Optical properties of semiconductors within the independent-quasiparticle approximation

R. Del Sole

Dipartimento di Fisica, Università di Roma "Tor Vergata," via della Ricerca Scientifica 1, I-00133 Roma, Italy

Raffaello Girlanda

Istituto di Struttura della Materia, Università di Messina, Salita Sperone 31, Casella Postale 57, I-98166 Sant'Agata-Messina, Italy

We develop a gauge-invariant formalism for calculating the optical properties of solids within the random-phase approximation and the GW approach to electron bands (the independent-quasiparticle approximation). We show that, within the scissors-operator approximation, it is enough to shift rigidly the optical spectrum, calculated according to the density-functional theory–local-density approximation, to higher energies by the gap correction. Moreover, a simple formula is given valid beyond the scissors-operator approximation.

I. INTRODUCTION

The *ab initio* calculation of the optical properties of solids is still an open problem. It is usually carried out within the local-density approximation (LDA) to the density-functional theory, which, however, underestimates the band gaps, and therefore transition energies, by about 0.5 or 1 eV in semiconductors, leading to redshifted spectra.[1] Semiempirical methods are devised in order to avoid this shortcoming, as using the Slater exchange[2] or other *ad hoc* potentials[3] in order to get better values for the gaps, or shifting the spectra by adjustable values to higher energies.[4] The use of electron bands calculated within the many-body GW formalism[5,6] is required in principle, together with the independent-electron approximation[7] [often called the random-phase approximation (RPA)] for the sake of simplicity. However this approach leads to additional problems. As a consequence of the spatial nonlocality of the self-energy operator, different results are obtained in the length and velocity gauges, namely in the calculation of the response to an external longitudinal or transverse perturbation, even in the long-wavelength limit, if the standard expression $\mathbf{p}/m$ is used for the velocity operator.[8,9]

In this paper we will show how to solve this problem. The gauge invariance is preserved by properly defining the velocity operator, according to the approach developed by Piparo and Girlanda[10] for the case of nonlocal potentials. Here we specialize such an approach to the case of the nonlocal self-energy of the GW approximation and carry out a tight-binding calculation of Si optical properties.

The velocity operator must be defined, in the Heisenberg representation, as the time derivative of the position operator, in terms of the quasiparticle (QP) time evolution operator. If the full many-body Hamiltonian is considered, the usual definition $\mathbf{p}/m$, where $\mathbf{p}$ is the momentum operator, is correct. If, however, the effective single-QP GW Hamiltonian is used, as in the RPA, a different definition of velocity should be used, in order to ensure charge-current conservation and therefore gauge

invariance. In this context, and neglecting the dynamical effects on the self-energy, a simple recipe is found, within the scissors-operator approximation, for computing optical spectra: it is enough to shift the calculated LDA spectrum to higher energies by the scissors operator, to get the RPA-GW spectrum.

A series of papers by Levine and Allan[11] have appeared in recent years, which show how to calculate, in the length gauge, the linear and nonlinear optical properties of solids within the GW scissors-operator approximation. An extension to the velocity gauge has been done for calculating the optical activity of the noncubic semiconductor selenium.[12] Our formulation is equivalent to that of Levine and Allan, when the exchange-correlation term K_{xc}, is there neglected; however, it is simpler and physically more transparent. Both approaches do not take into account dynamical effects in the self-energy, namely, its E dependence. Our formulation can also be applied to calculate $\varepsilon_2(\omega)$ without invoking the scissors-operator approximation. The ingredients of this calculation are LDA matrix elements and energies; only for the latter ones are GW values also needed.

Section II is devoted to the introduction of the gauge invariance in the framework of the independent-QP GW Hamiltonian. The formulas for calculating $\varepsilon_2(\omega)$ are derived in Sec. III. In Sec. IV we illustrate our findings by calculating the optical properties of Si within a tight-binding model approach. In Sec. V we introduce the scissors-operator approximation in our formulation and give simple recipes to calculate the GW optical properties using LDA results. A brief summary and final remarks are given in Sec. VI.

II. GAUGE INVARIANCE WITHIN THE RPA

An incident electromagnetic field is a transverse perturbation, so that the optical properties must be calculated using the transverse-response formalism, with the $\mathbf{A} \cdot \mathbf{p}$ perturbation (the velocity gauge). Since, however, the light wave vector is vanishingly small, the electric field can be equally well represented by a longitudinal pertur-

bation ($\mathbf{E} \cdot \mathbf{r}$, the length gauge): it can be shown that, in the limit $q \to 0$, both approaches yield the same result.[13] This property is referred to as the gauge invariance of the optical properties.

In the velocity gauge, the induced current is calculated as the response to a transverse perturbation; on the other hand, in the length gauge, the induced charge is calculated as the response to a perturbing potential. It is clear that the equivalence of the two approaches relies on the charge conservation equation

$$\text{div}\,\mathbf{j}(\mathbf{r},t) + \partial \rho(\mathbf{r},t)/\partial t = 0 , \tag{1}$$

which, in terms of spatial Fourier transforms, reads as

$$i\mathbf{k} \cdot \mathbf{j}(\mathbf{k},t) = -\partial \rho(\mathbf{k},t)/\partial t . \tag{1a}$$

In order to fulfill charge conservation, and therefore gauge invariance, it is enough that Eq. (1a) is fulfilled by current- and charge-density operators in the Heisenberg representation. By writing the charge- and current-density operators, respectively, as

$$\rho(\mathbf{k}) = \Sigma_i \exp(-i\mathbf{k} \cdot \mathbf{r}_i) \tag{2}$$

and

$$\mathbf{j}(\mathbf{k}) = \Sigma_i [\mathbf{v}_i \exp(-i\mathbf{k} \cdot \mathbf{r}_i) + \exp(-i\mathbf{k} \cdot \mathbf{r}_i)\mathbf{v}_i]/2 , \tag{3}$$

where i labels the electrons of the solid, and $\mathbf{r}_i$ and $\mathbf{v}_i$ are, respectively, the position and velocity operators of the i electron in the Heisenberg representation (the t dependence is understood), Eq. (1a) reads as

$$\Sigma_i [\mathbf{k} \cdot \mathbf{v}_i \exp(-i\mathbf{k} \cdot \mathbf{r}_i) + \exp(-i\mathbf{k} \cdot \mathbf{r}_i)\mathbf{k} \cdot \mathbf{v}_i]/2$$
$$= \Sigma_i [\exp(-i\mathbf{k} \cdot \mathbf{r}_i), H] . \tag{4}$$

If the Hamiltonian contains only local potentials, the commutator in the right-hand side of (4) can be expressed in terms of the momentum operator $\mathbf{p}_i$, and Eq. (4) is verified with the usual choice

$$\mathbf{v}_i = \mathbf{p}_i/m . \tag{5}$$

Problems arise when the total many-body Hamiltonian is approximated as a sum of effective single-QP Hamiltonians H_i,[14] which may contain, as it actually occurs within the GW approximation, a nonlocal self-energy. In this case the definition (5) of the velocity operator does not fulfill Eq. (4), even for $k \to 0$. The correct definition, which ensures the fulfillment of (4), is

$$\mathbf{v}_i = d\mathbf{r}_i/dt = i[H_i, \mathbf{r}_i]/\hbar . \tag{5a}$$

This can be easily checked by expressing the right-hand side of (4) in terms of the time derivative of $\rho(\mathbf{k})$:

$$\Sigma_i[\exp(-i\mathbf{k} \cdot \mathbf{r}_i), H] = i(\partial/\partial t)\Sigma_i\exp(-i\mathbf{k} \cdot \mathbf{r}_i)$$
$$= \Sigma_i \{ (\mathbf{k} \cdot d\mathbf{r}_i/dt)\exp(-i\mathbf{k} \cdot \mathbf{r}_i)$$
$$+ \exp(-i\mathbf{k} \cdot \mathbf{r}_i)(\mathbf{k} \cdot d\mathbf{r}_i/dt)\}/2 \tag{4'}$$

and using (5a).

In the presence of a nonlocal self-energy, (5a) yields

$$\mathbf{v}_i = \mathbf{p}_i/m + \partial \Sigma(\mathbf{r}_i, \mathbf{p}_i)/\partial \mathbf{p}_i , \tag{6}$$

with

$$\Sigma(\mathbf{r},\mathbf{p}) = \int d^3\mathbf{r}'\Sigma(\mathbf{r},\mathbf{r}')\exp[i(\mathbf{r}'-\mathbf{r})\cdot\mathbf{p}/\hbar] \tag{7}$$

related to the nonlocal self-energy $\Sigma(\mathbf{r},\mathbf{r}')$. (The index i will be dropped from now on, since we will deal only with one-electron quantities.)

The definition of velocity (5a) and its expression (6) ensure the charge and current conservation in the independent-QP approximation. When calculating the response to a transverse field in the velocity gauge, they must be used, instead of (5), both in the perturbation $H_{\text{int}} = -(e/2c)(\mathbf{A} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{A})$ (where $\mathbf{A}$ is the vector potential and $\mathbf{v}$ is the one-electron velocity operator) and in the definition of the induced current.

III. CALCULATION OF THE DIELECTRIC FUNCTION

The imaginary part of the dielectric function $\varepsilon(q,\omega)$, for $q = 0$, can be obtained, using the interaction Hamiltonian discussed in Sec. II, according to the golden rule:[15]

$$\varepsilon_2(\omega) = (4\pi^2 e^2/\omega^2)\Sigma_{vc}\Sigma_\mathbf{k} \left| \int d^3\mathbf{r}\,\Psi_{c,\mathbf{k}}^*(\mathbf{r})(\hat{\mathbf{e}} \cdot \mathbf{v})\Psi_{v,\mathbf{k}}(\mathbf{r}) \right|^2$$
$$\times \delta[E_c(\mathbf{k}) - E_v(\mathbf{k}) - \hbar\omega] , \tag{8}$$

where $\Psi_{v,\mathbf{k}}$ and $\Psi_{c,\mathbf{k}}$ are valence- and conduction-band QP wave functions, of energy $E_v(\mathbf{k})$ and $E_c(\mathbf{k})$, respectively, and $\hat{\mathbf{e}}$ is a unit vector indicating the polarization direction. This is the formula to be used to calculate the optical properties in the presence of nonlocal self-energy or pseudopotential; it contains, instead of $\mathbf{p}$ of the usual formula,[15] $m\mathbf{v}$, where the velocity operator $\mathbf{v}$ is given by (6). We do not explicitly write the occupation numbers in (8), since we restrict ourselves to the case of nonmetals at absolute zero.

To prove the gauge invariance of $\varepsilon_2(\omega)$ given by (8) we have to write the velocity operator in a more convenient form:

$$v_\alpha = \lim_{q \to 0}[H, e^{iqr_\alpha}]/\hbar q , \tag{9}$$

where v_α and r_α ($\alpha = x,y,z$) are Cartesian components of $\mathbf{v}$ and $\mathbf{r}$, respectively, and H is the QP Hamiltonian. Equation (9) is the generalization to nonlocal potentials and to periodic boundary conditions of the relation

$$\mathbf{p}/m = i[H,\mathbf{r}]/\hbar . \tag{10}$$

Since H_{LDA} contains only local potentials, we can write

$$p_\alpha/m = \lim_{q \to 0}[H_{\text{LDA}}, e^{iqr_\alpha}]/\hbar q . \tag{11}$$

(We assume for the sake of simplicity that the pseudopotential is local.)

Putting (9) into (8), we get

$$\varepsilon_2(\omega) = 4\pi^2 e^2 \lim_{q \to 0} (1/q^2) \Sigma_{vc} \Sigma_{\mathbf{k}}$$

$$\times \left| \int d^3 r \, \Psi^*_{c,\mathbf{k}+\mathbf{q}}(\mathbf{r}) \exp(i\mathbf{q} \cdot \mathbf{r}) \Psi_{v,\mathbf{k}}(\mathbf{r}) \right|^2$$

$$\times \delta[E_c(\mathbf{k}+\mathbf{q}) - E_v(\mathbf{k}) - \hbar\omega] . \tag{12}$$

This formula can be derived independently as the response to a longitudinal field for $q \to 0$, and can be found in many textbooks.[16] [No problem related to the nonlocal potential arises in the Coulomb gauge,[9] so that Eq. (12) remains valid also in the case of a nonlocal self-energy.] It is equivalent to the imaginary part of Eq. (17a) in Ref. 11(b). It demonstrates the gauge invariance of the optical properties stated in Sec. II. Such invariance is valid in general, beyond the independent-quasiparticle approximation, as shown by Ambegoakar and Kohn[13(a)] for cubic crystals and by Del Sole and Fiorino[13(b)] for less symmetric systems. In crystals of less-than-cubic symmetry, the $q \to 0$ limit in (12) may depend on the direction of $\mathbf{q}$. If $\mathbf{q}$ is chosen along the α-principal axis of the dielectric tensor, the $\alpha\alpha$ component of the latter is obtained, according to (11). Therefore all components are obtained by suitable choices of the $\mathbf{q}$ direction. In cubic crystals the three components are of course equal, so that the dielectric tensor becomes a scalar, and any direction of $\mathbf{q}$ yields the same limit. In the following we will consider only cubic crystals.

Taking into account that the GW wave functions are nearly identical to the LDA ones,[5,11] we can safely use in (12) the LDA wave functions $\Psi^0_{v,c}(\mathbf{r})$ instead of GW ones $\Psi_{v,c}(\mathbf{r})$. This allows us to express the matrix elements in (12), using (11), in terms of the matrix elements of $\mathbf{p}$:

$$\lim_{q \to 0} \langle \Psi^0_{c,\mathbf{k}+q} \hat{e}(r) | \exp(iq\hat{e} \cdot \mathbf{r}) | \Psi^0_{v,\mathbf{k}}(\mathbf{r}) \rangle / q$$

$$= (-i/m) \langle \Psi^0_{c,\mathbf{k}}(\mathbf{r}) | \hat{e} \cdot \mathbf{p} |$$

$$\times \Psi^0_{v,\mathbf{k}}(\mathbf{r}) \rangle / [E^0_c(\mathbf{k}) - E^0_v(\mathbf{k})] , \tag{13}$$

taking advantage of the fact that the LDA Hamiltonian contains only local potentials. $E^0_c(\mathbf{k})$ and $E^0_v(\mathbf{k})$ are the LDA eigenvalues.

In this way we obtain a simple but still correct expression for $\varepsilon_2(\omega)$:

$$\varepsilon_2(\omega) = (4\pi^2 e^2 \hbar^2/m^2) \Sigma_{vc} \Sigma_{\mathbf{k}}$$

$$\times \left| \int d^3 r \, \Psi^{0*}_{c,\mathbf{k}}(\mathbf{r}) \hat{e} \cdot \mathbf{p} \Psi^0_{v,\mathbf{k}}(\mathbf{r}) \right|^2$$

$$\times \delta[E_c(\mathbf{k}) - E_v(\mathbf{k}) - \hbar\omega]/[E^0_c(\mathbf{k}) - E^0_v(\mathbf{k})]^2 . \tag{14}$$

This is the main result of the present work. It shows that it is possible to use the $\mathbf{p}$ operator to compute the optical properties of systems with nonlocal potentials, provided that (14) is used instead of the usual formula [see (16) below]. Equation (14) has the advantage, with respect to (12), of involving $\mathbf{p}$, which is diagonal in a plane-wave basis, and of avoiding carrying out the $q \to 0$ limit numerically.

Equation (14) can be put in a slightly different form

taking advantage of the conservation of energy:

$$\varepsilon_2(\omega) = (4\pi^2 e^2/\omega^2 m^2) \Sigma_{vc} \Sigma_{\mathbf{k}}$$

$$\times \left| \int d^3 r \, \Psi^{0*}_{c,\mathbf{k}}(\mathbf{r}) \hat{e} \cdot \mathbf{p} \Psi^0_{v,\mathbf{k}}(\mathbf{r}) \right|^2$$

$$\times \delta[E_c(\mathbf{k}) - E_v(\mathbf{k}) - \hbar\omega]$$

$$\times \{[E_c(\mathbf{k}) - E_v(\mathbf{k})]/[E^0_c(\mathbf{k}) - E^0_v(\mathbf{k})]\}^2 . \tag{15}$$

This can be interpreted in terms of an effective (i.e., corrected for nonlocal effects) momentum operator $\mathbf{p}^{\text{eff}}$, whose matrix elements are

$$\langle c\mathbf{k} | \mathbf{p}^{\text{eff}} | v\mathbf{k} \rangle$$

$$= \langle c\mathbf{k} | \mathbf{p} | v\mathbf{k} \rangle \{[E_c(\mathbf{k}) - E_v(\mathbf{k})]/[E^0_c(\mathbf{k}) - E^0_v(\mathbf{k})]\} .$$

[This formula is equivalent to Eq. (15) of Ref. 11(a); recall that in the state $|n\mathbf{k}\rangle$ an occupied state appears, for which $E_n(\mathbf{k}) = E^0_n(\mathbf{k})$.]

A naive approach, not concerned about nonlocality and gauge invariance, would use the usual formula for $\varepsilon_2(\omega)$:[15]

$$\varepsilon_2^N(\omega) = (4\pi^2 e^2/\omega^2 m^2) \Sigma_{vc} \Sigma_{\mathbf{k}}$$

$$\times \left| \int d^3 r \, \Psi^{0*}_{c,\mathbf{k}}(\mathbf{r}) \hat{e} \cdot \mathbf{p} \Psi^0_{v,\mathbf{k}}(\mathbf{r}) \right|^2$$

$$\times \delta(E_c(\mathbf{k}) - E_v(\mathbf{k}) - \hbar\omega) , \tag{16}$$

underestimating the absorption coefficient by a factor $\{[E_c(\mathbf{k}) - E_v(\mathbf{k})]/[E^0_c(\mathbf{k}) - E^0_v(\mathbf{k})]\}^2$. Via the Kramers-Kronig transform, this reflects in a lower value of $\varepsilon_1(0)$, as shown also by Levine and Allan.

TABLE I. Upper part: First line, LDA band energies used to determine the tight-binding parameters (Refs. 18 and 19). Second line, LDA band energies computed within tight binding. Third line, GW band energies used to determine the tight-binding parameters (Refs. 18 and 19). Fourth line, GW band energies computed within tight binding. All energies (in eV) are relative to the valence-band top. Lower part: on-site and first-nearest-neighbor tight-binding parameters used in the calculations. All energies in eV.

	Γ_1^v	X_1^v	X_5^v	X_1^c	Γ_{15}^c	$\Gamma_2^{'c}$
LDA	-11.90	-7.75	-2.82	0.67	2.56	3.38
TB	-11.90	-8.59	-2.82	0.69	2.56	3.38
GW	-11.79	-7.77	-2.99	1.46	3.34	4.16
TB	-11.79	-8.60	-2.99	1.47	3.34	4.16

Tight-binding parameters		
	LDA	GW
E_s	-4.26	-3.82
E_p	1.28	1.67
E_{s^*}	6.25	6.64
V_{ss}	-1.91	-1.99
$V_{pp\sigma}$	2.37	2.75
$V_{pp\pi}$	-0.70	-0.75
$V_{sp\sigma}$	2.43	2.65
$V_{s^*p\sigma}$	2.69	2.66

IV. ILLUSTRATIVE CALCULATIONS FOR Si

To illustrate these findings, we have carried out calculations of the optical properties of Si using a semiempirical tight-binding (TB) method, with an sp^3s^* basis and first-neighbor interactions.[17] We start from LDA and GW bands taken from the literature.[18,19] We generate a set of TB parameters for each of the two sets of bands, using the band energies at Γ and X, according to the procedure of Ref. 17, as inputs to determine TB parameters (Table I). The quality of the fits is the same as in Ref. 17. Then we carry out calculations of $\varepsilon_2(\omega)$ within GW, namely, according to (8) or to (12). (We have checked that they give the same result. This confirms *a posteriori* the gauge invariance of the present formulation.) The calculation of the optical matrix elements [of the velocity operator (9) when using (8), of $\exp(i\mathbf{q\cdot r})$ when using (12)] between localized orbitals has been carried out as in Ref. 20.

The resulting $\varepsilon_2(\omega)$ is shown in Fig. 1. It shows a two-peak structure, similar to the experimental $\varepsilon_2(\omega)$ (see Fig. 2), but with a substantially reduced intensity. The lower-energy peak E_1 is due to transitions from the highest valence to the lowest conduction band along the Λ line, where the bands are approximately parallel. The higher-energy peak E_2 is due to transitions from X_{5v} to X_{1c} (see Table I). A static dielectric constant $\varepsilon_1(0)=8.5$ is obtained via the Kramers-Kronig transform. The discrepancy with respect to the experimental value 11.4 (Ref. 21) must be mostly ascribed to the intrinsic limitations of the empirical TB method [for instance, $\varepsilon_1(0)=7.6$ was obtained in the tight-binding calculation of Ref. 22]. We have carried out the calculation again using LDA wave functions. The result, also shown in Fig. 1, is nearly identical to the full GW result, yielding
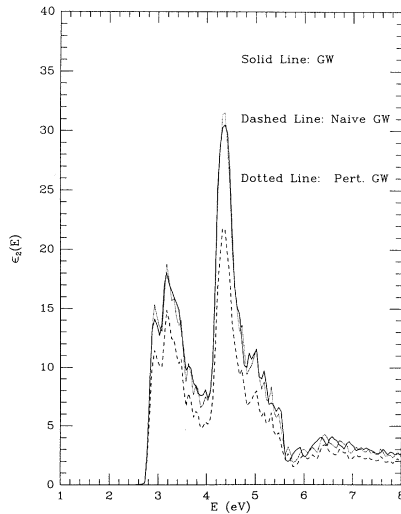

FIG. 1. The imaginary part of the dielectric constant of Si, calculated within tight binding, vs photon energy. Solid line: GW approximation, according to Eq. (8) or (12). Dotted line: the same, but using LDA wave functions. Dashed line: naive GW, according to Eq. (16).
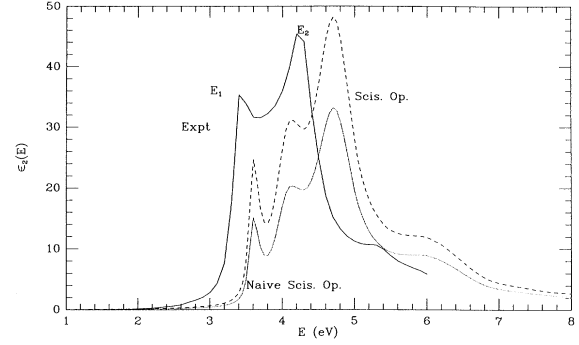

FIG. 2. The imaginary part of the dielectric constant of Si vs photon energy. Solid line: experimental data from Ref. 30. Dashed line: scissors-operator approximation, obtained according to Eq. (19), using the LDA calculation of Ref. 26 with a broadening of 50 meV. Dotted line: naive scissors-operator approximation, $\varepsilon_2^N(\omega)$, obtained from the LDA calculation of Ref. 26 according to Eqs. (18) and (19).

$\varepsilon_1(0)=8.4$. This shows the reliability of LDA wave functions for GW calculations. The naive GW calculation, according to (16), is also shown in Fig. 1. A substantial amount of oscillator strength is lost due to the incorrect calculation of matrix elements. The static dielectric constant is substantially decreased, to 6.6.

V. SCISSORS-OPERATOR APPROXIMATION

The relations between the GW, naive GW, and LDA dielectric constants become very simple when the GW conduction bands are rigidly shifted by an amount Δ with respect to the LDA conduction bands, leaving the wave functions unchanged (scissors-operator approximation[6,23]). The similarity between GW and LDA wave functions[5] and that between the lower-gap corrections over the Brillouin zone,[24] suggest the scissors operator as a quite realistic model. In this case, (14) becomes

$$\varepsilon_2(\omega)= [4\pi^2 e^2/m^2(\omega-\Delta/\hbar)^2]\Sigma_{vc}\Sigma_{\mathbf{k}}$$

$$\times \left| \int d^3r\, \Psi_{c,\mathbf{k}}^{0*}(\mathbf{r})\hat{\mathbf{e}}\cdot\mathbf{p}\Psi_{v,\mathbf{k}}^0(\mathbf{r}) \right|^2$$

$$\times\delta[E_c(\mathbf{k}-E_v(\mathbf{k})-\hbar\omega] \,, \qquad (17)$$

and the relation with the naive GW dielectric constant is

$$\varepsilon_2(\omega)=[\hbar\omega/(\hbar\omega-\Delta)]^2\varepsilon_2^N(\omega) \,. \qquad (18)$$

Furthermore, the comparison of (17) with its LDA counterpart, which is obtained neglecting Δ in the denominator of (17) and replacing $E_c(\mathbf{k})-E_v(\mathbf{k})$ by $E_c^0(\mathbf{k})-E_v^0(\mathbf{k})=E_c(\mathbf{k})-E_v(\mathbf{k})-\Delta$, simply yields

$$\varepsilon_2(\omega)=\varepsilon_2^{LDA}(\omega-\Delta/\hbar) \,. \qquad (19)$$

Equation (19) shows that it is sufficient to shift the LDA $\varepsilon_2(\omega)$ by $\Delta/\hbar$ along the frequency axis to get the correct result within the GW scissors-operator approximation. Although this might seem a trivial result, it is not, in view of the problems generated by the nonlocality of the self-energy. Furthermore, since (19) can be also derived

within the Coulomb gauge, as the response to a longitudinal field, it is valid even if the LDA wave functions are computed using nonlocal pseudopotentials.

Equation (19) allows one to compute the RPA dielectric constant by carrying out only LDA calculations, provided the magnitude of the scissors operator is known. Since the f-sum rule is respected within the LDA (if a local pseudopotential is used), it cannot be valid any longer after a rigid shift of the spectrum to higher energies. [We overestimate by 9% the f-sum rule in the calculation described below for Si, in fair agreement with the 12% overestimate of Ref. 11(b).[25]] This is an example of the breaking down of the f-sum rule in the presence of a nonlocal potential. The f-sum rule must be satisfied indeed if the complete many-body Hamiltonian is considered, or within a one-electron approximation with a local potential. Its validity is not guaranteed in the presence of a nonlocal potential, as it is the self-energy of the GW approximation.[11(b),26]

Some comments on the validity of other methods of calculating the optical properties are worth mentioning. The semiempirical tight-binding method is meant to reproduce the real bands, i.e., GW bands. Therefore the explicit use of the **p** operator, according to (9), is incorrect. If, however, its matrix elements between local orbitals are computed using the commutator relation (6), as in Ref. 20, **v** rather than **p**$/m$ is actually being used, so that the results are correct. The quality of the calculation is in this case set by the intrinsic limitations of semiempirical tight binding. Another method used to compute bulk and surface optical properties is the X_α method, which uses $\alpha \approx 1$ in the exchange potential, rather than the LDA, value $\frac{2}{3}$.[2] As a matter of fact, this method yields transition energies in good agreement with experimental and GW ones with a local potential. Therefore the use of **p** is allowed. The problem, however, is how similar are its wave functions to GW ones. This question can be answered by comparing X_α^2 and LDA (Ref. 4) calculations of the imaginary part of the dielectric function carried out for GaAs and GaP. The shapes and strengths are very close to each other, the only difference being a shift to higher frequencies of the X_α absorption spectrum, very similar to the scissors-operator effect. Therefore we conclude that, as far as the X_α method well describes transition energies, its accuracy in predicting bulk optical properties is comparable to that of the scissors-operator approximation.

At this point we can produce calculations of the optical properties of semiconductors, within the scissors-operator approximation, with little effort. We choose Si as a test case. We start from the LDA calculation of $\varepsilon_2(\omega)$ carried out by Engel and Farid,[26] and shift it to higher energies by the average gap correction, which is in the range 0.7–0.9 eV for Si.[11] We determine its value appropriate to the LDA calculation of Engel and Farid as follows: we know that the main peak in $\varepsilon_2(\omega)$, E_2, shows up at about one-tenth of an eV above the $X_{5v} \rightarrow X_{1c}$ transition energy, from which it originates. Since such energy is known to be 4.6 ± 0.2 eV from photoemission experiments,[27] we require the calculated E_2 peak to be at 4.7 eV. This yields $\Delta = 0.8$ eV. The resulting curve, after the

introduction of a Lorentzian broadening of 50 meV, is shown in Fig. 2 (dashed line). The narrow structure present at 2.8 eV in the original LDA calculation, and not obtained in other calculations,[28,29,4] survives, although strongly reduced, after the introduction of the broadening, and occurs now at 3.6 eV. The next peak, at 4.1 eV in Fig. 2, is also strongly reduced by the broadening with respect to Ref. 26 (see Fig. 5 therein). The low-frequency part of the spectrum shows different line shapes depending on the methods employed in the calculation (see, for instance, the tight-binding calculation of Fig. 1). Although the calculation of Ref. 26 appears to be the most accurate available in the literature, we cannot exclude that the true RPA line shape is different from that of Fig. 2, since it is very sensitive to the ingredients of the calculation, namely, type of pseudopotential, number of special points, cutoff, etc. We do not further investigate this point, since the accuracy of LDA calculations is not the object of the present work. Also shown in Fig. 2, for comparison, is the result of the naive scissors-operator approach, computed from the correct curve by means of (18). The reduction of the oscillator strength discussed above is clearly seen also in this case. The comparison of the correct calculation with the experimental data of Aspnes and Studna,[30] the solid line in Fig. 2, shows only qualitative agreement. In particular, the theoretical E_2 peak is 0.4 eV higher in energy than the experimental one. The most likely reason for the different positions of the E_2 peak are excitonic and local-field effects.[29] The experimental E_1 structure should correspond to the two low-energy peaks of the theoretical curve; its enhancement and distortion in the experiment may result from excitonic effects, but also from a more refined LDA calculation, as discussed before.

The calculated values of the static dielectric constant are quoted in Table II. The LDA value obtained by us applying the Kramers-Kronig transform to $\varepsilon_2(\omega)$ calculated by Engel and Farid[26] is 13.4, to be compared with 13.6 of Engel and Farid[26] and with 13.8 of Levine and Allan's LDA.[11] The small discrepancy with respect to Ref. 26 is easily explained in terms of different treatments of the Kramers-Kronig transform and of our introduction of a Lorentzian broadening. Within the scissors-operator approximation, we calculated $\varepsilon_1(0) = 11.2$, to be

TABLE II. Static dielectric constant of Si. The calculated values do not include excitonic and local-field effects. The low-temperature experimental value of Ref. 21 has been quoted, since the calculations are for zero temperature.

	LDA	Scis. Op.	Naive	Expt.
This work (LDA from Ref. 4)	12.1	10.7 ($\Delta = 0.58$ eV)	8.4	
This work (LDA from Ref. 26)	13.4	11.2 ($\Delta = 0.8$ eV)	8.1	
Levine and Allan Ref. 11(b)	13.8	11.7 ($\Delta = 0.9$ eV)	8.8	
Ref. 21				11.4

compared with the results of Levine and Allan,[11] 11.7 (without exchange-correlation and local-field effects) and 11.2 (including exchange-correlation within LDA and local-field effects, not shown in Table II), and with the experimental value 11.4,[21] which, however, includes excitonic and local-field effects. The naive scissors-operator calculation yields a noticeably smaller value, $\varepsilon_1(0) = 8.1$. Starting from a less accurate LDA calculation, using a local pseudopotential,[4] we get $\varepsilon_1(0) = 12.1$ within the LDA and 10.7 after applying the appropriate scissors operator, $\Delta = 0.58$ eV. The discrepancies between the correct scissors-operator RPA calculations with respect to the experimental value (11.7 and 11.2 versus 11.4) are smaller than that between the calculations themselves (0.5). This suggests that the effects not included in the present calculation, namely dynamical, excitonic, and local-field effects, have little influence on the static dielectric constant of semiconductors.

VI. SUMMARY AND CONCLUDING REMARKS

We have developed a simple method for the calculation of the RPA dielectric constant within the quasiparticle scheme. The reasons for the failure of the naive GW method are clarified. Within the scissors-operator approximation, the imaginary part of the GW dielectric constant can be obtained by simply shifting the LDA curve. GW-RPA calculations of linear optical properties are now feasible with little effort. The non-negligible discrepancy between our calculations and experiment suggests the occurrence of important dynamical (see below) and/or many-body effects in the optical spectrum of Si, in agreement with the previous findings of Hanke and Sham.[29]

In order to go beyond the scissors-operator approximation, one can use Eq. (12), within the length gauge, or Eq. (14), within the velocity gauge. The latter choice requires, however, that a local pseudopotential is used in the LDA calculations. If this is not the case, one must use, within the velocity gauge, Eq. (8).[31]

Our method is essentially equivalent to that of Levine and Allan,[11] when exchange-correlation and local-field effects are there neglected. Such effects cancel each other to some extent,[29] so that it is a reasonable approximation to leave them out. Our main purpose is to account for the nonlocality of the self-energy, that, if neglected, can produce relevant errors in the dielectric function. The simplicity, on the other hand, is the main advantage of the present formulation, which allows us to calculate the real and imaginary parts of the dielectric function with little effort. We start from the velocity gauge, which is the natural one in the case of light, and find a gauge-invariant formulation by properly defining the velocity operator.

Dynamical effects, namely the energy dependence of the self-energy, are neglected both in our approach and that of Levine and Allen. Such effects are difficult to treat, since they make the effective one-electron GW Hamiltonian time dependent: the time-evolution operator and therefore the velocity operator become difficult to define. Moreover, since dynamical effects reduce the QP amplitudes,[5,32] one can expect that they also reduce the absorption strength. This would destroy the best agreement between GW calculations and experiments, which is verified for the overall absorption strength and for the static dielectric constant. [Only qualitative agreement is obtained indeed for line shapes and peak positions (see Fig. 2).] Therefore, we believe that it is not worth including dynamical effects alone. The present approach, which neglects dynamical effects as well as exchange-correlation and local-field effects, provides a simple zeroth-order approximation to the optical spectra of semiconductors. The neglected effects should be included altogether in a more refined theory.

[1]W. E. Pickett, Comments Solid State Phys. **12**, 57 (1986).

[2]F. Manghi, R. Del Sole, A. Selloni, and E. Molinari, Phys. Rev. B **41**, 9935 (1990).

[3]M. Alouani, L. Brey, and N. E. Christensen, Phys. Rev. B **37**, 1167 (1988).

[4]C. S. Wang and B. M. Klein, Phys. Rev. B **24**, 3417 (1981).

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

[6]R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10 159 (1988).

[7]H. Ehrenreich and M. H. Cohen, Phys. Rev. **155**, 786 (1959).

[8]A. F. Starace, Phys. Rev. A **3**, 1242 (1971).

[9]Similar problems arise in the effective-mass approximation: see, for instance, A. Quattropani and R. Girlanda, Solid State Commun. **38**, 1041 (1981); La Rivista del Nuovo Cimento **6**, 1 (1983); and in the atomic case, when nonlocal pseudopotentials are used; see, for instance, A. J. Read and R. J. Needs, Phys. Rev. B **44**, 13 071 (1991).

[10]E. Piparo and R. Girlanda, Il Nuovo Cimento **14D**, 679 (1992).

[11]Z. H. Levine and D. C. Allan, (a) Phys. Rev. Lett. **63**, 1719 (1989); (b) Phys. Rev. B **43**, 4187 (1991); (c) **42**, 3567 (1990); (d) **44**, 12 781 (1991).

[12]H. Zhong, Z. H. Levine, D. C. Allan, and J. W. Wilkins, Phys. Rev. Lett. **69**, 379 (1992).

[13](a) V. Ambegoakar and W. Kohn, Phys. Rev. **117**, 423 (1960); (b) R. Del Sole and E. Fiorino, Phys. Rev. B **29**, 4631 (1984).

[14]Although this assumption is never explicitly done, it is, however, understood when the two-particle Green's function is approximated as a product of two one-particle Green's functions. It manifestly appears in the fact that the excitation energies are taken as differences of band energies, without electron-hole interaction effects.

[15]See, for instance, F. Bassani and G. Pastori Parravicini, in *Electronic States and Optical Transitions in Solids*, edited by R. A. Ballinger (Pergamon, London, 1975).

[16]See, for instance, J. M. Ziman, *Principles of the Theory of Solids* (Cambridge University Press, Cambridge, 1964), p. 129.

[17]P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem.

Solids **44**, 365 (1983).

[18]M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985); R. W. Godby, M. Schluter, and L. J. Sham, *ibid.* **56**, 2415 (1986); R. W. Jansen and O. F. Sankey, Phys. Rev. B **36**, 6520 (1987); X. Zhu, S. Fahy, and S. G. Louie, *ibid.* **39**, 7840 (1989); C. S. Wang and B. M. Klein, *ibid.* **24**, 3393 (1981); see also Refs. 5, 6, and 23. Averages between the different values are taken.

[19]We have not found in Ref. 18 the quasiparticle shift of the state X_1^v. We have determined it by linear interpolation between the shifts of X_5^v and Γ_1^v.

[20]A. Selloni, P. Marsella, and R. Del Sole, Phys. Rev. B **33**, 8885 (1986).

[21]H. H. Li, J. Chem. Phys. Ref. Data **9**, 561 (1980).

[22]D. J. Moss, E. Ghahramani, and J. E. Sipe, Phys. Status Solidi B **164**, 587 (1991).

[23]F. Gygi and A. Baldereschi, Phys. Rev. Lett. **62**, 2160 (1989).

[24]F. Bechstedt and R. Del Sole, Phys. Rev. B **38**, 7710 (1988).

[25]The "f-sum rule" mentioned in Ref. 11 [Eq. (2.37) of Ref. 11(b)] does not have the usual meaning. In our paper the f-sum rule has the usual meaning, associated with the plasma frequency. This explains why the "f-sum rule" is satisfied in Ref. 11 within both the LDA and the scissors-operator approximation. However their effective plasma frequency, given in Eq. (2.42) of Ref. 11(b) and equivalent to our f-sum rule, is different in the two approaches. The 12% overestimate quoted in the text results from the 6% overestimate of the effective plasma frequency reported in Table II of Ref. 11(b).

[26]G. E. Engel and B. Farid, Phys. Rev. B **46**, 15 812 (1992).

[27]See Ref. 5 and Ref. 59 therein.

[28]S. G. Louie, J. R. Chelikowski, and M. L. Cohen, Phys. Rev. Lett. **34**, 155 (1975).

[29]W. Hanke and L. J. Sham, Phys. Rev. B **21**, 4656 (1980).

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

[31]The matrix elements of $\mathbf{v}$, appearing in (8), can be calculated in a plane-wave basis set as

$$\langle \mathbf{k} | \mathbf{v} | \mathbf{k}' \rangle = (\hbar \mathbf{k}/m) \delta_{\mathbf{k},\mathbf{k}'}$$

$$+ (\partial/\partial \mathbf{k} + \partial/\partial \mathbf{k}') \langle \mathbf{k} | V_{NL} + \Sigma | \mathbf{k}' \rangle / \hbar ,$$

in terms of the matrix elements of the nonlocal pseudopotential V_{NL} and of Σ. However it would be impractical to perform the numerical derivatives of the Σ matrix elements. A computationally simpler possibility is to exploit the similarity of the LDA and GW wave functions, using (13). However, if the LDA Hamiltonian contains a nonlocal potential, $\mathbf{p}$ in the right-hand sides of (13) and (14) must be replaced by $m \mathbf{v}^{LDA} = mi [H^{LDA}, \mathbf{r}]/\hbar$. The matrix elements of this operator can be calculated using the expression above, without Σ. The matrix elements of V_{NL} are known in closed form, so that the gradients can be calculated without problems. In this way one still avoids having to perform numerically the $q \rightarrow 0$ limit of (12), but loses the other advantage of using the $\mathbf{A} \cdot \mathbf{p}$ perturbation, namely, the fact that $\mathbf{p}$ is diagonal in a plane-wave basis set.

[32]X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14 142 (1991).