First-Principles Theory of Quasiparticles: Calculation of Band Gaps in Semiconductors and Insulators

Mark S. Hybertsen and Steven G. Louie

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

(Received 15 April 1985)

We present a first-principles theory for the quasiparticle energies of semiconductors and insulators. The full dielectric matrix is used to evaluate the nonlocal, energy-dependent electron self-energy operator. Both local-field effects and dynamical screening are found to be essential for understanding quasiparticle energies. Results for the band gaps, optical transitions, and band dispersions for silicon and diamond are in excellent agreement with existing experimental data.

The problem of first-principles calculation of quasiparticle properties and optical constants of semiconductors and insulators is of longstanding interest. Its solution is of major importance both from the point of view of many-body theory and for practical studies of excited-state properties of solids. A major difficulty in the theory stems from the necessity of treating the dynamical interactions between electrons in a solid with an energy gap and significant charge inhomogeneity. The density-functional formalism in principle provides an exact formulation for ground-state properties,[1] and has had considerable success within the local-density approximation (LDA)[2] for many properties. Unfortunately, the eigenvalues from the effective one-electron equations in the density-functional theory are not formally interpretable as quasiparticle energies. Indeed, the failure of that interpretation for the gap in semiconductors and insulators is well documented; discrepancies with experiment are typically 30%–50%.[2] An explicit correction to the Kohn-Sham minimum gap has been proposed and calculated for simple models of semiconductors.[3] In addition, prompted by the failure of the one-electron theory to describe the gap, recent work[4,5] has focused on the calculation of the quasiparticle energies from the electron self-energy operator Σ.

In this Letter, we present a first-principles theory of the quasiparticle energies in semiconductors and insulators. The theory is based on evaluation of $\Sigma(\mathbf{r}, \mathbf{r}'; E)$ to first order in the dressed Green's function and the screened Coulomb interaction: the GW approximation.[6] We show that three critical elements are required for a quantitative theory for semiconductors and insulators: proper account of the nonlocality of Σ, inclusion of the full dielectric matrix in the screening of the bare Coulomb interaction, and adequate treatment of dynamical effects in the screening. We have applied the theory to the prototypical cases of silicon and diamond where relativistic effects are negligible and the cores are small. Results for the band gaps, optical transitions, and band dispersions are all within a few percent of the experimental values. One of the

crucial factors in this result is the effect of local fields, which can be evaluated through the screening from the off-diagonal elements of the dielectric matrix. This screening deepens the Coulomb-hole contribution to the electron self-energy for states with wave functions at regions of concentrated charge density as found in the bond chain.

To examine the role of local-field effects and of dynamical screening, we present the calculation in three stages starting with the Coulomb-hole–screened-exchange (COHSEX) approximation to Σ which essentially leaves out the effects of dynamical screening.[6] The trends are summarized in Fig. 1 for the indirect gap E_g and the E_1 optical transition $(L_{3'_v} \to L_{1c})$ of Si as compared to experiment.[7] Displayed are calculations using the LDA eigenvalues, the COHSEX approximation for Σ *without* local-field (LF) effects, the COHSEX approximation with the full
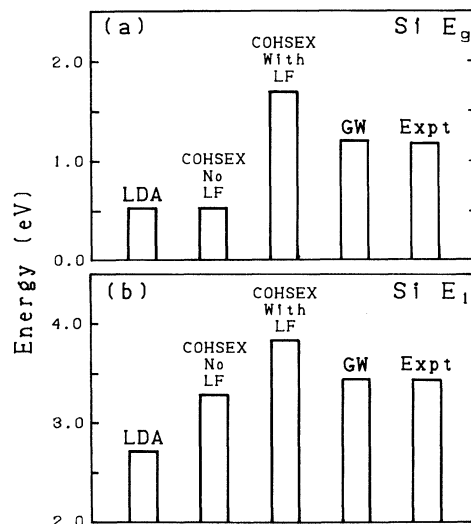

FIG. 1. Bar graphs comparing several calculations of (a) the minimum gap E_g and (b) the E_1 optical transition in Si to experimental results (Ref. 7). The calculations shown are discussed in the text.

dielectric matrix, and the full GW approximation for Σ as described below. Figure 1 demonstrates the importance of local fields and illustrates the excellent results from the full dynamical calculation.

The quasiparticle energies and wave functions are obtained by solving

$$(T + V_{\text{ext}} + V_H)\psi_{n\mathbf{k}}(\mathbf{r}) + \int d^3r' \, \Sigma(\mathbf{r}, \mathbf{r}'; E_{n\mathbf{k}})\psi_{n\mathbf{k}}(\mathbf{r}') = E_{n\mathbf{k}}\psi_{n\mathbf{k}}(\mathbf{r}). \tag{1}$$

The GW approximation for Σ is[6]

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

where δ is a positive infinitesimal. The dressed Green's function is then given in terms of the quasiparticle wave functions and energies by

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_{n\mathbf{k}} \frac{\psi_{n\mathbf{k}}(\mathbf{r})\psi_{n\mathbf{k}}^*(\mathbf{r}')}{E - E_{n\mathbf{k}} - i\delta_{n\mathbf{k}}}, \tag{3}$$

and W is the screened Coulomb interaction given by

$$W(\mathbf{r}, \mathbf{r}'; \omega)$$
$$= \Omega^{-1} \int d^3r'' \, \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; \omega) V_C(\mathbf{r}'' - \mathbf{r}'), \tag{4}$$

where ϵ^{-1} is the full time-ordered dielectric matrix. In principle, Σ, G, and W are coupled together with a vertex function Γ by a set of integral equations. We approximate $\Gamma = 1$ and, consistent with that, use the random-phase approximation (RPA) to obtain ϵ^{-1}. We propose good initial approximations for G and W and proceed to evaluate the quasiparticle energies.

Compared to 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 well as the full crystalline Green's function. In the present approach, $G(\mathbf{r}, \mathbf{r}'; E)$ is constructed with use of wave functions and band energies from a LDA calculation. The use of an energy-independent potential at this stage greatly simplifies the calculation and, as previously noted, is a good approximation.[6] The screened Coulomb interaction in Eq. (4) requires the full dielectric matrix. The static dielectric matrix, being a ground-state property, is obtainable within the density-functional theory.[2] We use a LDA band calculation with the standard Adler-Wiser formulation of the dielectric matrix in reciprocal space[8] to obtain $\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}; \omega = 0)$ from first principles in the RPA. We found that inclusion of exchange-correlation effects in the dielectric matrix within the LDA alters the final results by < 0.1 eV.

To extend the dielectric matrix to finite frequencies, we propose a generalized plasmon-pole model since $\text{Im}\,\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}, \omega)$ is generally a peaked function in ω. For each set of momentum components $(\mathbf{q}, \mathbf{G}, \mathbf{G}')$, $\text{Im}\,\epsilon^{-1}$ is taken to be

$$\text{Im}\,\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}; \omega) = A_{\mathbf{G}\mathbf{G}'}(\mathbf{q})[\delta(\omega - \tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})) - \delta(\omega + \tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}))]. \tag{5}$$

The full ω-dependent dielectric matrix can then be obtained once the matrices A and $\tilde{\omega}$ are determined. This is done exactly by use of the ab initio static dielectric matrices calculated above with the Kramers-Kronig relation,

$$\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}; 0)$$
$$= \delta_{\mathbf{G}\mathbf{G}'} + (2/\pi)\text{P}\int_0^\infty d\omega \, \omega^{-1} \text{Im}\,\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}; \omega), \tag{6}$$

and the Johnson sum rules,[9]

$$\int_0^\infty d\omega \, \omega \, \text{Im}\,\epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}; \omega)$$
$$= -\frac{\pi}{2}\omega_p^2 \frac{\rho(\mathbf{G} - \mathbf{G}')}{\rho(0)} \frac{(\mathbf{q} + \mathbf{G}) \cdot (\mathbf{q} + \mathbf{G}')}{|\mathbf{q} + \mathbf{G}|^2}, \tag{7}$$

where ω_p is the plasma frequency and $\rho(\mathbf{G})$ are the Fourier components of the crystalline charge density. There are no adjustable parameters. Comparisons to realistic calculations[10] of $\epsilon^{-1}(\mathbf{q}; \omega)$ are favorable over the important frequency range for Σ ($\omega \lesssim \tilde{\omega}$). The model moreover reproduces the ω and ω^{-1} moments

of the exact response function.[11]

All of our calculations are based on the ab initio pseudopotential[12] band-structure method carried out in a plane-wave basis. The basis gives full convergence for both Si and diamond with a plane-wave energy cutoff of 17 and 50 Ry, respectively. We use dielectric matrices of size (depending on $\mathbf{q}$) 140×140 for Si and 220×220 for diamond and note that they agree well with direct calculations performed as recently proposed.[13] Equation (1) is solved by expansion of the quasiparticle wave function in the basis of the LDA wave functions. We find a posteriori that the quasiparticle wave function is well represented by the LDA wave function; i.e., the mixing of a given basis function with others affects the quasiparticle energies by < 0.05 eV for Si and by < 0.1 eV for diamond. For the full dynamical calculation, the frequency integration required in Eq. (2) is easily done within the generalized plasmon-pole model. The self-energy operator in Eq. (1) must be evaluated at the quasiparticle energy, $E_{n\mathbf{k}}$. We have further iterated to include the new spec-
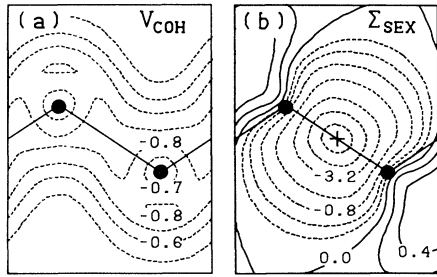
FIG. 2. Plots in the $(1\bar{1}0)$ plane of Si of the two parts of the electron self-energy operator calculated in the COHSEX approximation including the *full* dielectric matrix. The bond chain is indicated schematically. (a) The Coulomb-hole potential, $V_{\text{COH}}(\mathbf{r})$, with contour interval of 0.1 Ry. (b) The screened-exchange part $|\mathbf{r} - \mathbf{r}'|\Sigma_{\text{SEX}}(\mathbf{r}, \mathbf{r}')$, where $\mathbf{r}'$ is fixed in the bond center as indicated by the cross. The contours are chosen to increase by powers of 2 from 0.2 Ry-a.u./cell.

trum in the Green's function and found only small changes in the quasiparticle energies. Numerical results given here are estimated to have converged to better than 0.1 eV for Si and 0.2 eV for diamond.

The self-energy operator naturally breaks into two terms: Σ_{SEX} from the poles of G in the frequency integration in Eq. (2) which gives a dynamically screened exchange (SEX) contribution, and Σ_{COH} from the poles in W which gives a Coulomb-hole (COH) contribution. Alternatively, inclusion of correlation effects modifies bare exchange by the effect of virtual creation of electron-hole pairs and plasmons. The COHSEX approximation of Hedin[6] is obtained by taking the limit of zero exchange frequency. In this case the COH term is simply interpreted as the induced potential at the quasiparticle location $\mathbf{r}$ from the polarization of the surrounding electrons by that particle at $\mathbf{r}$. In this approximation, the COH term is a *local potential:* $\Sigma_{\text{COH}}(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') V_{\text{COH}}(\mathbf{r})$. Figure 2(a) shows V_{COH} in the $(1\bar{1}0)$ plane of Si. If only the diagonal terms in ϵ^{-1} are included or equivalently local fields are neglected and the screening in the solid is treated as isotropic and uniform, the result for the COH potential is uniform. The COH term has no influence on the gaps when local fields are neglected in this approximation. Within the COHSEX approximation the SEX term reduces to a static-screened-exchange operator $\Sigma_{\text{SEX}}(\mathbf{r}, \mathbf{r}')$, which is plotted in Fig. 2(b), for the case where $\mathbf{r}'$ is fixed in the center of a bond. The operator is manifestly nonlocal with its first node at roughly one bond length and significant lobes at larger distance.

Referring to Fig. 1, we see that the COHSEX approximation without local fields gives a small improvement as compared to the usual LDA eigenvalues. For the case of diamond the results are somewhat better, in agreement with previous calculation.[14] We under-

stand the improvement to arise from proper treatment of the nonlocality of the SEX term (i.e., the crystalline Green's function). Inclusion of the full static dielectric matrix gives a large change, especially for the indirect gap and bandwidth. This is largely due to the COH term illustrated in Fig. 2(a). Local fields describe the differences in polarizability at the various points in the unit cell. With the concentration of bonding charge in semiconductors, this is demonstrably an important effect. Finally, inclusion of dynamical effects reduces the gaps to good agreement with experiment. We understand the last effect as follows. Making the zero-exchange-frequency approximation for the matrix elements of Σ (the COHSEX approximation) overestimates the effect of virtual plasmon creation.

Detailed results comparing the GW calculation to experiment[7, 15, 16] are given in Table I for the best characterized transitions and valence-band features. The overall agreement is excellent. In particular we note that for Si, in addition to the features described in Fig. 1, the second indirect edge at L is well reproduced, although there is some disagreement between the optical measurement and the inverse-photoemission result. Results for the E_2 $(L_{3'_v} \rightarrow L_{3c})$ feature are excellent. In general, although the band gap is unambiguous, interpretation of the optical spectrum in terms of the quasiparticle energies can be subtle because of further many-body effects, i.e., the electron-hole interaction. The diamond results are also in good agreement

TABLE I. Comparison of results (in electronvolts) from the present calculation of the quasiparticle energies to the results of the LDA eigenvalues and to experiment.

	LDA	Present work	Expt.[a]
		Silicon	
E_g	0.52	1.21	1.17
$\Gamma_{1v} \rightarrow \Gamma_{25'v}$	11.93	11.84	12.5 ± 0.6
$\Gamma_{25'v} \rightarrow \Gamma_{15c}$	2.57	3.27	3.40
$L_{3'v} \rightarrow \Gamma_{25'v}$	1.21	1.26	1.2 ± 0.2, 1.5^b
$\Gamma_{25'v} \rightarrow L_{1c}$	1.51	2.18	$2.1,^c$ 2.4^b
$L_{3'v} \rightarrow L_{1c}$	2.73	3.44	3.45
$L_{3'v} \rightarrow L_{3c}$	4.58	5.40	5.50
		Diamond	
E_g	3.90	5.43	5.48
$\Gamma_{1v} \rightarrow \Gamma_{25'v}$	21.65	22.73	24.2 ± 1
$\Gamma_{25'v} \rightarrow \Gamma_{15c}$	5.51	7.38	7.3
$X_{4v} \rightarrow X_{1c}$	10.84	12.71	12.5

[a]Reference 7 except where noted.
[b]Reference 15.
[c]Reference 16.

with experiment.

The application of the electron self-energy to the understanding of excited-state properties of real materials has had a long history. Phillips proposed a screened-exchange approach based on a generalized Koopman's theorem and, with Kleinman, made estimates for Si.[17] Early calculations were done by Brinkman and Goodman[18] and by Kane[19] for Si and by Brener[14] for diamond using the COHSEX approximation and neglecting local-field effects. As is evident from Fig. 1, the neglect of local-field effects and dynamic screening compensates to some degree, leading to reasonable results in some cases. Wang and Pickett[4] use a LDA for the electron self-energy operator, finding systematic improvement over the LDA gaps. The work of Strinati, Mattausch, and Hanke[5] using a minimal-basis tight-binding approach is similar to the present calculation as regards the Green's-function formalism, although they go beyond the RPA in the dielectric matrix. Their calculated gaps in Si are consistently too large. The work of Horsch, Horsch, and Fulde,[20] using a local correlation approach, is a variational calculation. Our results for the direct gap in diamond agree with those of Strinati, Mattausch, and Hanke and Horsch, Horsch, and Fulde but other details of the quasiparticle energies differ substantially, e.g., for the direct gap at the zone edge noted in Table I. Sterne and Inkson[21] using an extreme tight-binding (flat band) model conclude that the Coulomb-hole contribution to the band-gap correction is negligible, in contrast to the present work.

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

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

[3]L. J. Sham and M. Schlüter, Phys. Rev. Lett. 51, 1888 (1983); J. P. Perdew and M. Levy, Phys. Rev. Lett. 51, 1884 (1983).

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

[5]G. Strinati, H. J. Mattausch, and W. Hanke, Solid State Commun. 51, 23 (1984), and references therein.

[6]L. Hedin, Phys. Rev. 139, A796 (1965); L. Hedin and S. Lundqvist, Solid State Phys. 23, 1 (1969).

[7]Landolt-Börnstein: Zahlenwerte und Funktionen aus Naturwissenschaften und Technik, edited by O. Madelung (Springer, New York, 1982), Group 3, Vol. 17, Part a.

[8]S. L. Adler, Phys. Rev. 126, 413 (1962); N. Wiser, Phys. Rev. 129, 62 (1963).

[9]D. L. Johnson, Phys. Rev. B 9, 4475 (1974).

[10]J. P. Walter and M. L. Cohen, Phys. Rev. B 5, 3101 (1972); S. G. Louie and M. L. Cohen, unpublished.

[11]A similar plasmon-pole model is used in a two–plane-wave model semiconductor calculation of M. Lannoo, M. Schlüter, and L. J. Sham (private communication). They used numerically calculated ω moments of ϵ instead of the Johnson sum rules.

[12]G. P. Kerker, J. Phys. C 13, L189 (1980).

[13]M. S. Hybertsen and S. G. Louie, unpublished; K. Kunc and E. Tosatti, Phys. Rev. B 29, 7045 (1984); R. Resta and A. Fleszar, Phys. Rev. B 31, 5305 (1985).

[14]N. Brener, Phys. Rev. B 11, 929 (1975).

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

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

[17]J. C. Phillips, Phys. Rev. 123, 420 (1961); J. C. Phillips and L. Kleinman, Phys. Rev. 128, 2098 (1962).

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

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

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

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