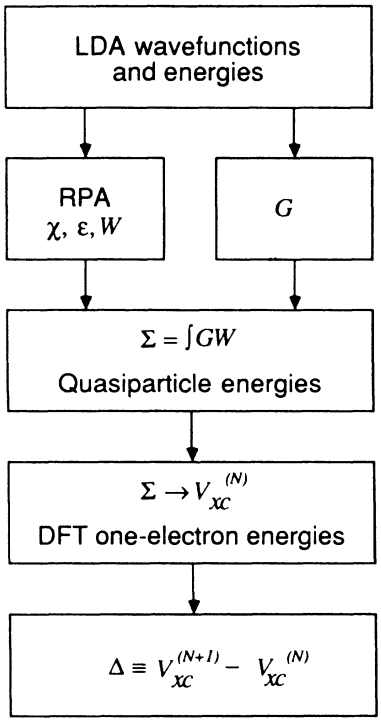

FIG. 2. Flow diagram of the calculation, starting from the self-consistent LDA pseudopotential calculation, forming the screened Coulomb interaction W and the Green's function G, combining them to form Σ, and then using Σ to obtain the quasiparticle energies E_i and also the true exchange-correlation potential V_{xc}.

to do this along the imaginary axis (Fig. 3) where the zero of real energy is the middle of the LDA minimum band gap. ϵ for imaginary ω has only weak structure: the details of the sampling of the Brillouin zone are much less crucial, and the number of frequencies at which it must be evaluated are fewer. Nevertheless, the calculation of the dielectric function represents a major computational effort in the whole work and simplifications without sacrificing accuracy are needed.[21,32] We note that the use of imaginary-axis integration has avoided the need to make a model ansatz for the frequency dependence of W, as was done, for example, in the plasmonpole models used in Refs. 16–20. This is especially important for our intention to use Σ to calculate V_{xc}, as this procedure involves an integration of Σ from $-\infty$ to the band gap, and the plasmon-pole models are expected[18,19] to break down at energies of the order of $-\omega_p$, where ω_p is the plasmon energy.

We then form *inverse dielectric function* ϵ^{-1} and the *screened Coulomb interaction* $W(\mathbf{r},\mathbf{r}',\omega)=\epsilon^{-1}V$, again in reciprocal space, by straightforward matrix inversion and multiplication for each $\mathbf{k}$. ϵ, ϵ^{-1}, and W are calculated from ten values of ω along the positive imaginary axis, chosen according to the requirements of the integration scheme to be used later for Σ (see the Appendix).

B. The self-energy operator

W is then used in the GW approximation[4,33] for the self-energy:

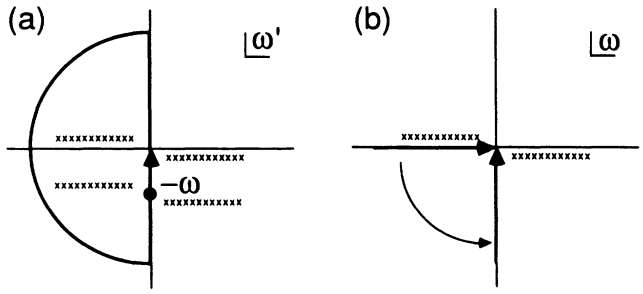

FIG. 3. (a) The contour of integration in the complex ω' plane for the frequency integral performed in calculating $X(\omega)$ (the "correlation" part of Σ: see the Appendix). The energy zero is chosen to be in the middle of the LDA band gap. The poles of the integrand are shown as crosses. The choice of contour—enclosing the poles to the left of the imaginary axis of both W and G—indicates the analytic continuation of the GW expression for Σ to imaginary frequencies. (b) The contour in the ω plane for the integral for Y (involved in calculating V_{xc}: see the Appendix). In this case the integrand has the poles as shown. The contour along the negative real axis up to zero is deformed to the negative imaginary axis as indicated. It is evident that in each case the chosen contour avoids the structure in the integrand as far as possible.

$$\Sigma(\mathbf{r},\mathbf{r}',\omega)=\frac{i}{4\pi}\int_{-\infty}^{\infty}e^{i\omega\delta}W(\mathbf{r},\mathbf{r}',\omega')G(\mathbf{r},\mathbf{r}',\omega+\omega')d\omega' \tag{14}$$

(where δ is a positive infinitesimal), which is the leading term in an expansion of Σ in powers of the relatively weak screened Coulomb interaction (as opposed to the strong bare Coulomb interaction) and corresponds to setting the so-called vertex function Γ to a δ function (see Ref. 4). The remaining terms in the series, which are termed *vertex corrections*, correspond to successive corrections to the vertex function.

Additionally, we approximate G, the many-body-theory one-particle Green's function, by its LDA counterpart,

$$G_{\mathrm{DFT}}^{\mathrm{LDA}}=2\sum_{n,\mathbf{k}}\frac{\psi_{n\mathbf{k},\mathrm{DFT}}^{\mathrm{LDA}}(\mathbf{r})\psi_{n\mathbf{k},\mathrm{DFT}}^{\mathrm{LDA}*}(\mathbf{r}')}{\omega-E_{n\mathbf{k},\mathrm{DFT}}^{\mathrm{LDA}}\pm i\delta}\ , \tag{15}$$

where the negative sign applies for occupied states, and n labels the bands. This procedure can be regarded as the first term in an iterative expansion

$$\Sigma=\int WG_{\mathrm{DFT}}^{\mathrm{LDA}}+\int WG_{\mathrm{DFT}}^{\mathrm{LDA}}(\Sigma-V_{\mathrm{xc}}^{\mathrm{LDA}})G_{\mathrm{DFT}}^{\mathrm{LDA}}+\cdots \tag{16}$$

(omitting the prefactor for clarity); as we shall see later, $G_{\mathrm{DFT}}^{\mathrm{LDA}}$ is so close to G that the truncation

$$\Sigma(\omega)=\frac{i}{4\pi}\int_{-\infty}^{\infty}e^{i\omega\delta}W(\omega')G_{\mathrm{DFT}}^{\mathrm{LDA}}(\omega+\omega')d\omega' \tag{17}$$

is justified.

The bare-exchange term

$$\Sigma_x(\mathbf{r},\mathbf{r}')=\frac{i}{4\pi}V(\mathbf{r}-\mathbf{r}')\int_{-\infty}^{\infty}e^{i\omega\delta}G_{\mathrm{DFT}}^{\mathrm{LDA}}(\mathbf{r},\mathbf{r}',\omega')d\omega' \tag{18}$$

is calculated separately because of the different frequency dependence of the integrand. In order to gain insight into the physics, we also calculate the "statically screened exchange" self-energy

$$\Sigma_{\rm ssx}=\frac{i}{4\pi}W(\omega=0)\int_{-\infty}^{\infty}e^{i\omega\delta}G_{\rm DFT}^{\rm LDA}(\omega')d\omega' \ . \tag{19}$$

This is the dominant term in another approximation for Σ, the Coulomb hole plus screened exchange (COHSEX) approximation,[33] the remainder of which is the Coulomb hole term which is obtained by setting ω equal to the one-electron energy in each term that comes from the contribution of the poles of W to the integral $\int WG$. This approximation has been found[16] to give a somewhat smaller difference (and opposite in sign) from the experimental band gap than the LDA.

In principle, both G and W have contributions from the core, dominated[34] by the core-valence exchange $V\int G_{\rm core}$. All core terms are omitted from these pseudopotential-based calculations, except insofar as they are included in the pseudopotential, which should have negligible effects, except in GaAs, where the gallium $3d$ core has significant overlap with the valence electrons. This is discussed further in Sec. III A.

The bare-exchange frequency integral $V\int G$ is evaluated analytically, and the remainder $\int(W-V)G$ (thus made well behaved at infinity) is evaluated numerically. As mentioned, we perform the frequency integral along the imaginary axis, in order to avoid the strong structure that occurs in W and G along the real axis, as described in the Appendix. In forming the product GW, we make use of the efficiency of fast Fourier transforms (FFT's) by observing that W, G, and Σ can all be decomposed into terms that are strictly periodic in the six-dimensional space $(\mathbf{r},\mathbf{r}')$. (The full quantities are periodic only in the three-dimensional vector $\mathbf{r}-\mathbf{r}'$.) The decomposition proceeds as follows:

$$\Sigma(\mathbf{r},\mathbf{r}',\mathbf{k},\omega)=\frac{i}{4\pi}\sum_{\mathbf{q}}\int_{-\infty}^{\infty}W(\mathbf{r},\mathbf{r}',\mathbf{q},\omega') \times G(\mathbf{r},\mathbf{r}',\mathbf{k}-\mathbf{q},\omega+\omega')d\omega' \ , \tag{20}$$

where (taking W as an example)

$$W(\mathbf{r},\mathbf{r}',\mathbf{q},\omega)=\frac{1}{\Omega^2}\sum_{\mathbf{G},\mathbf{G}'}e^{i\mathbf{G}\cdot\mathbf{r}}e^{-i\mathbf{G}'\cdot\mathbf{r}'}W_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) \ , \tag{21}$$

where $W_{\mathbf{G}\mathbf{G}'}$ are the reciprocal-space matrices described above. The quantities $W(\mathbf{r},\mathbf{r}',\mathbf{q},\omega)$, etc. can be calculated using the FFT. The calculation of Σ then consists of summing the products of pairs of these quantities [Eq. (20)], and taking the inverse FFT to obtain the $\mathbf{k}$-space representation of Σ, $\Sigma_{\mathbf{G}\mathbf{G}'}(\mathbf{k},\omega)$:

$$\Sigma_{\mathbf{G}\mathbf{G}'}(\mathbf{k},\omega)=\mathcal{F}(\Sigma(\mathbf{r},\mathbf{r}',\mathbf{k},\omega)) \ , \tag{22}$$

where $\mathcal{F}$ means to take the fast Fourier transform. The bare Coulomb $(k')^{-2}$ divergence as $k'\to 0$ in the bare-exchange term is handled in the obvious way by analytically integrating the divergence through a sphere of equivalent volume for the term $\mathbf{k}'=\mathbf{0}$.

C. The quasiparticle energies and true $V_{\rm xc}$

The calculated self-energy is then used in two ways. First, the quasiparticle energies E are found, using second-order perturbation theory in $\Sigma-V_{\rm xc}^{\rm LDA}$:

$$E_i=E_{i,\rm DFT}^{\rm LDA}+M_{ii}(E_i)+\sum_{j\neq i}\frac{|M_{ij}(E_i)|^2}{E_{i,\rm DFT}^{\rm LDA}-E_{j,\rm DFT}^{\rm LDA}} \ , \tag{23}$$

where the M_{ij} are the matrix elements of $\Sigma-V_{\rm xc}^{\rm LDA}$:

$$M_{ij}(E)=\langle\psi_{i,\rm DFT}^{\rm LDA}|\Sigma(E)-V_{\rm xc}^{\rm LDA}|\psi_{j,\rm DFT}^{\rm LDA}\rangle \ , \tag{24}$$

which are determined self-consistently within the Taylor expansion,

$$M_{ij}(E)\approx M_{ij}(0)+EM_{ij}'(0)+\tfrac{1}{2}E^2M_{ij}''(0)+\tfrac{1}{6}E^3M_{ij}'''(0) \ , \tag{25}$$

where the derivatives $M'\equiv dM/dE$, etc. are evaluated numerically along the *imaginary* axis from $E=0$ (the middle of the band gap). This analytic continuation is correct strictly only for the determination of band edges, i.e., energies outside the $E(k)$ spectrum. The second-order terms in (23) are very small (<0.03 eV), demonstrating the closeness of the LDA and quasiparticle wave functions. The calculated quasiparticle energies are found to be in excellent agreement with experiment (see Sec. III).

Second, we use Σ to calculate a DFT exchange-correlation potential,[35] $V_{\rm xc}$, for the semiconductor, using an exact relationship between Σ and the DFT $V_{\rm xc}$ that was derived by Sham and Schlüter.[22,36] This results from the fact that DFT is formulated so that the electron density of the system of effective noninteracting electrons is the same as that of the true, many-body system. Sham and Schlüter noted that the electron density is the integrated local density of states, which is, in turn, $(1/\pi){\rm Im}G(\mathbf{r},\mathbf{r})$, where G is the Green's function. They then used a Dyson equation to relate G, the many-body Green's function, to $G_{\rm DFT}$, the DFT Green's function, finally obtaining the integral equation

$${\rm Im}\int_{-\infty}^{\mu}[G_{\rm DFT}(\Sigma-V_{\rm xc})G]_{\mathbf{r},\mathbf{r}}d\omega=0 \ , \tag{26}$$

where $[\ ]_{\mathbf{r},\mathbf{r}}$ denotes matrix multiplication inside the square brackets, with the $\mathbf{r}=\mathbf{r}'$ matrix element then taken. In order to make this tractable, we replace both G and $G_{\rm DFT}$ by $G_{\rm DFT}^{\rm LDA}$. Tests indicate that this is of little consequence in the final $V_{\rm xc}$. Equation (26) can be written as a set of linear equations for $V_{\rm xc}$ of the form $\underline{M}\mathbf{V}_{\rm xc}=\mathbf{U}$, where $\underline{M}$ is a matrix and $\mathbf{U}$ a vector:

$$\int d\mathbf{r}'M(\mathbf{r},\mathbf{r}')V_{\rm xc}(\mathbf{r}')=U(\mathbf{r}) \ , \tag{27}$$

where

$$M(\mathbf{r},\mathbf{r}')={\rm Im}\int_{-\infty}^{\mu}d\omega\int d\mathbf{r}''G_{\rm DFT}^{\rm LDA}(\mathbf{r},\mathbf{r}'',\omega) \times G_{\rm DFT}^{\rm LDA}(\mathbf{r}'',\mathbf{r}',\omega) \ , \tag{28}$$

$$U(\mathbf{r})=\mathrm{Im}\int_{-\infty}^{\mu} d\omega \int d\mathbf{r}'' \int d\mathbf{r}''' G_{\mathrm{DFT}}^{\mathrm{LDA}}(\mathbf{r},\mathbf{r}'',\omega) \times \Sigma(\mathbf{r}'',\mathbf{r}''',\omega) G_{\mathrm{DFT}}^{\mathrm{LDA}}(\mathbf{r}''',\mathbf{r},\omega) \ . \tag{29}$$

M and U are evaluated in a basis of plane waves. We write $\Sigma=\Sigma_x+\Sigma_c$, where Σ_x is the ω-independent bare-exchange self-energy, and perform the frequency integrals for M and the Σ_x part of U analytically, and the integral for the remainder of U on the imaginary axis as described before for Σ (see Appendix). The inversion of the matrix M is somewhat sensitive to numerical noise in the calculation; however, this is reflected much more strongly in the detailed form of $V_{\mathrm{xc}}(\mathbf{r})$ in some parts of real space (where the uncertainty appears to be up to about 0.5 eV) than in its matrix elements $\langle \psi \mid V_{\mathrm{xc}} \mid \psi \rangle$ (which seem stable to about 0.05 eV). We again use second-order perturbation theory, now in $V_{\mathrm{xc}}-V_{\mathrm{xc}}^{\mathrm{LDA}}$, to calculate the DFT eigenvalues using this V_{xc}:

$$E_{i,\mathrm{DFT}}=E_{i,\mathrm{DFT}}^{\mathrm{LDA}}+\tilde{M}_{ii}+\sum_{\substack{i,j \\ j\neq i}} \frac{|\tilde{M}_{ij}|^2}{E_{i,\mathrm{DFT}}^{\mathrm{LDA}}-E_{j,\mathrm{DFT}}^{\mathrm{LDA}}} \ , \tag{30}$$

where the $\tilde{M}_{ij}$ are the matrix elements of $V_{\mathrm{xc}}-V_{\mathrm{xc}}^{\mathrm{LDA}}$:

$$\tilde{M}_{ij}=\langle \psi_{i,\mathrm{DFT}}^{\mathrm{LDA}} \mid V_{\mathrm{xc}}-V_{\mathrm{xc}}^{\mathrm{LDA}} \mid \psi_{j,\mathrm{DFT}}^{\mathrm{LDA}} \rangle \ . \tag{31}$$

As is obvious from its derivation, Eq. (26) does not determine the $\mathbf{G}=\mathbf{0}$ component (i.e., an additive constant) in V_{xc}. This is calculated separately by using the fact that the DFT chemical potential is correct: the constant is chosen so that $E_{i,\mathrm{DFT}}$ is identical to the corresponding quasiparticle energy E_i either (i) at the valence-band maximum, if $V_{\mathrm{xc}}^{(N)}$ is wanted, or (ii) at the conduction-band minimum, if $V_{\mathrm{xc}}^{(N+1)}$ is wanted. The results given later show $V_{\mathrm{xc}}^{(N)}$.

III. QUASIPARTICLE ENERGIES AND SELF-ENERGY OPERATORS

A. Quasiparticle energies

Table I gives the calculated quasiparticle energies for silicon and, for comparison, the LDA eigenvalues and the experimental values. The DFT eigenvalues with the calculated V_{xc} are also given, and will be discussed in Sec. IV. Tables II, III, and IV are the corresponding results for gallium arsenide, aluminum arsenide, and diamond, respectively.

The general trends are immediately obvious, and are the same for each of the four materials: The quasiparticle energies calculated using the *GW* self-energies are in excellent agreement with experiment (generally within 0.1 eV), and differ from the LDA eigenvalues mainly in a very approximately rigid vertical displacement of the conduction bands relative to the valence bands. In the remainder of this section we discuss the energies in more detail and relate them to the self-energy operators from which they are calculated.

In comparing the quasiparticle energies with experiment, it is important to bear in mind that optical experiments contain excitonic contributions (see Sec. I B). On the other hand, some of the experimental data come from photoemission (or inverse photoemission) in which the final (initial) state of the electron is essentially outside the crystal, so that the true quasiparticle energy is measured. We have not distinguished between the two in the "Expt." column of the tables, and the good agreement obtained with all measurements indicates that the corrections to the quasiparticle energy *differences* here are small.

In silicon (Table I) the *GW* energies are usually well within 0.1 eV of experiment. At L_{1c} and L_{3c}, where there are significant variations between different experi-

TABLE I. Quasiparticle energies (*GW*) at Γ, column X, and L, and the minimum gap in silicon, in eV. The experimental (Expt.) energies (at $T=0$ where known), and the DFT eigenvalues using (i) $V_{\mathrm{xc}}^{\mathrm{LDA}}$ ("LDA") and (ii) the V_{xc} calculated form Σ ("True DFT") are also shown.

	LDA[a]	True DFT[b]	*GW*	Expt.	*GW* − DFT
Γ'_{25v}	−0.07	0.00	0.00	0.00[c]	0.00
L'_{3v}	−1.29	−1.21	−1.19	−1.2±0.2[d], −1.5[e]	0.02
c-band minimum	0.45	0.66	1.24	1.17[d]	0.58
Γ_{15c}	2.50	2.68	3.30	3.40[d]	0.62
Γ'_{2c}	3.49	3.66	4.27	4.19[d]	0.61
L_{1c}	1.46	1.62	2.30	2.1[f], 2.4[e]	0.68
L_{3c}	3.30	3.49	4.11	4.3±0.2[d], 4.0[d,e]	0.62
Minimum gap	0.52	0.66	1.24	1.17[d]	0.58[g]

[a]Hedin RPA LDA (Ref. 33); the more accurate Ceperley-Alder LDA (Ref. 9) would increase the eigenvalues by 0.49 eV and leave the gaps unaffected.
[b]The $V_{\mathrm{xc}}(\mathbf{r})$ calculated from Σ_{GW}.
[c]Aligned with the quasiparticle valence-band maximum.
[d]Reference 37.
[e]Reference 38.
[f]Reference 39.
[g]Equal to Δ, the discontinuity $V_{\mathrm{xc}}^{(N+1)}-V_{\mathrm{xc}}^{(N)}$.

TABLE II. Quasiparticle and other energies in GaAs. (See Table I for details.)

	LDA[a,b]	True DFT[c,b]	GW[d]	Expt.	$GW-$DFT
Γ_{15v}	−0.13	0.00	0.00	0.00[e,f] (+0.11, −0.22)	0.00
X_{5v}	−2.70	−2.55	−2.64	−2.80[f,g] (+0.05, −0.02)	−0.09
L_{3v}	−1.20	−1.06	−1.11	−1.30[f,g] (±0.05)	−0.05
Γ_{1c}	0.54	0.91	1.58	1.63[h]	0.67
X_{1c}	1.36	1.59	2.19	2.09[h]	0.60
X_{3c}	1.53	1.74	2.41	2.49[g,h]	0.67
L_{1c}	1.02	1.30	1.93	1.93[h]	0.63
Minimum gap	0.67	0.91	1.58	1.63[h]	0.67[i]

[a]Hedin RPA LDA (Ref. 33); the more accurate Ceperley-Alder LDA (Ref. 9) would increase the eigenvalues by 0.49 eV and leave the gaps unaffected.
[b]Pseudopotential calculation omits core relaxation—see text.
[c]The $V_{xc}(\mathbf{r})$ calculated from Σ_{GW}.
[d]Core-valence exchange-correlation contribution to self-energy omitted—see text.
[e]Aligned with the quasiparticle valence-band maximum.
[f]Spin-orbit splitting (shown in parentheses) removed.
[g]Reference 37.
[h]Reference 40.
[i]Equal to Δ, the discontinuity $V_{xc}^{(N+1)}-V_{xc}^{(N)}$.

ments, the calculated energies are between the two.

In GaAs (Table II) it should be noted that two non-negligible but partly canceling effects have been omitted from the calculations: core relaxation (which has been found in LDA calculations[26] to decrease the band gaps by about 0.02–0.27 eV) and valence-core exchange and correlation (which we estimate from the symmetry of the wave functions and the atomic germanium calculations in Ref. 18 to increase the band gaps by about 0.1 eV). The partial cancellation of these terms leads to excellent agreement with experiment.

Another interesting but rarely discussed point concerns the ordering of the two lowest conduction bands at X in GaAs, which differ in energy by only about 0.2 eV. Experiments[44] in GaP on the valley-orbit splitting of different substitutional impurities suggest that the lower band has the character of anion s orbitals. Examination of the calculated wave functions in GaAs shows that the lower band indeed has s-like symmetry about the As atoms, while the upper band has a higher projection onto Ga s orbitals in both the LDA and GW calculations (and, incidentally, with the true V_{xc}). The calculated ordering is thus X_1, X_3 in GaAs and consistent with experiments on GaP. No similar experiments are known in GaAs.

AlAs is similar to GaAs, except in having an indirect rather than direct minimum band gap, and the quasiparticle energies are again mostly in good agreement with experiment, although less data are available. The notable exception is at the L point, where the conduction-band local minimum L_{1c} is placed 0.54 eV higher (relative to Γ_{15v}) by the GW calculations than by experiment, although the direct gap $L_{3v}\rightarrow L_{1c}$ is within 0.11 eV of experiment. We have proposed[2] that the resolution of this disagreement should be a revision of the "experimental" value, as investigation shows that the accepted indirect gap is based on no firmer evidence than a quadratic extrapolation in x, the alloy parameter in the alloy series $Ga_{1-x}Al_xAs$, which is already widely accepted to be in-

TABLE III. Quasiparticle and other energies in AlAs. (See Table I for details).

	LDA[a]	True DFT[b]	GW	Expt.	$GW-$DFT
Γ_{15v}	−0.04	0.00	0.00	0.00[c,d] (+0.09, −0.18)	0.00
X_{5v}	−2.21	−2.17	−2.30	−2.32[e,d] (+0.12, −0.06)	−0.13
L_{3v}	−0.85	−0.82	−0.89 ($=L_{1c}-4.01$)	$L_{1c}-4.00$	−0.07
Γ_{1c}	2.38	2.56	3.35	3.22[e]	0.79
X_{1c}	1.33	1.55	2.18	2.32[e]	0.63
L_{1c}	2.18	2.38	3.12	2.58[e]	0.74
Minimum gap	1.37	1.55	2.18	2.32[e,d] (+0.18, −0.09)	0.63[f]

[a]Hedin RPA LDA (Ref. 33); the more accurate Ceperley-Alder LDA (Ref. 9) would increase the eigenvalues by 0.49 eV and leave the gaps unaffected.
[b]The $V_{xc}(\mathbf{r})$ calculated from Σ_{GW}.
[c]Aligned with the quasiparticle valence-band maximum.
[d]Spin-orbit splitting (shown in parentheses) removed.
[e]Reference 37.
[f]Equal to Δ, the discontinuity $V_{xc}^{(N+1)}-V_{xc}^{(N)}$.

TABLE IV. Quasiparticle and other energies in diamond. (See Table I for details.)

	LDA[a]	True DFT[b]	*GW*	Expt.	*GW* − DFT
Γ'_{25v}	0.05	0.00	0.00	0.00[c]	0.00
X_{4v}	−6.31	−6.26	−6.71		−0.45
L'_{3v}	−2.79	−2.79	−2.98		−0.19
c-band minimum	3.85	4.21	5.33	5.48[d]	1.12
Γ_{15c}	5.57	5.72	7.26	7.12[d]	1.54
Γ'_{2c}	13.11	13.24	14.83	15.3±0.5[e]	1.59
X_{1c}	4.53	4.81	5.84 ($=X_{4v}+12.55$)	$X_{4v}+12.55$[d]	1.03
L_{3c}	8.38	8.48	9.63 ($=L'_{3v}+12.61$)	$L'_{3v}+12.0$[d]	1.15
L_{1c}	8.68	8.86	10.55		1.69
Minimum gap	3.90	4.21	5.33	5.48[d]	1.12[f]

[a]Hedin RPA LDA (Ref. 33); the more accurate Ceperley-Alder LDA (Ref. 9) would increase the eigenvalues by 0.52 eV and leave the gaps unaffected. Calculated using Refs. 18 and 41.
[b]The $V_{xc}(\mathbf{r})$ calculated from Σ_{GW}.
[c]Aligned with the quasiparticle valence-band maximum.
[d]Reference 37.
[e]Reference 42.
[f]Equal to Δ, the discontinuity $V_{xc}^{(N+1)}-V_{xc}^{(N)}$.

valid for the Γ_{1c} minimum. There is to our knowledge no direct measurement of L_{1c}.

Comparing our results for the valence-band top for bulk GaAs and bulk AlAs, we may infer a many-body correction to the GaAs/AlAs heterojunction band offset. From Tables II and III we obtain a lowering of AlAs with respect to GaAs of 0.09 eV, which is a small correction (probably within our error bars) to the LDA band offset.

We included diamond in this study as an example of a more covalent and wide-band-gap material. Table IV shows that the *GW* quasiparticle energies are again within about 0.1 eV of the experimental values, except that the calculated direct band gap at Γ is 0.6 eV greater than the value in Ref. 37.

It has been known for some time that the main discrepancies between the LDA energies in silicon and the experimentally measured quasiparticle energies can be corrected by simply adding a constant (about 0.7 eV) to the LDA energies above the band gap.[45] This procedure is often referred to as a *scissors operator,* because it can be thought of as cutting the band structure along the band gap and moving the conduction bands rigidly upwards. Now that it is possible to calculate the quasiparticle energies accurately from first principles, the quasiparticle and LDA energies may be compared at many more points in the Brillouin zone, so that the origins and range of applicability of the scissors operator can be studied. For this purpose we have plotted in Fig. 4 the difference between the quasiparticle and LDA energies as a function of $\mathbf{k}$ for the top valence band and bottom conduction band for each of the four materials. We shall consider the question of the alignment of the valence-band maxima later, but in this figure we have aligned them artificially for clarity (so that the plotted quantity is automatically zero at Γ in the valence band). Clearly, if the scissors operator is to be valid, the curves should be a $\mathbf{k}$-independent constant for the conduction band, and zero in the valence band. It is evident that the scissors operator is indeed accurate to about 0.1 eV in silicon for the bands shown, but that its accuracy is substantially diminished for the other materials, particularly diamond. (We note, however, that the validity of the scissors operator is substantially better in AlAs than comparison of the LDA results with the usual experimental values would indicate, because of the proposed correction of the quasiparticle L_{1c} energy mentioned above.)

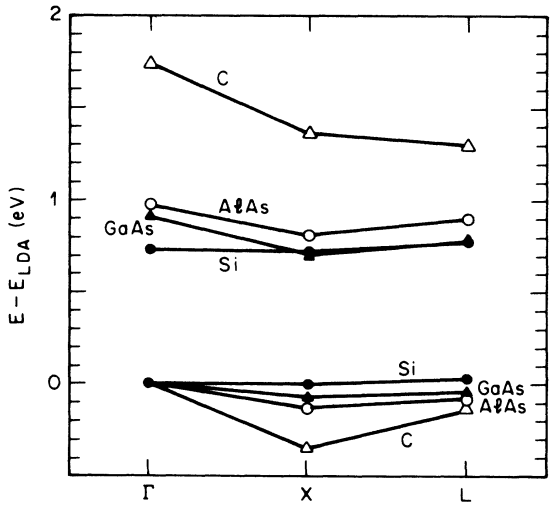

FIG. 4. The varying validity of the "scissors operator." The difference between the calculated quasiparticle energies and the LDA energies (relative to the valence-band maximum) in the lowest conduction band and highest valence band is plotted against $\mathbf{k}$ for each material. If the scissors operator is valid, this quantity should be independent of $\mathbf{k}$. It is evident that the scissors operator is accurate to about 0.1 eV in silicon, but only to 0.2, 0.2, and 0.4 eV, respectively, in GaAs, AlAs, and diamond.

When the ingredients of the plotted energy correction are considered (see next section), it becomes clear that for the scissors operator to be applicable there has to be a remarkable cancellation between separate terms of the equations, one which could not be expected to cancel automatically. In particular, the energy dependence of the self-energy leads to a strongly state-dependent energy correction, which is partially, though by no means completely, canceled by the difference between the nonlocal self-energy and the local exchange-correlation potential. Thus the natural expectation is that the net energy correction would have a strong state dependence within each band as well as across the band gap, so that the approximate validity of the scissors operator in silicon must be regarded as a coincidence.

Since Σ is determined in full by the GW expression, there is no need to align any two points in the quasiparticle and LDA band structures artificially. If both the self-energy operator and exchange-correlation potential were exact, the chemical potential [the quasiparticle energy at the valence-band maximum or the conduction-band minimum, depending on whether the N- or $(N+1)$-particle system is being considered] would be equal to the corresponding highest occupied DFT eigenvalue.[13] Here neither is exact, yet we note that the quasiparticle valence-band maxima are close to the LDA (RPA) valence-band maxima, suggesting that $V_{\rm xc}^{\rm LDA}$ is closer to $V_{\rm xc}^{(N)}$ than to $V_{\rm xc}^{(N+1)}$ (which differs by a constant, Δ).[46] The RPA LDA, which is essentially the GW approximation applied to the homogeneous electron gas, is known to be a constant 0.49 eV (0.52 eV) below the LDA used here [which we take from the accurate Monte Carlo calculations for $\Gamma_s=2$ ($\Gamma_s=1.3$) jellium in Ref. 9] over the range of densities found in these materials. Since the quasiparticle energy *differences* are already well reproduced by the GW self-energy operator, corrections to the GW approximation (the so-called vertex corrections) must merely raise all the quasiparticle energies by about 0.5 eV if the alignment of the valence-band maxima is to be maintained. (We note that although we go on to calculate a more accurate $V_{\rm xc}$ from Σ, we cannot use this $V_{\rm xc}$ to address the accuracy of the alignment of the valence-band maxima, as this alignment is perfect by construction.)

We end this section with a brief comment on the quasiparticle energies corresponding to the other self-energy operators calculated: the bare exchange (Hartree-Fock) operator Σ_x and the statically screened exchange operator $\Sigma_{\rm ssx}$, each of which is energy independent. In each of the four materials the bare-exchange gaps and widths of the individual bands are much too large: the minimum gaps are approximately 5, 6, 7, and 12 eV in Si, GaAs, AlAs, and diamond. Screening with the static dielectric matrix reduces the gaps (to 0.4, 1.2, 1.2, and 4.7 eV), but does not fully correct for the lack of energy dependence. Dynamical screening, as included in the GW calculations, is essential in calculating accurate quasiparticle energies.

B. Self-energy operators

As discussed in the Introduction, Σ is central to the quasiparticle properties of the system. We show in other sections of this paper that the GW approximation for Σ may be used successfully both to calculate quasiparticle energies and related properties, and to generate an exchange-correlation potential. However, it is also imperative to understand the underlying physics for a range of materials, as the size of the GW calculations makes them prohibitively expensive for complex systems, and it is likely that progress will be made through gaining a more thorough understanding of Σ. In this section, therefore, we concentrate on Σ itself, and examine the extent to which it may be understood in terms of simpler models.

We shall take as a point of reference the detailed calculations by Hedin[33] of the self-energy operator (within the same GW approximation) of the homogeneous electron gas. In order to focus on the physics, we consider the contributions to the first-order expression for the quasiparticle energy [obtained from Eq. (23)]:

$$E_i \approx E_i^{\rm LDA} + \langle i \mid \Sigma(\omega=E_i) \mid i \rangle - \langle i \mid V_{\rm xc}^{\rm LDA} \mid i \rangle \ . \qquad (32)$$

The last two terms, to which we shall refer as $\langle \Sigma(E_i) \rangle$ and $\langle V_{\rm xc}^{\rm LDA} \rangle$ for brevity, are plotted individually as a function of quasiparticle energy in Fig. 5(a) for silicon and Fig. 6(a) for diamond. Although each of the two quantities has a strong state dependence, the difference between them (which represents the energy correction) is much closer to a step function (the so-called scissors operator), especially in silicon [(Figs. 5(b) and 6(b)].

This simple form of the energy correction is not present if one compares $\langle V_{\rm xc}^{\rm LDA} \rangle$ with the self-energy evaluated at the midgap energy $\langle \Sigma(0) \rangle$ [also Figs. 5(b) and 6(b)], illustrating the importance of including the energy dependence of the self-energy. As can be seen, the effect of including the energy dependence is to alter greatly the dispersion of the individual bands, and to *reduce* the band gap slightly. So although the energy dependence is important, it is clearly not primarily the fact that Σ is energy dependent (while $V_{\rm xc}^{\rm LDA}$ is not) that leads to the band-gap correction. Figures 5(c) and 6(c) show the matrix elements for jellium corresponding to those for silicon and diamond in Figs. 5(b) and 6(b). Neither $\langle \Sigma \rangle$ and $\langle V_{\rm xc}^{\rm LDA} \rangle$ individually nor their difference reproduce details of the same quantities in silicon and diamond; in particular, the discontinuity in the difference across the band gap is not present at the Fermi energy in jellium.

The other way in which Σ is essentially different from $V_{\rm xc}$ is that Σ is a *nonlocal* operator, while $V_{\rm xc}^{\rm LDA}$ is a *local* potential. Since many years of experience with empirical potentials (which were never able to reproduce both the ground-state electron density and the quasiparticle band structure, as the true self-energy operator must) show that no local potential contains the entire physics of Σ, we conclude that the nonlocality is essential in determining the correct quasiparticle energies (and, in particular, the correct band gap).

To investigate the nonlocality of Σ further, we have plotted the self-energy operator in real space as a function of $\mathbf{r}'$, keeping $\mathbf{r}$ fixed at a particular point in the unit cell. This is shown in Fig. 7 for each of the four materials. Despite the complexity of the calculation of Σ, the

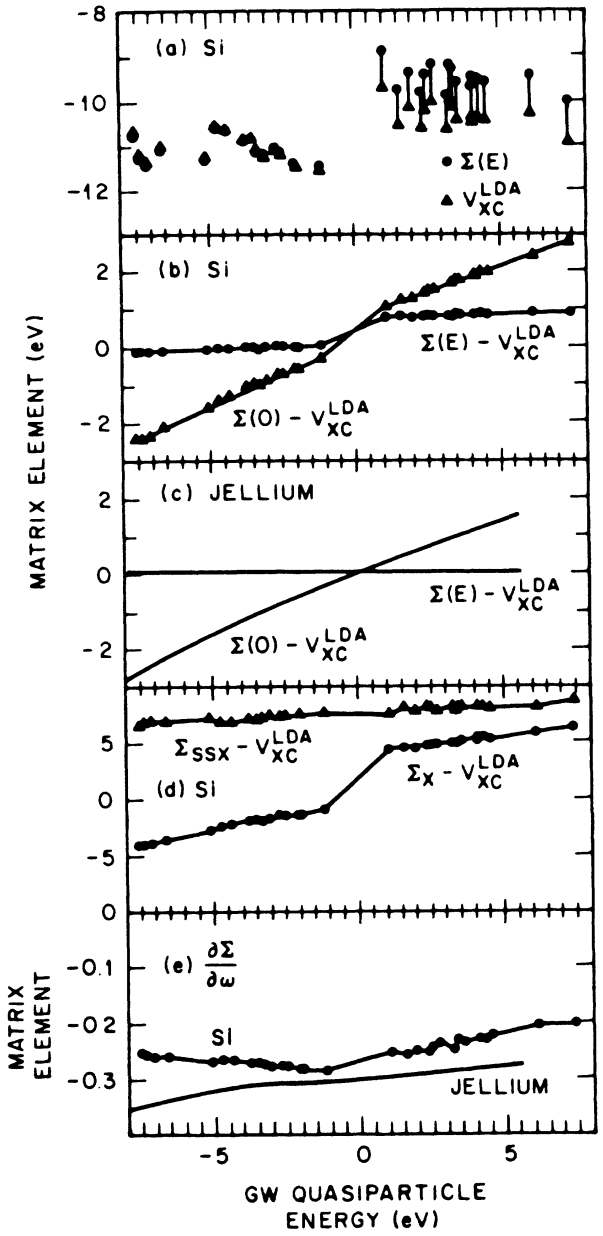

FIG. 5. (a) The real parts of the matrix elements of the *GW* self-energy operator of silicon $\langle \Sigma_{GW}(\omega=E) \rangle$ and the LDA exchange-correlation potential $\langle V_{xc}^{LDA} \rangle$, plotted against the quasiparticle energy E. (b) The differences $\langle \Sigma_{GW}(\omega=E) \rangle - \langle V_{xc}^{LDA} \rangle$ (essentially the quasiparticle-energy correction) and $\langle \Sigma_{GW}(\omega=0) \rangle - \langle V_{xc}^{LDA} \rangle$. The importance of the energy dependence of Σ can be seen. (c) For comparison, $\langle \Sigma_{GW}^{hom}(\omega=E) \rangle - \langle V_{xc}^{LDA} \rangle$ and $\langle \Sigma_{GW}^{hom}(\omega=E_F) \rangle - \langle V_{xc}^{LDA} \rangle$ are plotted against the quasiparticle energy E for jellium of the average density of silicon ($r_s=2.0$). (Data from Ref. 33.) (E_F is aligned with the middle of the band gap.) Both are relatively featureless and clearly do not share the sharp discontinuities present in (a). (d) The real parts of the matrix elements of the bare-exchange (Hartree-Fock) self-energy operator of silicon $\langle \Sigma_x \rangle$ and the statically screened exchange self-energy operator $\langle \Sigma_{ssx} \rangle$. $\langle V_{xc}^{LDA} \rangle$ is subtracted as in (b). (e) The real parts of the matrix elements in silicon and jellium ($r_s=2.0$) (Ref. 33) of the frequency derivative of the self-energy $\langle \partial\Sigma_{GW}(\omega=\text{midgap})/\partial\omega \rangle$.

operator has a relatively simple structure when viewed in this way. The operator is dominated by a single "hole," centered about $\mathbf{r}=\mathbf{r}'$ and approximately spherical. Moreover, the shape of the hole (though not its amplitude) is approximately independent of the location of $\mathbf{r}$. This suggests the following model form for Σ:

$$\Sigma(\mathbf{r},\mathbf{r}',\omega) \approx \tfrac{1}{2}[f(\mathbf{r},\omega)+f(\mathbf{r}',\omega)]g(|\mathbf{r}-\mathbf{r}'|), \quad (33)$$

where the form of the terms in the square brackets is chosen to reproduce the correct $\mathbf{r} \leftrightarrow \mathbf{r}'$ symmetry. This is an operator whose nonlocality is a universal function of $\mathbf{r}-\mathbf{r}'$. The function g is also very similar to the corre-

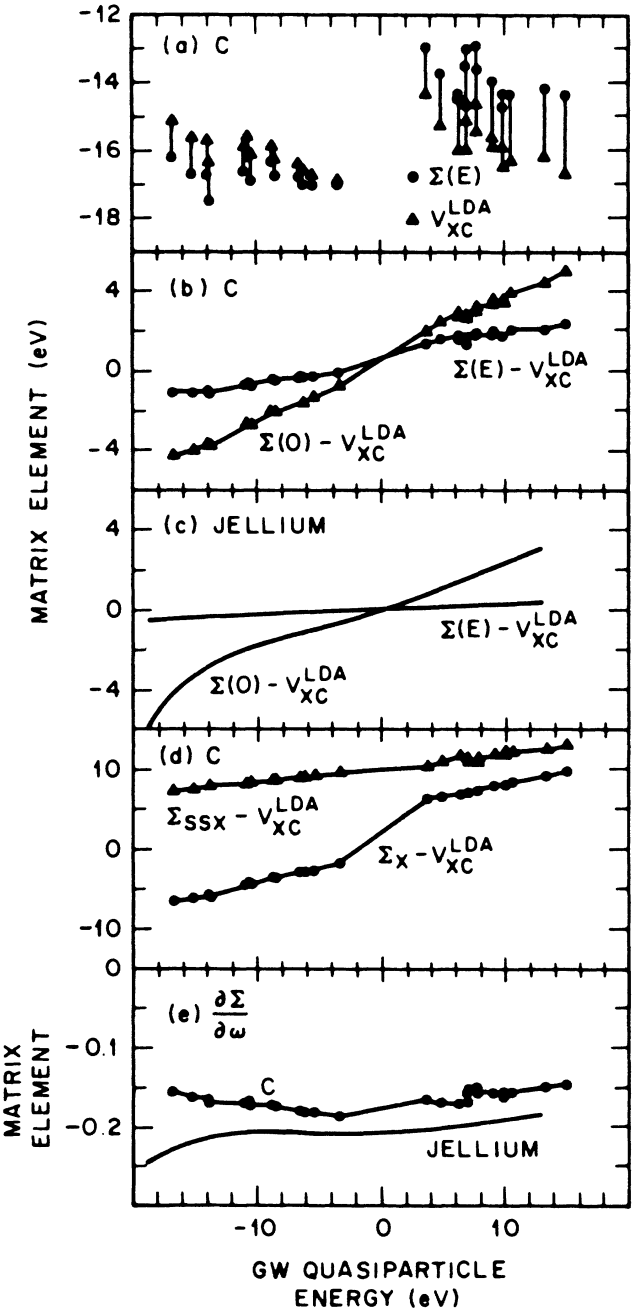

FIG. 6. Same as Fig. 5, but for diamond.

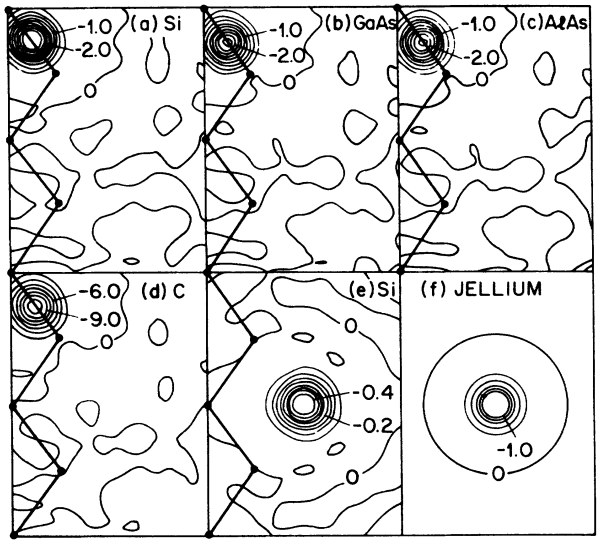

FIG. 7. Contour plots of the self-energy $\Sigma(\mathbf{r},\mathbf{r}', \omega = \text{midgap})$ in eV a.u.$^{-3}$ for $\mathbf{r}$ fixed at the bond center and $\mathbf{r}'$ shown in the (110) plane, for (a) silicon, (b) gallium arsenide, (c) aluminum arsenide, and (d) diamond. For silicon the corresponding plots with $\mathbf{r}$ fixed at the tetrahedral interstitial site is also shown [panel (e)]. For comparison, the self-energy operator of jellium with $r_s = 2.0$ (the average density of silicon) is shown in (f) (from Ref. 33).

sponding function extracted from the self-energy operator for jellium *of the average density of the semiconductor.* This is shown by comparing the "holes" in Figs. 7(a) and 7(e) (silicon) with the "hole" in Fig. 7(f) [jellium with $r_s = 2.0$ (Ref. 33)]. This "nonlocality hole" contains the essential nonlocality of Σ as far as the quasiparticle energies are concerned: for example, in silicon more than 99% of the matrix element $\langle \psi | \Sigma | \psi \rangle$ originates from $|\mathbf{r}-\mathbf{r}'| < r_h$, where r_h is the radius of the nonlocality hole.

The range (to the first zero) of the "nonlocality function" $g(r)$ is seen (by referring to the jellium calculations in Ref. 33 or by direct inspection of Fig. 7) to be approximately 2.0 r_s for the jellium of the corresponding average density; for example, 4.1 a.u. (2.1 Å) in silicon. This length is comparable with the wavelength of the wave functions[47] near the band gap,[1] thus providing further proof that the nonlocality of Σ is essential, since only if the range of Σ were much less than the wavelengths would the matrix elements $\langle \psi | \Sigma | \psi \rangle$ be preserved *if* Σ were approximated by a *local* operator of the correct weight. Thus our picture of the physical origins of the quasiparticle-energy corrections to the LDA energies is the following: The nonlocality of the *operator* $\Sigma - V_{xc}^{LDA}$ is an essential feature, and is well reproduced by that of the self-energy operator in jellium of the average density of the semiconductor, but the interaction of the *wave functions* with this nonlocality is *not* contained within the jellium calculations.

To examine this point in more detail, we consider a tight-binding approximation to the band structure of silicon. The states in the highest valence band and lowest conduction band are of quite distinct character in this picture: At Γ, for example, the valence-band maximum is bonding p, while the lowest conduction band is antibonding p. Thus the conduction-band wave functions have an extra node between the atoms and the effective wavelength in this important part of the unit cell (where the self-energy operator has greatest strength) is smaller. The matrix elements $\langle \psi | \Sigma | \psi \rangle$ are therefore less negative in the conduction band than they would be if Σ were of the same strength but local, and so the quasiparticle energies in the conduction bands are raised and the band gap increases. (This point is illustrated in Fig. 2 of Ref. 1, which shows the wave functions and self-energy operator concerned.) An increased band gap is therefore an immediate consequence of the essential nonlocality of Σ.

In Table V we show how well the nonlocality of Σ is reproduced by the self-energy operator of jellium of the *average* density of the semiconductor. The final column shows the corresponding "local-density" approximation for the range of Σ, in which the jellium with the local density at $\mathbf{r}$ is used. It is evident that this approximation severely overestimates the range of Σ in low-density regions and underestimates it in high-density regions. The approximation suggested in Ref. 13, in which Σ is approximated by the value of the self-energy operator of jellium at $\frac{1}{2}(\mathbf{r}+\mathbf{r}')$ (suitably aligned in energy), will clearly

TABLE V. The density dependence of the nonlocality of Σ. The rangel of Σ [defined as the spherically averaged distance to the first zero in the oscillatory function $\Sigma(\mathbf{r},\mathbf{r}', \omega = \text{midgap})$ with $\mathbf{r}$ fixed at the site shown] compared with the corresponding ranges of the self-energy operators at the Fermi energy in jellium (Ref. 33) with (i) the average density of the semiconductor and (ii) the local density at $\mathbf{r}$ in the semiconductor. The sites shown for $\mathbf{r}$ are the bond centers and the tetrahedral interstitial sites.

Material	Site	Range of Σ (a.u.)	Range of Σ in jellium (a.u.) Average density	Local density
Si	bond center	4.1	3.9	2.8
Si	interstitial	4.7	3.9	8.8
GaAs	bond center	4.2	4.0	2.8
GaAs	interstitial	4.3	4.0	9.0
AlAs	bond center	4.3	4.0	2.8
AlAs	interstitial	4.6	4.0	9.0
C	bond center	2.8	2.7	2.2
C	interstitial	3.5	2.7	4.7

be a better approximation here than the "local" column in the table because of the averaging effect of taking $\frac{1}{2}(\mathbf{r}+\mathbf{r}')$, but is nevertheless seen to be poor: in particular, the shape of the nonlocality hole will be far too anisotropic. This approximation is (as is stated in that paper) valid in the limit of slowly varying density, and these materials (in which the density varies appreciably over the range of the nonlocality hole) are outside the domain of its applicability.

Figures 8(a) and 8(b) indicate the origin of the simple form of the "nonlocality" hole. The self-energy is essentially G times W in real space, and its range is therefore that of the shorter ranged of the two. This is the screened Coulomb interaction, W, shown in Fig. 8(a) for silicon. G, in contrast, is long ranged and considerably more anisotropic, as demonstrated by Fig. 8(b), which shows the *bare*-exchange self-energy Σ_x, in which the screened W has been replaced by the long-ranged bare Coulomb interaction $1/|\mathbf{r}-\mathbf{r}'|$, so that Σ acquires the range of G. It is, therefore, the *nonlocality of the screening* (though not, of course, its strength) that is so transferrable from the jellium calculations.

In Figs. 5(d) and 6(d) we show the matrix elements of the other self-energy operators calculated for silicon and diamond: the bare-exchange (Hartree-Fock) and statically screened exchange self-energies. To assist comparison, the matrix elements of $V_{\rm xc}^{\rm LDA}$ are subtracted, and the quasiparticle energies used on the horizontal axis are the GW energies. There are some interesting features. First, as was reported in the section on the quasiparticle energies, the band gap is substantially increased when the bare exchange is used, because $\langle \Sigma_x \rangle$ has a large discontinuity across the band gap. Also, its state dependence within each band (none of which can come from its energy dependence since Σ_x is energy independent) is very similar to that of the GW self-energy *at fixed frequency* [Figs. 5(b) and 6(b)], but not to that of the GW self-energy evaluated at the correct quasiparticle energy, resulting in the Hartree-Fock bands being too wide. The statically screened self-energy has a greatly reduced discontinuity at the band gap, caused by the screening, which therefore improves the band gap. However, the dispersion of the individual bands is quite wrong, indicating the importance of including properly *dynamical* screening in the calculation.

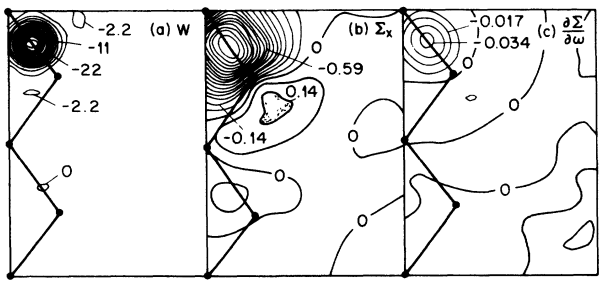

FIG. 8. Contour plots of quantities related to the self-energy of silicon for $\mathbf{r}$ fixed at the bond center and $\mathbf{r}'$ shown in the (110) plane: (a) the screened Coulomb interaction $W(\mathbf{r},\mathbf{r}',\omega=\text{midgap})$ in eV a.u.$^{-3}$, (b) the bare exchange $\Sigma_x(\mathbf{r},\mathbf{r}',\omega=\text{midgap})$ in eV a.u.$^{-3}$, and (c) $\partial\Sigma_{GW}(\mathbf{r},\mathbf{r}',\omega=\text{midgap})/\partial\omega$ in a.u.$^{-3}$.

The energy dependence of the calculated GW self-energy can also be modeled successfully using the corresponding quantity calculated for jellium. Figure 5(e) shows the matrix elements of $\partial\Sigma/\partial\omega$ for both silicon and jellium[33] ($r_s=2.0$), and while (as stated earlier) the energy dependence is significantly different from zero, it is approximately the same in each of the two cases. The same is true for diamond and its corresponding jellium [Fig. 6(e)]. [It is important to distinguish the matrix element of the energy derivative of the self-energy—which is considered here—from the variation from state to state of the matrix elements of the self-energy, which was plotted in Figs. 5(a), 5(b), 6(a), and 6(b), and is not at all well described by jellium.] The energy dependence is also examined in Fig. 8(c), which shows the energy derivative of the self-energy of silicon in real space. The fact that it resembles Fig. 7(a) shows that the energy dependence of Σ is, to a good approximation, separable. The complete model for Σ is therefore

$$\Sigma(\mathbf{r},\mathbf{r}',\omega)\approx\tfrac{1}{2}[f(\mathbf{r})+f(\mathbf{r}')]g(|\mathbf{r}-\mathbf{r}'|)h(\omega)\,, \qquad (34)$$

which we believe may serve well as an ansatz in more efficient methods for calculating Σ, since g and h may be conveniently taken from jellium calculations. The validity of this model is shown by fitting the calculated Σ_{GW} to this form, with g and h taken from jellium and f determined by optimizing the fit of the $\mathbf{k}$-averaged band *gap*. The resulting dispersion of the individual quasiparticle bands, which was *not* fitted, reproduces the GW results to within 0.05 eV in the bands within several eV of the gap.

The quantity plotted in Figs. 5(e) and 6(e) is also $1-Z^{-1}$, where Z is the well-known wave-function renormalization, which has a simple physical interpretation: the extent to which the single-particle-like quasiparticle excitations are valid descriptions of the true excitations. Z is the area under the peak in the spectral function, which is 1 if the excitation is exactly single-particle-like. Thus our results indicate that the renormalization constant in these materials is obtainable from jellium calculations in which the density is the average density of the semiconductor and the Fermi energy is aligned with the middle of the band gap.

IV. THE "TRUE" EXCHANGE-CORRELATION POTENTIAL

A. The potential and the DFT Kohn-Sham energies

The DFT eigenvalues using the calculated "true" $V_{\rm xc}$ are shown in Tables I–IV for each material, and in addition silicon's $V_{\rm xc}(\mathbf{r})$ is compared with $V_{\rm xc}^{\rm LDA}(\mathbf{r})$ in Fig. 9. In each case the true $V_{\rm xc}$ shown is that for the N-particle system—the semiconductor with its valence bands exactly filled. The corresponding potential and eigenvalues for the $(N+1)$-particle system would, as explained in Sec. I C, differ from those given by an r-independent constant Δ; that is, $E_g-E_g^{\rm DFT}$, where E_g is the calculated quasiparticle minimum gap and $E_g^{\rm DFT}$ is the gap when

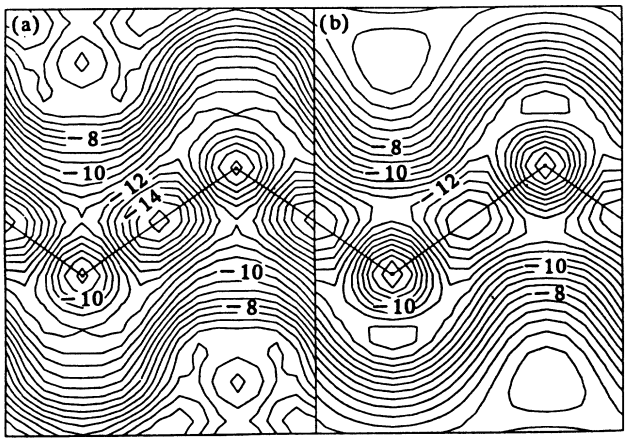

FIG. 9. The LDA and true exchange-correlation potentials, (a) V_{xc} and (b) V_{xc}^{LDA}, for silicon (the N-particle system), shown in eV in the (110) plane containing the bond chains as indicated.

V_{xc} is used.

In Tables I–IV and Fig. 9 the LDA exchange-correlation potential used is Hedin's RPA V_{xc}^{LDA}, which originates in GW calculations for jellium.[33] This makes our Σ and V_{xc} properly comparable, as in the homogeneous limit our procedure for determining V_{xc} reduces to the GW calculation for jellium. In particular, the overall alignment of the true DFT and LDA band structures (and also, at the top of the valence band, the quasiparticle band structure) can be seen to be very similar. For example, the LDA and DFT valence-band maxima are within 0.07, 0.13, 0.04, and 0.05 eV of each other in Si, GaAs, AlAs, and diamond, respectively, which is within our error bars. This alignment of the valence-band maximum clearly becomes perfect in the homogeneous limit since V_{xc}^{LDA} is then the true V_{xc}.

The similarity of the two potentials in Fig. 9 is striking, especially when one considers their quite different origins, and the fact that they have not been aligned in energy. The main difference is that the true V_{xc} is slightly deeper in the bond region, which leads to the slightly larger band gap (since the valence-band wave functions are concentrated there and are therefore lowered in energy). The curious but numerically insignificant structure in V_{xc} in the interstitial region is probably caused by numerical noise in solving the integral equation.

B. The discontinuity Δ

As discussed in Sec. II, the discontinuity Δ in V_{xc} upon addition of an electron to the N-electron system may be obtained by subtracting the calculated DFT minimum gap from the calculated GW minimum quasiparticle gap. Δ is shown in Table VI and graphically in Fig. 10. Although Δ increases with the band gap, it is a remarkably constant proportion of the error in the LDA band gap, showing that even diamond is sufficiently free-electron-like for the LDA to be a good approximation to the true DFT potential. The fact that Δ is in each case a substantial proportion of the band gap shows that attempts to calculate improved quasiparticle energies within DFT by going beyond the LDA are futile.

TABLE VI. Δ, the discontinuity in the exchange-correlation potential; E_g, the calculated minimum quasiparticle band gap; and the LDA gap error $E_g - E_g^{LDA}$ for the four materials.

Material	Δ (eV)	E_g (eV)	LDA gap error (eV)	Δ/(LDA gap error) (%)
Si	0.58	1.24	0.72	81
GaAs	0.67	1.58	0.91	74
AlAs	0.65	2.18	0.81	80
Diamond	1.12	5.33	1.43	78

C. Comparison with other non-LDA potentials

It is interesting to compare our V_{xc} with previous attempts to go beyond the LDA. The most successful such potentials that have stayed within the general framework of DFT have been the weighted-density approximation[48] (WDA), the average-density approximation[49] (ADA), and gradient-expansion techniques[50] (but only when explicitly designed to obey the important sum rule on the exchange-correlation hole[51]). The WDA has been applied to semiconductors by Kerker,[52] Manghi *et al.*,[53] and Hybertsen and Louie.[54] Since Ref. 52 makes an unreasonable further approximation,[53,54] we shall concentrate on Refs. 53 and 54.

In Ref. 53 the WDA was used to calculate the DFT band structure of GaAs, and in Ref. 54 that of silicon and germanium. (In Ref. 54, the calculations were performed for several lattice constants, and the total energy was also calculated; the agreement with the experimental lattice constant, bulk modulus, and cohesive energy was generally good, but no better than with the LDA.) In each case

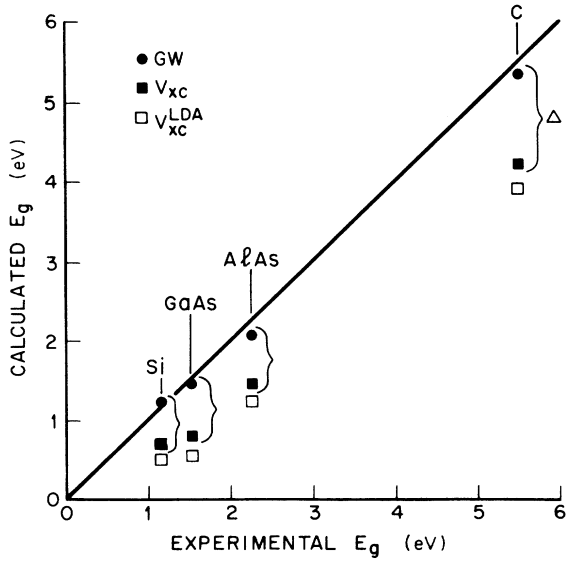

FIG. 10. The calculated minimum band gap in (i) the GW approximation, (ii) DFT, and (iii) the LDA plotted against the experimental band gap. The 45° line is a guide to the eye. Δ, the discontinuity in the exchange-correlation potential, is indicated.

the Kohn-Sham eigenvalues changed: the band gaps increased somewhat, especially when the exchange-correlation hole corresponding to the Levine-Louie model semiconductor dielectric function[55] (instead of jellium) was used in the WDA.

In Table VII we compare some sample band gaps in the LDA, WDA, and with our V_{xc}. The general trend is the same with each of the three non-LDA V_{xc}'s: the band gaps are increased by between 0.05 and 0.34 eV, so that the 55% error in the LDA minimum gap in silicon, for example, is reduced to 26–43 %. A detailed examination shows a surprising agreement between our V_{xc} and the WDA in silicon: they are within about 0.02 eV of each other. However, this excellent agreement is not obtained in silicon with the WDA using the Levine-Louie exchange-correlation hole, nor using the ordinary WDA in GaAs. The reason for the agreement is unclear, but one possible interpretation is that both the WDA and our V_{xc}, though calculated in quite different ways, are very close to the exact DFT V_{xc} in silicon.

V. CONCLUSIONS

We have shown how the self-energy operator of a semiconductor may be calculated within the GW approximation and then used to calculate both the quasiparticle energies and a density-functional-theory exchange-correlation potential. In the four materials studied (Si, GaAs, AlAs, and diamond) the quasiparticle energies are in good agreement with reliable experimental data. However, the one-electron density-functional-theory eigenvalues corresponding to the calculated exchange-correlation potential are significantly different, and are instead close to the eigenvalues obtained using the conventional local-density approximation for exchange and correlation. In particular, the DFT minimum band gap is substantially smaller than the quasiparticle gap, with the difference (equal to Δ, the discontinuity in the exchange-correlation potential upon addition of an extra electron) accounting for about 80% of the severe and well-known error in the LDA minimum gap. There are also considerable discrepancies in other parts of the energy spectrum. Thus, improving the density-functional-theory exchange-correlation potential beyond the local-density approximation will not lead to the correct quasiparticle energies, but the GW method, which goes outside density-functional theory, is capable of giving good quasiparticle properties. On the other hand, the LDA is seen to be an even better approximation to the true DFT exchange-correlation potential than it was previously known to be.

ACKNOWLEDGMENTS

One of us (R.W.G.) was supported in part by the Science and Engineering Research Council (United Kingdom). Another of us (L.J.S.) was supported in part by the U.S. National Science Foundation (NSF), Division of Materials Research, under Grant No. 85-14159, and in part by a NSF Supercomputer Supplement.

APPENDIX: THE FREQUENCY INTEGRALS

The frequency-dependent quantities ϵ, ϵ^{-1}, W, G, Σ, and U are calculated and integrated on the imaginary frequency axis, as explained in Sec. II A. The two frequency integrals to be performed numerically are (symbolically)

$$X(\omega)=\int_{-\infty}^{\infty}[W(\omega')-V]G(\omega+\omega')d\omega' \tag{A1}$$

(part of the self-energy Σ: see Sec. II B) and

$$Y=\mathrm{Im}\int_{-\infty}^{0}G(\omega)[\Sigma(\omega)-\Sigma_x]G(\omega)d\omega \tag{A2}$$

(part of the vector $\mathbf{U}$: see Sec. II C).

In the case of X, we first define the analytic continuation of $\Sigma(\omega)$ (and thus X) to complex frequencies as the

TABLE VII. The direct band gaps and the minimum gap of Si and GaAs, in eV, in (i) experiment (Expt.) and (ii) the LDA, and the corrections to the LDA values in (iii) the weighted-density approximation (using a jellium exchange-correlation hole) (WDA), (iv) the WDA using a Levine-Louie–model exchange-correlation hole [WDA(LL)], and (v) the V_{xc} calculated here from the self-energy (DFT).

Gap	Expt.	LDA	Corrections to LDA WDA	WDA(LL)	DFT
Si					
Γ	3.40[a]	2.57	+0.12[b]	+0.24[b]	+0.11
X	4.25[c]	3.53	+0.10[b]	+0.23[b]	+0.11
L	3.3[a,d,e]	2.75	+0.04[b]	+0.13[b]	+0.08
Minimum gap	1.17[a]	0.52	+0.15[b]	+0.34[b]	+0.14
GaAs					
Γ (=minimum gap)	1.63[a,f]	0.67[g]	+0.06[h]		+0.24
X	4.89[a,f]	4.06[g]	+0.13[h]		+0.08
L	3.23[a,f]	2.22[g]	+0.08[h]		+0.14

[a]Reference 37.
[b]Reference 54.
[c]Reference 43.
[d]Reference 38.
[e]Reference 39.
[f]Spin-orbit splitting omitted—see text.
[g]Core relaxation omitted—see text.
[h]Reference 53.

same expression, integrated along a contour that encloses the poles of W and G corresponding to those above the real axis when ω is real. For imaginary ω this is conveniently evaluated along the imaginary axis [Fig. 3(a)], since the presence of $W-V$ rather than W makes the contribution from the circle at infinity vanish,

$$X(\omega)=\int_{-i\infty}^{i\infty}[W(\omega')-V]G(\omega+\omega')d\omega' , \tag{A3}$$

which can be written as

$$X(\omega)=\int_{0}^{i\infty}[W(\omega')-V][G(\omega+\omega')+G(\omega-\omega')]d\omega' \tag{A4}$$

using the fact that

$$W(-\omega)=W(\omega) \quad (\mathrm{Im}\omega\neq 0) , \tag{A5}$$

which follows from its RPA formulation (11).

For Y, we deform the contour of integration from the negative real axis to the negative imaginary axis, since no poles are crossed and the integral over the circle at infinity again vanishes [Fig. 3(b)],

$$Y=\mathrm{Im}\int_{-i\infty}^{0}G(\omega)[\Sigma(\omega)-\Sigma_x]G(\omega)d\omega , \tag{A6}$$

which can be written as

$$Y=\mathrm{Im}\int_{0}^{i\infty}G(\omega^*)[\Sigma(\omega^*)-\Sigma_x]G(\omega^*)d\omega . \tag{A7}$$

Thus both X and Y have been transformed into integrals of the form

$$I=\int_{0}^{i\infty}f(\omega)d\omega . \tag{A8}$$

This is split into

$$\begin{aligned} I&=I_1+I_2 ,\\ I_1&=\int_{0}^{h}f(\omega)d\omega ,\\ I_2&=\int_{h}^{i\infty}f(\omega)d\omega \end{aligned} \tag{A9}$$

(where h is on the imaginary axis), which are transformed onto the range -1 to 1 using the transformations $u=2y/h-1$ and $v=2h/y-1$, respectively, so that standard Gaussian quadrature may be used with $n/2$ points in each of the two parts. We set $h=13.6i$ eV and use the $n=10$ and 30 sets of frequencies. For example, the $n=10$ set is $\omega=0.62i$, $3.13i$, $680i$, $10.47i$, $12.97i$, $14.28i$, $17.68i$, $27.20i$, $58.94i$, and $289.92i$ eV. As explained in the main text, G (which appears in each integral) is found to have a stronger variation along the imaginary ω axis than either W or Σ, so that while the $n=30$ is used for the main integration, Σ and W are calculated at the $n=10$ set of imaginary frequencies, and interpolated using splines to the $n=30$ set.

[1]R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).

[2]R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **35**, 4170 (1987).

[3]R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **36**, 6497 (1987).

[4]L. Hedin and S. Lundqvist, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1969), Vol. 23, p. 1.

[5]See, for example, *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N. H. March (Plenum, New York, 1983), and references therein.

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

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

[8]We use Hartree atomic units ($\hbar=e=m=4\pi\epsilon_0=1$) throughout.

[9]D. M. Ceperley and B. I. Alder, Phys. Rev. Lett. **45**, 566 (1980); parametrized in J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[10]G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. B **25**, 2867 (1982).

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

[12]W. E. Pickett and C. S. Wang, Phys. Rev. B **30**, 4719 (1984).

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

[14]S. Horsch, P. Horsch, and P. Fulde, Phys. Rev. B **28**, 5977 (1983); **29**, 1870 (1984).

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

[16]M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).

[17]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **32**, 7005 (1985).

[18]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[19]M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **58**, 1551 (1987).

[20]W. von der Linden and P. Horsch (unpublished).

[21]F. Gygi and A. Baldereschi (unpublished).

[22]L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983); L. J. Sham and M. Schlüter, Phys. Rev. B **32**, 3883 (1985).

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

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

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

[26]G. B. Bachelet and N. E. Christensen, Phys. Rev. B **31**, 879 (1985).

[27]G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[28]H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[29]We use matrixlike multiplications: $V\chi$ means $\int V(\mathbf{r},\mathbf{r}'')\chi(\mathbf{r}'',\mathbf{r}')d\mathbf{r}''$.

[30]F. Bassani and G. Pastori Parravicini, *Electronic States and Optical Transitions in Solids* (Pergamon, Oxford, 1975).

[31]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **35**, 5585 (1987); **35**, 5602 (1987), and references therein.

[32]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **37**, 2733 (1988).

[33]L. Hedin, Phys. Rev. **139**, A796 (1965).

[34]J. C. Phillips, Phys. Rev. **123**, 420 (1961).

[35]We use the phrase "true V_{xc}" to mean the V_{xc} which corresponds to our calculated Σ through the exact relationship (26); if Σ were exact, this would be the exact V_{xc}.

[36]L. J. Sham, Phys. Rev. B **2**, 876 (1985).

[37]*Physics of Group-IV Elements and III-V Compounds*, Landolt-Börnstein, Numerical Data and Functional Relation-

ships in Science and Technology, New Series, edited by O. Madelung, M. Schulz, and H. Weiss (Springer, New York, 1982) Group III, Vol. 17a.

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

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

[40]D. E. Aspnes, Phys. Rev. B **14**, 5331 (1976).

[41]M. S. Hybertsen and S. G. Louie (private communication).

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

[43]D. E. Aspnes and A. A. Studna, Phys. Rev. B **27**, 985 (1983).

[44]G. Martinez (private communication).

[45]G. A. Baraff and M. Schlüter, Phys. Rev. B **30**, 3460 (1984).

[46]We note, however, that this alignment of the LDA and quasiparticle valence-band maxima is not found in the calculations in Ref. 18, which use related but not indentical methods (see Sec. II A).

[47]Strictly, the pseudowave functions: the real wave functions have orthogonality wiggles whose interaction with the self-energy operator yields a constant and so is unimportant for the band gaps.

[48]O. Gunnarsson, M. Jonson and B. I. Lundqvist, Solid State Commun. **24**, 765 (1977).

[49]O. Gunnarsson, M. Jonson, and B. I. Lundqvist, Phys. Lett. **59A**, 177 (1976).

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

[51]O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B **13**, 4274 (1976).

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

[53]F. Manghi, G. Riegler, C. M. Bertoni, C. Calandra, and G. B. Bachelet, Phys. Rev. B **28**, 6157 (1983).

[54]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **30**, 5777 (1984).

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