PHYSICAL REVIEW B VOLUME 44, NUMBER 23 15 DECEMBER 1991-I

Optical second-harmonic generation in III-V semiconductors: Detailed formulation and computational results

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
and Department of Physics, The Ohio State University, Columbus, Ohio 43210-1106

(Received 13 May 1991)

In an earlier paper [Phys. Rev. Lett. **66**, 41 (1991)], we calculated both the dielectric constant (ϵ_∞) and the nonlinear optical susceptibilities for second-harmonic generation ($\chi^{(2)}$) in the static limit for AlP, AlAs, GaP, and GaAs in the local-density approximation with and without a self-energy correction in the form of a "scissors operator," including local-field effects. In this paper, we expand our presentation of this calculation. Agreement with experiment to within 15% for the nonlinear susceptibility is demonstrated where experiments are available (GaP and GaAs); the dielectric constants are in no worse than 4% agreement with experiment. The "virtual hole" contributions are reformulated to avoid large numerical cancellations in the case of near degeneracies. The "virtual electron" terms dominate over the "virtual hole" terms by about one order of magnitude. Local-field corrections are smaller than the main terms by about one order of magnitude. The formulas needed to apply a self-energy correction in the form of a "scissors operator" to this problem are presented. The addition of a self-energy correction requires a renormalization of the velocity operator; a failure to include the velocity-operator renormalization leads to a factor-of-2 correction to $\chi^{(2)}$, destroying the good agreement with experiment. The neglect of the short-wave charge induced at the second-harmonic frequency is justified. The f-sum rule and another, related sum rule for second-harmonic generation is well satisfied numerically. For well-converged results, a plane-wave-basis-set energy cutoff of 9–12 hartrees is required for GaAs, but only eigenfunctions with eigenvalues less than about 1–2 hartrees need be included. Using a special-points integration scheme, 10 points are not sufficient, 28 points are typically adequate, and for the material considered with the smallest band gap, GaAs, 60 special points are marginally desirable.

I. INTRODUCTION

Recently, we presented the results of a calculation of the nonlinear susceptibility for optical second-harmonic generation in III-V semiconductors.[1] This work followed the presentation by one of us of band theoretic formulas for optical second-harmonic generation including local-field effects.[2] Our work was distinguished from earlier work in that we predicted both linear and nonlinear optical susceptibilities from the band structure; i.e., we did not use linear-response parameters to predict nonlinear response. The III-V semiconductors AlP, AlAs, GaP, and GaAs are a good choice because they are among the simplest real substances for which second-harmonic generation in the bulk is allowed. Moreover, they are the parent compounds of certain celebrated semiconductor heterostructures. Extremely large values of the nonlinear susceptibility for second-harmonic generation have been observed in Al-Ga-As heterostructures recently.[3]

In Ref. 1, the response calculation was based on a Kohn-Sham local-density-approximation band structure, corrected for self-energy effects by a "scissors" operator.[4,5] That is, our Hamiltonian is

$$H_{\mathbf{k}} = H_{\mathbf{k}}^{\mathrm{LDA}} + \Sigma_{\mathbf{k}} = H_{\mathbf{k}}^{\mathrm{LDA}} + \Delta_{\mathbf{k}} P_{c\mathbf{k}} , \tag{1}$$

where $\mathbf{k}$ labels crystal momentum, $\Sigma_{\mathbf{k}}$ is a one-electron term intended to represent the electron self-energy, and $P_{c\mathbf{k}}$ is a projection operator onto the conduction bands. The parameter $\Delta_{\mathbf{k}}$, taken to be independent of $\mathbf{k}$, was taken from GW calculations in the literature.[6–8] Our calculation matched the experimental values of the linear optical susceptibility in Si, Ge, ADP, AlAs, GaP, and GaAs within 4%. For the second-order susceptibility, agreement was found with one set of experimental values[9] within 15% for the cases of GaP and GaAs. Various predictions were made, the most striking being that the change of the nonlinear susceptibility $\chi^{(2)}$ with the lattice constant a, $d \ln \chi^{(2)} / d \ln a$, is some 23 for GaAs in our theory. (Experimentally, the lattice constant may be changed by applying hydrostatic pressure.)

The purpose of the present work is to present the detailed formulation necessary to conduct this study, and to present numerical results which determine the precision of the present work. Moreover, a mistake was found in our computer program which affects the earlier results[1,4,5] by up to 3%; specifically, we were calculating $2^{3/2}\pi/a \approx 0.86$ (where a is the lattice constant in Bohr radii) times $\nabla_{\mathbf{k}} V_{\mathbf{k}}$ instead of $\nabla_{\mathbf{k}} V_{\mathbf{k}}$. Since the operator $\nabla_{\mathbf{k}} V_{\mathbf{k}}$ leads to terms nearly one order of magnitude smaller than those involving $p = -i\nabla$, this error, while regrettable, is not sufficiently large to change our understanding. Revised values for Ref. 1 are presented in this report. Qualitative conclusions are not affected by these modest numerical changes.

II. FORMULA FOR $\chi^{(2)}$ IN THE STATIC LIMIT

In Ref. 2, the nonlinear susceptibility for second-harmonic generation in the static limit is given by Eq. (5.15) with a correction in the erratum. This equation is reprinted below. In a slight notational shift, all operators and wave functions are assumed to have phase factors $e^{i\mathbf{k}\cdot\mathbf{r}}$ and $e^{i\mathbf{q}\cdot\mathbf{r}}$ omitted; i.e., $|n,\tilde{\mathbf{k}}\rangle$ of Ref. 2 is simply $|n\mathbf{k}\rangle$ here. Here, $\mathbf{q}$ is the wave vector of the applied scalar potential. The long- and short-wave potentials, given previously as $\phi_{1m}^{(1)}$ and $\phi_{1m}^{(1s)}$, are simply ϕ_1 and $i\phi$ here. The phase factor is included so that ϕ is purely real for frequencies less than the band gap. (Note that ϕ_1 here is ϕ_0 of Ref. 5 divided by the dielectric constant.) The formula is given in atomic units:

$$\begin{aligned}
\lim_{\omega\to 0}\chi^{(2)}(2\omega,\omega,\omega)=-\frac{i}{2}\frac{1}{(2\pi)^3}\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{v}&\phi_1^2(-12\,\mathrm{Im}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^4H_2|n\mathbf{k}\rangle+12\,\mathrm{Im}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^4H_1|n\mathbf{k}\rangle\\
&+6\,\mathrm{Im}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2H_1G_{n\mathbf{k}}^3H_1|n\mathbf{k}\rangle)\\
&+\phi_1(4\,\mathrm{Re}\langle n\mathbf{k}|H_2G_{n\mathbf{k}}^3\phi|n\mathbf{k}\rangle+4\,\mathrm{Re}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^3\phi|n\mathbf{k}\rangle\\
&\quad+10\,\mathrm{Re}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}\phi G_{n\mathbf{k}}^3H_1|n\mathbf{k}\rangle+2\,\mathrm{Re}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2H_1G_{n\mathbf{k}}^2\phi|n\mathbf{k}\rangle\\
&\quad+4\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2\phi_{n\mathbf{k}}^2G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle+10\,\mathrm{Re}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^3H_1G_{n\mathbf{k}}\phi|n\mathbf{k}\rangle\\
&\quad-14\langle n\mathbf{k}|\phi|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^4H_1|n\mathbf{k}\rangle-16\langle n\mathbf{k}|H_1|n\mathbf{k}\rangle\mathrm{Re}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^4\phi|n\mathbf{k}\rangle)\\
&+2\,\mathrm{Im}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}\phi G_{n\mathbf{k}}^2\phi|n\mathbf{k}\rangle+4\,\mathrm{Im}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2\phi G_{n\mathbf{k}}\phi|n\mathbf{k}\rangle\\
&+2\,\mathrm{Im}\langle n\mathbf{k}|\phi G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^2\phi|n\mathbf{k}\rangle-6\langle n\mathbf{k}|\phi|n\mathbf{k}\rangle\mathrm{Im}\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^3\phi|n\mathbf{k}\rangle\ . \qquad (2)
\end{aligned}$$

In the above formula, $H_1=\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}$, $H_2=\frac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}}$, and $G_{n\mathbf{k}}=(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})^{-1}$ is the static Green's operator. The pseudoinverse is intended, i.e., we project away from the one-or-more-dimensional subspace of eigenvalue $\varepsilon_{n\mathbf{k}}$. The final result of Eq. (2) is purely imaginary.

As it stands, Eq. (2) is not computationally tractable. Two occupied states with nearly equal energies will contribute terms of large magnitude which nearly cancel when summed. These states will contribute large values to the energy denominator; moreover, the denominator is raised to various powers up to the fifth in Eq. (2).

In discussing second-harmonic generation, Aspnes[10] notes that it is useful to divide each spectral sum into conduction and valence parts. In the present notation, we let

$$G_{n\mathbf{k}}=P_{c\mathbf{k}}G_{n\mathbf{k}}+P_{v\mathbf{k}}G_{n\mathbf{k}}\ , \qquad (3)$$

where $P_{v\mathbf{k}}$ is the projection operator onto the valence bands for crystal momentum $\mathbf{k}$. (Obviously, $1=P_{v\mathbf{k}}+P_{c\mathbf{k}}$.) For the formula Aspnes considered, which is equivalent to the second and third terms of Eq. (2), the terms were classified as *vvv*, *vvc*, or *ccv* according to whether they had zero, one, or two unoccupied states in the spectral sum. We extend this notation to include *vv* and *cv* terms for the cases where there is a single Green's operator in the matrix elements. Aspnes noted that the *vvv* terms vanished for his formulation; they and the *vv* terms vanish here as well. Here, we only give the essential relations needed to establish this relatively simple result. Time-reversal symmetry guarantees $|n\mathbf{k}\rangle=|n-\mathbf{k}\rangle^*$, $\varepsilon_{n\mathbf{k}}=\varepsilon_{n-\mathbf{k}}$, $\langle n\mathbf{k}|H_1|m-\mathbf{k}\rangle=-\langle m-\mathbf{k}|H_1|n-\mathbf{k}\rangle$, and $\langle n\mathbf{k}|H_2|m\mathbf{k}\rangle=\langle m-\mathbf{k}|H_2|n-\mathbf{k}\rangle$. Before summation the individual *vv* and *vvv* terms can be very large, so this is an important result numerically.

Additional analytic cancellations may be achieved by considering the *vvc* terms.[11] We just sketch this long and elementary derivation here. Aside from the first and fourth terms, each term in Eq. (2) has two Green's operators. To obtain the *vvc* term, one Green's operator should be restricted to the conduction band, and the other to the valence band. A double summation over the occupied valence states arises, with the restriction that $n\neq m$. Terms may be simplified by interchanging dummy indices. The relation

$$G_{n\mathbf{k}}-G_{m\mathbf{k}}=(-\varepsilon_{n\mathbf{k}}+\varepsilon_{m\mathbf{k}})G_{n\mathbf{k}}G_{m\mathbf{k}} \qquad (4)$$

is used repeatedly to cancel apparent divergences in the energy denominator. The 10th, 11th, and 15th (i.e., last) terms of Eq. (2) turn out to be just what are needed to release the restriction $n\neq m$ on the double summation. The result is

$$\begin{aligned}\lim_{\omega\to 0}\chi^{(2)}|_{vvc}(2\omega,\omega,\omega)=-\frac{i}{2}\sum_{nm}^{v}&(\varepsilon_{n\mathbf{k}}-\varepsilon_{m\mathbf{k}})\mathrm{Im}[\langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1(6G_{n\mathbf{k}}^4G_{m\mathbf{k}}^2+9G_{n\mathbf{k}}^3G_{m\mathbf{k}}^3+6G_{n\mathbf{k}}^2G_{m\mathbf{k}}^4)H_1|n\mathbf{k}\rangle]\\&+\mathrm{Re}[\langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1(-7G_{n\mathbf{k}}^3G_{m\mathbf{k}}-2G_{n\mathbf{k}}^2G_{m\mathbf{k}}^2-7G_{n\mathbf{k}}G_{m\mathbf{k}}^3)\phi|n\mathbf{k}\rangle]\\&+\mathrm{Re}[\langle n\mathbf{k}|\phi|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1(-5G_{n\mathbf{k}}^3G_{m\mathbf{k}}-4G_{n\mathbf{k}}^2G_{m\mathbf{k}}^2-5G_{n\mathbf{k}}G_{m\mathbf{k}}^3)H_1|n\mathbf{k}\rangle]\\&+(\varepsilon_{n\mathbf{k}}-\varepsilon_{m\mathbf{k}})\mathrm{Im}[\langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|\phi G_{n\mathbf{k}}^2G_{m\mathbf{k}}^2\phi|n\mathbf{k}\rangle]\\&+\mathrm{Im}[\langle n\mathbf{k}|\phi|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1(-2G_{n\mathbf{k}}^2G_{m\mathbf{k}}-4G_{n\mathbf{k}}G_{m\mathbf{k}}^2)\phi|n\mathbf{k}\rangle]\ .\end{aligned}\tag{5}$$

The full term $\lim_{\omega\to 0}\chi^{(2)}(2\omega,\omega,\omega)$ is the sum of the *vvc* terms given by Eq. (5) and Eq. (2) with each Green's operator restricted to the conduction bands which give the *cv* and *ccv* terms. We have implemented these equations in our computer program.

III. SUM RULE FOR SECOND-HARMONIC GENERATION

In linear response, the well-known f-sum rule for crystals[5] ensures that the static limit of the long-wavelength dielectric function is finite for an insulator. An analogous result for the nonlinear susceptibility for second-harmonic generation[2] is required to prevent a divergent response in those terms which are first order in the local field ϕ. (Terms which are zeroth and second order in ϕ are nondivergent in the static limit even for a metal.) Equation (2) is valid only for insulating systems.

The number density defined by

$$\rho_{\mathbf{k}}(\mathbf{r})=\sum_{n}^{v}\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|n\mathbf{k}\rangle\tag{6}$$

obeys

$$\begin{aligned}\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}[\mathbf{q}\cdot\nabla_{\mathbf{k}}\rho_{\mathbf{k}}(\mathbf{r})]=\sum_{n}^{v}&2\,\mathrm{Re}(\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_2|n\mathbf{k}\rangle)+2\,\mathrm{Re}(\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle)\\&+\langle n\mathbf{k}|H_1G_{n\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle-2\,\mathrm{Re}(\langle n\mathbf{k}|H_1|n\mathbf{k}\rangle\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle)\\&+\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle\end{aligned}\tag{7}$$

for values of **k** not on the Fermi surface; this class of point includes all **k** in the Brillouin zone in the case of an insulator. In deriving Eq. (7), the wave function is expanded to second order[12] via

$$\begin{aligned}|n\mathbf{k}+\mathbf{q}\rangle=&|n\mathbf{k}\rangle+G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle+G_{n\mathbf{k}}H_2|n\mathbf{k}\rangle\\&+G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle\\&-G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1|n\mathbf{k}\rangle\\&-\tfrac{1}{2}|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle+O(q^3)\end{aligned}\tag{8}$$

and the parts second order in q in the square of the wave function are collected. The penultimate term of Eq. (8) is associated with a perturbation-induced eigenvalue shift, and the final term in the expression above arises from the "wave-function renormalization constant."[12] Recall that H_1 is first order in q and H_2 is second order in q.

For an insulator, the divergence theorem causes the integral of Eq. (7) over the Brillouin zone to vanish. [An analogous argument on $\frac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}})$ is made to derive the f-sum rule for crystals.] Therefore,

$$\begin{aligned}0=\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{v}&2\,\mathrm{Re}(\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_2|n\mathbf{k}\rangle)+2\,\mathrm{Re}(\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle)\\&+\langle n\mathbf{k}|H_1G_{n\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle-2\,\mathrm{Re}(\langle n\mathbf{k}|H_1|n\mathbf{k}\rangle\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle)\\&-\langle n\mathbf{k}|\mathbf{r}\rangle\langle\mathbf{r}|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle\ .\end{aligned}\tag{9}$$

Recalling[5] that the operator for the application of the local field is defined to be

$$\phi=\int d\mathbf{r}|\mathbf{r}\rangle\phi(\mathbf{r})\langle\mathbf{r}|\ ,\tag{10}$$

regardless of the form of the scalar function $\phi(\mathbf{r})$, it is permissible to multiply both sides of Eq. (9) by $\phi(\mathbf{r})$ and to integrate over $\mathbf{r}$, yielding

$$\begin{aligned}0=\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{v}&2\,\mathrm{Re}\langle n\mathbf{k}|\phi G_{n\mathbf{k}}H_2|n\mathbf{k}\rangle+2\,\mathrm{Re}\langle n\mathbf{k}|\phi G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle+\langle n\mathbf{k}|H_1G_{n\mathbf{k}}\phi G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle\\&-2\langle n\mathbf{k}|H_1|n\mathbf{k}\rangle\mathrm{Re}\langle n\mathbf{k}|\phi G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle-\langle n\mathbf{k}|\phi|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1G_{n\mathbf{k}}^2H_1|n\mathbf{k}\rangle\ ;\end{aligned}\tag{11}$$

this sum rule is required to show that an otherwise divergent term vanishes for an insulator. The term in question is given by Eq. (5.16) of Ref. 2; the argument given here is the one promised in the erratum of Ref. 2.

To evaluate Eq. (11) numerically, it is useful to decompose it into its *vv*, *cv*, *vvv*, *vvc*, and *ccv* parts. Like similar formulas, the *vv* and *vvv* terms vanish identically. The *cv* and *ccv* terms are represented by the right side of Eq. (11) with the Green's operators restricted to the conduction bands (first term for *cv*, second and third terms for *ccv*); of course, under this restriction the right side does not sum to zero, in general. The *vvc* term of Eq. (11) is derived by taking each term in Eq. (11) which has two Green's operators with one Green's operator restricted to the valence band and another to the conduction band. Call this quantity Σ_{vvc}. As there are two ways to do this, each term contributes twice, leading to

$$\begin{aligned}
\Sigma_{vvc} &= (\varepsilon_{n\mathbf{k}}-\varepsilon_{m\mathbf{k}})^{-1}\int_{\mathrm{BZ}} d\mathbf{k} \sum_{\substack{nm\\ n\neq m}}^{v} \big(\langle n\mathbf{k}|\phi|m\mathbf{k}\rangle\langle n\mathbf{k}|H_1 G_{n\mathbf{k}} H_1|n\mathbf{k}\rangle + \langle n\mathbf{k}|\phi G_{n\mathbf{k}} H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1|n\mathbf{k}\rangle \\
&\qquad + \langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1 G_{n\mathbf{k}}\phi|n\mathbf{k}\rangle + \langle n\mathbf{k}|H_1 G_{n\mathbf{k}} H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|\phi|n\mathbf{k}\rangle \\
&\qquad + \langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|\phi G_{n\mathbf{k}} H_1|n\mathbf{k}\rangle + \langle n\mathbf{k}|H_1 G_{n\mathbf{k}}\phi|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1|n\mathbf{k}\rangle\big) \\
&\quad -2\langle n\mathbf{k}|H_1|n\mathbf{k}\rangle \mathrm{Re}\langle n\mathbf{k}|\phi G_{n\mathbf{k}}^2 H_1|n\mathbf{k}\rangle - \langle n\mathbf{k}|\phi|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1 G_{n\mathbf{k}}^2 H_1|n\mathbf{k}\rangle \\
&= (\varepsilon_{n\mathbf{k}}-\varepsilon_{m\mathbf{k}})^{-1}\int_{\mathrm{BZ}} d\mathbf{k} \sum_{\substack{nm\\ n\neq m}}^{v} [\langle n\mathbf{k}|\phi|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1(G_{n\mathbf{k}}-G_{m\mathbf{k}})H_1|n\mathbf{k}\rangle + \langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|\phi(G_{n\mathbf{k}}-G_{m\mathbf{k}})H_1|n\mathbf{k}\rangle \\
&\qquad + \langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1(G_{n\mathbf{k}}-G_{m\mathbf{k}})\phi|n\mathbf{k}\rangle] \\
&\quad -2\langle n\mathbf{k}|H_1|n\mathbf{k}\rangle \mathrm{Re}\langle n\mathbf{k}|\phi G_{n\mathbf{k}}^2 H_1|n\mathbf{k}\rangle - \langle n\mathbf{k}|\phi|n\mathbf{k}\rangle\langle n\mathbf{k}|H_1 G_{n\mathbf{k}}^2 H_1|n\mathbf{k}\rangle \\
&= \int_{\mathrm{BZ}} d\mathbf{k} \sum_{nm}^{v} [-\mathrm{Re}(\langle n\mathbf{k}|\phi|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1 G_{m\mathbf{k}} G_{n\mathbf{k}} H_1|n\mathbf{k}\rangle) - 2\,\mathrm{Re}(\langle n\mathbf{k}|H_1|m\mathbf{k}\rangle\langle m\mathbf{k}|H_1 G_{m\mathbf{k}} G_{n\mathbf{k}}\phi|n\mathbf{k}\rangle)]\ . \qquad (12)
\end{aligned}$$

The expression after the second equals sign is obtained from the expression after the first equals sign by interchanging the indices m and n in the second, fourth, and sixth terms, and then pairing terms 1 and 4, 2 and 5, and 3 and 6. The next equality proceeds using Eq. (4), $G_{n\mathbf{k}}G_{m\mathbf{k}}=G_{m\mathbf{k}}G_{n\mathbf{k}}$, additional interchanges of m and n, and noting that the final two terms are just what is needed to extend the sum from $\sum_{nm,n\neq m}^{v}$ to $\sum_{nm}^{v}$. The detailed argument to derive Eq. (5) from Eq. (2) has the same character as the derivation presented in Eq. (12), but is roughly ten times as long.

The quantity

$$\Sigma_{\rho}=2\frac{\Sigma_{vvc}+\Sigma_{ccv}+\Sigma_{cv}}{|\Sigma_{vvc}|+|\Sigma_{ccv}|+|\Sigma_{cv}|} \qquad (13)$$

should vanish; a numerical verification is presented later in the paper. The quantities Σ_{ccv} and Σ_{cv} are the *ccv* and *cv* parts of Eq. (11), just as Σ_{vvc}, given in Eq. (12), is its *vvc* part. The factor of 2 is included to make the deviations from 0 more comparable to the quantity $\Sigma_f - 1$.

IV. RESULTS

We present our main results in Tables I–IV. These results were presented and discussed earlier,[1] and are reprinted here largely because of corrections due to a minor problem found since the last publication. The main conclusions to be drawn from these tables are that we are able to predict both linear and nonlinear response in III-V semiconductors with accuracy of 4% and 15%, respectively. This is more accurate than all other published reports of which we are aware. Local-field corrections for second-harmonic generation are seen to be about −10% in these materials, which is only a bit larger (in magnitude) than the comparable effect for linear response.

We predict values for the changes in linear and second-harmonic response with respect to changes in lattice constant (i.e., pressure). For linear response, the values are in reasonable agreement with experiment, considering that the experiments themselves are not in agreement with each other within their stated uncertainties. The pressure derivative of the susceptibility for second-harmonic generation is far larger than had been suspected previously. We held the self-energy parameter $\Delta_{\mathbf{k}}$ constant while varying the lattice constant a. Such a procedure is supported by the results of a *GW* calculation in the case of silicon,[8], but this does represent an assumption on our part.

In Table V, we analyze the various terms which contribute to the nonlinear susceptibility for second-harmonic generation as a function of the power of the local field and the number of spectral sums (1 for *cv*, 2 for *ccv* and *vvc*). To our knowledge, all previous studies of the nonlinear optical susceptibility neglect the local-field corrections given in the columns marked ϕ^1 and ϕ^2, although an awareness of the local-field corrections goes back to the early days of the field.[13] These local-field corrections are the second-order analogues of the well-known local-field corrections of linear-response

TABLE I. The static dielectric constant of various III-V semiconductors. When the scissors operator is applied, $\Delta = 0.9$ eV for AlP, AlAs, and GaP, and $\Delta = 0.8$ eV for GaAs. The values of these lattice constants are given in Table II of Ref. 1. The experimental values are from the static limit of Fig. 1 in Ref. 24. An energy cutoff E_{cut} of 12 hartrees was used; for the integration over the irreducible Brillouin zone, we chose the 28-special-points (Ref. 19) scheme. Four hundred bands were retained, corresponding to an energy cutoff of the spectral sum, $E_{\text{cut}}^{(\text{sp})} \approx 9.2$ hartrees.

	Lattice constant	ϵ_∞ AlP	AlAs	GaP	GaAs
With scissors operators, no vel. renorm. (Ref. 16)	Expt.				11.3
Theory (Ref. 25)			10.2		13.0
LDA (Ref. 26)	LDA min.	8.4	9.3	9.7	12.0
LDA	LDA min.	8.3	9.5	9.8	12.0
LDA	Expt.	8.3	9.5	10.4	13.7
With scissors operators	LDA min.	7.2	8.0	8.4	10.1
With scissors operators	Expt.	7.2	8.1	8.8	11.2
Expt. (Ref. 24)	Expt.	7.4	8.2	9.0	10.8

TABLE II. The nonlinear susceptibility for optical SHG, $d = \frac{1}{2}\chi^{(2)}$, in the static limit for various III-V semiconductors. d_{123} suffices to specify the entire tensor in the zinc-blende system. We use the standard coordinates for the III-V semiconductors, i.e., the origin is placed at a bond center and the cation is in the positive octant. Both measurements used the 10.6-μm line; Miller's rule (Ref. 27) suggests a correction of less than 1% to extrapolate to zero frequency for GaP and GaAs which is not included here. Parameters of the calculation and the meaning of the lattice constants as in Table I. As indicated in Fig. 2 and Table VII, increasing these values is likely to increase the value of $\chi^{(2)}$ by a few percent.

	Lattice constant	$d = \frac{1}{2}\chi^{(2)}$ (pm/V) AlP	AlAs	GaP	GaAs
Bond-charge model (Ref. 2)		38	58	58	96
Two-band model (Ref. 23)		133	184	157	226
Semiempirical pseudopotential (Ref. 28)					±17
Semiempirical tight binding (Ref. 29)				80	201
With scissors operators, no vel. renorm. (Ref. 16)	Expt.				120
LDA	LDA min.	21	34	33	80
LDA	Expt.	23	39	60	174
With scissors operators	LDA min.	13	21	21	47
With scissors operators	Expt.	15	24	35	86
Expt. (Ref. 9)	Expt.			41±2	90±5
Expt. (Ref. 30)	Expt.			58±9	151±24
Linear combination of Gaussian orbitals, adjusted gap (Ref. 40)	Expt.	10	15	22	52

TABLE III. The effect of local fields on the linear response and SHG in our calculation. The designation "long wave" indicates the local-field corrections are ignored. The "total" designation indicates they are included. $d = \frac{1}{2}\chi^{(2)}$. The results are given for the experimental lattice constant, with the scissors (and all other) parameters chosen as in Table I.

	AlP	AlAs	GaP	GaAs
		ϵ_∞		
Long wave	7.80	8.65	9.30	11.70
Total	7.22	8.08	8.78	11.16
Local field (total)	−0.08	−0.07	−0.06	−0.05
		d (pm/V)		
Long wave	16.6	25.5	39.5	93.4
Total	14.9	23.6	35.1	86.4
Local field (total)	−0.12	−0.08	−0.13	−0.08

TABLE IV. The derivative of the static dielectric constant ϵ_∞ and nonlinear susceptibility for optical SHG in the static limit $\chi^{(2)}$ with respect to the lattice constant a for various III-V semiconductors, calculated at the experimental lattice constant. Convergence uncertainties in the case of GaAs are discussed in Table VIII. Values of Δ and other parameters used are given in Table I. Results for other models were calculated by the present authors taking $d\ln\chi^{(1)}/d\ln a$ from our calculation with the scissors operator.

	AlP	AlAs	GaP	GaAs
			$d\ln\epsilon_\infty/d\ln a$	
Theory (Ref. 25)				6.1
LDA	0.69	1.97	2.68	5.94
With scissors operators	0.37	1.36	1.91	4.19
Ref. 31			1.89±0.24	
Ref. 32				2.19±0.12
Ref. 33			1.59±0.05	3.13±0.09
			$d\ln\chi^{(2)}/d\ln a$	
LDA	11.3	15.0	20.4	29.9
With scissors operators	10.5	12.9	17.5	22.9
Miller's rule (Ref. 27)	1.6	4.6	6.6	13.6
Two-band model (Ref. 23)	3.2	4.9	6.1	10.2
Bond-charge model (Ref. 34)	2.9	2.8	2.9	2.9

theory[14,15] which were presented recently by one of us.[2]

The dominant term is the long-wave (or ϕ^0) *ccv* term. In agreement with the estimate of Aspnes,[10] this term is about one order of magnitude larger than its *vvc* counterpart. The dominance of the long-wave term is seen to arise from a cancellation of the *ccv* and *vvc* terms which are first order in the local field ϕ. The extent of this cancellation varies from system to system; the magnitude of the *vvc* term is 60%, 66%, 41%, and 39% in AlP, AlAs, GaP, and GaAs, respectively, indicating a rough inverse correlation with the band gap. The *cv* terms arise purely from the nonlocality of the pseudopotential; hence it is not surprising that these are quite small. They do play a role in obtaining the numerical agreement of the sum rule for second-harmonic generation, as seen in the final column of Table V. We expect the local-field correction

TABLE V. The terms contributing to the nonlinear susceptibility for second-harmonic generation (in pm/V) of Eqs. (2) and (5) are analyzed in terms of the contributing power of the local field ϕ and the type of states contributing to the spectral sum. The experimental lattice constant is used; parameters of the calculation are given in Table I. See the text for the definitions of *cv*, *ccv*, and *vvc*. The terms contributing to Σ'_p, the sum rule for second-harmonic generation, without the division by the normalizing denominator of Eq. (13) or the factor of 2; i.e., $\Sigma'_p = \Sigma_{cv} + \Sigma_{ccv} + \Sigma_{vvc}$, are given.

		ϕ^0 terms	ϕ^1 terms	ϕ^2 terms	$d=\frac{1}{2}\chi^{(2)}$	Σ'_p
AlP	*cv*	0.02	−0.00	0	0.02	−0.03
	ccv	15.10	−4.94	0.08	10.24	1.16
	vvc	1.48	2.96	0.16	4.60	−1.13
	total	16.60	−1.98	0.25	14.87	0.01
AlAs	*cv*	0.35	0.00	0	0.35	−0.01
	ccv	24.18	−6.47	0.12	17.83	1.30
	vvc	0.99	4.30	0.16	5.44	−1.25
	total	25.51	−2.17	0.28	23.62	0.04
GaP	*cv*	0.33	0.01	0	0.35	−0.20
	ccv	36.32	−7.91	0.02	28.44	2.04
	vvc	2.88	3.25	0.20	6.33	−1.89
	total	39.54	−4.64	0.22	35.12	−0.05
GaAs	*cv*	0.78	0.01	0	0.79	−0.15
	ccv	89.11	−12.17	0.08	77.02	1.91
	vvc	3.56	4.79	0.21	8.55	−1.68
	total	93.45	−7.36	0.28	86.37	0.09

TABLE VI. Illustration of the importance of the velocity renormalization operator. In the "naive" method, all energy denominators are shifted by the scissors operators—not just the ones that we have argued should be shifted. The result of this shift is to lead to an underestimate of the dielectric function and the nonlinear susceptibility for second-harmonic generation as well as violations of the sum rules for linear (Σ_f) and second-order (Σ_p) response. The experimental references are Ref. 24 for linear response and Ref. 9 for the second-harmonic response. All calculations are performed as in Table I, using the experimental lattice constant and the "scissors" parameter. The values of Σ_f for the LDA and scissors-operator cases are exactly equal.

	ϵ_∞	d (pm/V)	Σ_f	Σ_p
		GaP		
LDA	10.4	60	0.9996	−0.022
With scissors operators	8.8	35	0.9996	−0.022
"Naive"	6.3	20	0.8646	−0.263
Expt. or ideal	9.0	41	1	0
		GaAs		
LDA	13.7	174	0.9985	0.049
With scissors operators	11.2	86	0.9985	0.048
"Naive"	7.6	42	0.8663	−0.255
Expt. or ideal	10.8	90	1	0

to be more important for systems with larger unit cells. Such unit cells support longer-wavelength local fields, which are easier to excite. The near equality between the long-wave *ccv* term and the total answer is, we believe, a coincidence which will not hold true in other materials.

We illustrate the importance of the renormalization of the velocity operator [i.e., the inclusion of the term $\nabla_{\mathbf{k}}(\Delta_{\mathbf{k}} P_{c\mathbf{k}})$] on the calculation in Table VI. Earlier, we demonstrated that the "naive" use of the scissors operator (i.e., simply adjusting the every energy denominator in the perturbation theory expression) leads to a serious underestimate of the static dielectric constant in the cases of silicon and germanium overcorrecting the overestimate of the LDA; moreover, oscillator strength was lost.[4,5] The same effect is seen in Table VI in the cases of GaP and GaAs: the LDA overestimates ϵ_∞, and the "naive" scissors operator underestimates it. The "naive" scissors operator leads to a loss of oscillator strength. The same effects exist in the second-order coefficients: the LDA overestimates $\chi^{(2)}$, and the "naive" method underestimates it, compared to experiment. The second-harmonic sum rule is violated by the "naive" use of the scissors operator.

Huang and Ching[16] do not include velocity renormalization. Their results for ϵ_∞ and $\chi^{(2)}$ are given in Tables I and II, respectively. Our results suggest that our suggested modification of the velocity operator in their calculation would lead to a much larger reported value for both ϵ_∞ and $\chi^{(2)}$. As noted in Table VI, we find that ϵ_∞ is increased by 50% and $\chi^{(2)}$ is doubled when velocity renormalization is restored to the scissors operator. While we do not fully understand the origin of the difference between our plane-wave calculation and their orthogonalized linear combination of atomic-orbitals approach, we

TABLE VII. Direct gap (eigenvalues differences) at the Γ point for GaAs, calculated as in Fig. 1. The plane-wave cutoff is not stated directly for Ref. 35, but the Hamiltonian matrix is stated to be "of order 200×200;" using Eq. (14), $N_{pw}=200$ works out to 5.7 hartrees. Reference 7 has $E_{\rm cut}=6.25$ hartrees. We choose to present the version of the calculations of Refs. 36 and 17, which are closest to our own: scalar relativistic (i.e., without spin-orbit coupling) and without relaxation of the 3*d* gallium core. (More sophisticated versions of the calculation are presented in both of these references.) Reference 36 performed both a pseudopotential, plane-wave calculation and an LMTO (linear muffin-tin orbital) calculation. Reference 17 is the FLAPW calculation; the value given is their stated value of 0.31 eV including spin orbit plus their stated 0.11-eV gap reduction due to spin-orbit effects. $E_{\rm cut}$ is given in hartrees.

$E_{\rm cut}$	1.38	3	6	7.3	9	12	LCAO
Present	4.32	3.95	1.15		0.494	0.485	
Ref. 35			0.93				
Ref. 7			0.67				
Ref. 36				0.64		0.59 (LMTO)	
Ref. 37					0.48		
Ref. 16							1.04
Ref. 18							1.21
Ref. 17						0.42 (FLAPW)	

TABLE VIII. Numerical test of sum rules and their derivatives with respect to lattice constant. The values are given for each semiconductor calculated at the experimental lattice constant. The scissors parameter is chosen as in Table I; other parameters of the calculation are given there as well. Ideally, Σ_f is 1 and the other quantities vanish. The compounds are ordered with decreasing energy gap; the quality of the calculation degrades somewhat as the energy gap is reduced. The effect of varying the number of special points is illustrated in Tables VII and VIII for GaAs. AlP, AlAs, and GaP are expected to be less sensitive to variations in the number of special points.

	AlP	AlAs	GaP	GaAs
Σ_f	1.0004	0.9976	0.9996	0.9985
$d\ln\Sigma_f/d\ln a$	−0.001	−0.001	0.005	0.033
Σ_p	0.006	0.031	−0.022	0.048
$d\ln\Sigma_p/d\ln a$	0.028	−0.003	−0.66	1.56

believe that physical completeness may not have been achieved in their basis. In Table VII, we illustrate that our converged value for the energy gap is in agreement with many other calculations, particularly when the plane-wave energy cutoff is taken into account. Agreement with the all-electron full-potential linearized-augmented-plane-wave (FLAPW) calculation[17] is especially encouraging. In contrast, two LCAO calculations, those of Huang and Ching[16] and Wang and Klein,[18] are close to each other, but over 0.5 eV from the consensus value for the direct band gap in GaAs.

In Table VIII, we illustrate that the f-sum rule Σ_f and its derivatives with respect to the lattice constant are well satisfied for the four materials of this study. For the smaller band-gap materials, the lattice-constant derivative is somewhat more important. The second-harmonic analogue Σ_p is also satisfied, though not as well as Σ_f. Table IX illustrates that, in the case of GaAs—the most troublesome material studied—increasing the number of special integration points[19] from 28 to 60 improves the convergence of the Σ_p. (In a study of linear optical response, Hybertsen and Louie[20] also found an inverse relationship between the size of the band gap and the number of special integration points required in the series C, Si, and Ge.) The violation of Σ_f actually increases, but at a rather tiny level. The dielectric constant and the nonlinear susceptibility d are apparently within a few percent of their fully converged values with 28 special points. Table X tells a similar story about the effect on a variation in lattice constant: increasing the number of special points to 60 has a beneficial effect on the sum rules. The derivative $d\ln\epsilon_\infty/d\ln a$ is apparently within a few percent of its fully converged value. The second-harmonic analogue $d\ln\chi^{(2)}/d\ln a$ is a bit more uncertain. If the improvement of $d\Sigma_p/d\ln a$ is taken as a guide, the variation from 28 to 60 special points is probably comparable to the variation from 60 special points to the converged limit; hence we guess that the converged value for $d\ln\chi^{(2)}/d\ln a$ is about 26.

While Tables VII–X illustrate the degree of convergence of our program, they also argue for its *correctness* as well. The elementary operations (such as the application of the spectral sums or the H_1 and H_2 operators) are very similar to obtain the sum rules Σ_f and Σ_p as are required to obtain the final results ϵ_∞ and $\chi^{(2)}$. Given the long history of poor results in the area of the *ab initio* calculation of nonlinear optical susceptibilities, we find it reassuring to have a powerful (though not foolproof) check on our program's correctness which is completely mathematical, i.e., independent of appeal to experiment.

As a preliminary, we assessed the convergence of the total energy and the LDA energy gap (the eigenvalue difference at Γ) for GaAs, as shown in Fig. 1. While a basis-set energy cutoff, E_{cut}, of 6 hartrees may be ade-

TABLE IX. A convergence study of various parameters for GaAs as a function of the plane-wave energy cutoff E_{cut} and the number of special points for the irreducible zone integration. A scissors parameter of $\Delta_k = 0.8$ eV is chosen and the experimental lattice constant is used. The convergence of the parameters in GaAs is the slowest of the four semiconductors in this study, owing to its relatively small band gap. Convergence in the number of integration points in the LDA is somewhat worse than the data presented here. The ideal values for Σ_f and Σ_p are 1 and 0, respectively. The reasonably close agreement between the 10 and 28 special points values of $\chi^{(2)}$ is partially fortuitous: individual terms which are summed to form $\chi^{(2)}$ are in worse agreement, but sign cancellations occur. A 20% uncertainty is a more realistic estimate. E_{cut} is given in hartrees.

E_{cut}	10 points	28 points	60 points	E_{cut}	10 points	28 points	60 points
		ϵ_∞				$d=\frac{1}{2}\chi^{(2)}$ (pm/V)	
9	12.13	11.08	10.92	9	87.9	83.3	84.5
12	11.93	11.16		12	84.0	86.4	
		Σ_f				Σ_p	
9	1.022	1.002	0.993	9	0.331	0.075	−0.006
12	1.022	0.998		12	0.335	0.048	

TABLE X. Continuation of convergence study for the derivatives of the parameters for GaAs given in Table VII with respect to the lattice constant a. The derivatives were calculated using two-point finite differences at $\pm 0.4\%$ of the experimental lattice constant. Ideally, $d\Sigma_f/d\ln a$ and $d\Sigma_p/d\ln a$ vanish. $E_{\rm cut}$ is given in hartrees.

$E_{\rm cut}$	28 points	60 points	$E_{\rm cut}$	28 points	60 points
	$d\ln\epsilon_\infty/d\ln a$			$d\ln\chi^{(2)}/d\ln a$	
9	4.10	4.18	9	22.8	24.4
12	4.19		12	22.9	
	$d\Sigma_f/d\ln a$			$d\Sigma_p/d\ln a$	
9	0.029	0.004	9	1.33	0.70
12	0.033		12	1.56	

quate for convergence of the total energy, at least 9 hartrees is required for convergence of the direct gap, which is a prototype for the optical properties. We performed a more comprehensive convergence study, illustrated in Figs. 2 and 3. In addition to $E_{\rm cut}$, we introduce $E_{\rm cut}^{(\rm sp)}$, the eigenvalue cutoff of the spectral sums. (These eigenvalues are absolute values from the Hamiltonian; the valence-band maximum eigenvalue is 0.314 hartrees in GaAs with 28 special points and $E_{\rm cut}=12$ hartrees.) In Fig. 2, we increase $E_{\rm cut}$, keeping all the bands in the spectral sums. In Fig. 3 we choose one high value, $E_{\rm cut}=12$ hartrees, and vary the energy cutoff $E_{\rm cut}$ for the spectral sum. The number of plane waves used may be approximated by an elementary argument of counting states in the free-electron gas.[21] The result is

$$N \approx \frac{\sqrt{2}}{3\pi^2} V E_{\rm cut}^{3/2} , \tag{14}$$

where V is the unit-cell volume in a_0^3 and $E_{\rm cut}$ is given in hartrees. $V=305a_0^3$ for GaAs at the experimental lattice constant. The same formula gives the number of states retained in the spectral sum as a function of $E_{\rm cut}^{(\rm sp)}$. Whereas a high value for the energy cutoff (perhaps 9 hartrees) is required to get good convergence in the values of ϵ_∞ and $\chi^{(2)}$, a much lower value, as little as 1 hartree, suffices for $E_{\rm cut}^{(\rm sp)}$. Indeed, for the dielectric constant, almost the whole contribution comes from the first state or two included in the spectral sum. Our suggested value of $E_{\rm cut}^{(\rm sp)}$ of 1–2 hartrees is in agreement with the findings of the convergence study of Huang and Ching;[16] in the end, Huang and Ching[16] suggest 28 eV or 1.03 hartrees for a reasonably converged value of what we call $E_{\rm cut}^{(\rm sp)}$.

Increasing the energy cutoff $E_{\rm cut}$ has two roles: it increases the number of eigenstates, and it leads to an im-

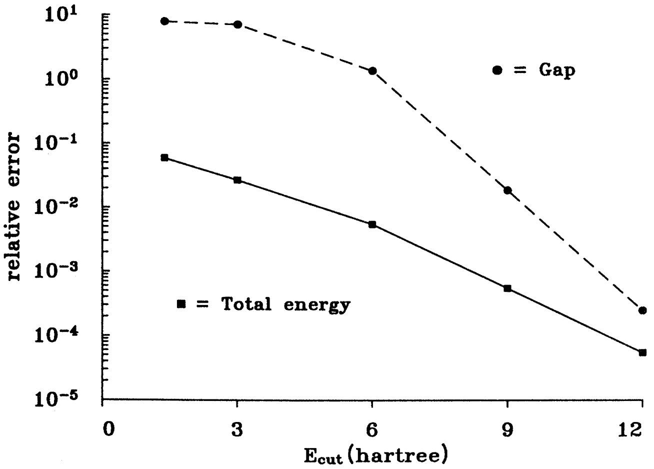

FIG. 1. The relative error in the energy gap and total energy for GaAs calculated at the experimental lattice constant for various values of the energy cutoff for the plane-wave basis $E_{\rm cut}$. The self-consistent potential is found using the 28-special-points quadrature (Ref. 19). We extrapolate to infinite energy cutoff by applying the Shanks procedure (Refs. 38 and 39) to the three data points with largest values of $E_{\rm cut}$.

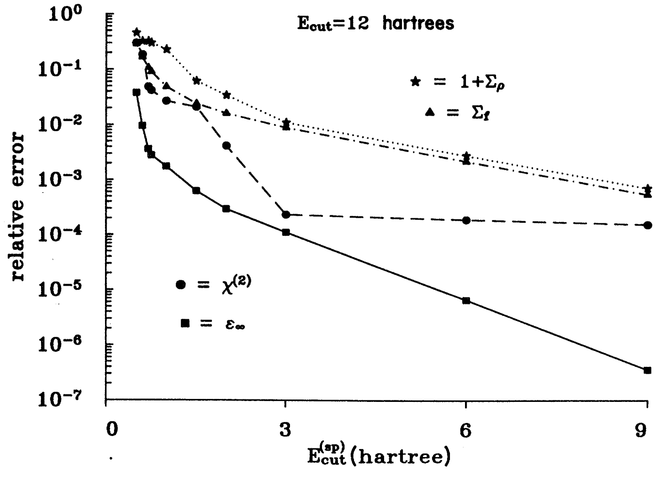

FIG. 2. Values of $\epsilon_\infty(E_{\rm cut})/\{\epsilon_\infty\}-1$ (solid line), $\Sigma_f(E_{\rm cut})/\{\Sigma_f\}-1$ (dotted line), $\chi^{(2)}(E_{\rm cut})/\{\chi^{(2)}\}-1$ (dashed line), and $[1+\Sigma_p(E_{\rm cut})]/\{1+\Sigma_p\}-1$ (dash-dotted line) for the case of GaAs with the 28-special-points integration scheme. The asymptotic value was estimated by applying the Shanks extrapolation procedure (Refs. 38 and 39) to the final three data points; the notation { } represents this extrapolated value. We extrapolate only on $E_{\rm cut}$; to find the converged value within the theory, it would be necessary to make an extrapolation on the number of integration points as well.

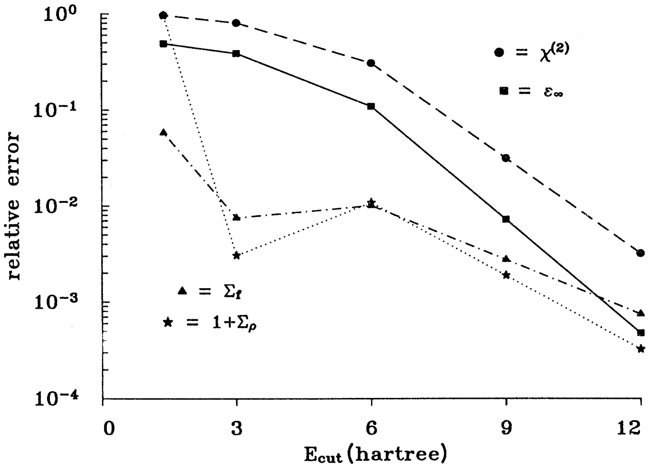

FIG. 3. Values of $\epsilon_\infty(E_{\text{cut}}^{(\text{sp})})/\{\epsilon_\infty\}$ (solid line), $\Sigma_f(E_{\text{cut}}^{(\text{sp})})/\{\Sigma_f\}-1$ (dotted line), $\chi^{(2)}(E_{\text{cut}}^{(\text{sp})})/\{\chi^{(2)}\}-1$ (dashed line), and $[1+\Sigma_\rho(E_{\text{cut}}^{(\text{sp})})]/\{1+\Sigma_\rho\}-1$ (dash-dotted line) for the case of GaAs with the 28-special-points integration scheme. The asymptotic limit was estimated as in Fig. 2, except here we extrapolate on the $E_{\text{cut}}^{(\text{sp})}$ variable; we vary $E_{\text{cut}}^{(\text{sp})}$ at fixed E_{cut} and a fixed number of integration points.

proved description of the low-lying eigenstates. The latter effect is more important: the calculation is best performed with a relatively limited number of well-converged eigenstates. Symbolically, it is sensible to consider both E_{cut} and $E_{\text{cut}}^{(\text{sp})}$ because in practice it is reasonable to compute under the condition $E_{\text{cut}}^{(\text{sp})} \ll E_{\text{cut}}$.

Notice that we are not able to converge the value of $\chi^{(2)}$ as a function of $E_{\text{cut}}^{(\text{sp})}$ to better than a few parts in 10^{-4}. Our extrapolation procedure does not give an obviously improved result on the data at hand for $\chi^{(2)}$. The terms contributing to $\chi^{(2)}$ have a somewhat complicated structure, being the sum of eight different terms of various signs; these are illustrated in Table V. Moreover, each term may be composed of terms of varying sign, in contrast to the simpler situation for ϵ_∞. For example, the largest term (which is of the type ϕ^0 and ccv) itself comes to a maximum near $E_{\text{cut}} = E_{\text{cut}}^{(\text{sp})}$ and then comes down by a fraction of a percent for the study shown in Fig. 3, as well as corresponding studies we carried out at $E_{\text{cut}}=6$ and 9 hartrees.

V. CONCLUSIONS

We have performed a series of highly accurate calculations for the linear optical susceptibility and nonlinear susceptibilities for second-harmonic generation in a series of III-V semiconductors. A preliminary report of this work has appeared.[1] We find that correcting the Kohn-Sham local-density approximation by a self-energy term in the form of a "scissors operator" is an accurate and effective way to describe both linear and nonlinear response. This extends our earlier work[4,5] on linear optical response to nonlinear properties. The present numerical work confirms our thesis that the introduction of a self-energy term in the Hamiltonian leads to a required renormalization of the velocity operator.

We demonstrate computationally preferred forms for the nonlinear susceptibility for second-harmonic generation in the static limit. Specifically, the so-called "*vvc*" or virtual-hole terms are combined into formulas which avoid numerical cancellations which might otherwise be present. The formalism required to apply a self-energy correction in the form of a "scissors" operator in second-harmonic generation is presented in detail.

We numerically verified the f-sum rule and a second-harmonic analogue derived recently by one of us.[2] In verifying these sum rules, we help verify the overall correctness of our computer program and help to estimate the errors in the computation associated with, for example, the number of integration points chosen. We find that 28 special points are usually adequate, but that materials with small band gaps tend to require more points.

We tested the result as a function of various convergence parameters. We find that having a couple dozen very-well-resolved states at each $\mathbf{k}$ point is sufficient to converge the second-order susceptibility $\chi^{(2)}$. A few states is sufficient for ϵ_∞, but states of energies up to perhaps 1 or 2 hartrees are required for $\chi^{(2)}$. We find that the total energy converges more quickly than the direct gap; underconverged calculations will tend to have energy gaps which are too large compared to the fully converged LDA results.

The most physically striking prediction we make is that $\chi^{(2)}$ should change by about 25% for each 1% change in the lattice constant of GaAs. Our prediction is much larger than that of simple models proposed earlier. We hope that a measurement of this quantity will be forthcoming.

ACKNOWLEDGMENTS

We thank Siqing Wei and Hua Zhong for assistance in discovering the mistake in our computer code reported on in this document. We thank John Wilkins for his continued encouragement during the course of this project. This work was supported by the Department of Energy–Basic Energy Science, Division of Materials Research, and the Cornell National Supercomputer Facility. Hospitality by the Laboratory of Atomic and Solid State Physics of Cornell University is also gratefully acknowledged.

APPENDIX A: SCISSORS FORMALISM FOR SHG

In this appendix, the formulas needed in this calculation which are specific to the "scissors" Hamiltonian of Eq. (1) are derived and presented. We may write

$$\begin{aligned} H_1 &= H_1^{\text{LDA}} + \mathbf{q}\cdot\nabla_{\mathbf{k}}(\Delta_{\mathbf{k}} P_{c\mathbf{k}}) \\ &= H_1^{\text{LDA}} + (\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}} - \Delta_{\mathbf{k}} \sum_n^v G_{n\mathbf{k}} H_1 P_{n\mathbf{k}} + P_{n\mathbf{k}} H_1 G_{n\mathbf{k}} , \end{aligned} \tag{A1}$$

where $P_{n\mathbf{k}} = |n\mathbf{k}\rangle\langle n\mathbf{k}|$ is a projection operator. Consider the restriction of H_1 to the valence bands:

$$P_{v\mathbf{k}}H_1P_{v\mathbf{k}}$$
$$=P_{v\mathbf{k}}H_1^{\mathrm{LDA}}P_{v\mathbf{k}}-\sum_n^v P_{v\mathbf{k}}G_{n\mathbf{k}}H_1P_{n\mathbf{k}}+P_{n\mathbf{k}}H_1G_{n\mathbf{k}}P_{v\mathbf{k}}$$
$$=P_{v\mathbf{k}}H_1^{\mathrm{LDA}}P_{v\mathbf{k}}-\sum_{\substack{nm\\ n\neq m}}^v P_{m\mathbf{k}}G_{n\mathbf{k}}H_1P_{n\mathbf{k}}+P_{m\mathbf{k}}H_1G_{m\mathbf{k}}P_{m\mathbf{k}}$$
$$=P_{v\mathbf{k}}H_1^{\mathrm{LDA}}P_{v\mathbf{k}}\ , \tag{A2}$$

using the relation

$$\langle m\mathbf{k}|G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle=-\langle m\mathbf{k}|H_1G_{m\mathbf{k}}|n\mathbf{k}\rangle\ . \tag{A3}$$

The final cancellation involves swapping the dummy indices m and n. Also, $P_{c\mathbf{k}}P_{v\mathbf{k}}=0$, $P_{n\mathbf{k}}P_{v\mathbf{k}}=P_{n\mathbf{k}}$, and $G_{n\mathbf{K}}P_{n\mathbf{k}}=0$ are used. Equation (2.27) of Ref. 5 demonstrated

$$P_{c\mathbf{k}}G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle=P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle\ . \tag{A4}$$

Since $P_{v\mathbf{k}}G_{n\mathbf{k}}=P_{v\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}$, and $P_{v\mathbf{k}}G_{n\mathbf{k}}=G_{n\mathbf{K}}P_{v\mathbf{k}}$, Eq. (18) may be generalized to

$$G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle=P_{c\mathbf{k}}G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle+P_{v\mathbf{k}}G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle$$
$$=P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle+G_{n\mathbf{k}}^{\mathrm{LDA}}P_{v\mathbf{k}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$=G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle\ . \tag{A5}$$

Equation (A5) was stated, but not proved, in Eq. (2.28b) of Ref. 5. Equation (A5) may be applied to Eq. (A1) to arrive at the relation

$$H_1=H_1^{\mathrm{LDA}}+(\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}}-\Delta_{\mathbf{k}}\sum_n^v G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{n\mathbf{k}}$$
$$+P_{n\mathbf{k}}H_1^{\mathrm{LDA}}G_{n\mathbf{k}}^{\mathrm{LDA}}\ . \tag{A6}$$

Equation (A6) is written in terms of known LDA operators and $\Delta_{\mathbf{k}}$ and so is computationally tractable.

When H_1 appears in the "middle position" of a matrix element, it is useful to make the decomposition

$$H_1=P_{c\mathbf{k}}H_1P_{c\mathbf{k}}+P_{v\mathbf{k}}H_1P_{c\mathbf{k}}+P_{c\mathbf{k}}H_1P_{v\mathbf{k}}+P_{v\mathbf{k}}H_1P_{v\mathbf{k}}\ . \tag{A7}$$

The terms in $P_{v\mathbf{k}}H_1P_{v\mathbf{k}}$ form parts of the vvv terms which all vanish; this is a generalization of a result due to Aspnes.[10] Clearly,

$$P_{c\mathbf{k}}H_1P_{c\mathbf{k}}=P_{c\mathbf{k}}(H_1^{\mathrm{LDA}}+\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}}\ . \tag{A8}$$

Less clearly, but equally truthfully,

$$P_{v\mathbf{k}}H_1P_{c\mathbf{k}}=\sum_n^v P_{n\mathbf{k}}(H_1^{\mathrm{LDA}}-\Delta_{\mathbf{k}}H_1^{\mathrm{LDA}}G_{n\mathbf{k}}^{\mathrm{LDA}})P_{c\mathbf{k}}$$
$$=\sum_n^v P_{n\mathbf{k}}H_1^{\mathrm{LDA}}G_{n\mathbf{k}}^{\mathrm{LDA}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})P_{c\mathbf{k}}\ , \tag{A9}$$

with a similar relation for $P_{c\mathbf{k}}H_1P_{v\mathbf{k}}$. Equation (A1) is used to write Eqs. (A8) and (A9).

H_2 terms. The projection operator $P_{n\mathbf{k}+\mathbf{q}}=|n\mathbf{k}+\mathbf{q}\rangle\langle n\mathbf{k}+\mathbf{q}|$ may be expanded as $P_{n\mathbf{k}+\mathbf{q}}=P_{n\mathbf{k}}+P_{n\mathbf{k}}^{(1)}+P_{n\mathbf{k}}^{(2)}+O(q^3)$, where

$$P_{n\mathbf{k}}^{(1)}=\mathbf{q}\cdot\nabla_{\mathbf{k}}P_{n\mathbf{k}}=G_{n\mathbf{k}}H_1P_{n\mathbf{k}}+P_{n\mathbf{k}}H_1G_{n\mathbf{k}} \tag{A10}$$

and

$$P_{n\mathbf{k}}^{(2)}=\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}P_{n\mathbf{k}})=G_{n\mathbf{k}}H_2P_{n\mathbf{k}}+P_{n\mathbf{k}}H_2G_{n\mathbf{k}}+G_{n\mathbf{k}}H_1P_{n\mathbf{k}}H_1G_{n\mathbf{k}}+G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1P_{n\mathbf{k}}+P_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}$$
$$-P_{n\mathbf{k}}H_1G_{n\mathbf{k}}^2H_1P_{n\mathbf{k}}-P_{n\mathbf{k}}H_1P_{n\mathbf{k}}H_1G_{n\mathbf{k}}^2-G_{n\mathbf{k}}^2H_1P_{n\mathbf{k}}H_1P_{n\mathbf{k}}\ , \tag{A11}$$

are the terms which are first and second order in $\mathbf{q}$, respectively. In writing Eq. (A11), we use the expansion of the wave function given in Eq. (8). The expression

$$H_2=\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}H_{\mathbf{k}})=H_2^{\mathrm{LDA}}+\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\Sigma_{\mathbf{k}}) \tag{A12}$$

for the scissors form of the self-energy operator $\Sigma_{\mathbf{k}}=\Delta_{\mathbf{k}}P_{c\mathbf{k}}$ leads us to consider the three terms in

$$\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\Sigma_{\mathbf{k}})=\tfrac{1}{2}[\mathbf{q}\cdot\nabla_{\mathbf{k}}(\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})]P_{c\mathbf{k}}$$
$$+(\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}}^{(1)}+\Delta_{\mathbf{k}}P_{c\mathbf{k}}^{(2)}\ . \tag{A13}$$

Inspection of the equations for the charge density induced at the second-harmonic frequency indicates that only the combination $H_2|n\mathbf{k}\rangle$ is required. Applying the terms in the expansion of the projection operator to $|n\mathbf{k}\rangle$ leads to

$$P_{c\mathbf{k}}|n\mathbf{k}\rangle=0\ , \tag{A14}$$

$$P_{c\mathbf{k}}^{(1)}|n\mathbf{k}\rangle=-\sum_m^v(G_{m\mathbf{k}}H_1P_{m\mathbf{k}}+P_{m\mathbf{k}}H_1G_{m\mathbf{k}})|n\mathbf{k}\rangle$$
$$=-G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$-\sum_{m\neq n}^v P_{m\mathbf{k}}H_1^{\mathrm{LDA}}G_{m\mathbf{k}}^{\mathrm{LDA}}|n\mathbf{k}\rangle\ , \tag{A15}$$

making use of Eqs. (A5) and (A10), and

$$P_{c\mathbf{k}}^{(2)}|n\mathbf{k}\rangle=(-G_{n\mathbf{k}}H_2-G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1+P_{n\mathbf{k}}H_1G_{n\mathbf{k}}^2H_1+G_{n\mathbf{k}}^2H_1P_{n\mathbf{k}}H_1)|n\mathbf{k}\rangle$$
$$+\sum_{m\neq n}^v(-P_{n\mathbf{k}}H_2-G_{m\mathbf{k}}H_1P_{m\mathbf{k}}H_1-P_{n\mathbf{k}}H_1G_{m\mathbf{k}}H_1+P_{n\mathbf{k}}H_1P_{n\mathbf{k}}H_1G_{m\mathbf{k}})G_{m\mathbf{k}}|n\mathbf{k}\rangle\ , \tag{A16}$$

making use of Eq. (A11). The relation $P_{c\mathbf{k}}^{(\alpha)}=-\sum_n^v P_{n\mathbf{k}}^{(\alpha)}$, $\alpha=1,2$, is used here.

The vanishing of the vv terms implies that we may further restrict our discussion to terms of the form $P_{c\mathbf{k}}H_2|n\mathbf{k}\rangle$:

$$P_{c\mathbf{k}}H_2|n\mathbf{k}\rangle = P_{c\mathbf{k}}H_2^{\mathrm{LDA}}|n\mathbf{k}\rangle - (\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}}G_{n\mathbf{k}}H_1|n\mathbf{k}\rangle$$
$$+\Delta_{\mathbf{k}}P_{c\mathbf{k}}(-G_{n\mathbf{k}}H_2 - G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1 + G_{n\mathbf{k}}^2H_1P_{n\mathbf{k}}H_1)|n\mathbf{k}\rangle - \Delta_{\mathbf{k}}P_{c\mathbf{k}}\sum_{m\neq n}^{v} G_{m\mathbf{k}}H_1P_{m\mathbf{k}}H_1G_{m\mathbf{k}}|n\mathbf{k}\rangle\ , \tag{A17}$$

making use of Eqs. (A12)–(A16). Since the wave functions are not changed by the introduction of the scissors operator, it is permissible to replace the projection operators by their LDA counterparts. If this substitution is made, and, moreover, the relation

$$P_{c\mathbf{k}}(1-\Delta_{\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}) = P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}}) \tag{A18}$$

is used, Eq. (A17) may be written as

$$P_{c\mathbf{k}}H_2|n\mathbf{k}\rangle = P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})H_2^{\mathrm{LDA}}|n\mathbf{k}\rangle - (\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$-\Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle + \Delta_{\mathbf{k}}P_{c\mathbf{k}}(G_{n\mathbf{k}}^{\mathrm{LDA}})^2H_1^{\mathrm{LDA}}P_{n\mathbf{k}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$+\Delta_{\mathbf{k}}P_{c\mathbf{k}}\sum_{m\neq n}^{v} G_{m\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{m\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle\ . \tag{A19}$$

The third term of Eq. (A19) may be divided into two parts, as follows:

$$-\Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle = -\Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle - \Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{v\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle\ . \tag{A20}$$

The final term of Eq. (A20) and the final two terms in Eq. (A19) may be combined into a simple expression, namely

$$-\Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{v\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$+\Delta_{\mathbf{k}}P_{c\mathbf{k}}(G_{n\mathbf{k}}^{\mathrm{LDA}})^2H_1^{\mathrm{LDA}}P_{n\mathbf{k}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle + \Delta_{\mathbf{k}}P_{c\mathbf{k}}\sum_{m\neq n}^{v} G_{m\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{m\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$=\Delta_{\mathbf{k}}P_{c\mathbf{k}}\sum_{m\neq n}^{v}(G_{m\mathbf{k}}^{\mathrm{LDA}}-G_{n\mathbf{k}}^{\mathrm{LDA}})H_1^{\mathrm{LDA}}P_{m\mathbf{k}}(\varepsilon_{n\mathbf{k}}-\varepsilon_{m\mathbf{k}})^{-1}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle + \Delta_{\mathbf{k}}P_{c\mathbf{k}}(G_{n\mathbf{k}}^{\mathrm{LDA}})^2H_1^{\mathrm{LDA}}P_{n\mathbf{k}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$=\Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}\sum_{m}^{v} G_{m\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{m\mathbf{k}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle\ , \tag{A21}$$

exploiting the relation $(G_{m\mathbf{k}}-G_{n\mathbf{k}})(\varepsilon_{n\mathbf{k}}-\varepsilon_{m\mathbf{k}})^{-1}=G_{n\mathbf{k}}G_{m\mathbf{k}}$. Equation (A21) may be used to simplify Eq. (A19) to

$$P_{c\mathbf{k}}H_2|n\mathbf{k}\rangle = P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})H_2^{\mathrm{LDA}}|n\mathbf{k}\rangle - (\mathbf{q}\cdot\nabla_{\mathbf{k}}\Delta_{\mathbf{k}})P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle$$
$$-\Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle + \Delta_{\mathbf{k}}P_{c\mathbf{k}}G_{n\mathbf{k}}^{\mathrm{LDA}}\sum_{m}^{v} G_{m\mathbf{k}}^{\mathrm{LDA}}H_1^{\mathrm{LDA}}P_{m\mathbf{k}}H_1^{\mathrm{LDA}}|n\mathbf{k}\rangle\ . \tag{A22}$$

This version is more robust computationally than Eq. (A19); it is quite desirable to have the same analytic formula whether or not eigenstates are degenerate. We regard the first two terms of Eq. (A22) as *cv*, the third as *ccv*, and the final term as *vvc*. To date, our numerical work has considered only the **k**-independent scissors operator, i.e., $\nabla_{\mathbf{k}}\Delta_{\mathbf{k}}=0$.

APPENDIX B: NEGLECT OF SHORT-WAVE INDUCED CHARGE IN SECOND-HARMONIC GENERATION

In the presence of a nonlinear polarization $\mathbf{P}(\mathbf{r};\Omega)$, the electric field $\mathbf{E}(\mathbf{r};\Omega)$ is given by[22]

$$-\nabla^2\mathbf{E}(\mathbf{r};\Omega)-\frac{\Omega^2}{c^2}\epsilon\mathbf{E}(\mathbf{r};\Omega)=\frac{4\pi}{c^2}\mathbf{P}(\mathbf{r};\Omega)\ . \tag{B1}$$

Defining the matrix

$$\mathbf{M}_{\mathbf{G}\mathbf{G}'}=G^2\delta_{\mathbf{G}\mathbf{G}'}-\frac{\Omega^2}{c^2}\epsilon_{\mathbf{G}\mathbf{G}'}\ , \tag{B2}$$

the "forced-wave" solution may be written as

$$\mathbf{E}_{\mathbf{G}}=\frac{4\pi}{c^2}\sum_{\mathbf{G}'}\mathbf{M}_{\mathbf{G}\mathbf{G}'}^{-1}\mathbf{P}_{\mathbf{G}'}\ . \tag{B3}$$

The electric field may also have a "free-wave" solution depending upon the boundary conditions.[23]

Measurements of the second-harmonic-generation coefficient, defined schematically by $\mathbf{P}(2\omega)=\chi^{(2)}:\mathbf{E}(\omega)\mathbf{E}(\omega)$, involves measuring macroscopic electric fields at $\Omega=2\omega$, which are generated by interference between the free- and forced-wave solutions of Eq. (B1).[23] In principle, there is no distinction between the electric field at the second-harmonic frequency generated from long-wave (or macroscopic) polarization field or its short-wave (or local-field) components. Here, we ask the question: do the short-wave components of the polarization at the second-harmonic frequency affect the macroscopic electric fields to a significant degree? Inspection of Eq. (B3) indicates that this is equivalent to asking: how large is the off-diagonal coupling of the matrix **M**?

The diagonal element splitting of $\mathbf{M}_{GG}-\mathbf{M}_{00}$ is nearly $(2\pi/a)^2$ or greater, where a is the lattice constant of the crystal. The off-diagonal elements are of the order $\varepsilon_{GG'}(\Omega^2/c^2)=\varepsilon_{GG'}(2\pi/\lambda)^2$ where λ is the free space wavelength of light at angular frequency Ω; typically $\varepsilon_{GG'}$ is of order 1 or less for $\mathbf{G}\neq\mathbf{G}'$. Since **M** is diagonally dominant, the off-diagonal coupling is of the order of the ratio

$$\frac{\mathbf{M}_{0G}}{\mathbf{M}_{GG}-\mathbf{M}_{00}}=\epsilon_{0G}\frac{a^2}{\lambda^2}\ . \tag{B4}$$

Such a term should be neglected: the long-wavelength response formalism is predicated on the neglect of terms of order a/λ, which is about 10^{-3} for simple crystals and visible light.

---

[1] Z. H. Levine and D. C. Allan, Phys. Rev. Lett. **66**, 41 (1991).

[2] Z. H. Levine, Phys. Rev. B **42**, 3567 (1990); **44**, 3567(E) (1991).

[3] P. Boucaud, F. H. Julien, D. D. Yang, J.-M. Lourtioz, E. Rosencher, P. Bois, and J. Nagle, Appl. Phys. Lett. **57**, 215 (1990); P. Boucaud, F. H. Julien, D. D. Yang, J.-M. Lourtioz, E. Rosencher, and P. Bois, Opt. Lett. **16**, 199 (1991).

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

[5] Z. H. Levine and D. C. Allan, Phys. Rev. B **43**, 4187 (1991).

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

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

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

[9] B. F. Levine and C. G. Bethea, Appl. Phys. Lett. **20**, 272 (1972).

[10] D. E. Aspnes, Phys. Rev. B **6**, 4648 (1972).

[11] We have evaluated Eq. (2) taking into account (a) just *vv* cancellations, (b) *vvv* and *vv* cancellations, and (c) *vvv* and *vv* cancellations, with the *ccv*, *vvc*, and *cv* terms computed separately. We find that numerical results using method (a) are almost always grossly incorrect, and method (b) is an unreliable method, but does yield the correct answer sometimes. Method (c) is strongly recommended.

[12] See, for example, G. Baym, *Lectures on Quantum Mechanics* (Benjamin, Reading, MA, 1969), Chap. 11. Specifically, Eqs. (11-20), (11-23), and (11-35) give an expansion for the wave function in second order. Baym's V is H_1+H_2 here.

[13] N. Bloembergen, *Nonlinear Optics* (Benjamin, New York, 1965), p. 69.

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

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

[16] M.-Z. Huang and W. Y. Ching (unpublished).

[17] B. I. Min, S. Massidda, and A. J. Freeman, Phys. Rev. B **38**, 1970 (1988).

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

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

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

[21] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976), Chap. 2.

[22] Y. R. Shen, *The Principles of Non-Linear Optics* (Wiley, New York, 1984), p. 68.

[23] D. A. Kleinman, Phys. Rev. **128**, 1761 (1962).

[24] B. Monemar, Phys. Scr. **3**, 193 (1971).

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

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

[27] R. C. Miller, Appl. Phys. Lett. **5**, 17 (1964).

[28] C. Y. Fong and Y. R. Shen, Phys. Rev. B **12**, 2325 (1975).

[29] D. J. Moss, J. E. Sipe, and H. M. van Driel, Phys. Rev. B **36**, 9708 (1987).

[30] M. M. Choy and R. L. Byer, Phys. Rev. B **14**, 1693 (1976).

[31] K. Strössner, S. Ves, and M. Cardona, Phys. Rev. B **32**, 6614 (1985).

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

[33] G. A. Samara, Phys. Rev. B **27**, 3494 (1983).

[34] B. F. Levine, Phys. Rev. Lett. **22**, 787 (1969); **25**, 440 (1970).

[35] R. W. Ransen and O. F. Sankey, Phys. Rev. B **36**, 6520 (1987).

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

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

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

[40] E. Ghahramani, D. J. Moss, and J. E. Sipe, Phys. Rev. B **43**, 8990 (1991); **43**, 9269 (1991); **43**, 9700 (1991).