PHYSICAL REVIEW B VOLUME 43, NUMBER 5 15 FEBRUARY 1991-I

Quasiparticle calculation of the dielectric response of silicon and germanium

Zachary H. Levine
Department of Physics, The Ohio State University, Columbus, Ohio 43210-1106

Douglas C. Allan
Applied Process Research, SP-PR-22, Corning Incorporated, Corning, New York 14831

(Received 22 May 1990)

We calculate the dielectric function $\epsilon_1(\omega)$ of silicon and germanium for frequencies below the direct band gap at the experimental lattice constant as well as the pressure dependence of the static dielectric constant $\epsilon_1(P)$. Our theory is an extension of the local-density approximation (LDA). We include self-energy effects by adding an operator to the usual LDA Hamiltonian; i.e., $H_{\mathbf{k}}=H_{\mathbf{k}}^{\mathrm{LDA}}+\Delta_{\mathbf{k}}P_{c\mathbf{k}}$. This form leads to a Ward-identity replacement $\mathbf{p}\to(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}}^{\mathrm{LDA}})^{-1}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})\mathbf{p}$. Our theory is in agreement with experiment for $\epsilon_1(\omega)$ at the level of a few percent. For $\epsilon_1(P)$, we are in agreement with experiments for silicon and some of the conflicting experiments for germanium. A prediction is made for the pressure dependence of the dielectric function of silicon beyond the linear regime.

I. INTRODUCTION

In the past few years, a number of highly accurate local-density-approximation[1] (LDA) calculations of the dielectric constants of semiconductors have appeared,[2–5] supplementing earlier work[6–9] and recent work which makes additional geometric assumptions.[10,11] LDA overestimates the macroscopic dielectric constant by at least 10% in most cases;[12] for germanium, there is an error of 25% (Ref. 3) or more.[13] From one point of view,[11] this failure is surprising: the static dielectric response of a system is a "ground-state" property, and therefore it should be well predicted by density-functional theory, which in turn is thought to be well represented by the LDA.[14]

From a second point of view, there is no surprise. The overestimate of the dielectric constant in LDA is related to the underestimate of the optical-absorption energies by the Kohn-Sham eigenvalues.[15] To take a simple model, if there is only a single optical excitation energy E_g, then the f-sum rule and the Kramers-Kronig relations suffice to impose the relationship

$$\epsilon(\omega)=1+\frac{\omega_p^2}{E_g^2-\omega^2}\ , \tag{1.1}$$

where $\epsilon(\omega)$ is the dielectric function and ω_p is the plasma frequency. Clearly, an underestimate of E_g will lead to an overestimate of $\epsilon(\omega)$ for small ω. LDA's underestimate of band gaps plays just this role. Density-functional approaches beyond LDA such as weighted-density,[16,17] Levine-Louie,[18] or nonlocal exchange[19] improve the calculation only to a limited extent. The principal improvements must come from outside of density-functional theory. Recently, Godby, Schlüter, and Sham have presented clear evidence that the density-functional formalism is incapable of correctly predicted the band gaps; the local of a local approximation contributes an additional error in energies of about 20%.[14]

The GW approximation[20] has proved capable of predicting the optical excitation frequencies in semiconductors[14,21] and semiconductor surfaces[22,23] to within 0.1 eV. In brief, the GW approximation is the Hartree-Fock approximation with the bare Coulomb interaction between the electrons ($1/r$) replaced by the screened Coulomb interaction ($1/\epsilon r$). In addition to the computational success, tentative estimates of corrections to GW are seen to be small for these systems.[24] Happily, the quasiparticle wave functions calculated in the GW approximation[14,21] are in excellent agreement (wave-function overlaps exceeding 0.999) with the LDA wave functions.[25] This is consistent with arguments that the excited-state exchange-correlation functional differs from the ground-state functional only by a position-independent constant.[26–28]

The highly successful GW theory was flawed in that it depended crucially upon dielectric screening, yet attempts to use the improved eigenvalues of GW actually degraded the values of ϵ compared to experiment.[3] In practice, the LDA dielectric function is used or the local-density approximation and random-phase approximation (LDA/RPA) dielectric function. (In LDA/RPA, exchange-correlation corrections to the local field are neglected.) Our thesis is that previous attempts to use the GW energies were performed naively because the velocity operator was not renormalized; such renormalization yields an excellent description of the dielectric response of semiconductors.

Motivated by the GW results, we propose that the many-body effect of a quasiparticle energy shift in the optical response may be included in one-electron theory at the level of a scissors operator[14,29] $\Delta_{\mathbf{k}}P_{c\mathbf{k}}$, i.e., a $\mathbf{k}$-dependent energy shift in the conduction bands which does not change their wave functions. $P_{c\mathbf{k}}$ is the projection operator onto all conduction bands at wave vector $\mathbf{k}$ and $\Delta_{\mathbf{k}}$ is some energy shift. That is, we adopt the Hamiltonian

$$H_{\mathbf{k}}=H_{\mathbf{k}}^{\mathrm{LDA}}+\Sigma_{\mathbf{k}} \tag{1.2}$$

with the self-energy $\Sigma_{\mathbf{k}}$ chosen to have the scissors form

$$\Sigma_{\mathbf{k}}=\Delta_{\mathbf{k}}P_{c\mathbf{k}} \ . \tag{1.3}$$

We perform realistic calculations in our theory in the cases of silicon and germanium. As discussed below, the origin of "**p**" in "**q**·**p**" perturbation theory is the velocity operator $\nabla_{\mathbf{k}}H_{\mathbf{k}}$. The self-energy term modifies both the eigenvalues and the velocity operator; moreover, even if $\Delta_{\mathbf{k}}$ has no **k** dependence, the **k** dependence of the projection operator $P_{c\mathbf{k}}$ is critical. The modification of the velocity operator by a self-energy is known as a Ward identity.[30] The magnitude of the shift can be found semiempirically[31] from a quasiparticle local-density calculation[32] or other simple models[28,29,33] or from *GW* calculations.[14,21,34] A preliminary report of this work has already appeared.[5]

Our results reproduce the experimental dispersion (i.e., frequency dependence of the dielectric function) and the static dielectric constant as a function of pressure to within a few percent. The scissors operator yields consistently better results than LDA, although the LDA also gives a reasonable picture. In the case of the pressure dependence in the linear regime, we favor certain of the mutually contradictory experiments over others. We make a true prediction, i.e., a calculation in advance of the experiments, for the value of the pressure dependence of silicon in the nonlinear regime.

II. THEORY

We will modify the time-dependent local-density approximation (TDLDA) to take into account scissors effects at the level of a scissors operator. The goal is to calculate the macroscopic dielectric function of semiconductors (specifically silicon and germanium in this work) using a plane-wave pseudopotential approach.

Our presentation is organized as follows. The TDLDA equations will be presented and specialized to the case of periodic solids in a plane-wave basis, and more specifically to the semiconducting case. The long-wavelength limit is of particular physical interest. The equations are worked out in detail in the appendixes. The modifications of the equations required to accommodate the self-energy operator in the scissors form are presented next. The principal result is given by Eq. (2.27). In a key example, we show that the net effect on a well-known formula[35] for the static dielectric constant without local-field correction is to modify only one of three energy denominators. The discussion concludes with a section on the methods we chose to solve the equations. Specifically, we have adopted Richardson iteration[36] to find induced wave functions and the Shanks transformation[37,38] to aid in determining the self-consistent fields.

We follow the suggestion of Martin[39] and consider the response of the system to an externally imposed potential $\phi_0 e^{i\mathbf{q}\cdot\mathbf{r}}$ in the limit $\mathbf{q}\rightarrow\mathbf{0}$. In this way, we avoid the usual practice of writing ill-defined quantities such as $\langle n\mathbf{k}|\mathbf{r}|n\mathbf{k}\rangle$ and then converting to more sensible symbols before any calculation takes place.

A. TDLDA—general formulation

The TDLDA has been presented earlier,[40,41] so only the basic equations will be given here. With the exception of the exchange-correlation term of Eq. (2.6), the formulation has been given earlier for extended systems by several authors[35,42,43] and applied in realistic calculations.[7,9] Consider the linear response of a system to a time-dependent scalar potential $\phi^{\mathrm{ext}}(\mathbf{r};\omega)$. In linear-response theory, a number density will be induced according to

$$\delta n(\mathbf{r};\omega)=\int d\mathbf{r}'\chi(\mathbf{r},\mathbf{r}';\omega)\phi^{\mathrm{ext}}(\mathbf{r}';\omega) \ , \tag{2.1}$$

where $\chi(\mathbf{r},\mathbf{r}';\omega)$ is the susceptibility associated with the fully interacting system. To make progress, the susceptibility is replaced by the independent-particle susceptibility $\chi_0(\mathbf{r},\mathbf{r}';\omega)$, and the external field is replaced by a self-consistent field $\phi^{\mathrm{SCF}}(\mathbf{r};\omega)$ in the expression

$$\delta n(\mathbf{r};\omega)=\int d\mathbf{r}'\chi_0(\mathbf{r},\mathbf{r}';\omega)\phi^{\mathrm{SCF}}(\mathbf{r}';\omega) \ . \tag{2.2}$$

The self-consistent field is related to the external field by

$$\phi^{\mathrm{SCF}}(\mathbf{r};\omega)=\phi^{\mathrm{ext}}(\mathbf{r};\omega)+\phi^{\mathrm{ind}}(\mathbf{r};\omega) \ . \tag{2.3}$$

The induced field is taken to have two parts

$$\phi^{\mathrm{ind}}(\mathbf{r};\omega)=\phi^{C}(\mathbf{r};\omega)+\phi^{\mathrm{xc}}(\mathbf{r};\omega) \ , \tag{2.4}$$

an induced Coulomb potential

$$\phi^{C}(\mathbf{r};\omega)=e^2\int d\mathbf{r}'\frac{\delta n(\mathbf{r}';\omega)}{|\mathbf{r}-\mathbf{r}'|} \ , \tag{2.5}$$

and an induced exchange-correlation potential

$$\phi^{\mathrm{xc}}(\mathbf{r};\omega)=\left.\frac{\delta V_{\mathrm{xc}}(\mathbf{r})}{\delta n(\mathbf{r})}\right|_{n(\mathbf{r})=n_0(\mathbf{r})}\delta n(\mathbf{r};\omega) \ . \tag{2.6}$$

The choice of Eq. (2.5) is characteristic of the Coulomb gauge.[44]

The use of Eq. (2.6) implies neglect of the nonlocality and energy dependence of the electron's self-energy.[20] In the zero-frequency limit, the TDLDA involves no assumptions beyond that of the LDA and linear-response theory. Runge and Gross have proposed a time-dependent density-functional theory which agrees with the TDLDA in the static limit.[45] The corrections to the TDLDA in a local-density-functional version of this theory[46] appear to be small.[46–48] Note that the theory of Runge and Gross is a density-functional theory of the electron gas, whereas the displacement of the Kohn-Sham eigenvalues[15] from the many-body quasiparticle energies both lies outside of the realm of density-functional theory and depends upon the nature of exchange and correlation in insulators.[26,27]

The independent-particle susceptibility is given by

$$\chi_0(\mathbf{r},\mathbf{r}';\omega)=\sum_{ij}(f_i-f_j)\frac{\varphi_i^*(\mathbf{r})\varphi_j(\mathbf{r})\varphi_j^*(\mathbf{r}')\varphi_i(\mathbf{r}')}{\varepsilon_i+\hbar\omega-\varepsilon_j+i\eta} \ , \tag{2.7}$$

where the sum is taken over all single-particle states, the

f_i are orbital occupation factors, and η is a positive infinitesimal. In the following, the single-particle states will be taken to have occupancies of 0 or 1. Let $G_c(\mathbf{r},\mathbf{r}';E)$ be the Green's function associated with the subspace of unoccupied single-particle orbitals. The induced electron density from Eq. (2.2) may be written

$$\delta n(\mathbf{r};\omega)=\sum_i^{\text{occ}}\int d\mathbf{r}'\varphi_i^*(\mathbf{r})G_c(\mathbf{r},\mathbf{r}';\varepsilon_i+\hbar\omega)\varphi_i(\mathbf{r}')\phi^{\text{SCF}}(\mathbf{r}';\omega)+\varphi_i(\mathbf{r})G_c^*(\mathbf{r},\mathbf{r}';\varepsilon_i-\hbar\omega)\varphi_i^*(\mathbf{r}')\phi^{\text{SCF}}(\mathbf{r};\omega)\ . \tag{2.8}$$

It is convenient to introduce the quantities

$$u_i^{(+)}(\mathbf{r};\varepsilon_i+\hbar\omega)=\int d\mathbf{r}'G_c(\mathbf{r},\mathbf{r}';\varepsilon_i+\hbar\omega)\varphi_i(\mathbf{r}')\phi^{\text{SCF}}(\mathbf{r}';\omega) \tag{2.9a}$$

and

$$u_i^{(-)*}(\mathbf{r};\varepsilon_i-\hbar\omega)=\int d\mathbf{r}'G_c^*(\mathbf{r},\mathbf{r}';\varepsilon_i-\hbar\omega)\varphi_i^*(\mathbf{r}')\phi^{\text{SCF}}(\mathbf{r};\omega) \tag{2.9b}$$

in terms of which

$$\delta n(\mathbf{r};\omega)=\sum_i^{\text{occ}}\varphi_i^*(\mathbf{r})u_i^{(+)}(\mathbf{r};\varepsilon_i+\hbar\omega)+u_i^{(-)*}(\mathbf{r};\varepsilon_i-\hbar\omega)\varphi_i(\mathbf{r})\ . \tag{2.10}$$

The quantity $u_i^{(+)}(\mathbf{r};\varepsilon_i+\hbar\omega)$ obeys

$$(\varepsilon_i+\hbar\omega-H)u_i^{(+)}(\mathbf{r};\varepsilon_i+\hbar\omega)=P_c\varphi_i(\mathbf{r})\phi^{\text{SCF}}(\mathbf{r};\omega) \tag{2.11a}$$

and $u_i^{(-)}(\mathbf{r};\varepsilon_i-\hbar\omega)$ obeys

$$(\varepsilon_i-\hbar\omega-H)u_i^{(-)}(\mathbf{r};\varepsilon_i-\hbar\omega)=P_c\varphi_i(\mathbf{r})\phi^{\text{SCF}*}(\mathbf{r};\omega)\ , \tag{2.11b}$$

where P_c is the projection operator onto the unoccupied states, and it is understood that the $u_i^{(\pm)}(\mathbf{r};\varepsilon_i\pm\hbar\omega)$ are orthogonal to the occupied states. The induced monopole is given by $\int d\mathbf{r}\,\delta n(\mathbf{r};\omega)$. This quantity vanishes, because the $u_i^{(\pm)}(\mathbf{r};\varepsilon_i\pm\hbar\omega)$ are orthogonal to the occupied states $\varphi_i(\mathbf{r})$. Solving for the linear response in this formulation has been discussed earlier by other authors.[41,49,50]

B. TDLDA for periodic solids in a plane-wave basis

Assume that a periodic single-particle Hamiltonian exists which defines the states $\varphi_i(\mathbf{r})$. For simplicity, the Hamiltonian is taken to be spin independent. The wave function may be expressed in terms of the reciprocal-lattice vectors $\mathbf{G}$ as a sum

$$\varphi_{n\mathbf{k}}(\mathbf{r})=\sum_{\mathbf{G}}\varphi_{n\mathbf{k}\mathbf{G}}e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}\ , \tag{2.12}$$

where the index i has been replaced by the band index n and the crystal momentum $\mathbf{k}$. Similarly, the potential may be taken to be of the form

$$\phi_{\mathbf{q}}(\mathbf{r})=\sum_{\mathbf{G}}\phi_{\mathbf{q}\mathbf{G}}e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}\ . \tag{2.13}$$

If the external potential has this form, the induced and self-consistent potentials will retain it. Similarly the functions $u_{n\mathbf{k}\mathbf{q}}^{(\pm)}(\mathbf{r};\varepsilon_{n\mathbf{k}}\pm\hbar\omega)$ may be represented as a function which has the periodicity of the unit cell times a phase factor $e^{i(\mathbf{k}\pm\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}$,

$$u_{n\mathbf{k}\mathbf{q}}^{(\pm)}(\mathbf{r};\varepsilon_{n\mathbf{k}}\pm\hbar\omega)=\sum_{\mathbf{G}}u_{n\mathbf{k}\mathbf{q}\mathbf{G}}^{(\pm)}e^{i(\mathbf{k}\pm\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}\ . \tag{2.14}$$

The problem has been reduced to solving the system of inhomogeneous linear equations

$$\sum_{\mathbf{G}'}\langle\mathbf{k}\pm\mathbf{q}+\mathbf{G}|\varepsilon_{n\mathbf{k}}\pm\hbar\omega-H|\mathbf{k}\pm\mathbf{q}+\mathbf{G}'\rangle u_{n\mathbf{k}\mathbf{q}\mathbf{G}'}^{(\pm)}=f_{n\mathbf{k}\pm\mathbf{q}\mathbf{G}}\ , \tag{2.15}$$

where the $f_{n\mathbf{k}\pm\mathbf{q}\mathbf{G}}$ are expansion coefficients in

$$P_c\varphi_{n\mathbf{k}}(\mathbf{r})\phi_{\pm\mathbf{q}}(\mathbf{r})=\sum_{\mathbf{G}}f_{n\mathbf{k}\pm\mathbf{q}\mathbf{G}}e^{i(\mathbf{k}\pm\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}\ . \tag{2.16}$$

As long as the photon energy $\hbar\omega$ is less than the band gap of the material (as calculated in the LDA), the operator $(\varepsilon_{n\mathbf{k}}\pm\hbar\omega-H)P_c$ will have only negative eigenvalues. (Here, $\varepsilon_{n\mathbf{k}}$ is the eigenvalue of an occupied state.) As discussed below, this is a favorable circumstance for the use of iterative solution methods. This motivates retaining the operator P_c in the formulas. In contrast, Ref. 40 eliminates the need for it with an identity to simplify their representation of the one-particle Green's function.

The long-wavelength limit $(q\rightarrow 0)$ is of particular interest because visible light has a wavelength large compared to the crystal unit cell, and because it is only in this limit that the use of the longitudinal potential is justified for a crystal.[51] The long-wavelength limit is derived in detail in the appendixes. In reciprocal space, the induced Coulomb potential is given by

$$\phi^{C}(\mathbf{q};\omega)=\frac{4\pi e^2\delta n(\mathbf{q};\omega)}{q^2}\ . \tag{2.17}$$

Fixing the direction of $\mathbf{q}$ to be $\hat{q}$, a power-series expansion in the scalar q may be written as

$$\delta n(\mathbf{q};\omega)=\delta n^{(0)}+q\delta n^{(1)}+q^2\delta n^{(2)}+O(q^3)\ . \tag{2.18}$$

The first term represents an induced monopole, and hence vanishes by the argument given at the end of Sec. II, as well Eq. (A9). The second term also vanishes, by Eq. (A10). The third term is given by Eq. (A14) or Eq. (A17) for a long-wave perturbation and Eq. (A13) for a short-wave perturbation source. Note that Eqs. (A14) and (A17) are manifestly second order, but Eq. (A13) is first order. Since the long-wavelength perturbation produces a response only in first order by Eqs. (A8) and (A11), the short-wave potentials are first order, so only one additional order in $\mathbf{q}$ is required to avoid the q^{-2} divergence manifest in Eq. (2.17). The long- and short-wave potentials must differ by one order in $\mathbf{q}$ so that the respective electric fields remain of the same order of magnitude. The situation is summarized in Fig. 1. The angular dependence of the surviving terms in Eq. (2.18) is discussed in Appendix C.

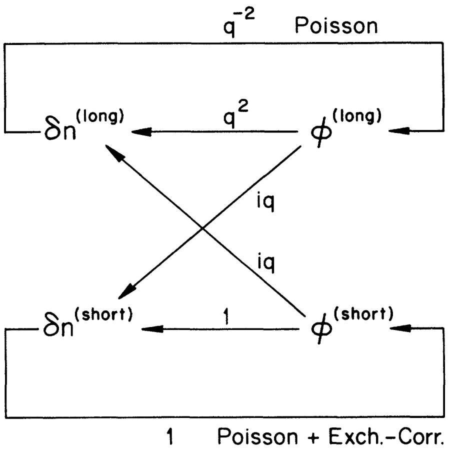

FIG. 1. Schematic representation of self-consistent-field calculation in crystals. The indicated powers of q are gained at each stage in the cycle. The indicated factors of I are appropriate below the absorption threshold. The top stroke is performed analytically, as discussed in Appendix C. See also Sec. II B.

C. Self-energy operator

In the Introduction, we motivated the generalization of the LDA Hamiltonian

$$H_{\mathbf{k}}^{\text{LDA}}=\tfrac{1}{2}(\mathbf{p}+\mathbf{k})^2+V_{\mathbf{k}}\ , \tag{2.19}$$

consisting of the usual kinetic- and potential-energy terms to

$$H_{\mathbf{k}}=\tfrac{1}{2}(\mathbf{p}+\mathbf{k})^2+V_{\mathbf{k}}+\Sigma_{\mathbf{k}} \tag{2.20}$$

which has an additional self-energy term $\Sigma_{\mathbf{k}}$ included; $\Sigma_{\mathbf{k}}$ is, in our approximation, a one-electron operator which does not change the problem symmetry. The potential energy $V_{\mathbf{k}}$ has $\mathbf{k}$ dependence to allow for nonlocal pseudopotentials.[52,53] When the scissors form Eq. (1.3), $\Sigma_{\mathbf{k}}=\Delta_{\mathbf{k}}P_{c\mathbf{k}}$, is taken, $\Sigma_{\mathbf{k}}$ has $\mathbf{k}$ dependence through both $\Delta_{\mathbf{k}}$ and $P_{c\mathbf{k}}$.

The various formulas for the induced charge density, such as Eqs. (A11), (A13), and (A17), contain the term

$$(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle=\mathbf{q}\cdot(\mathbf{p}+\mathbf{k}+\nabla_{\mathbf{k}}V_{\mathbf{k}}+\nabla_{\mathbf{k}}\Sigma_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\ , \tag{2.21}$$

where the tilde over a term denotes the part that is periodic in the Brillouin zone or the unit cell, as appropriate. This leads us to consider the term

$$(\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}}P_{c\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle=\Delta_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}P_{c\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\ . \tag{2.22}$$

The periodic part of a wave function of a nondegenerate Hamiltonian may be expressed as[54]

$$|n,\tilde{\mathbf{k}}+\mathbf{q}\rangle=|n,\tilde{\mathbf{k}}\rangle+G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle+O(q^2)\ , \tag{2.23}$$

where $G_{n\mathbf{k}}=(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})^{-1}$ and the pseudoinverse is intended; i.e., $G_{n\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle=0$. The identity $I=P_{v\mathbf{k}}+P_{c\mathbf{k}}$ implies $0=\nabla_{\mathbf{k}}P_{v\mathbf{k}}+\nabla_{\mathbf{k}}P_{c\mathbf{k}}$, where $P_{v\mathbf{k}}$ is the projection operator onto the valence bands with wave vector $\mathbf{k}$ and I is the identity operator. These considerations allow us to write

$$\mathbf{q}\cdot\nabla_{\mathbf{k}}P_{c\mathbf{k}}=-\sum_{n}^{\text{occ}}G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\langle n,\tilde{\mathbf{k}}| + |n,\tilde{\mathbf{k}}\rangle\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}\ , \tag{2.24}$$

leading to

$$P_{c\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}P_{c\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle=-P_{c\mathbf{k}}G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle \tag{2.25}$$

for an occupied state $|n,\tilde{\mathbf{k}}\rangle$, where $P_{n\mathbf{k}}$ is the projection operator $|n,\tilde{\mathbf{k}}\rangle\langle n,\tilde{\mathbf{k}}|$. Putting this result into Eq. (2.21) and taking the scissors form Eq. (1.3) for the self-energy $\Sigma_{\mathbf{k}}$ leads to

$$P_{c\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle=P_{c\mathbf{k}}[(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\text{LDA}}) -\Delta_{\mathbf{k}}G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle\ . \tag{2.26}$$

After a little algebra,[55]

$$P_{c\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle =G_{n\mathbf{k}}^{\text{LDA}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})P_{c\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\text{LDA}})|n,\tilde{\mathbf{k}}\rangle\ . \tag{2.27}$$

Equation (2.27) is the main theoretical result of this paper. The order of magnitude of $G_{n\mathbf{k}}^{\text{LDA}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})$ is the ratio of the true direct band gap to the LDA direct band gap; for silicon and related semiconductors, this represents a 20% correction.[14,21] Standard TDLDA is recovered for $\Delta_{\mathbf{k}}=0$ since $G_{n\mathbf{k}}^{\text{LDA}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}}^{\text{LDA}})P_{c\mathbf{k}}=(I-P_{n\mathbf{k}})P_{c\mathbf{k}}=P_{c\mathbf{k}}$. The presence of $P_{c\mathbf{k}}$ in these formulas causes no difficulty, as discussed in the first paragraph of Appendix B. Using the equations in this section, Eq. (2.27) may be generalized to

$$(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})=(\mathbf{q}\cdot\nabla_{k}H_{\mathbf{k}}^{\text{LDA}})+(\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}} -\Delta_{\mathbf{k}}\sum_{n}^{\text{occ}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\text{LDA}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\text{LDA}})P_{n\mathbf{k}} +P_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\text{LDA}})G_{n\mathbf{k}}^{\text{LDA}}P_{c\mathbf{k}}\ , \tag{2.28a}$$

a result given only for reference. This equation may be used to generalize Eq. (2.27) slightly to

$$G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle=G_{n\mathbf{k}}^{\text{LDA}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\text{LDA}})|n,\tilde{\mathbf{k}}\rangle\ . \tag{2.28b}$$

Equation (2.27) may be applied to all of the results in the appendixes such as Eqs. (A11), (A13), (A14), and (A17). Although we primarily consider the real part of the dielectric response in this paper, Eq. (2.27) applies equally well to the imaginary part of the dielectric function. We exhibit only one application of Eq. (2.27): the static limit of Eq. (A17) for the long-wave charge induced by a long-wave potential in a semiconductor determines the macroscopic dielectric constant[35] without the local-field correction from

$$\hat{\mathbf{q}}\cdot\overleftrightarrow{\epsilon}\cdot\hat{\mathbf{q}}=1+\lim_{q\to0}\frac{8\pi e^2}{q^2}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}^{3}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\ . \tag{2.29}$$

A naive appreciation of the self-energy correction might lead one to correct the LDA formula

$$\hat{\mathbf{q}}\cdot\overleftrightarrow{\epsilon}^{\mathrm{LDA}}\cdot\hat{\mathbf{q}}=1+\lim_{q\to0}\frac{8\pi e^2}{q^2}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\mathrm{LDA}})(G_{n\mathbf{k}}^{\mathrm{LDA}})^{3}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\mathrm{LDA}})|n,\tilde{\mathbf{k}}\rangle \tag{2.30}$$

to

$$\hat{\mathbf{q}}\cdot\overleftrightarrow{\epsilon}^{N}\cdot\hat{\mathbf{q}}=1+\lim_{q\to0}\frac{8\pi e^2}{q^2}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\mathrm{LDA}})G_{n\mathbf{k}}^{3}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\mathrm{LDA}})|n,\tilde{\mathbf{k}}\rangle\ . \tag{2.31}$$

This would be equivalent to replacing the energy denominators of an Adler-Wiser expansion of the dielectric function with energy denominators corrected by the scissors operator Eq. (1.3). Although this would appear a sensible correction, we show just below that this would result in a large overcorrection for the LDA error. The correct approach is to replace only one of the three energy denominators with a scissors-operator-corrected version. Application of Eq. (2.27) to Eq. (2.29) leads to

$$\hat{\mathbf{q}}\cdot\overleftrightarrow{\epsilon}\cdot\hat{\mathbf{q}}=1+\lim_{q\to0}\frac{8\pi e^2}{q^2}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\mathrm{LDA}})G_{n\mathbf{k}}^{\mathrm{LDA}}G_{n\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}^{\mathrm{LDA}})|n,\tilde{\mathbf{k}}\rangle\ . \tag{2.32}$$

Thus, in effect the scissors form of the self-energy operator Eq. (1.3) leads to the replacement of only one of three energy denominators. We understand this in that one of the energies is a true many-body energy, but two are associated with the "$\mathbf{q}\cdot\mathbf{p}$" expansion of the LDA wave functions, which are, empirically, nearly identical to more sophisticated *GW* quasiparticle wave functions. Hybertsen and Louie[3] have, in effect, previously considered the attractions and drawbacks of Eq. (2.31).

D. *f*-sum rule and plasma frequency

In the case of a semiconductor or insulator the *f*-sum rule for crystals, Eq. (A16), states

$$\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|[(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})+\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle=0\ . \tag{2.33}$$

Let the quantity N_{occ} be the number of occupied bands (times two for spin); in a semiconductor, this value is the same at every point in the Brillouin zone. Since

$$\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\tfrac{1}{2}q^2|n,\tilde{\mathbf{k}}\rangle=\tfrac{1}{2}q^2\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle=\tfrac{1}{2}q^2N_{\mathrm{occ}}\ , \tag{2.34}$$

the quantity

$$\Sigma_f=\frac{2}{q^2N_{\mathrm{occ}}}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}-\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle+\langle n,\tilde{\mathbf{k}}|[\tfrac{1}{2}q^2-\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle \tag{2.35}$$

will be equal to unity. Here Σ_f is the sum of the oscillator strengths, or *f* sum. We may transform the term in $\tfrac{1}{2}q^2-\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})$ as follows:

$$\begin{aligned}
&\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{q}\cdot\nabla_{\mathbf{k}}[\mathbf{q}\cdot\nabla_{\mathbf{k}}(V_{\mathbf{k}}+\Sigma_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle\\
&\quad=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\mathbf{q}\cdot\nabla_{\mathbf{k}}\langle n,\tilde{\mathbf{k}}|[\mathbf{q}\cdot\nabla_{\mathbf{k}}(V_{\mathbf{k}}+\Sigma_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle-\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle\mathbf{q}\cdot\nabla_{\mathbf{k}}n,\tilde{\mathbf{k}}|[\mathbf{q}\cdot\nabla_{\mathbf{k}}(V_{\mathbf{k}}+\Sigma_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle\\
&\qquad-\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|[\mathbf{q}\cdot\nabla_{\mathbf{k}}(V_{\mathbf{k}}+\Sigma_{\mathbf{k}})]|\mathbf{q}\cdot\nabla_{\mathbf{k}}n,\tilde{\mathbf{k}}\rangle\\
&\quad=-\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}[\mathbf{q}\cdot\nabla_{\mathbf{k}}(V_{\mathbf{k}}+\Sigma_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle-\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|[\mathbf{q}\cdot\nabla_{\mathbf{k}}(V_{\mathbf{k}}+\Sigma_{\mathbf{k}})]G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle.
\end{aligned} \tag{2.36}$$

The first step is simply the rule for differentiation of a product; in the second step, the first term vanishes by application of the divergence theorem, and the second and third terms make use of Eq. (2.23) in the form

$$|\mathbf{q}\cdot\nabla_{\mathbf{k}} n,\tilde{\mathbf{k}}\rangle = G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}} H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle \ . \tag{2.37}$$

Applying the substitution in Eq. (2.36) to Eq. (2.35) leads to

$$\Sigma_f = \frac{2}{q^2 N_{\mathrm{occ}}}\bar{\Omega}_0 \times \int_{\mathrm{BZ}} d\mathbf{k} \sum_n^{\mathrm{occ}} \mathrm{Re}\langle n,\tilde{\mathbf{k}}|\mathbf{q}\cdot\mathbf{p}G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}} H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle \ . \tag{2.38}$$

In the case of the self-energy operator $\Sigma_{\mathbf{k}} = \Delta_{\mathbf{k}} P_{c\mathbf{k}}$, Eq. (2.27) may be applied to yield the expression

$$\Sigma_f = \frac{2}{q^2 N_{\mathrm{occ}}}\bar{\Omega}_0 \times \int_{\mathrm{BZ}} d\mathbf{k} \sum_n^{\mathrm{occ}} \mathrm{Re}\langle n,\tilde{\mathbf{k}}|\mathbf{q}\cdot\mathbf{p}G_{n\mathbf{k}}^{\mathrm{LDA}}(\mathbf{q}\cdot\nabla_{\mathbf{k}} H_{\mathbf{k}}^{\mathrm{LDA}})|n,\tilde{\mathbf{k}}\rangle \ ; \tag{2.39}$$

i.e., the scissors form of the self-energy operator has no effect on the f sum. We evaluate Eq. (2.39) in our computer program, replacing the Brillouin-zone integral by a special-points quadrature.[56] The results test the quality of the Brillouin-zone integration, as discussed in Sec. III A.

At high frequencies, the dielectric function may be derived from Eq. (A14) with the aid of Eqs. (C12) and (C13). The formula is

$$\epsilon(\omega) = 1 - \lim_{q\to 0}\frac{4\pi e^2}{q^2}\bar{\Omega}_0 \times \int_{\mathrm{BZ}} d\mathbf{k} \sum_n^{\mathrm{occ}} \langle n,\tilde{\mathbf{k}}|\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}} H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle + O(\omega^{-4}) \ . \tag{2.40}$$

The long-wave formula Eq. (A14) leads to corrections of order ω^{-4}; local-field corrections enter in order ω^{-6}. The usual high-frequency formula for the dielectric function

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2} + O(\omega^{-4}) \tag{2.41}$$

with $\omega_p^2 = 4\pi N_{\mathrm{occ}}\Omega_0^{-1}e^2$ is recovered from Eq. (2.40) in the standard case $\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}} H_{\mathbf{k}}) = q^2$ via Eq. (2.34). In the present theory, the coefficient of the ω^{-2} term in Eq. (2.40) forms an "effective plasma frequency"

$$(\omega_p^{\mathrm{eff}})^2 = \omega_p^2 \left[1 + \frac{2}{q^2 N_{\mathrm{occ}}}\bar{\Omega}_0 \int_{\mathrm{BZ}} d\mathbf{k} \sum_n^{\mathrm{occ}} \mathrm{Re}\langle n,\tilde{\mathbf{k}}|[(\mathbf{q}\cdot\nabla_{\mathbf{k}} H_{\mathbf{k}}) - \mathbf{q}\cdot\mathbf{p}]G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}} H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle \right] \ . \tag{2.42}$$

The physical origin of this deviation may be taken from the following observation. First, note that the sum rule of Eq. (A16) may be seen as connecting total oscillator strength with the effective plasma frequency discussed here. Every level has one unit of oscillator strength in its transitions to other levels. This is, however, a signed quantity, being positive for absorption transitions and negative for emission ones. Transitions between two occupied states sum to zero net oscillator strength. In the pseudopotential construction, the core electrons are removed. The valence electrons only have higher states to make transitions to, leading to an excess of oscillator strength. The corrections to the valence-electron plasma frequency have been discussed by Smith and coworkers.[57,58] The effect of a nonlocal potential on the f-sum rule, as well as the relationship of the "length" and "velocity" forms of a matrix element in the presence of a nonlocal potential, has been discussed in the context of finite systems by Starace.[59]

E. Algorithmic considerations

Richardson iteration[37] is a convenient method to solve the system of inhomogeneous linear equations (2.15), or the related equations Eqs. (A7), (A11), (A13), and (A17). Consider the equation

$$L|u\rangle = |f\rangle \tag{2.43}$$

in which L is a general linear operator $[(\varepsilon_{n\mathbf{k}} \pm \hbar\omega - H)P_c$ in our case], $|u\rangle$ is the desired solution vector, and $|f\rangle$ is a known vector. In Richardson iteration, an approximate solution undergoes the process

$$|u^{(n+1)}\rangle = |u^{(n)}\rangle - \alpha L_{ap}^{-1}(L|u^{(n)}\rangle - |f\rangle) \ , \tag{2.44}$$

where L_{ap}^{-1} is an approximate inverse operator. The sequence converges if all of the eigenvalues of $L_{ap}^{-1}L$ have the same sign. Assuming M is the maximum eigenvalue of $L_{ap}^{-1}L$ and m is the minimum, then $\alpha_{\mathrm{opt}} = 2/(M+m)$, and the magnitude of the error is reduced by a factor of $(M-m)/(M+m)$ on every iteration. We choose the diagonal elements of $\varepsilon_{n\mathbf{k}} \pm \hbar\omega - H$ for L_{ap}, with a least negative value of $\varepsilon_{n\mathbf{k}} \pm \hbar\omega - \varepsilon_{c\mathbf{k}}$, where $\varepsilon_{c\mathbf{k}}$ is the minimum conduction-band eigenvalue. The low-lying eigenvectors are rich in small-G plane waves. The action of $(\varepsilon_{n\mathbf{k}} \pm \hbar\omega - \varepsilon_{c\mathbf{k}})P_c$ on a single small-G plane wave is primarily to multiply it by a factor of order $\varepsilon_{n\mathbf{k}} \pm \hbar\omega - \varepsilon_{c\mathbf{k}}$ to $\varepsilon_{n\mathbf{k}} \pm \hbar\omega - \varepsilon_{c'\mathbf{k}}$, where $\varepsilon_{c'\mathbf{k}}$ is another "typical" low-lying conduction-band eigenvalue. This factor is substantially removed by L_{ap}^{-1}. For large G, $\varepsilon_{n\mathbf{k}} \pm \hbar\omega - H$ is diagonally dominant, and so the approximate inverse is excellent. Indeed, the difficulty of finding a solution (i.e., number of iterations required) depends primarily upon the small-G components, and does not grow with the number of stars of reciprocal-lattice vectors.

If a few low-lying conduction-band states $|m\mathbf{k}\rangle$ are known, then $|f\rangle$ may be broken into the parts parallel and orthogonal to these states, with the total solution being the sum of the two subspace solutions. In the parallel subspace, the solution is immediate:

$$|u_{\parallel}\rangle = \sum_{m}^{\substack{\text{known}\\ \text{conduction-band}\\ \text{states}}} |m\mathbf{k}\rangle(\varepsilon_{n\mathbf{k}} \pm \hbar\omega - \varepsilon_{m\mathbf{k}})^{-1}\langle m\mathbf{k}|f\rangle \ . \tag{2.45}$$

In the orthogonal subspace, the rate of convergence is improved because the excluded states are the least diagonally dominant. The cutoff state c should be the one associated with the lowest-lying state not projected out by Eq. (2.35). If we wish to study absorption processes, it is necessary to project out all states with energies below $\varepsilon_{v\mathbf{k}}+\omega$ to preserve the single signed character of $L_{ap}^{-1}L$, and hence the convergence of the iterative scheme. Here $\varepsilon_{v\mathbf{k}}$ is the maximum eigenvalue in the valence band for the $\mathbf{k}$ point. For frequencies near the absorption threshold, projection is a practical necessity: in the language of the previous paragraph, if $\varepsilon_{n\mathbf{k}} \pm \hbar\omega - \varepsilon_{c\mathbf{k}}$ is nearly vanishing, it is not of the same order of magnitude as $\varepsilon_{n\mathbf{k}} \pm \hbar\omega - \varepsilon_{c'\mathbf{k}}$, leading to large values for M, while m remains of order unity. In practice, we usually used six known conduction-band states. For the systems involved in this study (silicon with a plane-wave energy cutoff of 9 hartrees and germanium with a 10-hartree cutoff) standard matrix diagonalization is superior to the method described above. For larger systems, we expect the iterative methods described above to be more efficient.

To determine the short-wave self-consistent induced field ϕ^{SCF}, we used the Shanks transformation.[37,38] The Shanks transformation acts on three members of a sequence A_{n-1}, A_n, and A_{n+1} via the expression

$$A = \frac{A_{n+1}A_{n-1} - A_n^2}{A_{n+1} + A_{n-1} - 2A_n} \ . \tag{2.46}$$

If the sequence is of the form $A_n = A_\infty + cq^n$, this transformation is exact (i.e., if A is formed from any three consecutive elements of a geometric sequence, $A = A_\infty$). We apply the Shanks transformation (2.46) independently to each Fourier component of the short-wave induced charge density. We are motivated to do this to reduce the "charge sloshing" problem common in electronic-structure calculations. The transformation leads to an acceleration of the convergence of a factor of about 3 compared to a simple mixing scheme. We find about ten iterations are sufficient to achieve a reasonable degree of convergence (10^{-5} relative accuracy in the potential), or seven iterations if a guess from a nearby case is available. Overall, it takes about 2–3 times as long to solve the SCF short-wave equation (A7) as the non-SCF long-wave equations (A11) and (A17) because we do not fully converge the short-wave-induced wave functions before updating the short-wave potential. [Equations (A11) and (A17) rely on the same inhomogeneous equation, and so are in effect solved together. The loop is closed with Eq. (A13) which takes very little time after Eqs. (A7), (A11), and (A17) have been solved.]

III. RESULTS

We have extended the ground-state LDA program of Allan and Teter[60] to calculate linear optical response. This program minimizes the total energy of the system with a new iterative method.[61] The exchange-correlation functional is chosen to be the Perdew-Zunger[62] parametrization of the exchange-correlation potential derived from the calculation of Ceperley and Alder[63] of the exchange-correlation energy of the electron gas. Our pseudopotential is a norm-conserving separable potential of the form introduced by Kleinman and Bylander.[52] The pseudopotential itself is generated by Hamann's program.[53] We choose the s part of the pseudopotential to be local in the Kleinman-Bylander construction, and include nonlocal p and d pseudopotentials. Choosing the d part to be local is adequate for silicon but not for germanium. We work in a plane-wave basis. We chose the relatively high plane-wave energy cutoffs of 9 hartrees for silicon and 10 hartrees for germanium, but did not systematically vary these limits. We performed the calculations with 60 special points. The linear calculation makes use of the ground-state program in two ways: to calculate ground-state information, and as a subroutine library for the linear-response program. In particular, the Hamiltonian used for the ground state is identical to the LDA part of the Hamiltonian used for linear response.

A. Static response and sum rules

We calculated the static dielectric constant of silicon and germanium in the LDA, the naive model, and the present theory at the experimental lattice constants at zero temperature, given in Table I. Without the local-field correction, these three theories correspond to Eqs. (2.30), (2.31), and (2.32), respectively. We present the results with and without the local-field corrections in Table II, and compare to other work.

The present results differ from our earlier report[5] in that (a) the Hamann[53] pseudopotential is used for both silicon and germanium instead of just germanium, (b) the energy cutoff for silicon has been raised to 9 hartrees from 6 hartrees, (c) we choose 0.9 eV rather than 0.7 eV for the scissors operator based on the recent study of Zhu, Fahy, and Louie,[64] and (d) we present the germanium results for two values of the self-energy operator 0.7 and 0.8 eV (see Table III). We feel 0.8 eV is more reasonable based on the work of Hybertsen and Louie.[21] By presenting both values, we give some indication of the sensitivity of the dielectric constant to this parameter—about 3% in germanium for the 0.1-eV variation. In silicon, the sensitivity is expected to be somewhat lower but was not investigated explicitly.

We find excellent agreement between the present linear-response program and the results calculated from

TABLE I. Experimental values of lattice constant a_0, bulk modulus B_0, and pressure derivative of bulk modulus B_0' for Si and Ge. These were the values used to convert between lattice constant variations and pressure variations.

	Si	Ge
a_0 (Å)	5.429	5.655
B_0 (GPa)	99	77
B_0'	4.2	4.6

TABLE II. Experimental and theoretical values of ϵ and related quantities for silicon from the literature and the present work. ϵ_{00} omits the local-field correction. ϵ_1 is sometimes called ϵ_M. The f sum Σ_f is defined by Eq. (2.38) and ω_p^{eff} is defined by Eq. (2.42). $\omega_p = 16.6$ eV for silicon. Energy cutoffs are 7 hartrees for Refs. 2 and 3 and 9 hartrees for the present study. Values shown for 60 special integration points in the irreducible wedge of the Brillouin zone (Ref. 56).

	Silicon			
	ϵ_{00}	ϵ_1	f sum (Σ_f)	$\omega_p^{\text{eff}}/\omega_p$
LDA[a]	13.4	12.7		
LDA[b]	13.6	13.0		
Present LDA	14.2	13.5	1.013	0.972
Present "naive" approximation $\Delta = 0.9$ eV	8.1	7.7	0.856	0.978
Present $\Delta = 0.9$ eV	11.7	11.2	1.013	1.060
Experiment[c] (0 K)		11.4		
Experiment[c] (300 K)		11.7		

[a] Reference 2.

[b] Reference 3.

[c] Reference 69.

the ground-state program by applying a finite field and taking finite differences. We considered two cases: a short-wave perturbation which is commensurate with one of the reciprocal-lattice vectors and a long-wave perturbation. For the long-wave perturbation, we extrapolated the result from a series of supercell calculations with up to 12 standard unit cells (i.e., 24 silicon atoms). The results presented in this paper compare well with the results of Baroni and Resta[2] and Hybertsen and Louie.[3] Agreement is excellent with de Gironcoli, Baroni, and Resta[4] in the cases of AlP, AlAs, GaP, and GaAs: specifically, the local-field-corrected static macroscopic dielectric constants calculated at the lattice constant corresponding to the minimum of the LDA total energy are 8.4, 9.3, 9.7, and 12.0, respectively, for Ref. 4 compared to 8.4, 9.4, 9.8, and 11.8 in the present calculation. Our Brillouin-zone integration yields an f sum, defined by Eq. (2.38), of 1.013 for silicon and 0.993 for germanium (where 1 is the ideal value) when 60 special points are chosen. Omission of the nonlocal term increases the value of ϵ by about 15% in the case of silicon, in agreement with Hybertsen.[65] Omitting the nonlocal operator leads to f-sum values of 0.959 for silicon and 1.070 for germanium.

Because we have been relying primarily upon published results for the magnitude of the scissors operator, we have not attempted to include any $\mathbf{k}$ dependence in $\Delta_{\mathbf{k}}$ (although our theory and program is capable of including such effects). The magnitude of the $\mathbf{k}$ dependence of the scissors operator has been estimated to be 0.1 eV in silicon and 0.2 eV in germanium.[14] Since we are calculating Brillouin-zone averages this probably represents a small effect, although it may be significant in the case of photon frequencies near the band edge.

As seen in Table II, the self-energy operator is seen to improve the agreement with experiment dramatically compared to the LDA results. The "naive" model leads to a large overcorrection as suggested earlier.[3] Agreement at the level of a few percent is achieved. The local-field correction is found to be about a 5% effect. It is not clear why our local-field correction is substantially smaller than that of Ref. 3 in the case of germanium but nearly equal in the case of silicon.

The accuracy of the f sum at the level of about 1% suggests uncertainties in the integration at the level of perhaps 4%, since the static dielectric constant depends upon the inverse third power of the energy gap, rather than the inverse first power. The "naive" model leads to substantial f-sum-rule violations. The "effective plasma frequency" from Eq. (2.42) is also given in Table II. The

TABLE III. Experimental and theoretical values of ϵ and related quantities for germanium from the literature and the present work. See Table II for details. $\omega_p = 15.6$ eV for germanium. Energy cutoffs are 10 hartrees for Ref. 3 and present work. Integration on 60 special points.

	Germanium			
	ϵ_{00}	ϵ_1	f sum (Σ_f)	$\omega_p^{\text{eff}}/\omega_p$
LDA[a]	21.9	20.7		
Present LDA	22.0	21.3	0.993	1.039
Present "naive" approximation $\Delta = 0.7$ eV	10.8	10.4	0.857	1.035
Present $\Delta = 0.7$ eV	17.0	16.5	0.993	1.119
Present $\Delta = 0.8$ eV	16.6	16.0	0.993	1.121
Experiment[b] (0 K)		15.3		
Experiment[b] (300 K)		16.0		

[a] Reference 3.

[b] Reference 69.

considerations in Sec. II D suggest $\omega_p^{\text{eff}} > \omega_p$. This condition is violated for the LDA in the case of silicon. Both the magnitude of this correction and the trend are in line with rough expectations based on calculations of alkali-metal atoms.[57]

B. Frequency dependence

We calculated the frequency-dependent dielectric function for frequencies below the direct band gaps, which are 3.4 eV for silicon and 1.0 eV for germanium.[66] Following Wemple and DiDominico,[67] we plot the results in the form of $(\epsilon - 1)^{-1}$ versus ω^2. If the response follows that of a single oscillator, given by Eq. (1.1), then such a plot will be a straight line. The experimental data fit this simple relationship to a remarkable degree for a wide range of semiconductors and insulators below their band gaps including silicon and germanium.[67] Our theory also yields a nearly linear relation in these variables. Corrections will lead to larger values of ϵ, i.e., deviations below a straight line in our plot.

The role of local fields is illustrated in Figs. 2 and 3. The local-field correction is seen to be a modest, featureless correction throughout in the range of this study. At excitation frequencies near transition energies, the local-field effects can be striking.[68]

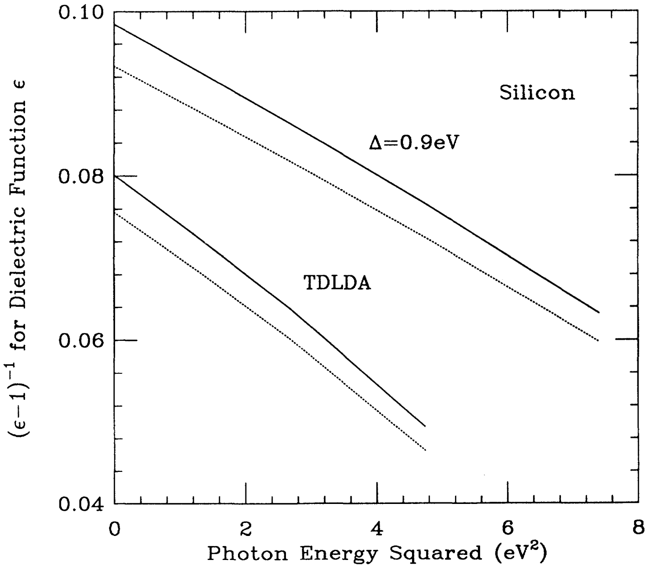

FIG. 2. Dielectric function of silicon as a function of frequency calculated in the TDLDA and present theory with the self-energy operator in the scissors form. Solid (dotted) lines include (exclude) local-field effects.

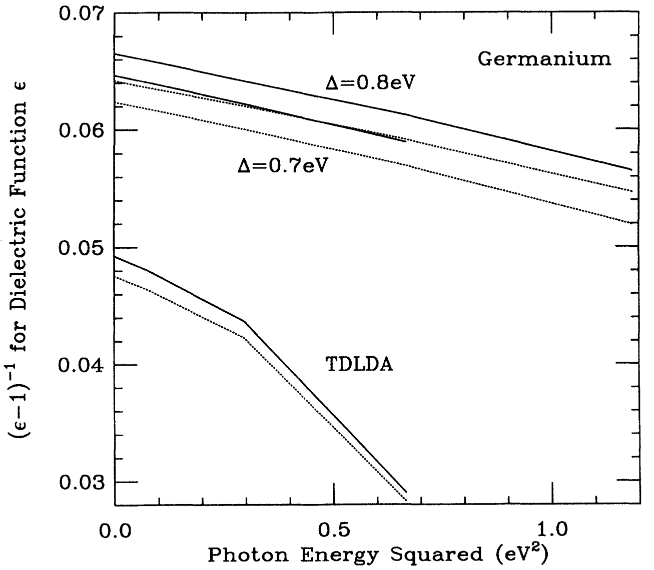

FIG. 3. Dielectric function of germanium as a function of frequency calculated in the TDLDA and present theory with the self-energy operator in the scissors form. Solid (dotted) lines include (exclude) local-field effects. The kink in the TDLDA curve is due to the limited number of points at which we calculated the frequency.

Our results for silicon, with the self-energy operator taking the value of $\Delta = 0.9$ eV are shown in Fig. 4, along with experimental data. The data are given for 100 and 300 K but the parameters of our theoretical calculation would be suitable for the zero-temperature case. Presumably, the lower temperature will bring the curves into closer agreement perhaps at the level of no worse than 10% over the range plotted. At lower temperatures, agreement with the LDA will be somewhat worse than for the room-temperature data of Fig. 4. The small mismatch between the data compiled by Li[69] and the data of Aspnes and Struda[70] is, we feel, more likely to be due to the use of different samples (or possibly measurement technique) than a physical effect, due to, say, the onset of phonons coincident with the indirect band gap at 1.2 eV. Measurements of the dielectric function by two other

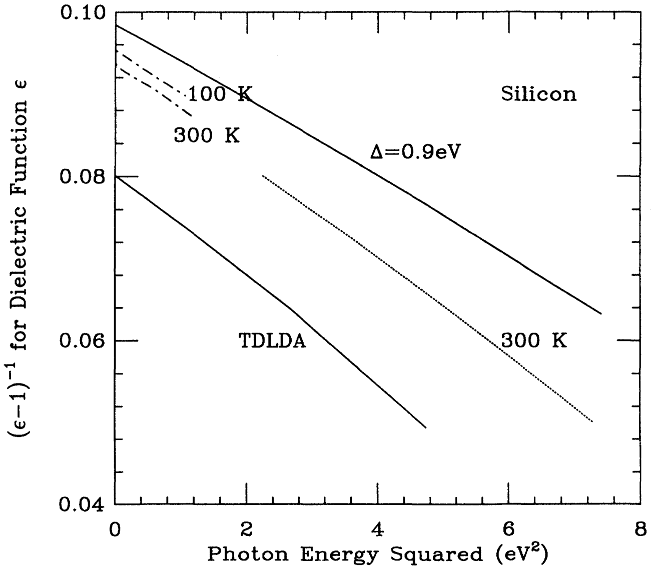

FIG. 4. Dielectric function of silicon as a function of frequency. Comparison of the TDLDA and present theory (solid lines) with experiments. Theoretical values are given in Table IV. Dashed-dotted lines are from Ref. 69; dotted lines are from Ref. 70.

groups[71,72] across the indirect band gap show no sign of deviation from the simple linear relation between $1/(\epsilon-1)$ and ω^2.

The germanium results are shown in Fig. 5. The range plotted is smaller than for silicon because we only calculate below the direct band gap. The TDLDA curves are calculated omitting any imaginary part of the response above its threshold. Overall agreement with experiment is quite good. At zero temperature, the removal of thermal effects will reduce ϵ_1 below the value shown in the figure [i.e., $(\epsilon^{-1}-1)^{-1}$ will increase]. Disagreement with the experiment is at the level of a few percent. Since the sign of the error is different in silicon compared with germanium, it is difficult *a priori* to name a particular effect which is responsible for the observed discrepancies. At very low frequencies the data show a small but sudden increase. This effect, representing a shift in ϵ from 16.0 to 15.8, is due to the excited carriers with a plasma frequency of order 100 μeV; this is outside the scope of the present theory.

Our calculated results are given in Table IV. The final two significant figures presented are intended for use in relative comparisons of the data to itself, e.g., finite differences. In Table V, the initial slopes are fit to single-oscillator parameters in the form

$$\epsilon_1(\omega)=1+\frac{f\omega_p^2}{\omega_0^2-\omega^2}\ , \tag{3.1}$$

where f is an oscillator strength, and ω_p is the plasma frequency associated with the valence electrons. These parameters do not have a great deal of physical meaning other than to roughly identify the position and strength of the main absorption peak in the two semiconductors, and to illustrate the expected result that shifting the spectrum by the self-energy operator is reflected in the average oscillator position by about the same amount. We include these parameters to permit comparison of several calculations and experiment within the oversimplified single-oscillator model.

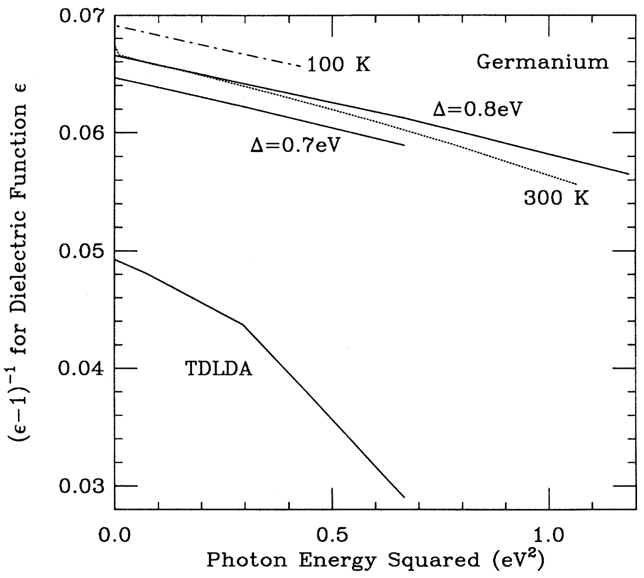

FIG. 5. Dielectric function of germanium as a function of frequency. Comparison of the TDLDA and present theory (solid lines) with experiments. Theoretical values are given in Table IV. Dashed-dotted lines are from Ref. 69; dotted lines are from Ref. 100.

C. Pressure dependence

The pressure or lattice-constant dependence of the real part of the electronic dielectric constant at zero frequency, $\epsilon_1(P)$, plays a fundamental role in theories of covalent bonding in crystals.[73–75] The fractional change in ϵ_1 with pressure, at zero pressure, $d\ln[\epsilon_1(P)]/dP$, is about an order of magnitude larger in Ge than in Si. One finds several measurements for $d\ln[\epsilon_1(P)]/dP$ for Si and Ge in the literature, with only moderate agreement among the measurements (values and references given in Table VI). There are also a number of theoretical descriptions of $d\ln[\epsilon_1(P)]/dP$ from simple models[73,76] to semiempirical treatments[77] to more sophisticated calculations.[78] The simple models do not reproduce the order-of-magnitude difference between the size of $d\ln[\epsilon_1(P)]/dP$ in Si and Ge

TABLE IV. Dielectric function for silicon and germanium. ϵ_{00} excludes the local-field correction, but ϵ_1 includes it. Frequency values are exact multiples of 0.01 hartree. Plane-wave energy cutoff of 9 hartrees for silicon and 10 hartrees for germanium; integration on 60 special points.

	Silicon				Germanium					
	LDA		Δ=0.9 eV		LDA		Δ=0.7 eV		Δ=0.8 eV	
ω (eV)	ϵ_{00}	ϵ_1	ϵ_{00}	ϵ_1	ϵ_{00}	ϵ_1	ϵ_{00}	ϵ_1	ϵ_{00}	ϵ_1
0.00	14.227	13.484	11.715	11.160	22.027	21.293	17.034	16.468	16.577	16.029
0.27					22.548	21.807	17.180	16.612	16.722	16.164
0.54	14.521	13.759	11.861	11.297	24.645	23.879	17.650	17.072	17.117	16.585
0.82					36.347	35.444	18.559	17.966	17.901	17.328
1.09	15.503	14.679	12.330	11.737			20.256		19.291	18.696
1.63	17.606	16.648	13.223	12.575						
2.18	22.487	21.234	14.794	14.049						
2.72			17.738	16.814						

TABLE V. Oscillator parameters from Eq. (3.1) for silicon and germanium computed from the frequency-dependent results of Table IV. Parameters from Ref. 67 are computed from $f=E_dE_0\omega_p^{-2}$ and $\omega_0=E_0$, where E_d and E_0 are defined in the reference.

	Silicon		Germanium	
	f	ω_0 (eV)	f	ω_0 (eV)
Present LDA	0.62	3.7	0.22	1.6
Present $\Delta=0.7$ eV			0.50	2.8
Present $\Delta=0.8$ eV			0.51	2.9
Present $\Delta=0.9$ eV	0.85	4.8		
Semiempirical[a]	0.64	4.0	0.45	2.7
Experiment[b] 100 K	0.80	4.6	0.49	2.9
Experiment[b] 300 K	0.56	3.8	0.35	2.4

[a] Reference 67.

[b] Reference 69.

and the more sophisticated calculations do not agree well in their predicted values either (Table VI), nor do they agree particularly well with experiment. Thus it is appropriate to use the current parameter-free theory to produce reliable values for $d\ln[\epsilon_1(P)]/dP$ for Si and Ge for comparison with the different experimental and theoretical results and for gaining insight into the large difference in the value for Si and Ge.

In our calculations we actually vary the lattice constant and compute values of ϵ_1. The connection to pressure is accomplished via an equation of state. This can be done using the theoretical equation of state or almost equivalently by using the Murnaghan equation of state[79] with the experimentally measured values of the bulk modulus and its pressure derivative.[80] Our theoretical equilibrium lattice constants are less than 1% too low

TABLE VI. Experimental and theoretical values of $d\ln[\epsilon_1(P)]/dP$ for Si and Ge from the literature and the present work.

	Si	Ge
$\frac{d\ln[\epsilon_1(P)]}{dP}$ (10^{-12}/Pa), expt.	$-2.8\pm12\%$[a]	$-34\pm4\%$[b]
	$-2.3\pm5\%$[c]	$-23\pm6\%$[d]
	$-6\pm67\%$[e]	$-14\pm29\%$[e]
	$-4\pm25\%$[e]	$-12\pm25\%$[e]
		$-50.6\pm4\%$[f]
$\frac{d\ln[\epsilon_1(P)]}{dP}$ (10^{-12}/Pa), theory	-6.1[g]	
	-3.1[h]	-4.1[h]
	-3.0[i]	-22[i]
	-6.2[j]	-17.8[j]
		$-6.0, -16$[k]
Present LDA	-4.1	-47
Present $\Delta=0.9$ eV (Si), 0.8 eV (Ge)	-2.6	-31

[a] Reference 84.

[b] Reference 83.

[c] Reference 96.

[d] Reference 85.

[e] Reference 97.

[f] Reference 98.

[g] Reference 73.

[h] Reference 76.

[i] Reference 77.

[j] Reference 99.

[k] Reference 78, a smaller number is quoted in the paper while a larger number is derived from their Table VII.

compared with experiment and our theoretical bulk modulus and its pressure derivative come out within a few percent of experiment, as is typical for LDA calculations. For the present work we use the pressure–lattice-constant relation from the Murnaghan equation of state,

$$P=\frac{B_0}{B_0'}\left[\left[\frac{a_0}{a}\right]^{3B_0'}-1\right], \tag{3.2}$$

where B_0 is the bulk modulus at zero pressure, B_0' is the pressure derivative of the bulk modulus at zero pressure, a_0 is the equilibrium lattice constant at zero pressure, and a is the lattice constant at pressure P. The experimental values for a_0, B_0, and B_0' are given in Table I. The present study uses the experimental (0 K) values for the zero-pressure lattice constant. We could have used the equilibrium lattice constant predicted by the LDA, which is about 1% smaller than experiment. This would result in less than a percent decrease in ϵ_1 for Si and about 7% decrease in ϵ_1 for Ge, based on our results for $d\ln[\epsilon_1(P)]/dP$. The conversion between $d\ln[\epsilon_1(a)]/d\ln a$ and $d\ln[\epsilon_1(P)]/dP$ at $P=0$ is given simply by

$$\frac{d\ln[\epsilon_1(P)]}{dP}=-\frac{1}{3B_0}\frac{d\ln[\epsilon_1(a)]}{d\ln a} \tag{3.3}$$

which comes from

$$\frac{d\ln a}{dP}=-\frac{1}{3B_0}\ . \tag{3.4}$$

One simple model is to assume the Clausius-Mossotti form for the dielectric function

$$\frac{\epsilon_1-1}{\epsilon_1+1}=\frac{4\pi}{3}n\alpha\ , \tag{3.5}$$

where n is the number of polarizable bonds (or ions) per unit volume and α is a polarizability. If α is assumed to be constant with pressure one may very simply compute a resulting $d\ln[\epsilon_1(P)]/dP$, which is positive due to the increase in n at higher density under compression and which therefore has the opposite sign to that observed experimentally. Apparently the effective α must decrease with compression, which is expected for a covalent material when bonding-antibonding splitting becomes larger with compression. In an effort to include the changing polarizability with compression, Harrison[76] treated the change in linear susceptibility χ_1 with a lattice constant in the bond orbital approximation. His result, given in Table VI, comes out with the correct sign and agrees fairly well with experiment for Si but is far too small for Ge. This results from oversimplification of the difference in the band structures of Si and Ge. Van Vechten's theory,[73] which also greatly simplifies the band structures of covalent semiconductors, also gets the sign of $d\ln[\epsilon_1(P)]/dP$ correct but cannot make quantitative predictions. Part of the problem here is the fitting of unknown parameters to describe the simplified models. An insight gained from the Van Vechten model is that the sign of $d\ln[\epsilon_1(P)]/dP$ is negative because the average gap increases faster with compression than the plasma frequency increases.

Our calculations for $\epsilon_1(a)$ at lattice constants away from the experimental lattice constant are carried out as described above, in particular using the same fixed self-energy shift Δ. Evidence for Δ remaining approximately constant under compression comes from successful LDA predictions of energy-gap variations compared with experiment[81,82] and directly from comparison of pressure-dependent energy gap in the LDA with quasiparticle predictions.[64] Our self-energy operator differs from the "false Darwin shift" or "spike" used by Alouani, Brey, and Christensen[78] in an important respect. Both methods are introduced to correct the LDA underestimate of the energy gap. The self-energy operator, however, does not modify the wave functions but merely shifts the conduction-band states by a constant amount, approximating the behavior observed in quasiparticle calculations.[25] The spike, which actually modifies the potential at the origin for each atom, does change the wave functions and therefore represents a departure from the accurate wave-function description available from the LDA. We obtain $d\ln[\epsilon_1(a)]/d\ln a$ by computing ϵ_1 at the experimental lattice constant a_0 and $\pm0.4\%$ of a_0 (about ±1 GPa equivalent pressure) and taking a numerical derivative. We examine the higher-pressure behavior of $\epsilon_1(P)$ at lattice constants approximating 3, 6, and 10 GPa as shown in Fig. 6 and Tables VII and VIII.

The first result we see is a rather close agreement of our computed $d\ln[\epsilon_1(P)]/dP$ with the most precise relatively low-pressure measurements of Vetter.[83,84] (Compare the first and last rows of Table VI.) In particular we reproduce the order-of-magnitude difference between Si and Ge even in the uncorrected LDA, and with the self-

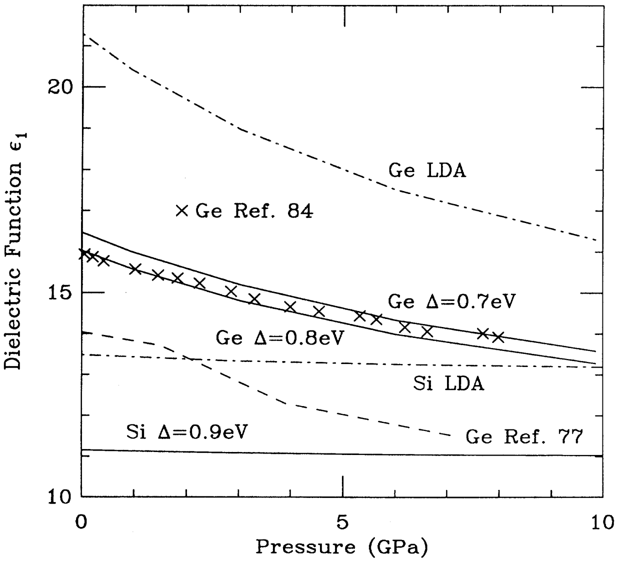

FIG. 6. Pressure-dependent dielectric function of Ge and Si from several sources. ×'s are experimental data for Ge from Ref. 85; the dashed line is the theoretical value for Ge from Ref. 78; the double-dashed lines are straight LDA for Ge and Si; solid lines are the LDA with a self-energy correction in the form of a scissors operator for Ge and Si. The LDA and self-energy corrected results are also given in Tables VII and VIII.

TABLE VII. Si: theoretical ϵ_1 at various lattice constants in the LDA and the LDA modified by the scissors form of the self-energy operator ($\Delta=0.9$ eV) versions of theory (plotted against equivalent pressure in Fig. 6).

a (Bohr)	a/a_0	P (GPa)	ϵ_1 (LDA)	ϵ_1 ($\Delta=0.9$ eV)
10.300 34	1.004	-1.16	13.554	11.197
10.259 3	1.000	0	13.484	11.160
10.218 26	0.996	1.22	13.422	11.128
10.166 97	0.991	2.84	13.335	11.095
10.074 63	0.982	6.06	13.258	11.051
9.972 04	0.972	10.1	13.187	11.028

energy operator corrected version of the theory we obtain quantitative agreement with experiment. We explain the order-of-magnitude-larger effect in Ge than Si with a picture similar to that proposed by Brust and Liu.[77] The dominant contributions to ϵ_1 in Si and Ge come from electronic transitions along the $\Gamma \rightarrow L$ line ($\Lambda_3 \rightarrow \Lambda_1$). Ge differs from Si in two respects: it has a far smaller direct gap [0.98 eV for Ge versus 3.4 eV for Si (Ref. 66)] so that the relative variation in the gap with presure is larger, and the pressure dependence of the direct band gap E_0 is much larger in Ge than in Si [dE_0/dP is -120 (10^{-12} eV/Pa) for Ge (Ref. 85) but only -5 to -6 (10^{-12} eV/Pa) for Si (Ref. 64)]. These two effects taken together explain the much larger pressure dependence of ϵ_1 in Ge than in Si. The simple models without sufficient details about the differences in the conduction bands of Si and Ge cannot reproduce this larger dependence.

The second result is quantitative agreement for the high-presure behavior of $\epsilon_1(P)$ for Ge in comparison with the measurements of Goñi *et al.*[85] In particular, our present theory reproduces the observed nonlinear dependence of $\epsilon_1(P)$ at higher pressures (above about 4 GPa), as shown in Fig. 6. On the basis of the present calculations we predict similar behavior for Si at pressures above about 4 GPa (Fig. 6 and Tables VII and VIII), albeit with variations in ϵ_1 about ten times less than for Ge. To our knowledge the analogous experiment has not yet been carried out for Si.

These calculations show that the scissors form of the self-energy operator cannot only correct the LDA to give reliable results for the electronic part of the dielectric constant ϵ_1, but can yield very good values of ϵ_1 as a function of lattice constant or pressure variations.

IV. CONCLUSION

We have extended the local-density approximation by adding a self-energy operator to the usual LDA Hamiltonian, for the purpose of improving the description of the dielectric function in solids. Without such a modification, the LDA predicts a dielectric function which is too large at low frequencies. We were motivated to try this extension of LDA because the highly successful *GW* calculations indicated that the LDA wave functions were a good description of the quasiparticles, although the excitation energies are given improperly by the Kohn-Sham eigenvalues. Although *GW* depends upon the dielectric function in a critical way, early attempts to use the improved eigenvalues of *GW* did not improve the dielectric function. We trace this to the need to modify the velocity operator (a Ward identity) as well as the eigenvalues in the expressions for the dielectric function.

In the usual point of view, the dielectric response in the static limit is regarded as known in-principle through density-functional theory; moreover, LDA is now thought to be an excellent representation of the exact density-functional theory. We find that the static dielectric constant may be found only through consideration of quasiparticle excitations which are usually considered not to be ground-state properties of the system. This is, at the very least, a puzzle.

The resulting frequency- and pressure-dependent $\epsilon_1(\omega)$ or $\epsilon_1(P)$ is improved remarkably over the LDA and agrees with experiment at the level of a few percent. Calculated results for $\epsilon_1(\omega)$ and $\epsilon_1(P)$ for Si and Ge are presented and compared with experiment. The results for $\epsilon_1(P)$ for Si constitute a prediction in that no measurements are known beyond about 1 GPa. The detailed expressions for linear response are presented in the appendixes. The present approach completely avoids matrix diagonalization and explicit summation over conduction-band states in solving for ϵ_1.

ACKNOWLEDGMENTS

We thank David DiVincenzo, Mark Hybertsen, Richard M. Martin, and John Rehr for helpful discus-

TABLE VIII. Ge: theoretical ϵ_1 at various lattice constants in the LDA and the LDA modified by the scissors form of the self-energy operator ($\Delta=0.7$, 0.8 eV) versions of theory (plotted against equivalent pressure in Fig. 6).

a (Bohr)	a/a_0	P (GPa)	ϵ_1 (LDA)	ϵ_1 ($\Delta=0.7$ eV)	ϵ_1 ($\Delta=0.8$ eV)
10.728 74	1.004	-0.90	22.280	16.981	16.511
10.686	1.000	0	21.293	16.468	16.029
10.643 26	0.996	0.95	20.412	15.996	15.584
10.557 77	0.988	3.03	18.973	15.194	14.799
10.450 91	0.978	6.01	17.514	14.333	13.990
10.333 36	0.967	9.86	16.294	13.580	13.278

sions. We thank Stefano de Gironcoli, Mark Hybertsen, and Steven Louie for making unpublished results available to us. We thank Michael Teter, John Wilkins, and Ken Wilson for support and encouragement. Support for this work was provided by the National Science Foundation under Grant No. DMR-8702002, the Cornell Electronic Structure Theory Group, and the Cornell National Supercomputer Facility. We thank the systems consultants of the CNSF for their continued aid throughout the project.

APPENDIX A: LONG-WAVELENGTH LIMIT

The expression for the induced electron density, Eq. (2.8), may be specialized to the case of a periodic solid replacing the label i by the band index n and the crystal momentum $\mathbf{k}$. Introducing Dirac notation, the electron density may be written

$$\begin{aligned}\delta n(\mathbf{r};\omega)=N\overline{\Omega}_0\int_{\mathrm{BZ}} d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n\mathbf{k}|\mathbf{r}\rangle &\\ \times\langle\mathbf{r}|G(\varepsilon_{n\mathbf{k}}+\omega)\phi^{\mathrm{SCF}}(\omega)|n\mathbf{k}\rangle &\\ +\langle n-\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|G(\varepsilon_{n\mathbf{k}}-\omega)\phi^{\mathrm{SCF}}(\omega)|n-\mathbf{k}\rangle & \end{aligned} \tag{A1}$$

using time-reversal symmetry ($(|n\mathbf{k}\rangle)^*=|n-\mathbf{k}\rangle$ and $\varepsilon_{n\mathbf{k}}=\varepsilon_{n-\mathbf{k}}$). N is the number of unit cells in the crystal. The constant $\overline{\Omega}_0=\Omega_0/(2\pi)^3$, where Ω_0 is the unit-cell volume. The domain of integration over $\mathbf{k}$ is a Brillouin zone.[86] The spin index is implicit in the sum on occupied states. The operator $\phi^{\mathrm{SCF}}(\omega)$ is defined to be $\int d\mathbf{r}'|\mathbf{r}'\rangle\phi^{\mathrm{SCF}}(\mathbf{r}';\omega)\langle\mathbf{r}'|$. For simplicity, the spin-orbit term is not considered in this discussion. The Green's operator G appearing in Eq. (A1) includes the occupied states in contrast to the Green's function $G_c(\mathbf{r},\mathbf{r}';\varepsilon_{n\mathbf{k}}\pm\omega)$ defined before Eq. (2.8); these forms are equivalent, as the reader may verify. The dummy variable $\mathbf{k}$ may be changed to $-\mathbf{k}$, leading to the expression

$$\begin{aligned}\delta n(\mathbf{r};\omega)=N\overline{\Omega}_0\int_{\mathrm{BZ}} d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n\mathbf{k}|\mathbf{r}\rangle &\\ \times\langle\mathbf{r}|G(\varepsilon_{n\mathbf{k}}\pm\hbar\omega)\phi^{\mathrm{SCF}}(\omega)|n\mathbf{k}\rangle\ , &\end{aligned} \tag{A2}$$

where the sum over ± simply represents adding the terms with the two indicated signs.

By Bloch's theorem, the ground-state wave functions are in the form of a plane wave with wave vector $\mathbf{q}$ times a function which is periodic in the Bravais lattice, referred to here as the "Bloch form." If the external potential has the form of a plane wave $\phi^{\mathrm{ext}}(\mathbf{r};\omega)=\phi_0 e^{i\mathbf{q}\cdot\mathbf{r}}\cos(\omega t)$, the electron density and the self-consistent potential will be in this form as well. To see this, consider the independent-particle susceptibility defined in Eq. (2.7) for the case of single-particle orbitals which are Bloch states. It may be seen that $\chi_0(\mathbf{r},\mathbf{r}';\omega)=\chi_0(\mathbf{r}+\mathbf{R},\mathbf{r}'+\mathbf{R};\omega)$ for Bravais lattice vectors $\mathbf{R}$. If $\phi^{\mathrm{SCF}}(\mathbf{r};\omega)$ is in the Bloch form, then from Eq. (2.2) the induced electron density has the property

$$\begin{aligned}\delta n(\mathbf{r}+\mathbf{R};\omega)&=\int d\mathbf{r}'\chi_0(\mathbf{r}+\mathbf{R},\mathbf{r}'+\mathbf{R};\omega)\phi^{\mathrm{SCF}}(\mathbf{r}'+\mathbf{R};\omega)\\ &=e^{i\mathbf{q}\cdot\mathbf{R}}\int d\mathbf{r}'\chi_0(\mathbf{r},\mathbf{r}';\omega)\phi^{\mathrm{SCF}}(\mathbf{r}';\omega)\\ &=e^{i\mathbf{q}\cdot\mathbf{R}}\delta n(\mathbf{r};\omega)\ ,\end{aligned} \tag{A3}$$

which shows $\delta n(\mathbf{r};\omega)$ has the Bloch form. If $\delta n(\mathbf{r};\omega)$ has the Bloch form, the induced Coulomb potential will have it [by a similar argument applied to Eq. (2.5)]. The functional derivative $\delta V_{\mathrm{xc}}(\mathbf{r})/\delta n(\mathbf{r})|_{n(\mathbf{r})=n_0(\mathbf{r})}$ is a local function of the ground-state electron density, and hence is periodic in the Bravis lattice. Hence the induced exchange-correlation potential defined by Eq. (2.6) inherits the Bloch form from $\delta n(\mathbf{r};\omega)$. The self-consistent potential is the sum of three terms each of which has the Bloch form. A tilde will denote the part of the corresponding function which is periodic in the Bravais lattice.

From Eq. (A2), the electron density will be in the form $\delta n(\mathbf{r};\omega)=\Omega_0^{-1}e^{i\mathbf{q}\cdot\mathbf{r}}\delta\tilde{n}(\mathbf{r};\omega)$ with the periodic part of the electron density given by

$$\delta\tilde{n}(\mathbf{r};\omega)=\overline{\Omega}_0\int_{\mathrm{BZ}} d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|(\varepsilon_{n\mathbf{k}}\pm\hbar\omega-H_{\mathbf{k}+\mathbf{q}})^{-1}\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\ , \tag{A4}$$

where $H_{\mathbf{k}}$ is the "$\mathbf{q}\cdot\mathbf{p}$" Hamiltonian[87,88]

$$H_{\mathbf{k}}=\tfrac{1}{2}(\mathbf{p}+\mathbf{k})^2+V_{\mathbf{k}}+\Sigma_{\mathbf{k}} \tag{A5}$$

in atomic units. Here, $\mathbf{p}=-i\nabla$, $V_{\mathbf{k}}$ is some external potential ($\mathbf{k}$ dependence in the Fourier transform will usually arise if a nonlocal external potential is used[2,3]), and $\Sigma_{\mathbf{k}}$ is the self-energy; at this stage no assumptions of the form of the self-energy have been made other than that it is a one-electron operator and is consistent with Bloch's theorem. The states $|n,\tilde{\mathbf{k}}\rangle$ are periodic functions in the Bravias lattice, related to the states $|n\mathbf{k}\rangle$ by

$$\langle\mathbf{r}|n\mathbf{k}\rangle=(N\Omega_0)^{-1/2}e^{i\mathbf{k}\cdot\mathbf{r}}\langle\mathbf{r}|n,\tilde{\mathbf{k}}\rangle\ .$$

The operator $\tilde{\phi}^{\mathrm{SCF}}(\omega)$ is defined analogously to $\phi^{\mathrm{SCF}}(\omega)$, i.e.,

$$\tilde{\phi}^{\mathrm{SCF}}(\omega)=\Omega_0^{-1}\int_0 d\mathbf{r}|\mathbf{r}\rangle\tilde{\phi}^{\mathrm{SCF}}(\mathbf{r};\omega)\langle\mathbf{r}|\ ,$$

with the integral to be taken over unit cell; the tilde again refers to the periodic part: $\phi^{\mathrm{SCF}}(\mathbf{r};\omega)=e^{i\mathbf{q}\cdot\mathbf{r}}\tilde{\phi}^{\mathrm{SCF}}(\mathbf{r};\omega)$. By our convention, $\Omega_0^{-1}\int_0 d\mathbf{r}|\mathbf{r}\rangle\langle\mathbf{r}|$ is an identity operation. The operator $(\varepsilon_{n,\tilde{\mathbf{k}}}\pm\hbar\omega-H_{\mathbf{k}+\mathbf{q}})^{-1}$ is a factor of $N\Omega_0$ larger than $(\varepsilon_{n\mathbf{k}}\pm\hbar\omega-H)^{-1}$ in its normalization. The integral implicit in the "bra-ket" of quantities involving the "$\mathbf{q}\cdot\mathbf{p}$" Hamiltonian is taken to be $\Omega_0^{-1}\int_0 d\mathbf{r}$.

The long-wavelength limit may be obtained via the Taylor expansion

$$(\varepsilon_{n\mathbf{k}}\pm\hbar\omega-H_{\mathbf{k}+\mathbf{q}})^{-1}=G_{n\mathbf{k}}(\pm\omega)+G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)+G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)$$
$$+G_{n\mathbf{k}}(\pm\omega)\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)+O(q^3) \qquad \text{(A6)}$$

with $G_{n\mathbf{k}}(\pm\omega)=(\varepsilon_{n\mathbf{k}}\pm\hbar\omega-H_{\mathbf{k}})^{-1}$. Usually, a "double commutator" is invoked at this stage of the argument, or the introduction of ill-defined symbols such as $\langle n\mathbf{k}|\mathbf{r}|n\mathbf{k}\rangle$. We find the present line simpler. Had we desired to use a Green's operator restricted to the unoccupied states, the expansion in Eq. (A6) would have required an expansion of the projection operator onto the conduction bands $P_{c\mathbf{k}+\mathbf{q}}$ to second order in q. In the present derivation, we will require only first-order expansion and then only because we explicitly consider a self-energy operator in the scissors form.

1. Zeroth order in q

To zeroth order in **q**, the induced electron density is given by

$$\delta\tilde{n}(\mathbf{r};\omega)=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle \times\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\ . \qquad \text{(A7)}$$

Consider the electron density induced by the external potential of amplitude ϕ_0 (or some other potential lacking a "local field") in zeroth order. In this case,

$$\begin{aligned}\delta\tilde{n}(\mathbf{r};\omega)&=\phi_0\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)|n,\tilde{\mathbf{k}}\rangle\\&=\phi_0\omega^{-1}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\pm\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|n,\tilde{\mathbf{k}}\rangle\\&=0\ .\end{aligned} \qquad \text{(A8)}$$

Hence, a long-wave perturbation induces no electron density in zeroth order in **q**.

The long-wave or $\mathbf{G}=0$ component of $\delta\tilde{n}(\mathbf{r};\omega)$ is found by integrating over a unit cell,

$$\begin{aligned}\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)&=\Omega_0^{-1}\int_0 d\mathbf{r}\,\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\\&=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\\&=\omega^{-1}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\pm\langle n,\tilde{\mathbf{k}}|\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\\&=0\ .\end{aligned} \qquad \text{(A9)}$$

This is the same result as mentioned at the end of Sec. II, that no monopole may be induced. In the present language, no perturbation induces a long-wave electron density in zeroth order in **q**.

However, a short-wave perturbation induces a short-wave electron density in zeroth order in **q**, by Eq. (A7), or equivalently by Eq. (2.15) with $\mathbf{q}=0$.

2. First order in q

Consider next the long-wave response to a long-wave perturbation to first order in **q**. The zeroth-order term has been shown to vanish, and so it suffices to replace $(\varepsilon_{n\mathbf{k}}\pm\hbar\omega-H_{\mathbf{k}+\mathbf{q}})^{-1}$ by the first-order term given in Eq. (A6). The general expression of Eq. (A4) reduces to

$$\begin{aligned}\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)&=\phi_0\Omega_0^{-1}\int_0 d\mathbf{r}\,\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)|n,\tilde{\mathbf{k}}\rangle\\&=2\phi_0\omega^{-2}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\\&=2\phi_0\omega^{-2}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\mathbf{q}\cdot\nabla_{\mathbf{k}}\langle n,\tilde{\mathbf{k}}|H_{\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle\\&=0\ .\end{aligned} \qquad \text{(A10)}$$

The Hellmann-Feynman theorem[89] is invoked in proceedings from line 2 to line 3 in Eq. (A10). The final cancellation occurs by noting $\langle n,\tilde{\mathbf{k}}|H_{\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle=\varepsilon_{n\mathbf{k}}$ and considering $\pm\mathbf{k}$ pairs. These considerations indicate the long-wave response to a long-wave perturbation vanishes to first order in $\mathbf{q}$.

For a long-wave perturbation, the short-wave response is given by

$$\begin{aligned}\delta\tilde{n}(\mathbf{r};\omega)&=\phi_0\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)|n,\tilde{\mathbf{k}}\rangle\\&=\phi_0\omega^{-1}\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\pm\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\\&=-2\phi_0\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\ .\end{aligned}\tag{A11}$$

In proceeding to the last line of the equation, the identity

$$G_{n\mathbf{k}}(\omega)-G_{n\mathbf{k}}(-\omega)=-2\omega G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tag{A12}$$

is exploited. The use of this identity eliminates a numerical finite difference in the second line of Eq. (A11); moreover, the static limit is manifest. The term in $\mathbf{q}\cdot\mathbf{k}$ implicit in $(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})$ reduces to $\phi_0\omega^{-2}\mathbf{q}\cdot\mathbf{k}|\langle\mathbf{r}|n,\tilde{\mathbf{k}}\rangle|^2$, so will sum to zero in $\pm\mathbf{k}$ pairs. The other terms are nonvanishing. So a long-wave perturbation creates a short-wave response in first order in $\mathbf{q}$.

The long-wave response to a short-wave perturbation in first order is given by the expression

$$\begin{aligned}\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)&=\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\\&=\omega^{-1}\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\pm\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\\&=-2\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\ ;\end{aligned}\tag{A13}$$

again, Eq. (A12) is used. The term in $\mathbf{q}\cdot\mathbf{k}$ reduces to $\omega^{-2}\mathbf{q}\cdot\mathbf{k}\langle n,\tilde{\mathbf{k}}|\tilde{\phi}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle$, and therefore sums to zero in ± pairs. The other terms survive, so a short-wave perturbation creates a long-wave response to first order in $\mathbf{q}$.

3. Second order in q

It remains only to consider the long-wavelength response to a long-wavelength perturbation. In the previous sections the zeroth- and first-order terms were seen to vanish. Hence, we may apply only the second-order term in Eq. (A6) to the general expression Eq. (A4). The result is

$$\begin{aligned}\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)&=\phi_0\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|G_{n\mathbf{k}}(\pm\omega)[(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})+\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})]G_{n\mathbf{k}}(\pm\omega)|n,\tilde{\mathbf{k}}\rangle\\&=\phi_0\omega^{-2}\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|[(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})+\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle\ .\end{aligned}\tag{A14}$$

To understand the static limit, we derive the f-sum rule for crystals[90] in the present context. One may expand the eigenvalue $\varepsilon_{n\mathbf{k}+\mathbf{q}}$ to second order in $\mathbf{q}$ using either perturbation theory or a Taylor expansion, leading to the equation

$$\langle n,\tilde{\mathbf{k}}|[(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})+\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle=\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}})\ ,\tag{A15}$$

where $G_{n\mathbf{k}}=(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})^{-1}$, and the pseudoinverse is intended. Because $\varepsilon_{n\mathbf{k}}$ is periodic in the reciprocal lattice its derivatives have zero value when integrated over a complete Brillouin zone. In the case of a semiconductor or insulator, this implies

$$\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|[(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})+\tfrac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})]|n,\tilde{\mathbf{k}}\rangle=0\ .\tag{A16}$$

For the metal case, see Ref. 91. Application of Eq. (A16) to Eq. (A14) leads to

$$\begin{aligned}\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)&=\phi_0\omega^{-2}\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\{(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})[G_{n\mathbf{k}}(\pm\omega)-G_{n\mathbf{k}}](\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})\}|n,\tilde{\mathbf{k}}\rangle\\&=2\phi_0\overline{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{k}H_{\mathbf{k}})G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\end{aligned}\tag{A17}$$

utilizing the relation

$$G_{n\mathbf{k}}(\omega)+G_{n\mathbf{k}}(-\omega)-2G_{n\mathbf{k}}=2\omega^2 G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega) \tag{A18}$$

which is similar to Eq. (A12). The use of Eq. (A18) eliminates a second-order numerical finite difference in the first line of Eq. (A17) and makes the static limit manifest. Note that $\frac{1}{2}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})$ is removed from the problem analytically.

APPENDIX B: TIME-REVERSAL SYMMETRY

In deriving the long-wavelength limit, Eqs. (A7), (A11), (A13), and (A17), we chose to define $G_{n\mathbf{k}}(\pm\omega)$ to include all states—occupied and unoccupied. However, it is an elementary exercise to exploit time-reversal symmetry [specifically, $\langle \mathbf{r}|n,\tilde{\mathbf{k}}\rangle=\langle \mathbf{r}|n-\tilde{\mathbf{k}}\rangle^*$, $\varepsilon_{n\mathbf{k}}=\varepsilon_{n-\mathbf{k}}$, and the odd parity of $\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle$] to show that it is permissible to exclude all occupied states from the Green's function in these formulas. In our implementation, we exclude the occupied states from these sums.

Time-reversal symmetry may be used to simplify the main results, Eqs. (A7), (A11), (A13), and (A17). Below the absorption threshold, the case of interest here, the short-wave charge density induced by the long-wave potential, Eq. (A11), becomes

$$\delta\tilde{n}(\mathbf{r};\omega)=-2i\phi_0\,\mathrm{Im}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle \mathbf{r}|G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\ , \tag{B1}$$

taking ϕ_0 to be a real quantity. Neither the Coulomb interaction of Eq. (2.5) nor the exchange-correlation term in Eq. (2.6) introduce any additional phase. Hence, the short-wave potential induced by the long-wave potential is 90° out of phase. Assuming that $\tilde{\phi}^{\mathrm{SCF}}(\omega)$ is purely imaginary, the short-wave charge density induced by the short-wave potential, described by Eq. (A7) becomes

$$\delta\tilde{n}(\mathbf{r};\omega)=i\,\mathrm{Re}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle \mathbf{r}|G_{n\mathbf{k}}(\pm\omega)[\mathrm{Im}\tilde{\phi}^{\mathrm{SCF}}(\omega)]|n,\tilde{\mathbf{k}}\rangle\ . \tag{B2}$$

As Eq. (B2) involves no phase change for the purely imaginary $\tilde{\phi}^{\mathrm{SCF}}(\omega)$, the assumption of a purely imaginary $\tilde{\phi}^{\mathrm{SCF}}(\omega)$ is internally consistent. The long-wave charge density induced by the short-wave potential, described by Eq. (A13), becomes

$$\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)=2\,\mathrm{Im}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)[\mathrm{Im}\tilde{\phi}^{\mathrm{SCF}}(\omega)]|n,\tilde{\mathbf{k}}\rangle\ . \tag{B3}$$

Similarly, the long-wave charge density induced by the long-wave potential, given in Eq. (A17), becomes

$$\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)=2\phi_0\mathrm{Re}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle\ . \tag{B4}$$

Between Eqs. (B3) and (B4) it may be seen that the total induced long-wave charge is in phase with the long-wave potential. The Coulomb interaction of Eq. (2.5) does not change the phase relation. The exchange-correlation term does not enter the long-wave self-consistency procedure as it affects the long-wave potential in the second order in q, whereas (by convention) the zeroth order is the leading term. The situation is summarized graphically in Fig. 1.

As argued above, in the absence of absorptions, a potential chosen by convention to be real for long wavelengths will be purely imaginary for short waves. The physical significance of the imaginary part is seen by considering an external field of the form $\phi^{\mathrm{ext}}(\mathbf{r};\omega)=\phi_0\cos\mathbf{q}\cdot\mathbf{r}$. The electron density induced from Eqs. (B3) and (B4) will be of the form $\delta n(\mathbf{r};\omega)=\phi_0[c(\omega)\cos\mathbf{q}\cdot\mathbf{r}+\tilde{f}(\mathbf{r};\omega)\sin\mathbf{q}\cdot\mathbf{r}]$, where $c(\omega)$ is a real function and $\tilde{f}(\mathbf{r};\omega)$ is a real function which is periodic in the lattice whose average value is zero. For the long-wave response, the charge density is greatest at the minimum of the potential, whereas the short-wave response, which is a polarization, is largest where the potential gradient is largest.

APPENDIX C: TENSOR PROPERTIES

The dielectric function transforms as a symmetric second-rank tensor as the direction of the electric field ($\hat{q}$ in our longitudinal formulation) is varied. The purpose of this section is to show that the angular dependence of the equations of Appendix A, Eqs. (A7), (A11), (A13), and (A17), together with Poisson's equation (2.5) and the induced exchange-correlation potential, Eq. (2.6), lead to a dielectric matrix with this transformation property. A computationally convenient formula for the macroscopic dielectric matrix is presented.

The short-wave electron density induced by a long-wave perturbation, Eq. (A11), transforms like the vector $\mathbf{q}$. That is, we may introduce the quantities

$$\delta\tilde{n}_{\mu}^{(l)}(\mathbf{r};\omega)=-2\phi_0\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\hat{x}_{\mu}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle \tag{C1}$$

in terms of which

$$\delta\tilde{n}(\mathbf{r};\omega)=\sum_{\mu}q_{\mu}\delta\tilde{n}_{\mu}(\mathbf{r};\omega)\ . \tag{C2}$$

The superscript (l) refers to a long-wave source; (s) will refer to a short-wave source. The $\hat{x}_{\mu}$ form a set of three orthogonal unit vectors. The solution of the short-wave SCF equations (A7), (2.5), and (2.6) does not have any additional reference to $\mathbf{q}$; as these equations are linear in $\delta\tilde{n}(\mathbf{r};\omega)$, both the short-wave SCF electron density and potentials will transform like a vector under transformations of $\mathbf{q}$. It is convenient (if counterintuitive) to introduce the quantities $\tilde{\phi}_{\mu}^{\mathrm{SCF}}(\mathbf{r};\omega)$, a vector of short-wave scalar potentials (i.e., potentials with zero average) related to the short-wave part of $\tilde{\phi}^{\mathrm{SCF}}(\mathbf{r};\omega)$ by $\tilde{\phi}^{\mathrm{SCF}}(\mathbf{r};\omega)=\sum_{\mu}q_{\mu}\tilde{\phi}_{\mu}^{\mathrm{SCF}}(\mathbf{r};\omega)$. (Of course, an arbitrarily imposed short-wave potential need have no special transformation properties.) Equation (A7), the angular dependence of the short-wavelength electron density induced from a short-wavelength potential, may be recast as

$$\delta\tilde{n}_{\mu}^{(s)}(\mathbf{r};\omega)=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{\mu}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\ , \tag{C3}$$

where the relationship given by Eq. (C2) holds. The total short-wave electron density is given by $\delta\tilde{n}_{\mu}(\mathbf{r};\omega)=\delta\tilde{n}_{\mu}^{(l)}(\mathbf{r};\omega)+\delta\tilde{n}_{\mu}^{(s)}(\mathbf{r};\omega)$. Each $\tilde{\phi}_{\mu}^{\mathrm{SCF}}(\mathbf{r};\omega)$ is defined by a single $\delta\tilde{n}_{\mu}(\mathbf{r};\omega)$ through Poisson's equation (2.5) and the linearized exchange-correlation interaction of Eq. (2.6).

Consider the induced long-wavelength electron density. This electron density may be induced directly from the long-wavelength perturbation via Eqs. (A14) or indirectly via the short-wave response [Eqs. (A11), (2.5), (2.6), (A7), and (A13)]. The scheme is summarized in Fig. 1. If induced directly, the interaction is of the form

$$\delta\tilde{n}_{\mu\nu}^{(l)}(\omega)=\phi_0\omega^{-2}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n\pm}^{\mathrm{occ}}[\langle n,\tilde{\mathbf{k}}|(\hat{x}_{\mu}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\pm\omega)(\hat{x}_{\nu}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle+\tfrac{1}{2}\delta_{\mu\nu}] \tag{C4}$$

with

$$\Omega_0^{-1}\int_0 d\mathbf{r}\,\delta\tilde{n}(\mathbf{r};\omega)=\sum_{\mu,\nu}q_{\mu}q_{\nu}\delta\tilde{n}_{\mu\nu}(\omega) \tag{C5}$$

as may be gleaned by inspection of Eq. (A14). [As Eq. (A14) rather than Eq. (A17) is used, the formula applies to metals and insulators alike.] If induced indirectly, i.e., via Eq. (A13), the interaction may be seen to be of the form

$$\delta\tilde{n}_{\mu\nu}^{(s)}(\omega)=-2\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\langle n,\tilde{\mathbf{k}}|(\hat{x}_{\mu}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}_{\nu}^{\mathrm{SCF}}(\omega)|n,\tilde{\mathbf{k}}\rangle\ , \tag{C6}$$

where the relationship given by Eq. (C5) holds. The total long-wave number density is given by $\delta\tilde{n}_{\mu\nu}(\omega)=\delta\tilde{n}_{\mu\nu}^{(l)}(\omega)+\delta\tilde{n}_{\mu\nu}^{(s)}(\omega)$. Hence, the long-wave-induced number density transforms like a second-rank tensor under transformations of $\mathbf{q}$. The tensors of Eqs. (C4) and (C6) are symmetric: $\delta\tilde{n}_{\mu\nu}^{(l)}(\omega)$ is manifestly symmetric, and replacement of $\tilde{\phi}_{\mu}^{\mathrm{SCF}}(\omega)$ in the expression for $\delta\tilde{n}_{\mu\nu}^{(s)}(\omega)$ leads to an infinite sum of manifestly symmetric operators.

1. Macroscopic dielectric functions

The self-consistent potential may be written as in terms of the external potential and the induced charge as

$$\tilde{\phi}^{\mathrm{SCF}}(\omega)=\tilde{\phi}^{\mathrm{ext}}+v\sum_{\mu,\nu}q_{\mu}q_{\nu}\delta\tilde{n}_{\mu\nu}(\omega)\ , \tag{C7}$$

where $v=4\pi e^2/q^2$ is the (isotropic) Coulomb interaction. (The induced exchange-correlation potential need not be considered, as it is zeroth order in $\mathbf{q}$, whereas the Coulomb part is of order -2.) Equivalently, we may write

$$\sum_{\mu,\nu}q_{\mu}q_{\nu}\tilde{\phi}_{\mu\nu}^{\mathrm{SCF}}(\omega)=\sum_{\mu,\nu}q_{\mu}q_{\nu}[\tilde{\phi}_{\mu\nu}^{\mathrm{ext}}+v\delta\tilde{n}_{\mu\nu}(\omega)]\ , \tag{C8}$$

where $\tilde{\phi}(\omega)=\sum_{\mu,\nu}q_{\mu}q_{\nu}\tilde{\phi}_{\mu\nu}(\omega)$, for both the SCF and external potentials, and $\tilde{\phi}_{\mu\nu}^{\mathrm{ext}}=\phi_0\delta_{\mu\nu}$. As the vector $\mathbf{q}$ is arbitrary, it is permissible to write an equation for the reduced quantities

$$\tilde{\phi}_{\mu\nu}^{\mathrm{SCF}}(\omega)=\tilde{\phi}_{\mu\nu}^{\mathrm{ext}}+v\delta\tilde{n}_{\mu\nu}(\omega)\ . \tag{C9}$$

Note that the short-wave SCF potential transforms with $\mathbf{q}$ like a vector, but the long-wave SCF potential transforms like a symmetric matrix. The symmetries reported here are identical to those presented by others for the SCF case[42] and the exact case.[51,92]

Switching just for this paragraph to a notation in which the symbols represent long-wave quantities with 3×3 matrices, the long-wave SCF equation may be solved explicitly as

$$\tilde{\phi}^{\mathrm{SCF}}=\sum_{n=0}^{\infty}(v\chi_0)^n\tilde{\phi}^{\mathrm{ext}}\ . \tag{C10}$$

Summation of the geometric series leads to an expression for the inverse dielectric matrix

$$\epsilon^{-1}=(I-v\chi_0)^{-1} \tag{C11}$$

which implies

$$\epsilon=I-v\chi_0 \,. \tag{C12}$$

These relationships are the matrix generalizations of those holding for the electron gas in the RPA.[93] In the present context, the macroscopic χ_0 matrix[94] is most easily found from the relation

$$\delta\tilde{n}^{\,\text{non-SCF}}=\chi_0\tilde{\phi}^{\,\text{ext}} \,. \tag{C13}$$

Computationally, we find $\delta\tilde{n}^{\,\text{non-SCF}}$ from the external potential as the sum of direct and indirect (i.e., local field) contributions. The macroscopic dielectric matrix is then determined immediately from Eq. (C12). We must screen the induced short-wave charge densities and potentials numerically, but the long-wave screening is done analytically. As we have formulated the problem with 3×3 matrices, this holds true even in low symmetry crystals for which **D** is not parallel to **E**.

2. Point-group symmetry

We wish to consider only the symmorphic symmetry operators of the point group of the Hamiltonian.[95] We define the irreducible analogous of Eqs. (C1), (C3), (C4), and (C6) to be $\delta\tilde{n}_{\mu}^{(l,\text{RBZ})}(\mathbf{r};\omega)$, $\delta\tilde{n}_{\mu}^{(s,\text{RBZ})}(\mathbf{r};\omega)$, $\delta\tilde{n}_{\mu\nu}^{(l,\text{RBZ})}(\omega)$, and $\delta\tilde{n}_{\mu\nu}^{(s,\text{RBZ})}(\omega)$, where the domain of integration is restricted to the reduced Brillouin zone (RBZ). We use the notation RBZ rather than irreducible Brillouin zone (IBZ) to emphasize that the nonsymmorphic symmetry operators are not considered. It is our opinion that the inclusion of nonsymmorphic elements in the integrations of terms involving $\tilde{\phi}^{\,\text{SCF}}(\omega)$ reduces computational efficiency, because apparently a separate term must be defined for every distinct value of $S\tau$, where S is any rotation operator associated with the group and τ is a nonprimitive translation from the nonsymmorphic group. (The situation is different if only the momentum operator is symmetrized.) The use of time-reversal symmetry as discussed in Appendix B reduces the computational requirements. In the case of the diamond structure, omitting the nonsymmorphic symmetry operators but including time reversal leads to as many symmetry operators as considering nonsymmorphic symmetry operators explicitly. Under these terms, it is straightforward to show

$$\delta\tilde{n}_{\mu}(\mathbf{r};\omega)=\sum_{S\nu} S_{\mu\nu}^{-1}\,\delta\tilde{n}_{\nu}^{(\text{RBZ})}(S^{-1}\mathbf{r};\omega) \tag{C14}$$

and

$$\delta\tilde{n}_{\mu\nu}(\omega)=\sum_{S\mu'\nu'} S_{\mu\mu'}^{-1}S_{\nu\nu'}^{-1}\,\delta\tilde{n}_{\mu'\nu'}^{(\text{RBZ})}(\omega) \tag{C15}$$

for both the short- and long-wave parts of the charge density.

---

[1] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

[2] S. Baroni and R. Resta, Phys. Rev. B **33**, 7017 (1986).

[3] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **35**, 5585 (1987).

[4] S. de Gironcoli, S. Baroni, and R. Resta, Phys. Rev. Lett. **62**, 2853 (1989).

[5] Z. H. Levine and D. C. Allan, Phys. Rev. Lett. **63**, 1719 (1989).

[6] J. P. Walter and M. L. Cohen, Phys. Rev. B **2**, 1821 (1970).

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

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

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

[10] R. Resta and K. Kunc, Phys. Rev. B **34**, 7146 (1986).

[11] W. R. L. Lambrecht and B. Segall, Phys. Rev. B **40**, 7793 (1989).

[12] Indium compounds are apparently an exception; however, Ref. 4 suggests that this is an artifact due to freezing the In 4*d* electrons.

[13] See Ref. 8 in Z. H. Levine and D. C. Allan, Phys. Rev. Lett. **63**, 1719 (1989).

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

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

[16] O. Gunnarsson and R. O. Jones, Phys. Scr. **21**, 394 (1980).

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

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

[19] Y.-T. Shen, D. M. Bylander, and L. Kleinman, Phys. Rev. B **36**, 3465 (1987).

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

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

[22] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **38**, 4033 (1988).

[23] F. Bechstedt, R. Del Sole, and F. Manghi, J. Phys. Condens. Matter **1**, SB75 (1989).

[24] R. Daling and W. van Haeringen, Phys. Rev. B **40**, 11 659 (1989).

[25] M. S. Hybertsen and S. G. Louie, Ref. 21, p. 5937.

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

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

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

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

[30] G. D. Mahan, *Many-Particle Physics* (Plenum, New York, 1981), p. 618.

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

[32] C. S. Wang and W. E. Pickett, Phys. Rev. Lett. **51**, 597 (1983); W. E. Pickett and C. S. Wang, Phys. Rev. B **30**, 4719 (1984); Int. J. Quantum Chem. **20**, 299 (1986).

[33] A. E. Carlsson, Phys. Rev. B **31**, 5178 (1985).

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

[35] H. Ehrenreich and M. H. Cohen, Phys. Rev. **115**, 786 (1959).

[36] S. A. Orszag, J. Comput. Phys. **37**, 70 (1980); in particular, see Sec. 6.

[37] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978), p. 369.

[38] F. S. Acton, *Numerical Methods That Work* (Harper & Row, New York, 1970), p. 216. Acton uses the term "Aitken extrapolation" rather than "Shanks transformation."

[39] R. M. Martin, Phys. Rev. B **9**, 1998 (1974). Martin has emphasized the importance of using the long-wavelength limit in calculations of electric polarization in crystals. Many authors incorrectly assume $\mathbf{P}=\Omega_0^{-1}\int_{\text{cell}}\mathbf{r}\rho(\mathbf{r})d\mathbf{r}$, a formula which does not hold for extended systems despite its successful application for molecular aggregates in the Clausius-Mossotti relation.

[40] A. Zangwill and P. Soven, Phys. Rev. A **21**, 1561 (1980).

[41] A. Zangwill, in *Atomics Physics 8*, edited by I. Lindgren, A. Rosén, and S. Svanberg (Plenum, New York, 1983), p. 339.

[42] S. L. Adler, Phys. Rev. **126**, 413 (1962).

[43] N. Wiser, Phys. Rev. **129**, 62 (1963).

[44] J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975), p. 221.

[45] E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).

[46] E. K. U. Gross and W. Kohn, Phys. Rev. Lett. **55**, 2850 (1985).

[47] N. Iwamoto and E. K. U. Gross, Phys. Rev. B **35**, 3003 (1987).

[48] P. Gies and R. R. Gerhardts, Phys. Rev. B **36**, 4422 (1987).

[49] G. D. Mahan, Phys. Rev. A **22**, 1780 (1980).

[50] S. Baroni and P. Giannozzi, Phys. Rev. Lett. **58**, 1861 (1987).

[51] V. Ambegoakar and W. Kohn, Phys. Rev. **117**, 423 (1960).

[52] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[53] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[54] C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963), p. 187.

[55] Combine terms in $(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})|n,\tilde{\mathbf{k}}\rangle$. Multiply through by $\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}}$. Make the replacement $H_{\mathbf{k}}^{\text{LDA}}=H_{\mathbf{k}}-\Delta_{\mathbf{k}}P_{c\mathbf{k}}$. Apply $G_{n\mathbf{k}}^{\text{LDA}}$.

[56] D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973); H. J. Monkhorst and J. D. Pack, *ibid.* **13**, 5188 (1976).

[57] D. Y. Smith and D. L. Dexter, *Progress in Optics 10*, edited by W. Wolf (North-Holland, Amsterdam, 1972), Sect. 3.1.

[58] D. Y. Smith and E. Shiles, Phys. Rev. B **17**, 4689 (1978); D. Y. Smith, in *Handbook of Optical Constants of Solids*, edited by E. D. Palik (Academic, New York, 1985), Chap. 3.

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

[60] D. C. Allan and M. P. Teter, Phys. Rev. Lett. **59**, 1136 (1987).

[61] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12 255 (1989).

[62] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[63] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1981).

[64] X. Zhu, S. Fahy, and S. G. Louie, Phys. Rev. B **39**, 7840 (1989); **40**, 5821(E) (1989).

[65] M. S. Hybertsen (private communication).

[66] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors* (Springer-Verlag, Berlin, 1988), pp. 86 and 96.

[67] S. H. Wemple and M. DiDomenico, Jr., Phys. Rev. Lett. **23**, 1156 (1969); Phys. Rev. B **3**, 1338 (1971).

[68] Z. H. Levine and P. Soven, Phys. Rev. A **29**, 625 (1984); **35** 3964(E) (1987); Z. H. Levine, *ibid.* **36**, 1181 (1987).

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

[70] D. E. Aspnes and A. A. Studna, Phys. Rev. B **27**, 985 (1983); reprinted in D. F. Edwards, in *Handbook of Optical Constants of Solids*, edited by E. D. Palik (Academic, New York, 1985), p. 547. Edwards's Ref. 20 apparently was intended to be Aspnes and Studna rather than Aspnes and Theetnen.

[71] H. W. Verleur, J. Opt. Soc. Am. **58**, 1356 (1968).

[72] G. K. M. Thatupalli and S. G. Thomlin, J. Phys. C **10**, 467 (1977).

[73] J. A. Van Vechten, Phys. Rev. **182**, 891 (1969).

[74] J. C. Phillips, Rev. Mod. Phys. **42**, 317 (1970).

[75] V. Heine and R. O. Jones, J. Phys. C **2**, 719 (1969).

[76] W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Fransisco, 1980), p. 115.

[77] D. Brust and L. Liu, Phys. Rev. **154**, 647 (1967).

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

[79] F. D. Murnaghan, Proc. Natl. Acad. Sci. U.S.A. **30**, 244 (1944).

[80] H. J. McSkimin and P. Andreatch, Jr., J. Appl. Phys. **34**, 651 (1963); **35**, 2161 (1964).

[81] K. J. Chang, S. Froyen, and M. L. Cohen, Solid State Commun. **50**, 105 (1984).

[82] C. O. Rodriguez, E. L. Peltzer, Y. Blanca, and O. M. Cappannini, Solid State Commun. **56**, 575 (1985).

[83] N. J. Trappeniers, R. Vetter, and H. A. R. de Bruin, Physica (Utrecht) **45**, 619 (1970).

[84] R. Vetter, Phys. Status Solidi A **8**, 443 (1971).

[85] A. R. Goñi, K. Syassen, K. Strössner, and M. Cardona, Semiconductor. Sci. Technol. **4**, 246 (1989); A. R. Goñi, K. Syassen, and M. Cardona, Phys. Rev. B **41**, 10 104 (1990).

[86] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976), p. 143.

[87] J. Bardeen, J. Chem. Phys. **6**, 367 (1938).

[88] F. Wooten, *Optical Properties of Solids* (Academic, New York, 1972), Appendix H; Ref. 86, p. 140.

[89] H. Hellmann, *Einfuhrung in die Quantenchemie* (Leibzig, Deuticke, 1937); R. P. Feynman, Phys. Rev. **56**, 340 (1939).

[90] F. Wooten, Ref. 8, Appendix D.

[91] For a metal, at least one band is partially occupied, so the ω^{-2} term of Eq. (A14) has a nonzero coefficient in general. The coefficient is given by

$$\begin{aligned}\sum_{n\pm}^{\text{occ}}\bar{\Omega}_0\int_{\text{BZ}}d\mathbf{k}\,\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}) &= \tfrac{1}{2}\sum_{n\pm}^{\text{occ}}\bar{\Omega}_0\oint_{\text{FS}}dS_{\mathbf{k}}(\mathbf{q}\cdot\hat{n})(\mathbf{q}\cdot\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}})\\ &= \tfrac{1}{2}\sum_{n\pm}^{\text{occ}}\bar{\Omega}_0\oint_{\text{FS}}dS_{\mathbf{k}}|\mathbf{q}\cdot\hat{n}|^2|\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}|dS\\ &\geq 0\,,\end{aligned}$$

where the volume integral over **k** includes only occupied states, FS represents the Fermi surface, and $\hat{n}$ is an unit vector normal to the surface and directed outward. The unit normal and $\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}$ must be aligned, by the definition of the Fermi surface. Hence the coefficient of the divergent term is positive for all metals, except if the pathological condition $|\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}|=0$ obtains on the entire Fermi surface. The limit $q\rightarrow 0$ is taken before $\omega\rightarrow 0$ in this derivation.

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

[93] D. Pines, *Elementary Excitation in Solids* (Benjamin, New York, 1964), pp. 137–138.

[94] The matrix χ_0 appearing here, the long-wave part of the mi-

croscopic response function, is related to the usual macroscopic susceptibility $\overline{\chi}$ which appears in the constitutive relation $\mathbf{P}=\overline{\chi}\mathbf{E}$ by $\overline{\chi}=-(e^2/q^2)\chi_0$.

[95] J. F. Cornwell, *Group Theory and Electronic Energy Bands in Solids* (Wiley, New York, 1969).

[86] D. K. Biegelsen, Phys. Rev. Lett. **32**, 1196 (1974); Phys. Rev. B **12**, 2427 (1975).

[97] M. Cardona, W. Paul, and H. Brooks, J. Phys. Chem. Solids **8**, 204 (1959).

[98] R. L. Adams and D. A. Pinnow, J. Appl. Phys. **41**, 2765 (1970).

[99] Y. F. Tsay, S. S. Mitra, and B. Bendow, Phys. Rev. B **10**, 1476 (1974).

[100] N. P. Barnes and M. S. Piltch, J. Opt. Soc. Am. **69**, 178 (1979); reprinted in R. F. Potter, in *Handbook of Optical Constants of Solids,* edited by E. D. Palik (Academic, New York, 1985), p. 465.