PHYSICAL REVIEW B VOLUME 42, NUMBER 6 15 AUGUST 1990-II

One-electron formalism for second-harmonic generation in crystalline semiconductors

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

(Received 13 November 1989)

The one-electron, self-consistent-field formalism for harmonic generation is presented. The solution is expanded in a Fourier series, assuming that the time-dependent perturbation is weak. The equations of harmonic response form a set of inhomogeneous Schrödinger equations which, if solved order by order, are coupled only through self-consistent-field effects. Special attention is paid to the case of second-harmonic generation in crystalline semiconductors with a longitudinal field present. The f-sum rule for crystals is exploited to eliminate certain terms which apparently diverge as ω^{-2}, in perfect analogy to the linear-response theory. Additionally, terms which are in the form of numerical first-, second-, and third-order finite differences of the frequency are combined into numerically superior forms.

I. INTRODUCTION

The calculation of linear-response coefficients for semiconductors has recently become quite accurate. Well-converged Hohenberg-Kohn-Sham local-density approximation (LDA) calculations of the dielectric constant of various semiconductors,[1-4] phonon frequencies,[5] and piezoelectric coefficients[4] have appeared. The systematic overestimate of the static dielectric constant by the LDA has been shown to be correctable by the inclusion of a "scissors operator" in the LDA Hamiltonian; deviations from experiment occur at the level of 4% in the cases of silicon and germanium.[3]

There has not yet been a successful *ab initio* calculation of second-order nonlinear optical response in a semiconductor to my knowledge. Previous work includes a semiempirical pseudopotential calculation[6] and an empirical tight-binding calculation;[7] despite the use of fitting parameters these works disagree with experiments, the static limit is underestimated by a factor of 10–100 in Ref. 6 and overestimated by a factor of 1–4 in Ref. 7. It is a measure of the difficulty of the problem that the literature is riddled with errors even at the formal level.[8] It is only a few years since the first correct local-density-approximation calculation of the third-harmonic generation coefficients in rare-gas atoms;[9,10] and even these calculations overestimate the experimental hyperpolarizabilities by about a factor of 2. These calculations have been extended to alkali-halide crystals in the "spherical-solid" approximation.[11] Second-harmonic generation at a jellium surface has also been studied recently using the LDA.[12]

The goal of the present work is to set the stage for the calculation of a hyperpolarizability $\chi^{(2)}(2\omega;\omega,\omega)$ in a semiconducting crystal such as GaAs in a framework of *ab initio* one-electron theory and self-consistent induced fields.[1-4,9,13-17] The one-electron response formalism is set down as a set of inhomogeneous Schrödinger equations; the inhomogeneous Schrödinger equations have been used previously for the calculation of atomic polarizabilities[15-17] and hyperpolarizabilities[9,10] and more recently have been introduced into the solid-state context.[5] Certain exact cancellations are applied to these equations. The resulting formulas are, it is hoped, computationally tractable.

The formulation is restricted to longitudinal response; optical experiments will typically measure transverse response. In linear electronic response, the long-wavelength longitudinal dielectric function is equal to the corresponding transverse quantity.[18,19] However, in the case of phonons, the longitudinal- and transverse-optical frequencies are not equal in polar semiconductors; instead they are related through the Lyddane-Sachs-Teller relation.[20] So far as I have been able to determine, the relationship between transverse and longitudinal response is not known for the hyperpolarizabilities. Presumably, elements of the approach developed here will be useful in a deriving the more complete current-current response formalism.[13]

II. GENERAL ONE-ELECTRON FORMALISM

Consider the time-dependent Schrödinger equation

$$\left[-i\frac{\partial}{\partial t}-H_0\right]|\psi(t)\rangle=V(t)|\psi(t)\rangle \; ; \tag{2.1}$$

here, H_0 is a time-independent Hamiltonian, and the standard phase convention is associated with the equation for the "bra" form $\langle\psi(t)|$. We will restrict consideration to the case where $V(t)$ is a small interaction of the form

$$V(t)=\sum_{j=1}^{\infty}\lambda^j \sum_{l=-j}^{j}{}^{[2]}\, V_{l\omega}^{(j)}e^{il\omega t} \; . \tag{2.2}$$

Here, the $V_{l\omega}^{(j)}$ are permitted to be arbitrary time-independent operators, including scalar and vector potentials. Time-reversal symmetry imposes the restriction $V_{l\omega}^{(j)}=V_{-l\omega}^{(j)\dagger}$. The notation $\sum_{l=-j}^{j[2]}$ means the index l is incremented by 2's. We will seek solutions order by order in λ, which ultimately will be set to unity.[21] In practice, we will typically be concerned only with $j \le 3$, i.e., up to third-order processes. (An eye-catching exception to this is the recent observation of odd harmonics as high as 33rd order in rare-gas atoms.[22])

We will seek a solution in the form

$$|\psi(t)\rangle = \sum_{j=0}^{\infty} \lambda^j \sum_{l=-j}^{j}{}^{[2]} |\psi_l^{(j)}\rangle e^{i(\varepsilon + l\omega)t} \tag{2.3}$$

with a temporary normalization $\langle \psi_0^{(0)} | \psi_0^{(0)} \rangle = 1$. (The correct normalization is considered in Sec. III.) Substituting the form (2.3) in Eq. (2.1), we arrive at

$$\sum_{j=0}^{\infty} \lambda^j \sum_{l=-j}^{j}{}^{[2]} (\varepsilon + l\omega - H_0) |\psi_l^{(j)}\rangle e^{il\omega t} = \sum_{j'=1}^{\infty} \lambda^{j'} \sum_{l'=-j'}^{j'}{}^{[2]} V_{l'\omega}^{(j')} e^{il'\omega t} \sum_{j''=0}^{\infty} \lambda^{j''} \sum_{l''=-j''}^{j''}{}^{[2]} |\psi_{l''}^{(j'')}\rangle e^{il''\omega t} \; . \tag{2.4}$$

This equation must be satisfied separately for each frequency $l\omega$ and each power of λ. Let $\Theta_j = 1$ for $j \ge 0$ and $\Theta_j = 0$ for $j < 0$. Hence, for $|l| \le j$ with $j - l$ even,

$$(\varepsilon + l\omega - H_0) |\psi_l^{(j)}\rangle = \sum_{j'=1}^{j} \sum_{l'=-j'}^{j'}{}^{[2]} \Theta_{j-j'+l-l'} \Theta_{j-j'-l+l'} V_{l'\omega}^{(j')} |\psi_{l-l'}^{(j-j')}\rangle \; . \tag{2.5}$$

Only terms in $V_{l\omega}^{(j)}$ and $|\psi_l^{(j)}\rangle$ with the appropriate parity indices enter Eq. (2.5). Since $j' \ge 1$, if the $V_{l\omega}^{(j)}$ are known, the functions $|\psi_l^{(j)}\rangle$ may be determined independently of each other in order of ascending values of j. In practice, we will seek a self-consistent solution, i.e., the $V_{l\omega}^{(j)}$ will depend upon the $|\psi_l^{(j')}\rangle$ for $j' \le j$; this represents only a moderate increase in the difficulty of the problem. The amount of $|\psi_0^{(0)}\rangle$ to be included in $|\psi_0^{(j)}\rangle$ is not determined. However, as the value of $\langle \psi_0^{(j)} | \psi_0^{(0)} \rangle$ can affect only the phase of the normalized wave function, we take $\langle \psi_0^{(j)} | \psi_0^{(0)} \rangle = 0$, $j > 0$.

We expand Eq. (2.5) for process from zeroth order to third order in the perturbation:

$$j=0,\ l=0:\ (\varepsilon - H_0)|\psi_0^{(0)}\rangle = 0 \; ; \tag{2.6a}$$

$$j=1,\ l=\pm 1:\ (\varepsilon \pm \omega - H_0)|\psi_{\pm 1}^{(1)}\rangle = V_{\pm\omega}^{(1)} |\psi_0^{(0)}\rangle \; ; \tag{2.6b}$$

$$j=2,\ l=\pm 2:\ (\varepsilon \pm 2\omega - H_0)|\psi_{\pm 2}^{(2)}\rangle = V_{\pm 2\omega}^{(2)} |\psi_0^{(0)}\rangle + V_{\pm\omega}^{(1)} |\psi_{\pm 1}^{(1)}\rangle \; ; \tag{2.6c}$$

$$j=2,\ l=0:\ (\varepsilon - H_0)|\psi_0^{(2)}\rangle = V_0^{(2)} |\psi_0^{(0)}\rangle + V_{\omega}^{(1)} |\psi_{-1}^{(1)}\rangle + V_{-\omega}^{(1)} |\psi_1^{(1)}\rangle \; ; \tag{2.6d}$$

$$j=3,\ l=\pm 3:\ (\varepsilon \pm 3\omega - H_0)|\psi_{\pm 3}^{(3)}\rangle = V_{\pm 3\omega}^{(3)} |\psi_0^{(0)}\rangle + V_{\pm 2\omega}^{(2)} |\psi_{\pm 1}^{(1)}\rangle + V_{\pm\omega}^{(1)} |\psi_{\pm 2}^{(2)}\rangle \; ; \tag{2.6e}$$

$$j=3,\ l=\pm 1:\ (\varepsilon \pm \omega - H_0)|\psi_{\pm 1}^{(3)}\rangle = V_{\pm\omega}^{(3)} |\psi_0^{(0)}\rangle + V_{\pm 2\omega}^{(2)} |\psi_{\mp 1}^{(1)}\rangle + V_0^{(2)} |\psi_{\pm 1}^{(1)}\rangle + V_{\pm\omega}^{(1)} |\psi_0^{(2)}\rangle + V_{\mp\omega}^{(1)} |\psi_{\pm 2}^{(2)}\rangle \; . \tag{2.6f}$$

The formalism presented here is similar to the formulation of Subbaswamy and others.[9,10]

III. WAVE-FUNCTION NORMALIZATION

In Sec. I, the normalization $\langle \psi_0^{(0)} | \psi_0^{(0)} \rangle = 1$ was introduced. The correct wave function is defined by $|\bar{\psi}(t)\rangle = \bar{Z}^{-1/2} |\psi(t)\rangle$ and $\langle \bar{\psi}(t) | \bar{\psi}(t) \rangle = 1$. The normalization constant $\bar{Z}$ is given by $\bar{Z} = \langle \psi(t) | \psi(t) \rangle$. The normalization constant has no time dependence because probability is conserved in systems with a Hermitian Hamiltonian. Let us define expansion coefficients

$$\bar{Z} = \sum_{j=0}^{\infty} \lambda^j \bar{Z}^{(j)} \; . \tag{3.1}$$

From the expansion for the wave function in Eq. (2.3),

$$\bar{Z} = \sum_{j=0}^{\infty} \lambda^j \sum_{j'=0}^{j} \sum_{l'=-j'}^{j'}{}^{[2]} \sum_{l=-j+j'-l'}^{j-j'-l'} \langle \psi_{l'}^{(j')} | \psi_{l+l'}^{(j-j')} \rangle e^{il\omega t} \; . \tag{3.2}$$

Although Eq. (3.2) is apparently time dependent, the conservation of probability requires the time-dependent terms to vanish. (See the Appendix for an expanded discussion.) We may identify terms separately for each j. Hence,

$$\bar{Z}^{(j)} = \sum_{j'=0}^{j} \sum_{l'=-j'}^{j'}{}^{[2]} \Theta_{j-j'+l'} \Theta_{j-j'-l'} \langle \psi_{l'}^{(j')} | \psi_{l'}^{(j-j')} \rangle \; . \tag{3.3}$$

For $j \le 3$, the only nontrivial term is

$$\bar{Z}^{(2)} = \langle \psi_1^{(1)} | \psi_1^{(1)} \rangle + \langle \psi_{-1}^{(1)} | \psi_{-1}^{(1)} \rangle \; . \tag{3.4}$$

Also, $\bar{Z}^{(0)} = 1$. All terms with j odd vanish. In particular, the wave-function normalization constant does not affect linear response.

Since $\bar{Z}$ is of the form

$$\bar{Z} = 1 + \lambda^2 \bar{Z}^{(2)} + O(\lambda^4) \; , \tag{3.5}$$

the normalization constant has the expansion

$$\bar{Z}^{-1/2} = 1 - \tfrac{1}{2} \lambda^2 \bar{Z}^{(2)} + O(\lambda^4) \; . \tag{3.6}$$

The normalized wave function may be written as

$$\begin{aligned} |\bar{\psi}(t)\rangle = {} & [\,|\psi_0^{(0)}\rangle (1 - \tfrac{1}{2}\lambda^2 \bar{Z}^{(2)}) + |\psi_0^{(2)}\rangle \lambda^2] e^{i\varepsilon t} \\ & + \sum_{\pm} [\,|\psi_{\pm 1}^{(1)}\rangle (\lambda - \tfrac{1}{2}\lambda^3 \bar{Z}^{(2)}) + |\psi_{\pm 1}^{(3)}\rangle \lambda^3] e^{i(\varepsilon \pm \omega)t} \\ & + |\psi_{\pm 2}^{(2)}\rangle \lambda^2 e^{i(\varepsilon \pm 2\omega)t} + |\psi_{\pm 3}^{(3)}\rangle \lambda^3 e^{i(\varepsilon \pm 3\omega)t} + O(\lambda^4) \; . \end{aligned} \tag{3.7}$$

IV. INDUCED ELECTRON DENSITY

The ground-state electron number density is given by $n_0(\mathbf{r})=\langle\psi_0^{(0)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(0)}\rangle$. The electron number density induced by the interaction $V(t)$ is given by

$$\delta n(\mathbf{r},t)=\langle\bar{\psi}(t)|\mathbf{r}\rangle\langle\mathbf{r}|\bar{\psi}(t)\rangle-n_0(\mathbf{r})\ . \tag{4.1}$$

The induced electron density has an expansion of the form

$$\delta n(\mathbf{r},t)=\sum_{j=1}^{\infty}\lambda^j\sum_{l=-j}^{j}{}^{[2]}\delta n_l^{(j)}(\mathbf{r})e^{il\omega t}\ . \tag{4.2}$$

Since $\delta n(\mathbf{r},t)$ is real, $\delta n_l^{(j)}=\delta n_{-l}^{(j)*}$. Explicit formulas for the low-order $\delta n_l^{(j)}(\mathbf{r})$ may be obtained from Eq. (3.7). For $j\le 3$, these are the following:

$$\delta n_{\pm1}^{(1)}(\mathbf{r})=\langle\psi_0^{(0)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm1}^{(1)}\rangle+\langle\psi_{\mp1}^{(1)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(0)}\rangle\ , \tag{4.3a}$$

$$\delta n_{\pm2}^{(2)}(\mathbf{r})=\langle\psi_0^{(0)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm2}^{(2)}\rangle+\langle\psi_{\mp1}^{(1)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm1}^{(1)}\rangle+\langle\psi_{\mp2}^{(2)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(0)}\rangle\ , \tag{4.3b}$$

$$\delta n_{0}^{(2)}(\mathbf{r})=\langle\psi_0^{(0)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(2)}\rangle+\langle\psi_1^{(1)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_1^{(1)}\rangle+\langle\psi_{-1}^{(1)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{-1}^{(1)}\rangle+\langle\psi_0^{(2)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(0)}\rangle-\bar{Z}^{(2)}n_0(\mathbf{r})\ , \tag{4.3c}$$

$$\delta n_{\pm3}^{(3)}(\mathbf{r})=\langle\psi_0^{(0)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm3}^{(3)}\rangle+\langle\psi_{\mp1}^{(1)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm2}^{(2)}\rangle+\langle\psi_{\mp2}^{(2)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm1}^{(1)}\rangle+\langle\psi_{\mp3}^{(3)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(0)}\rangle\ , \tag{4.3d}$$

$$\begin{aligned}\delta n_{\pm1}^{(3)}(\mathbf{r})=&\ \langle\psi_0^{(0)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm1}^{(3)}\rangle+\langle\psi_{\pm1}^{(1)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm2}^{(2)}\rangle+\langle\psi_{\mp1}^{(1)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(2)}\rangle\\&+\langle\psi_0^{(2)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\pm1}^{(1)}\rangle+\langle\psi_{\mp2}^{(2)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_{\mp1}^{(1)}\rangle+\langle\psi_{\mp1}^{(3)}|\mathbf{r}\rangle\langle\mathbf{r}|\psi_0^{(0)}\rangle-\bar{Z}^{(2)}\delta n_{\pm1}^{(1)}(\mathbf{r})\ .\end{aligned} \tag{4.3e}$$

The integrated induced charge density must vanish. This will be demonstrated for the formulas in Eq. (4.3). We may invoke the completeness relation

$$I=\int d\mathbf{r}\,|\mathbf{r}\rangle\langle\mathbf{r}|\ . \tag{4.4}$$

Applied to the ground-state charge density, we compute

$$\begin{aligned}\int d\mathbf{r}\,n_0(\mathbf{r})&=\int d\mathbf{r}\,\langle\psi_0^{(0)}|\mathbf{r}\rangle\,\langle\mathbf{r}|\psi_0^{(0)}\rangle\\&=\langle\psi_0^{(0)}|\psi_0^{(0)}\rangle\\&=1\ .\end{aligned} \tag{4.5}$$

The integrated first-order response is

$$\begin{aligned}\int d\mathbf{r}\,\delta n_{\pm1}^{(1)}(\mathbf{r})&=\langle\psi_0^{(0)}|\psi_{\pm1}^{(1)}\rangle+\langle\psi_{\mp1}^{(1)}|\psi_0^{(0)}\rangle\\&=\bar{Z}_{\pm1}^{(1)}\\&=0\ .\end{aligned} \tag{4.6}$$

(The symbol $\bar{Z}_{\pm1}^{(1)}$ is defined in the Appendix.) For the second-order induced number density, application of the identity (4.4) to Eq. (4.3), the results of Eqs. (3.4) and (4.5), and the relation $\langle\psi_0^{(0)}|\psi_0^{(2)}\rangle=0$ lead to the conclusion that the integrated induced charge density vanishes. Moreover, there would have been no effect on the induced density if $\langle\psi_0^{(0)}|\psi_0^{(2)}\rangle$ had not been required to vanish. More specifically, cancellation occurs between the ordinary terms and the normalization corrections in the expressions for $\delta n_0^{(2)}(\mathbf{r})$ and $\delta n_{\pm1}^{(3)}(\mathbf{r})$ of Eq. (4.3). (In this case, the normalization constant would be $\bar{Z}^{(2)}=\langle\psi_0^{(0)}|\psi_0^{(2)}\rangle+\langle\psi_1^{(1)}|\psi_1^{(1)}\rangle+\langle\psi_{-1}^{(1)}|\psi_{-1}^{(1)}\rangle+\langle\psi_0^{(0)}|\psi_0^{(2)}\rangle$.) For the other terms in Eq. (4.3) the completeness relation of Eq. (4.4) reduces the expression to the time-dependent terms in the wave-function normalization which all vanish, as discussed in the Appendix.

V. LONGITUDINAL RESPONSE IN INSULATING CRYSTALS

Consider the case in which H_0 represents a crystalline solid. For definiteness, we restrict attention to a longitudinal perturbation of the form

$$\phi(\mathbf{r},t)=\sum_{j=1}^{\infty}\lambda^j\sum_{l=-j}^{j}{}^{[2]}\sum_{m=-j}^{j}{}^{[2]}\tilde{\phi}_{lm}^{(j)}(\mathbf{r})e^{i(l\omega t+m\mathbf{q}\cdot\mathbf{r})}\ , \tag{5.1}$$

where each $\tilde{\phi}_{lm}^{(j)}(\mathbf{r})$ is periodic in the unit cell. As indicated in the Introduction, the response to light should be done with in a transverse or current-response formalism, but this harder problem is deferred for now. Let the long-wave part of this potential be denoted by $\tilde{\phi}_{lm}^{(j)}=\Omega_0^{-1}\int_0 d\mathbf{r}\,\tilde{\phi}_{lm}^{(j)}(\mathbf{r})$, where the subscript 0 refers to an integral over a unit cell, and Ω_0 is the unit-cell volume. Reality of $\phi(\mathbf{r},t)$ requires $\tilde{\phi}_{lm}^{(j)}(\mathbf{r})=\tilde{\phi}_{-l,-m}^{(j)*}(\mathbf{r})$, but otherwise leaves these coefficients unrestricted. In particular, for $j=1$, $\tilde{\phi}_{11}^{(1)}(\mathbf{r})=\tilde{\phi}_{-1,-1}^{(1)*}(\mathbf{r})$ and $\tilde{\phi}_{1,-1}^{(1)}(\mathbf{r})=\tilde{\phi}_{-1,1}^{(1)*}(\mathbf{r})$, but reality imposes no relationship between $\tilde{\phi}_{11}^{(1)}(\mathbf{r})$ and $\tilde{\phi}_{1,-1}^{(1)}(\mathbf{r})$.

We may expand the wave function and the charge density in the wave vector as well as the frequency in expansions similar to Eq. (5.1). The normalization constant $\bar{Z}$ has no spatial dependence, hence no wave-vector expan-

TABLE I. Lowest nonvanishing order in q for various quantities in linear response (ω) and second-harmonic generation (2ω).

		ϕ	$\mathbf{E},\mathbf{P}$	δn
ω	long wave	1	q	q^2
ω	short wave	q	q	q
2ω	long wave	q	q^2	q^3
2ω	short wave	q^2	q^2	q^2

sion. Specifically,

$$|\psi(t)\rangle = \sum_{j=0}^{\infty} \lambda^j \sum_{l=-j}^{j}{}^{[2]} \sum_{m=-j}^{j}{}^{[2]} |\tilde{\psi}_{lm}^{(j)}\rangle e^{i[(\varepsilon+l\omega)t+m\mathbf{q}\cdot\mathbf{r}]} \tag{5.2}$$

and

$$\delta n(\mathbf{r},t) = \Omega_0^{-1} \sum_{j=1}^{\infty} \lambda^j \sum_{l=-j}^{j}{}^{[2]} \sum_{m=-j}^{j}{}^{[2]} \delta\tilde{n}_{lm}^{(j)}(\mathbf{r}) e^{i(l\omega t+m\mathbf{q}\cdot\mathbf{r})} ; \tag{5.3}$$

the functions $\langle \mathbf{r}|\tilde{\psi}_{lm}^{(j)}\rangle$ and $\delta\tilde{n}_{lm}^{(j)}(\mathbf{r})$ are periodic in the unit cell. Let $\delta\tilde{n}_{lm}^{(j)} = \Omega_0^{-1}\int_0 d\mathbf{r}\, \delta\tilde{n}_{lm}^{(j)}(\mathbf{r})$. The superscript s will refer to a "short-wave part," i.e., $\delta\tilde{n}_{lm}^{(js)}(\mathbf{r}) = \delta\tilde{n}_{lm}^{(j)}(\mathbf{r}) - \delta\tilde{n}_{lm}^{(j)}$.

Explicit formulas for the expansion coefficients of Eq. (5.3) are presented below for linear response ($j=1$) and for second-harmonic generation ($j=2$, $|l|=2$, and $|m|=2$). (The presentation is limited in this section because the formulas are lengthy. The $m=0$ terms are a form of "three-wave mixing," and hence are excluded from the presentation.) The results presented here beyond linear response were obtained with the aid of the software package MATHEMATICA.[23]

For optical response, we are interested in the long-wavelength limit, i.e., $\mathbf{q}\rightarrow\mathbf{0}$.[24] The lowest nonvanishing terms are summarized in Table I. This pattern of these lowest non-vanishing orders is most easily understood by considering the electric fields.

A. Linear response

In linear response,[3,25] the long-wave-induced charge density is given by

$$\delta\tilde{n}_{lm}^{(1)} = \overline{\Omega}_0 \int_{\mathrm{BZ}} d\mathbf{k} \sum_{n,\pm}^{\mathrm{occ}} \omega^{-2}\tilde{\phi}_{lm}^{(1)} [\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1 G_{n\mathbf{k}}(\pm\omega) H_1|n,\tilde{\mathbf{k}}\rangle + \mathcal{R}\langle n,\tilde{\mathbf{k}}|H_2|n,\tilde{\mathbf{k}}\rangle] \pm i\omega^{-1}\mathcal{J}\langle n,\tilde{\mathbf{k}}|H_1 G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{lm}^{(1s)}|n,\tilde{\mathbf{k}}\rangle , \tag{5.4}$$

where $H_{\mathbf{k}}$ is the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, and H_1 is the leading term of the expansion of $H_{\mathbf{k}+\mathbf{q}}$ in $\mathbf{q}$ (typically $\mathbf{q}\cdot\mathbf{p}$). H_2 is the second term in this expansion (typically $\frac{1}{2}q^2$). The "typical" identifications are not made in this work in anticipation of nonlocal potentials, such as nonlocal pseudopotentials[26] or the "scissors operator."[3] The one-electron Green's operator is

$$G_{n\mathbf{k}}(\omega) = (\varepsilon_{n\mathbf{k}} + \omega - H_{\mathbf{k}})^{-1} . \tag{5.5}$$

In Eq. (5.4), the Green's operators may be (a) unrestricted, (b) projected away from the state $|n\tilde{\mathbf{k}}\rangle$, (c) projected away from all states of energy $\varepsilon_{n\mathbf{k}}$, or (d) projected away from all occupied states. Let $G_{n\mathbf{k}} = (\varepsilon_{n\mathbf{k}} - H_{\mathbf{k}})^{-1}$, projected away from the subspace of wave functions with energy $\varepsilon_{n\mathbf{k}}$. The kets $|n,\tilde{\mathbf{k}}\rangle$ are periodic in the unit cell, and obey $H_{\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle = \varepsilon_{n\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle$. Also, n is a band index, and the symbol "occ" limits the sum to the occupied states. The wave functions themselves carry an additional phase $e^{i\mathbf{k}\cdot\mathbf{r}}$, i.e., $|n,\mathbf{k}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|n,\tilde{\mathbf{k}}\rangle$. The notation BZ refers to the Brillouin zone at the domain of integration, and $\overline{\Omega}_0$ is the unit-cell volume divided by $(2\pi)^3$. The superscript s implies that only the short-wave part of the term is used; i.e., $\tilde{\phi}_{1m}^{(1s)}(\mathbf{r}) = \tilde{\phi}_{1m}^{(1)}(\mathbf{r}) - \tilde{\phi}_{1m}^{(1)}$. Inside a matrix element, $\tilde{\phi}_{lm}^{(js)}$ is an operator which is local in real space. The range of the summation over n and possibly ± extends to the end of the expression; terms without an explicit ± dependence occurring in the range of a ± summation simply sum to twice the expression written. Time-reversal symmetry has been used to derive Eq. (5.4), and most of following formulas. Specifically, we take $\langle\mathbf{r}|n,\tilde{\mathbf{k}}\rangle = \langle\mathbf{r}|n,-\tilde{\mathbf{k}}\rangle^*$, and both states have the same energy and occupancy. The operator H_1 has odd parity, and H_2 and $\tilde{\phi}_{1m}^{(1s)}$ have even parity. The symbol $\mathcal{R}$ operating on a matrix element is defined as

$$\mathcal{R}\langle n,\tilde{\mathbf{k}}|O|n,\tilde{\mathbf{k}}\rangle = \mathrm{Re}\langle n,\tilde{\mathbf{k}}|\mathrm{Re}O|n,\tilde{\mathbf{k}}\rangle + i\,\mathrm{Re}\langle n,\tilde{\mathbf{k}}|\mathrm{Im}O|n,\tilde{\mathbf{k}}\rangle .$$

Similarly, the symbol $\mathcal{J}$ operating on a matrix element is defined as

$$\mathcal{R}\langle n,\tilde{\mathbf{k}}|O|n,\tilde{\mathbf{k}}\rangle = \mathrm{Im}\langle n,\tilde{\mathbf{k}}|\mathrm{Re}O|n,\tilde{\mathbf{k}}\rangle + i\,\mathrm{Im}\langle n,\tilde{\mathbf{k}}|\mathrm{Im}O|n,\tilde{\mathbf{k}}\rangle$$

for any operator O.

Non-Hermitian operators enter the discussion in several ways: the product of Hermitian operators is not, in general, Hermitian; $G_{n\mathbf{k}}(\omega)$ may have an imaginary part if ω is above the absorption threshold; moreover, $\tilde{\phi}$ is, in general, complex. Below the absorption threshold, $\tilde{\phi}_{1m}^{(1)}$ is purely real, but $\tilde{\phi}_{1m}^{(1s)}(\mathbf{r})$ is purely imaginary. Above the absorption threshold, there is no general phase relation for $\tilde{\phi}_{1m}^{(1)}(\mathbf{r})$.

Application of the f-sum rule (discussed in Sec. VI) for insulating crystals leads to

$$\begin{aligned}\delta\tilde{n}_{1m}^{(1)} = \overline{\Omega}_0 \int_{\mathrm{BZ}} d\mathbf{k} \sum_{n,\pm}^{\mathrm{occ}} \omega^{-2}\tilde{\phi}_{1m}^{(1)} [\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1 G_{n\mathbf{k}}(\pm\omega) H_1|n,\tilde{\mathbf{k}}\rangle - \mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1 G_{n\mathbf{k}} H_1|n,\tilde{\mathbf{k}}\rangle] \\ \pm i\omega^{-1}\mathcal{J}\langle n,\tilde{\mathbf{k}}|H_1 G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle ,\end{aligned} \tag{5.6}$$

a formula which does not hold for metals. (The terms "semiconductor" and "insulator" are used interchangeably in this work.) Computationally, Eq. (5.6) is superior to Eq. (5.4) because a cancellation implicit in a Brillouin-zone integral is performed analytically, not numerically. The situation may be further improved by noting that

$$G_{n\mathbf{k}}(\omega)+G_{n\mathbf{k}}(-\omega)-2G_{n\mathbf{k}}=2\omega^2 G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega) \tag{5.7}$$

and, for reference,

$$G_{n\mathbf{k}}(\omega)-G_{n\mathbf{k}}(-\omega)=-2\omega G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\ . \tag{5.8}$$

Applying Eq. (5.7) to Eq. (5.6) yields

$$\delta\tilde{n}_{1m}^{(1)}=2\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\tilde{\phi}_{1m}^{(1)}\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle-i\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle\ , \tag{5.9}$$

which eliminates the manifest finite differences of Eq. (5.6). Formulas like Eq. (5.6) are typical of those generated from a diagrammatic analysis.[8,27] The "product form" of Eq. (5.9) eliminates a "subtle cancellation" noted in Ref. 8.

The short-wave-induced charge density is given by

$$\begin{aligned}\delta\tilde{n}_{1m}^{(1s)}(\mathbf{r})&=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n,\pm}^{\mathrm{occ}}\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle\pm i\omega^{-1}\tilde{\phi}_{1m}^{(1)}\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm\omega)H_1|n,\tilde{\mathbf{k}}\rangle\\&=2\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+i\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle\end{aligned} \tag{5.10}$$

for metals and insulators alike, making use of Eq. (5.8). Below the absorption threshold, $\delta\tilde{n}_{1m}^{(1s)}(\mathbf{r})$ is purely imaginary, and $\delta\tilde{n}_{1m}^{(1)}$ is purely real. The Green's operator may be restricted in Eq. (5.8) the same four ways as for Eq. (5.4).

The $\omega\rightarrow 0$ limits of these formulas are, in the case of insulators,

$$\delta\tilde{n}_{1m}^{(1)}=2\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\tilde{\phi}_{1m}^{(1)}\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^3H_1|n,\tilde{\mathbf{k}}\rangle-i\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle \tag{5.11}$$

and

$$\delta\tilde{n}_{1m}^{(1s)}(\mathbf{r})=2\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle-i\tilde{\phi}_{1m}^{(1)}\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2H_1|n,\tilde{\mathbf{k}}\rangle\ . \tag{5.12}$$

For metals, a leading term of $O(\omega^{-2})$ would enter the long-wave response, Eq. (5.11); the short-wave response, Eq. (5.12), remains correct.

B. Second-harmonic generation

The process of second-harmonic generation in the one-electron, self-consistent-field picture has three steps: (1) an external field at frequency ω is imposed, and screened with a self-consistent linear response formalism, (2) the screened fields at frequency ω induce a bare charge density at frequency 2ω, and finally (3) the bare charge density at frequency 2ω generates forced and free electromagnetic waves at frequency 2ω, first discussed by Kleinman.[28] A large response is generated when the "pulse matching" condition is achieved between the forced and free waves.[28] This subsection is devoted to step (2)—the bare charge density induced at frequency 2ω; step (1) has been dealt with in the preceding subsection, and elsewhere.[1–3,13,14,25]

The long-wave-induced charge densities are given by

$$\begin{aligned}\delta\tilde{n}_{2(2m)}^{(2)}=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n,\pm}^{\mathrm{occ}}&\,i\omega^{-3}(\tilde{\phi}_{1m}^{(1)})^2[\pm\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm 2\omega)H_2|n,\tilde{\mathbf{k}}\rangle\mp 2\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm\omega)H_2|n,\tilde{\mathbf{k}}\rangle\\&\qquad\pm\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm 2\omega)H_1G_{n\mathbf{k}}(\pm\omega)H_1|n,\tilde{\mathbf{k}}\rangle\\&\qquad+\tfrac{1}{2}\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(-\omega)H_1G_{n\mathbf{k}}(\omega)H_1|n,\tilde{\mathbf{k}}\rangle]\\&+\omega^{-2}\tilde{\phi}_{1m}^{(1)}[2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_2G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm 2\omega)H_1G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle\\&\qquad+\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm 2\omega)\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}(\pm\omega)H_1|n,\tilde{\mathbf{k}}\rangle+\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm\omega)H_1G_{n\mathbf{k}}(\mp\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle]\\&+i\omega^{-1}[\pm\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm 2\omega)\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+\tfrac{1}{2}\mathcal{I}\langle n,\tilde{\mathbf{k}}|\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}(-\omega)H_1G_{n\mathbf{k}}(\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle]\end{aligned} \tag{5.13}$$

for $|m|=1$. In Eq. (5.13), the Green's operators may be (a) unrestricted, (b) projected away from the state $|n,\tilde{\mathbf{k}}\rangle$, or (c) projected away from all states of energy $\varepsilon_{n\mathbf{k}}$. Terms involving more than one Green's operator may not be projected away from all occupied states as in case (d) for linear response. If this projection is employed one recovers what

Aspnes[29] calls "vcc" terms or "virtual-electron" terms. The difference between these and the full expression are called "cvv" terms or "virtual-hole" terms. At zero wave vector, the long-wave-induced charge $\delta\tilde{n}_{20}^{(2)}=0$ vanishes at all frequencies and to all orders in the expansion in powers of q; this is a derived result which is required to avoid a net induced charge.

The short-wave-induced charge densities are given by

$$\begin{aligned}\delta\tilde{n}_{2(2m)}^{(2s)}(\mathbf{r})=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n,\pm}^{\mathrm{occ}}\omega^{-2}(\tilde{\phi}_{1m}^{(1)})^2[&\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm2\omega)H_2|n,\tilde{\mathbf{k}}\rangle+\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm2\omega)H_1G_{n\mathbf{k}}(\pm\omega)H_1|n,\tilde{\mathbf{k}}\rangle\\&+\tfrac{1}{2}\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)H_1|n,\tilde{\mathbf{k}}\rangle]\\&+i\omega^{-1}\tilde{\phi}_{1m}^{(1)}[\mp\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\pm\omega)|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\mp\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle\pm\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm2\omega)H_1G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle\\&\pm\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm2\omega)\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}(\pm\omega)H_1|n,\mathbf{k}\rangle]\\&+[\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\pm2\omega)\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}(\pm\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+\tfrac{1}{2}\mathcal{R}\langle n,\tilde{\mathbf{k}}|\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle]\ .\end{aligned}\tag{5.14}$$

The rules for restricting the Green's operator are the same as for Eq. (5.13). Below threshold, $\delta\tilde{n}_{2(2m)}^{(2s)}(\mathbf{r})$ is purely real and $\delta\tilde{n}_{2(2m)}^{(2)}$ is purely imaginary. This pattern is the reverse of that obtained for the corresponding linear-response terms.

In the static limit, the induced charge densities are

$$\begin{aligned}\delta\tilde{n}_{2(2m)}^{(2)}=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}i(\tilde{\phi}_{1m}^{(1)})^2[&-12\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^4H_2|n,\tilde{\mathbf{k}}\rangle+12\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^4H_1|n,\tilde{\mathbf{k}}\rangle\\&+6\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^2H_1G_{n\mathbf{k}}^3H_1|n,\tilde{\mathbf{k}}\rangle]\\&+\tilde{\phi}_{1m}^{(1)}[4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_2G_{n\mathbf{k}}^3\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^3\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+10\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}^3H_1|n,\tilde{\mathbf{k}}\rangle\\&+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^2H_1G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}^2H_1|n,\mathbf{k}\rangle\\&+10\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^3H_1G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle]\\&+i[-2\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle-4\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle\\&-2\mathcal{I}\langle n,\tilde{\mathbf{k}}|\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle]\end{aligned}\tag{5.15}$$

for long waves, assuming that an apparently diverging term vanishes, i.e.,

$$\begin{aligned}0=\omega^{-2}\tilde{\phi}_{1m}^{(1)}\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}&4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_2G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle\\&+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\ ,\end{aligned}\tag{5.16}$$

in analogy with the f-sum rule needed in the linear case. The proof of Eq. (5.16) is presented in Sec. VI.

For short waves in the static limit, the induced charge density is given by

$$\begin{aligned}\delta\tilde{n}_{2(2m)}^{(2s)}(\mathbf{r})=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}(\tilde{\phi}_{1m}^{(1)})^2[&8\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^3H_2|n,\tilde{\mathbf{k}}\rangle+8\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^3H_1G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\\&+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^3|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle+4\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2H_1G_{n\mathbf{k}}^2H_1|n,\tilde{\mathbf{k}}\rangle\\&-\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^2|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2H_1|n,\tilde{\mathbf{k}}\rangle+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^3H_1|n,\tilde{\mathbf{k}}\rangle]\\&+i\tilde{\phi}_{1m}^{(1)}[-2\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle-4\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\\&+2\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}^2|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle-2\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}^2H_1|n,\tilde{\mathbf{k}}\rangle\\&-2\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1G_{n\mathbf{k}}^2\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle-4\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}^2H_1G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle]\\&+[2\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle+\mathcal{R}\langle n,\tilde{\mathbf{k}}|\tilde{\phi}_{1m}^{(1s)}G_{n\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}\tilde{\phi}_{1m}^{(1s)}|n,\tilde{\mathbf{k}}\rangle]\end{aligned}\tag{5.17}$$

assuming that

$$\begin{aligned}0=\omega^{-2}(\tilde{\phi}_{1m}^{(1)})^2\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}}&2\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_2|n,\tilde{\mathbf{k}}\rangle+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\\&+\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\ ,\end{aligned}\tag{5.18}$$

an assumption which is similar to Eq. (5.16), and is justified in Sec. VI.

A subtlety arises in this limit owing to the existence of terms of the form $\delta\tilde{n}\,^{(2)}_{0q}$; these terms should be included if a strictly dc phenomenon is considered, but not for a "low-frequency" phenomenon. The expressions in Eqs. (5.15) and (5.17) are simply the $\omega\rightarrow 0$ limits of Eqs. (5.13) and (5.14), without a contribution from $\delta\tilde{n}\,^{(2)}_{0q}$. For metals, terms of $O(\omega^{-2})$ arise in the charge density for both long and short waves in second-harmonic generation. In the more familiar linear response, the terms of $O(\omega^{-2})$ are only associated with the long-wave response to a long-wave perturbation.

The "diagrammatic"-type formulas of Eqs. (5.13) and (5.14) may be put into a "product" form using formulas similar to Eqs. (5.7) and (5.8). Recasting these formulas eliminates numerical finite differences of up to third order. The long-wave charge is given by

$$\begin{aligned}
\delta\tilde{n}\,^{(2)}_{2(2m)}=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\sum_{n}^{\mathrm{occ}} i(\tilde{\phi}\,^{(1)}_{1m})^2[& -12\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_2|n,\tilde{\mathbf{k}}\rangle \\
& +12\mathcal{I}\langle n\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})^3\,G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega) \\
& \qquad\times H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& +6\mathcal{I}\langle n\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})^2\,G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega) \\
& \qquad\times H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})H_1|n,\tilde{\mathbf{k}}\rangle \\
& -24\omega^2\mathcal{I}\langle n\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega) \\
& \qquad\times H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1|n,\tilde{\mathbf{k}}\rangle] \\
+\tilde{\phi}\,^{(1)}_{1m}[4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_2G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
+4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
+8\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
-8\omega^2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle & \\
+4\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle & \\
+8\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})H_1|n,\tilde{\mathbf{k}}\rangle & \\
-8\omega^2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle & \\
+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
-2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
-2\omega^2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle] & \\
+i[-4\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
-2\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle & \\
-2\mathcal{I}\langle n,\tilde{\mathbf{k}}|\tilde{\phi}\,^{(1s)}_{1m}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle]\;. &
\end{aligned}\tag{5.19}$$

In deriving Eq. (5.19) the right-hand side of Eq. (5.16) was subtracted from the $\tilde{\phi}\,^{(1)}_{1m}$ term of Eq. (5.13). Hence, this equation is restricted to insulators. Similarly, to obtain the "product" form for the short-wave induced charge at the second-harmonic frequency, it is necessary to subtract the right-hand side of Eq. (5.18) from the $(\tilde{\phi}\,^{(1)}_{1m})^2$ term of Eq. (5.14); hence Eq. (5.20) is also restricted to insulators. The result is

$$\begin{aligned}
\delta\tilde{n}\,^{(2s)}_{2(2m)}(\mathbf{r}) = \bar{\Omega}_0 \int_{\mathrm{BZ}} d\mathbf{k} \sum_n^{\mathrm{occ}} (\tilde{\phi}\,^{(1)}_{1m})^2 [& 8\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_2|n,\tilde{\mathbf{k}}\rangle \\
& +2\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& +4\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& +8\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& -8\omega^2\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& +2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& -\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& -\omega^2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle] \\
+i\tilde{\phi}\,^{(1)}_{1m}[& -2\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle \\
& +2\mathcal{I}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle \\
& -2\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1)}_{1m}|n,\tilde{\mathbf{k}}\rangle \\
& -4\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)H_1(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle \\
& -2\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle \\
& -4\mathcal{I}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(2\omega)G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})G_{n\mathbf{k}}(\omega)G_{n\mathbf{k}}(-\omega)H_1|n,\tilde{\mathbf{k}}\rangle] \\
+[& \mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle+\mathcal{R}\langle n,\tilde{\mathbf{k}}|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(-2\omega)\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(-\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle \\
& +\mathcal{R}\langle n,\tilde{\mathbf{k}}|\tilde{\phi}\,^{(1s)}_{1m}G_{n\mathbf{k}}(-\omega)|\mathbf{r}\rangle\langle\mathbf{r}|G_{n\mathbf{k}}(\omega)\tilde{\phi}\,^{(1s)}_{1m}|n,\tilde{\mathbf{k}}\rangle] \,. \qquad (5.20)
\end{aligned}$$

The static limits in Eqs. (5.15) and (5.17) may be obtained from Eqs. (5.19) and (5.20) by inspection. (In practice, the static limits were obtained by series expansion, so recovering the static limits is a check.) Note that in Eqs. (5.9), (5.10), (5.19), and (5.20), the Green's operators occur in $\pm\omega$ or $\pm 2\omega$ pairs. The induced charge density is seen to be an even function of frequency.

Historically, the discovery of optical second-harmonic generation in crystals in 1961 (Ref. 30) sparked a theoretical description as a second-order time-dependent perturbation of one-electron bands.[31,32] The difficulty in this development of a manifest divergence in the static limit was resolved to a certain extent a decade later by Aspnes,[29] who showed (a) for a vector-potential perturbation, cubic symmetry could be used to eliminate divergent terms, and (b) for a scalar time-dependent perturbation, assuming only local potentials in the ground-state Hamiltonian and neglecting short-wave induced potentials, the divergent terms could be eliminated by a summation over the Brillouin zone. Aspnes's final result, the sum of his Eqs. (2.19) and (2.20), is similar to Eq. (5.15) of the present work if the short-wave terms are neglected and $H_2=\frac{1}{2}q^2$, which permits the simplification $\langle n,\tilde{\mathbf{k}}|H_1G^4_{n\mathbf{k}}H_2|n,\tilde{\mathbf{k}}\rangle=\frac{1}{2}q^2\langle n,\tilde{\mathbf{k}}|H_1G^4_{n\mathbf{k}}|n,\tilde{\mathbf{k}}\rangle=0$. The local-field corrections to $\chi^{(2)}$ have not been reported previously to my knowledge. Simplified local field corrections of the "point dipole" type are given by Bloembergen.[33]

VI. *F*-SUM RULE FOR SECOND-HARMONIC GENERATION

By use of Goldstone's theorem,[34] the energy of a system with applied perturbation $H'(\mathbf{k},\mathbf{q},\beta)$ is

$$\varepsilon_{n,\mathbf{k}+\mathbf{q}}(\beta)=\varepsilon_{n\mathbf{k}}+\left\langle n,\tilde{\mathbf{k}}\middle|H'\sum_{j=0}^{\infty}(G_{n\mathbf{k}}H')^j\middle|n,\tilde{\mathbf{k}}\right\rangle . \qquad (6.1)$$

Here, the perturbing Hamiltonian is a combination of $\mathbf{k}\cdot\mathbf{p}$ terms (such as H_1 and H_2) and other terms, denoted collectively by β. The Green's function $G_{n\mathbf{k}}(\varepsilon_{n\mathbf{k}}-H_{\mathbf{k}})^{-1}$ always yields a function orthogonal to the ground state $|n,\tilde{\mathbf{k}}\rangle$. A second expression for the energy comes from the Taylor series,

$$\begin{aligned}
\varepsilon_{n,\tilde{\mathbf{k}}+\mathbf{q}}(\beta)= & \ \varepsilon_{n\mathbf{k}}(\beta)+\mathbf{q}\cdot\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}(\beta) \\
& +\tfrac{1}{2}\mathbf{q}\cdot\nabla_{\mathbf{k}}\mathbf{q}\cdot\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}(\beta)+O(q^3) \,. \qquad (6.2)
\end{aligned}$$

The function $\varepsilon_{n\mathbf{k}}(\beta)$ may be expanded in powers of β as well. Now, $0=\int_{\mathrm{BZ}}d\mathbf{k}\,\nabla_{\mathbf{k}}\tilde{f}(\mathbf{k})$ for any function $\tilde{f}$ which is periodic in the Brillouin zone. Since all terms arising in the expansion (6.2) are of the form $\nabla_{\mathbf{k}}\tilde{f}(\mathbf{k})$, their integrals over the Brillouin zone will vanish.

To obtain the f-sum rule for crystals,[35] let $H'=H_1+H_2$. The auxiliary variables β are not needed at this stage. Equating the terms of order in q^2 in Eqs.

(6.1) and (6.2),

$$0=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\langle n,\tilde{\mathbf{k}}|H_2|n,\tilde{\mathbf{k}}\rangle+\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\ . \tag{6.3}$$

If β is a single scalar variable, then it is permissible to expand $\varepsilon_{n\mathbf{k}}(\beta)-\varepsilon_{n\mathbf{k}}=\beta(\partial\varepsilon_{n\mathbf{k}}(0)/\partial\beta)+O(\beta^2)$. Let $H'=H_1+H_2+\beta\bar{H}$, where $\bar{H}$ is some perturbation which is independent of $\mathbf{q}$ and β. Equating terms in Eqs. (6.1) and (6.2) that are the coefficients of terms of order $q^2\beta$, we find

$$\begin{aligned}0=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\langle n,\tilde{\mathbf{k}}|H_2G_{n\mathbf{k}}\bar{H}|n,\tilde{\mathbf{k}}\rangle+\langle n,\tilde{\mathbf{k}}|\bar{H}G_{n\mathbf{k}}H_2|n,\tilde{\mathbf{k}}\rangle+\langle n\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}\bar{H}|\mathbf{n},\tilde{\mathbf{k}}\rangle\\+\langle n\tilde{\mathbf{k}}|\bar{H}G_{n\mathbf{k}}H_1G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle+\langle n\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}\bar{H}G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\ .\end{aligned} \tag{6.4}$$

This result may be simplified if we impose the parity requirements of Sec. V A, namely H_1 is odd in $\mathbf{k}$ and H_2 is even. If we further assume $\bar{H}$ is even in $\mathbf{k}$, then Eq. (6.4) reduces to

$$0=\bar{\Omega}_0\int_{\mathrm{BZ}}d\mathbf{k}\,2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_2G_{n\mathbf{k}}\bar{H}|n,\tilde{\mathbf{k}}\rangle+2\mathcal{R}\langle n,\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}H_1G_{n\mathbf{k}}\bar{H}|n,\tilde{\mathbf{k}}\rangle+\mathcal{R}\langle n\tilde{\mathbf{k}}|H_1G_{n\mathbf{k}}\bar{H}G_{n\mathbf{k}}H_1|n,\tilde{\mathbf{k}}\rangle\ . \tag{6.5}$$

If we specify $\bar{H}=\tilde{\phi}\,{}^{(1s)}_{1m}$, Eq. (6.5) reduces to Eq. (5.16) as a special case. Similarly, $\bar{H}=|\mathbf{r}\rangle\langle\mathbf{r}|$ reduces Eq. (6.5) to Eq. (5.18). Taking $\bar{H}$ as odd in $\mathbf{k}$ reduces Eq. (6.4) to a tautology, as each term will cancel in ±**k** pairs.

VII. DETERMINATION OF $\chi^{(2)}$

The second-harmonic generation tensor is defined by the relation

$$P_\alpha=\chi^{(2)}_{\alpha\beta\gamma}E_\beta E_\gamma\ , \tag{7.1}$$

where **P** is the polarization and **E** is the electric field. Throughout this section, the variables refer to the corresponding macroscopic electrodynamic quantities (called "long-wave parts" elsewhere in this paper). The Einstein summation convention is used for the Cartesian indices α, β, and γ. The frequency labels 2ω for **P** and ω for **E** are suppressed. In our bulk system, the polarization will be in the form

$$\mathbf{P}(\mathbf{r})=\mathbf{P}e^{2i\mathbf{q}\cdot\mathbf{r}}\ . \tag{7.2}$$

Equation (5.13), (5.15), or (5.19) determines the macroscopic induced electron number density,

$$\delta n(\mathbf{r})=\Omega_0^{-1}\delta\tilde{n}\,{}^{(2)}_{2m}\,e^{2i\mathbf{q}\cdot\mathbf{r}} \tag{7.3}$$

using the definition, Eq. (5.3). The polarizability is related to the induced charge density by

$$\nabla\cdot\mathbf{P}(\mathbf{r})=e\,\delta n(\mathbf{r})\ , \tag{7.4}$$

where $-e$ is the charge on the electron. From Eqs. (7.2)–(7.4), it may be seen that

$$2i\mathbf{q}\cdot\mathbf{P}=\Omega_0^{-1}e\,\delta\tilde{n}\,{}^{(2)}_{2m}\ . \tag{7.5}$$

The longitudinal electric field is given by

$$\mathbf{E}(\mathbf{r})=\frac{1}{e}\nabla\phi(\mathbf{r})=\frac{i}{e}\mathbf{q}\phi_0e^{i\mathbf{q}\cdot\mathbf{r}}\ , \tag{7.6}$$

where the unusual coefficient arises because the potential is defined to be an energy, following the convention of quantum mechanics rather than electrodynamics. Combining Eqs. (7.1), (7.5), and (7.6),

$$\delta\tilde{n}\,{}^{(2)}_{2m}=-\frac{2i}{e^3}\Omega_0\chi^{(2)}_{\alpha\beta\gamma}q_\alpha q_\beta q_\gamma\phi_0^2\ . \tag{7.7}$$

By picking different directions for the wave vector **q**, the 10 linear combinations of $\chi^{(2)}_{\alpha\beta\gamma}$ that have distinct angular dependence may be determined, namely χ_{111}, χ_{222}, χ_{333}, $\chi_{112}+\chi_{121}+\chi_{211}$, $\chi_{113}+\chi_{131}+\chi_{311}$, $\chi_{221}+\chi_{212}+\chi_{122}$, $\chi_{331}+\chi_{313}+\chi_{133}$, $\chi_{223}+\chi_{232}+\chi_{322}$, $\chi_{332}+\chi_{323}+\chi_{233}$, and $\chi_{123}+\chi_{231}+\chi_{312}+\chi_{132}+\chi_{213}+\chi_{321}$. (The superscript on χ has been suppressed.) For second-harmonic generation, $\chi_{\alpha\beta\gamma}=\chi_{\alpha\gamma\beta}$; there are up to 18 independent quantities to be determined. For certain crystal classes, there are additional relations among the tensor elements. Tables for all crystal classes have been given by Shen[27] and Nye.[36] The zinc-blende structure (with symmetry[37] $\bar{4}3m$ or T_d) has but a single independent tensor element, so one direction of **q** suffices to determine $\chi^{(2)}$ in this case. (Here, any direction other than $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, or $\hat{\mathbf{z}}$ will suffice for $\hat{\mathbf{q}}$.)

Knowledge of the induced charge density is not sufficient to determine the polarization in a crystal.[24] Hence, the present developments do not allow for determination of all the components of the tensor $\chi^{(2)}$ in the most general situation. If the induced currents were known, the polarization could be determined from the relation $\mathbf{J}=\mathbf{J}_{\mathrm{dc}}+\partial\mathbf{P}/\partial t$; such a formulation would always permit the determination of all tensor components. For atoms,[9] other finite systems, and the jellium surface,[12] calculation of induced charges without induced currents is sufficient. Calculations in the "spherical-solid" approximation[11] have also found the polarizability by taking moments on an induced charge, which represents a neglect of intercell charge transfer;[24] Johnson *et al.* argue that the charge is sufficiently localized in the alkali halides that the neglect of the intercell contributions is a good approximation.[11]

VIII. SUMMARY

Formulas for second-order nonlinear longitudinal response in semiconductors have been set down in a one-electron, self-consistent-field formalism. The analogous linear-response formalism has been shown to predict the dielectric constant of silicon and germanium to some 4%.[3,25] The key features of the formulation are (1) use of the inhomogeneous Schrödinger equation to simply the formulation, (2) the use of the f-sum rule for crystals to secure certain analytic cancellations in the response formulas, and (3) a "product" form (product of Green's operators) for the response that avoids numerical finite differences which might otherwise be present in the formulas. The formulation is sufficient to predict certain linear combinations of elements of the second-harmonic generation tensor. For certain crystal structures, including the case of zinc blende, this information is sufficient to determine the complete tensor. More generally, a current-current response theory should be formulated. It is hoped that an accurate *ab initio* calculation of the second-harmonic generation coefficient $\chi^{(2)}$ for a semiconductor such as GaAs will be forthcoming.

ACKNOWLEDGMENTS

I take pleasure in thanking Douglas Allan and Hua Zhong for key technical comments, Chris Myers for assistance with the software package Mathematica, and Douglas Allan, Michael Teter, and John Wilkins for helping to create a context in which this research is relevant. Support for this work was provided by National Science Foundation Grant No. DMR-87-02002. Much of this work was done at Cornell University's Laboratory of Atomic and Solid State Physics; hospitality of the Laboratory and specifically the Cornell Electronic Structure Theory Group is gratefully acknowledged.

APPENDIX: WAVE-FUNCTION NORMALIZATION

In Sec. III, the probability conserving property of a Hermitian Hamiltonian was exploited to simplify the wave function normalization constant $\bar{Z}$. A somewhat more pedantic approach is to define an expansion

$$\bar{Z}(t)=\sum_{j=0}^{\infty}\lambda^j\bar{Z}^{(j)}(t)=\sum_{j=0}^{\infty}\lambda^j\sum_{l=-j}^{j}{}^{[2]}\,\bar{Z}_l^{(j)}e^{il\omega t}\ . \tag{A1}$$

Since $\bar{Z}(t)$ is real, $\bar{Z}_l^{(j)}=\bar{Z}_{-l}^{(j)*}$. We may identify terms separately for each pair of j and l in Eq. (3.2). Hence,

$$\bar{Z}_l^{(j)}=\sum_{j'=0}^{j}\sum_{l'=-j'}^{j'}{}^{[2]}\,\Theta_{j-j'+l+l'}\Theta_{j-j'-l-l'}\langle\psi_{l'}^{(j')}|\psi_{l+l'}^{(j-j')}\rangle\ . \tag{A2}$$

To bolster the general argument that $\bar{Z}$ is not time-dependent, some of the low order terms will be seen to vanish by direct argument.

First we consider

$$\bar{Z}_{\pm1}^{(1)}=\langle\psi_0^{(0)}|\psi_{\pm1}^{(1)}\rangle+\langle\psi_{\pm1}^{(1)}|\psi_0^{(0)}\rangle\ . \tag{A3}$$

From the $j=1$ line of Eq. (2.6),

$$\pm\omega\langle\psi_0^{(0)}|\psi_{\pm1}^{(0)}\rangle=\langle\psi_0^{(0)}|V_{\pm\omega}^{(1)}|\psi_0^{(0)}\rangle\ . \tag{A4}$$

Since $V_{l\omega}^{(j)}=V_{-l\omega}^{(j)\dagger}$, $\langle\psi_0^{(0)}|\psi_1^{(1)}\rangle=-\langle\psi_0^{(0)}|\psi_{-1}^{(1)}\rangle^*$. From Eqs. (A3) and (A4), $\bar{Z}_{\pm1}^{(1)}=0$, as desired.

For the terms

$$\bar{Z}_{\pm2}^{(2)}=\langle\psi_0^{(0)}|\psi_{\pm2}^{(2)}\rangle+\langle\psi_{\mp1}^{(1)}|\psi_{\pm1}^{(1)}\rangle+\langle\psi_{\mp2}^{(2)}|\psi_0^{(0)}\rangle \tag{A5}$$

we consider only the case in which $V_{\pm2\omega}^{(2)}=0$; if this term is nonzero, the argument given for $\bar{Z}_{\pm1}^{(1)}$ may be applied to show that this term makes no contribution to $\bar{Z}_{\pm2}^{(2)}$. The wave functions in Eq. (A5) may be written using Eq. (2.6) in terms of the Green's operator $G(\omega)=(\varepsilon+\omega-H_0)^{-1}$. If ω is nonzero and nonresonant, this inverse exists. Moreover, $G(\omega)$ is Hermitian. The terms in the induced wave function may be written as

$$\begin{aligned}|\psi_{\pm1}^{(1)}\rangle&=G(\pm\omega)V_{\pm\omega}^{(1)}|\psi_0^{(0)}\rangle\ ,\\ |\psi_{\pm2}^{(2)}\rangle&=G(\pm2\omega)V_{\pm\omega}^{(1)}G(\pm\omega)V_{\pm\omega}^{(1)}|\psi_0^{(0)}\rangle\ .\end{aligned} \tag{A6}$$

Application of Eq. (A6) to Eq. (A5) and the time reversal relation $V_{l\omega}^{(j)}=V_{-l\omega}^{(j)\dagger}$ yields the expression

$$\begin{aligned}\bar{Z}_{\pm2}^{(2)}=\ &\langle\psi_0^{(0)}|G(\pm2\omega)V_{\pm\omega}^{(1)}G(\pm\omega)V_{\pm\omega}^{(1)}|\psi_0^{(0)}\rangle\\ &+\langle\psi_0^{(0)}|V_{\pm\omega}^{(1)}(G\mp\omega)G(\pm\omega)V_{\pm\omega}^{(1)}|\psi_0^{(0)}\rangle\\ &+\langle\psi_0^{(0)}|V_{\pm\omega}^{(1)}(G\mp\omega)V_{\pm\omega}^{(1)}G(\mp2\omega)|\psi_0^{(0)}\rangle\ .\end{aligned} \tag{A7}$$

Now, $G(\omega)|\psi_0^{(0)}\rangle=\omega^{-1}|\psi_0^{(0)}\rangle$ and

$$G(\omega_1)G(\omega_2)=\frac{1}{\omega_2-\omega_1}[G(\omega_1)-G(\omega_2)]\ . \tag{A8}$$

These relations are sufficient to show $\bar{Z}_{\pm2}^{(2)}=0$. Arguments of this type may be extended to the higher terms.

---

[1] S. Baroni and R. Resta, Phys. Rev. B **33**, 7017 (1986).
[2] M. S. Hybersten and S. G. Louie, Phys. Rev. B **35**, 5585 (1987).
[3] Z. H. Levine and D. C. Allan, Phys. Rev. Lett. **63**, 1719 (1989).
[4] S. de Gironcoli, S. Baroni, and R. Resta, Phys. Rev. Lett. **62**, 2853 (1989).
[5] S. Baroni and P. Giannozzi, Phys. Rev. Lett. **58**, 1861 (1987).
[6] C. Y. Fong and Y. R. Shen, Phys. Rev. B **12**, 2325 (1975).
[7] D. J. Moss, J. E. Sipe, and H. M. van Driel, Phys. Rev. B **36**, 9708 (1987). The method is extended to $\chi^{(3)}$ in D. J. Moss, E. Ghahramani, J. E. Sipe, and H. M. van Driel, *ibid.* **41**, 1512 (1990); also see E. Ghahramani, D. J. Moss, and J. E. Sipe, Phys. Rev. Lett. **64**, 2815 (1990).]
[8] D. P. DiVinenzo, Phys. Rev. B **37**, 1245 (1988), especially the Appendix; I hasten to add this paper indicates errors in earlier work but does not, to my knowledge, contain them.
[9] K. R. Subbaswamy and G. D. Mahan, J. Chem. Phys. **84**, 3317 (1986).
[10] G. Senatore and K. R. Subbaswamy, Phys. Rev. A **34**, 3619 (1986).
[11] M. D. Johnson, K. R. Subbaswamy, and G. Senatore, Phys.

Rev. B **36**, 9202 (1987); M. D. Johnson and K. R. Subbaswamy, *ibid.* **39**, 10 275 (1989).

[12] M. Weber and A. Liebsch, Phys. Rev. B **35**, 7411 (1987); **36**, 6411 (1987); **37**, 1019(E) (1988); A. Liebsch and W. L. Schaich, *ibid.* **40**, 5401 (1989).

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

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

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

[16] M. J. Stott and E. Zaremba, Phys. Rev. A **21**, 12 (1980).

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

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

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

[20] See, for example, N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Reinhart and Winston, New York, 1976), Chap. 27.

[21] To a certain extent, we follow the notation and presentation of G. Baym, *Lectures on Quantum Mechanics* (Benjamin, Reading, MA, 1969), Chaps. 11–13. Our $\bar{Z}$ is Baym's Z^{-1}.

[22] X. F. Li, A. L'Huillier, M. Ferray, L. A. Lompré, and G. Mainfray, Phys. Rev. A **39**, 5751 (1989).

[23] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer* (Addison-Wesley, Redwood City, CA, 1988).

[24] 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 that does not hold for extended systems despite its successful application for molecular aggregates in the Clausius-Mossotti relation.

[25] Z. H. Levine and D. C. Allan (unpublished).

[26] For a recent development in nonlocal pseudopotential theory with references to earlier work, see D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[27] Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984), Chap. 2.

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

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

[30] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich, Phys. Rev. Lett. **7**, 118 (1961).

[31] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Phys. Rev. **127**, 1918 (1962).

[32] R. Loudon, Proc. Phys. Soc. London **80**, 952 (1962).

[33] N. Bloembergen, *Nonlinear Optics* (Benjamin, New York, 1965), pp. 68-70.

[34] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many Particle Systems* (McGraw-Hill, New York, 1971), p. 111.

[35] F. Wooten, *Optical Properties of Solids* (Academic, New York, 1972), Appendix D; Ref. 20, p. 140.

[36] J. F. Nye, *Physical Properties of Crystals* (Oxford, London, 1957), Chap. 7.5. The piezoelectric tensor has the same symmetry properties as the second-harmonic-generation tensor.

[37] R. W. G. Wyckoff, *Crystal Structures,* 2nd ed. (Wiley, New York, 1963), Vol. 1, p. 108.