Nonlinear optical susceptibilities of semiconductors: Results with a length-gauge analysis

Claudio Aversa* and J. E. Sipe
Department of Physics, and Ontario Laser and Lightwave Research Centre, University of Toronto, Toronto, Ontario, Canada M5S-1A7
(Received 12 April 1995; revised manuscript received 25 July 1995)

We present a simple prescription for the derivation of electronic contributions to the nonlinear optical response of crystals in the independent particle approximation. Semiconductor Bloch equations are found that include previously neglected effects of intraband motion. Applying perturbation theory to clean, cold semiconductors we find expressions for the susceptibilities lacking the unphysical divergences at zero-frequency that have plagued other calculations. For these materials we present well-behaved, general expressions for $\chi^{(2)}$ and $\chi^{(3)}$ for arbitrary frequency mixings, and give an explicit demonstration of the finite zero-frequency value of $\chi^{(3)}$. We further show how second-order photogalvanic effects are contained in certain *physical* zero frequency divergences of $\chi^{(2)}$, and consider the corresponding *physical* zero-frequency divergences of $\chi^{(3)}$.

I. INTRODUCTION

A comprehensive understanding of the nonlinear optical properties of bulk semiconductors is crucial to the efficient design and analysis of many modern optical devices,[1–3] and will aid in the study of more complex systems involving extended states, such as heterostructures and amorphous materials.[4,5] However, even at the simplest level of approximation — treating electrons as independent particles interacting with the electromagnetic field in the long-wavelength limit and neglecting local field effects — the calculation of nonlinear optical coefficients of semiconductors has been troublesome. For example, semiconductor susceptibility expressions often appear to diverge as all light frequencies approach zero;[6,7] such divergences are not expected on physical grounds for clean, cold semiconductors. This was discussed previously by Sipe and Ghahramani[8] (a paper we henceforth refer to as SG), who presented a formalism, in the spirit of Genkin and Mednis,[9] which naturally provides well behaved expressions. In this paper, we present an alternative derivation of such semiconductor susceptibilities at this simplest level of approximation.[10] Our scheme is less abstract, involves fewer mathematical details, and better highlights the relevant physical issues.

We build on the insight provided by previous results[8–14] to show how standard length gauge perturbation theory can be rigourously applied to crystals without being mathematically cumbersome. We show that the popular semiconductor Bloch equations, which are often used in nonperturbative analyses of optical response, neglect effects due to intraband motion; we present "corrected" semiconductor Bloch equations in the independent particle approximation. But our main focus here is on perturbative results. For clean, cold semiconductors, we find expressions lacking the unphysical divergences at zero frequency that appear in realistic velocity-gauge calculations if the inevitable electronic structure approximations are made. Our scheme allows one to easily classify the distinct physical contributions to the crystal optical response, thus more clearly highlighting how the insulator optics has signatures of *both* atomic *and* free electron character. We give general expressions for $\chi^{(2)}$ and $\chi^{(3)}$ for arbitrary frequency mixings, which have not, to our knowledge, appeared before in the literature. From this follows what we believe to be the first explicit demonstration of the finite zero-frequency value of $\chi^{(3)}$. Finally, since in this formalism the *unphysical* divergences are absent from the start, we can turn to the significance of the *physical* divergences associated with mixing frequencies that sum to zero.

The general formalism and the relationship between the optical responses of crystals, atoms, and electron gases is described in Sec. II. In Sec. III, we then present the general second- and third-order susceptibilities for clean, cold semiconductors, and discuss their zero-frequency limits. Finally, in Sec. IV, we describe the relationship of our formalism to other approaches.

II. GENERAL FORMALISM

A. Gauge and the position operator

The independent particle approximation allows one to describe the system using a scaled one electron density operator ρ, with $\mathrm{Tr}(\rho)$ giving the total number of electrons present; this approach leads to the same results as a many body formulation for all expectation values of single particle operators.[15] In the Schrodinger representation, ρ obeys the dynamical equation

$$i\hbar\frac{d\rho}{dt} = [H, \rho], \tag{1}$$

where H is the total single electron Hamiltonian. In the absence of an electromagnetic field, the bulk crystal Hamiltonian is $H_0 = \mathbf{p}^2/2m + V$, where V is the periodic crystal potential; H_0 has eigenvalues $\hbar\omega_n(\mathbf{k})$ and eigenstates $|n\mathbf{k}\rangle$ (Bloch states) labeled by a band index n and crystal momentum $\mathbf{k}$. The Bloch states are chosen so that $\langle \mathbf{r}|n\mathbf{k}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$, with cell-periodic $u_{n\mathbf{k}}(\mathbf{r})$ and orthonormality $\langle n\mathbf{k}|m\mathbf{k}'\rangle = \delta_{nm}\delta(\mathbf{k}-\mathbf{k}')$. With the electromagnetic field, the Hamiltonian is exactly obtained in the velocity-gauge by replacing $\mathbf{p}$ by $\mathbf{p} - e\mathbf{A}/c$ in H_0, where $\mathbf{A} = \mathbf{A}(t)$ in the long-wavelength limit of interest here. In this velocity-gauge, a perturbative treatment of the density operator dynamics, due to the effective interaction $-(e/mc)\mathbf{p}\cdot\mathbf{A}$, will lead to the standard optical susceptibility expressions. As discussed in SG and further in Sec. IV, such expressions have certain unattractive features; we instead consider an alternative formulation. At least formally, we can recast this problem using a unitary transformation[16] to find the length gauge,

$$H = H_0 - e\mathbf{r}\cdot\mathbf{E}, \tag{2}$$

involving the electric field $\mathbf{E} = -\dot{\mathbf{A}}/c$.

A common problem with a perturbation theory for solids in the length gauge is the treatment of the position operator $\mathbf{r}$ in view of the extended Bloch states. However, for the crystals of interest, such a treatment has been described previously.[14] The matrix elements of $\mathbf{r}$ are most conveniently described by distinguishing between its *intraband* part $\mathbf{r}_i$ and *interband* part $\mathbf{r}_e$, where $\mathbf{r} = \mathbf{r}_i + \mathbf{r}_e$ and

$$\langle n\mathbf{k}|\mathbf{r}_i|m\mathbf{k}'\rangle = \delta_{nm}\left[\delta(\mathbf{k}-\mathbf{k}')\xi_{nn} + i\nabla_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}')\right], \tag{3}$$

$$\langle n\mathbf{k}|\mathbf{r}_e|m\mathbf{k}'\rangle = (1-\delta_{nm})\delta(\mathbf{k}-\mathbf{k}')\xi_{nm}. \tag{4}$$

In the vectors ξ_{nm}, and often throughout this paper, $\mathbf{k}$ dependences are implicit,

$$\xi_{nm} \equiv \frac{(2\pi)^3 i}{\Omega}\int_\Omega \mathbf{d}^3\mathbf{r}\; u^*_{n\mathbf{k}}(\mathbf{r})\nabla_{\mathbf{k}}u_{m\mathbf{k}}(\mathbf{r}), \tag{5}$$

where Ω is a unit cell volume. It is the intraband part $\mathbf{r}_i$, which leads to the troublesome nature of the position operator in crystals; not only is the $\nabla_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}')$ in Eq. (3) highly singular, but the ξ_{nn} are not unique.[14] This contrasts with the momentum operator for which analogous intraband and interband parts, $\mathbf{p} = \mathbf{p}_i + \mathbf{p}_e$, can both be considered *simple* operators, since $\langle n\mathbf{k}|\mathbf{p}|m\mathbf{k}'\rangle = \delta(\mathbf{k}-\mathbf{k}')\mathbf{p}_{nm}$, valid regardless of n and m; a *simple* operator is defined as one whose Bloch state matrix elements involve only $\delta(\mathbf{k}-\mathbf{k}')$. In addition, the $\mathbf{k}$ dependent $\mathbf{p}_{nm}$ are uniquely defined by

$$\mathbf{p}_{nm} = \frac{-(2\pi)^3 i\hbar}{\Omega}\int_\Omega \mathbf{d}^3\mathbf{r}\; \psi^*_{n\mathbf{k}}(\mathbf{r})\nabla_{\mathbf{r}}\psi_{m\mathbf{k}}(\mathbf{r}). \tag{6}$$

These differences have led many researchers to avoid the position operator in crystals by using the velocity gauge in theoretical studies of optical properties. Nevertheless, we now show that the use of $\mathbf{r}$ need not be troublesome, and later illustrate that, as often noted for atomic systems,[16,17] there are advantages to adopting the length gauge.

A familiarity with $\mathbf{r}$ can be developed by first considering the interband part; defining $\mathbf{r}_{nm} \equiv (1-\delta_{nm})\xi_{nm}$, so that $\mathbf{r}_{nm} = 0$ for $n = m$, the commutator $[H_0, \mathbf{r}] = \hbar\mathbf{p}/im$ leads to the familiar useful relation

$$\mathbf{r}_{nm} = \frac{\mathbf{p}_{nm}}{im\omega_{nm}}, \tag{7}$$

where $\omega_{nm} = \omega_n - \omega_m$.[18] Second, while the highly singular nature of the intraband $\mathbf{r}_i$ is cumbersome, it can nonetheless be handled; past work applied to various specific systems has proceeded in this respect.[19,13] However, it is simpler to deal with $\mathbf{r}_i$ than appears to have been appreciated. When $\mathbf{r}_i$ appears in commutators involving a *simple* operator S, one can verify that

$$\langle n\mathbf{k}|[\mathbf{r}_i, S]|m\mathbf{k}'\rangle = i\delta(\mathbf{k}-\mathbf{k}')(S_{nm})_{;\mathbf{k}}, \tag{8}$$

where the $;\mathbf{k}$ operator represents a generalized derivative,

$$(S_{nm})_{;\mathbf{k}} \equiv \frac{\partial S_{nm}}{\partial\mathbf{k}} - iS_{nm}(\xi_{nn} - \xi_{mm}). \tag{9}$$

Equation (8) shows that, for a simple operator S, the commutator $[\mathbf{r}_i, S]$ is itself simple. This eliminates the need for careful limiting procedures, which $\mathbf{r}_i$ and its highly singular matrix element $\nabla_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}')$ otherwise demand. Further, quantities such as $(S_{nm})_{;\mathbf{k}}$ will be seen to be convenient entities, since the $\partial S_{nm}/\partial k$ never appears without the $S_{nm}(\xi_{nn}-\xi_{mm})$. Indeed, the $\partial S_{nm}/\partial\mathbf{k}$ and ξ_{nn} are not individually uniquely defined,[14] while the $(S_{nm})_{;\mathbf{k}}$ is.[20]

The use of the superoperator $[\mathbf{r}_i, \cdot]$ permits the easy derivation of many useful expressions. For example, consider the commutation relation $[r^a, p^b] = [r^a_i, p^b] + [r^a_e, p^b] = i\hbar\delta^{ab}$, where the superscripts indicate Cartesian components; matrix elements of this lead to the useful relations

$$\frac{1}{\hbar}\frac{\partial p^b_{nn}}{\partial k^a} = \delta^{ab} - \frac{m}{\hbar}\sum_l \omega_{ln}(r^a_{nl}r^b_{ln} + r^b_{nl}r^a_{ln}), \tag{10}$$

$$(r^b_{nm})_{;k^a} = \frac{r^a_{nm}\Delta^b_{mn} + r^b_{nm}\Delta^a_{mn}}{\omega_{nm}} + \frac{i}{\omega_{nm}}\sum_l(\omega_{lm}r^a_{nl}r^b_{lm} - \omega_{nl}r^b_{nl}r^a_{lm}). \tag{11}$$

The final forms of these have been obtained by using Eqs. (7,9) and by defining $(\omega_{nm})_{;k^a} \equiv (\omega_n)_{;k^a} - (\omega_m)_{;k^a}$; in this sense, we use $S = H_0$ in Eq. (8) so that $(\omega_{nm})_{;k^a} = (p^a_{nn} - p^a_{mm})/m \equiv \Delta^a_{nm}$. Equation (10) is just the familiar expansion for the inverse effective mass tensor $(m_n^{-1})_{ab}$. Equation (11) is equivalent to Eq. (3.36) in SG, which was obtained there through much more algebra; it will later be seen to be quite useful. We can similarly carefully take matrix elements of $[r^a, r^b] = 0$, again by decomposing each position operator into intraband and interband parts, to obtain

$$(r^b_{nm})_{;k^a} - (r^a_{nm})_{;k^b} = i\sum_l(r^a_{nl}r^b_{lm} - r^b_{nl}r^a_{lm}), \tag{12}$$

$$\frac{\partial \xi^b_{nn}}{\partial k^a} - \frac{\partial \xi^a_{nn}}{\partial k^b} = i \sum_m (r^a_{nm} r^b_{mn} - r^b_{nm} r^a_{mn}). \quad (13)$$

Generally, each of Eqs. (10–13) can be interpreted as a "sum rule," and higher order commutators, such as $[r^c,[r^b,p^a]]=0$, can be used similarly to obtain still more sum rules.

The distinction between intraband and interband operators helps expose the relationship between the crystal and atomic and free electron systems. That is, the crystal problem strongly parallels the (bound) atomic one if we neglect $\mathbf{p}_i$ and $\mathbf{r}_i$, while it mirrors the free electron system if $\mathbf{p}_e$ and $\mathbf{r}_e$ are ignored. The ξ_{nn} in Eq. (3) somewhat complicate this picture, since for crystals they lack simple interpretation,[14,21] and yet have clear analogs in bound state problems lacking a center of inversion symmetry. This latter point, and the $\delta(\mathbf{k}-\mathbf{k}')$ appearing with the ξ_{nn} in Eq. (3), might suggest that this part of $\mathbf{r}_i$ would be better grouped with the $\mathbf{r}_e$. However, in the crystal problem it is more convenient to keep $\mathbf{r}_i$ intact, again because, in contrast to $\partial \mathbf{r}_{nm}/\partial \mathbf{k}$, the $(\mathbf{r}_{nm})_{;\mathbf{k}}$ is a uniquely defined unambiguous quantity; this point has apparently not been well appreciated.

B. Physical contributions to the polarization

A macroscopic polarization density $\mathbf{P}$ for a crystal is commonly obtained from the associated macroscopic current density, as in

$$\frac{d\mathbf{P}}{dt} = \overline{\mathrm{Tr}}(\mathbf{j}_\mu \rho), \quad (14)$$

where $\mathbf{j}_\mu$ is the microscopic current operator and the bar over the trace indicates a macroscopic spatial average. The microscopic current operator appearing above and defined in Eq. (A2) of the Appendix involves the velocity operator, $\mathbf{v}=\mathbf{p}/m$ for the length gauge. We can then rewrite Eq. (14) using the cyclic invariance of the trace,

$$\frac{d\mathbf{P}}{dt} = \frac{e}{m}\overline{\mathrm{Tr}}(\mathbf{p}\rho_\mu), \quad (15)$$

which simply shifts the microscopic information onto the density operator, but is convenient here. Now using $\mathbf{p}/m = (1/i\hbar)[\mathbf{r},H]$, the cyclic invariance of the trace, and Eq. (1), while decomposing $\mathbf{r}=\mathbf{r}_i+\mathbf{r}_e$ and insuring that $\mathbf{r}_i$ only appears in commutators, Eq. (15) leads to

$$\frac{d\mathbf{P}}{dt} = \frac{e}{i\hbar}\overline{\mathrm{Tr}}([\mathbf{r}_i,H]\rho_\mu) + e\frac{d}{dt}\overline{\mathrm{Tr}}(\mathbf{r}_e\rho_\mu), \quad (16)$$

the two terms above defining $\mathbf{J}_\sigma$ and $\mathbf{P}_\chi$, respectively, in $d\mathbf{P}/dt \equiv \mathbf{J}_\sigma + d\mathbf{P}_\chi/dt$. The expression for $\mathbf{J}_\sigma$ can be further rewritten, using $H=H_0-e\mathbf{r}\cdot\mathbf{E}$ and $[\mathbf{r}_i,H_0]=i\hbar\mathbf{p}_i/m$, so that

$$\mathbf{J}_\sigma = \frac{e}{m}\overline{\mathrm{Tr}}(\mathbf{p}_i\rho_\mu) + \frac{e}{i\hbar}\overline{\mathrm{Tr}}([\mathbf{r}_i,-e\mathbf{r}\cdot\mathbf{E}(t)]\rho_\mu). \quad (17)$$

This decomposition, $d\mathbf{P}/dt = \mathbf{J}_\sigma + d\mathbf{P}_\chi/dt$, is similar to one revealed in SG; our expression is obtained much more easily. Further, note that the contribution in Eq. (17) involving $[\mathbf{r}_i,-e\mathbf{r}_i\cdot\mathbf{E}(t)]$, while not explicitly appearing in SG, does not vanish. Indeed, it is this commutator that leads to Eq. (13), and it has been previously identified and referred to as an anomalous velocity.[22]

Considering our interpretation of the interband and intraband effects, one is tempted to regard $\mathbf{P}_\chi$ as that part of the crystal polarization analogous to the total polarization in a bound atomic problem, and the first term in Eq. (17) for $\mathbf{J}_\sigma$ as analogous to the current expected in free electron systems. But, these interpretations are premature; we show in the next section that to higher than linear order ρ itself involves a mixture of intraband and interband effects.

The macroscopic averaging required in both Eq. (15) and Eq. (16) is simple and described in Appendix A. In the long-wavelength limit, this leads to a macroscopic polarization vector component P^a from Eq. (15), given by

$$\frac{dP^a}{dt} = \frac{e}{m}\sum_{nm\mathbf{k}} p^a_{mn}\rho_{nm}. \quad (18)$$

Alternatively, we could use Eq. (16) to obtain

$$\frac{dP^a}{dt} = \frac{e}{m}\sum_{n\mathbf{k}} p^a_{nn}\rho_{nn} - \frac{e^2}{\hbar}\mathbf{E}(t)\cdot\sum_{nm\mathbf{k}}\mathbf{D}^a_{mn}\rho_{nm} + e\frac{d}{dt}\sum_{nm\mathbf{k}} r^a_{mn}\rho_{nm}, \quad (19)$$

where the vector $\mathbf{D}^a_{mn} = (\mathbf{r}_{mn})_{;k^a} + \delta_{nm}[\partial\xi_{nn}/\partial k^a - \nabla_{\mathbf{k}}\xi^a_{nn}]$ includes the anomalous velocity. Of course, the sums over $\mathbf{k}$ in Eqs. (18,19) imply the standard Brillouin zone integrals.

C. Dynamics and the perturbation series

In studies of optically excited semiconductors, the length gauge is sometimes employed, but with the intraband piece $(-e\mathbf{r}_i\cdot\mathbf{E})$ of the optical perturbation ignored from the start.[23,24] Such work suggests that states with different $\mathbf{k}$ values remain uncoupled under the optical perturbation in the independent particle approximation. The electron dynamics is not correctly described by such an approach. This is perhaps best illustrated by considering the dynamical equation for the density operator matrix elements, which can be obtained from Eqs. (1), (2), and (8),

$$i\hbar\frac{d\rho_{nm}}{dt} = \hbar\omega_{nm}\rho_{nm} - e\mathbf{E}(t)\cdot\sum_l(\mathbf{r}_{nl}\rho_{lm}-\rho_{nl}\mathbf{r}_{lm}) - ie\mathbf{E}(t)\cdot(\rho_{nm})_{;\mathbf{k}}. \quad (20)$$

These equations for ρ_{nm}, $n=m$ describing populations and $n\neq m$ coherences, are our versions of the popular "semiconductor Bloch equations" in the limit of noninteracting electrons. They differ from the usual equations by the last term, which arises from the previously neglected $\mathbf{r}_i$ and now leads to a coupling between the

different crystal momenta. Equation (20) might instead be considered collisionless multiband Boltzmann equations. They are intuitively satisfying in the one band limit at zero frequency, reducing to the usual collisionless Boltzmann equation adopted in transport theory;[25] the standard semiconductor Bloch equations do not satisfy this limit. The band structure effects associated with the $(\rho_{nm})_{;\mathbf{k}}$ in Eq. (20) can be important, as implicitly discussed previously;[12,26] SG has shown that their neglect would lead to qualitatively incorrect predictions for second harmonic generation. Of course, in most applications of the semiconductor Bloch equations the inclusion of scattering effects is crucial, and Eq. (20) would have to be extended to include those. Nonetheless, we emphasize that the intraband effects in Eq. (20), which cannot *a priori* be neglected, do not appear in the usual semiconductor Bloch equations.

While semiconductor Bloch equations are most often applied if the incident light has a photon energy near the band gap, where the sum in Eq. (20) can, in practice, be limited to a few bands, perturbative calculations are the norm if the optical response is to be characterized over a wide-frequency range. Here, at least to lowest approximation, scattering effects can be, and are often, neglected.[6,8,27] This we do as well. In some instances, however, the perturbation expansion itself indicates that scattering effects are essential, as we discuss at the end of Sec. III A. To develop a perturbation expansion, we adopt an interaction representation, where for $U = e^{iH_0 t/\hbar}$, we have operators $\tilde{\theta}(t)$ defined as $\tilde{\theta}(t) = U\theta U^{+}$. Equation (1) then leads to a dynamical equation with the solution

$$i\hbar\tilde{\rho}(t) = i\hbar\rho_0 + \int_{-\infty}^{t} [-e\tilde{\mathbf{r}}(t')\cdot\mathbf{E}(t'), \tilde{\rho}(t')]\, dt'. \quad (21)$$

The standard iterative solution for this provides the $(N+1)$th order correction $\tilde{\rho}^{(N+1)}$ in terms of $\tilde{\rho}^{(N)}$; the series is generated by the unperturbed density operator $\tilde{\rho}^{(0)} \equiv \rho_0$, assumed to be the diagonal Fermi-Dirac distribution, $\langle n\mathbf{k}|\rho_0|n\mathbf{k}\rangle = f_n(\hbar\omega_n)$. Equation (21) shows that the position operator appears only in commutators, and so the intraband $\mathbf{r}_i$ is handled easily. It is convenient to decompose $\mathbf{r}$ in Eq. (21) into $\mathbf{r}_i$ and $\mathbf{r}_e$, and to consider a matrix element, giving

$$\tilde{\rho}_{nm}^{(N+1)}(t) = \frac{ie}{\hbar}\int_{-\infty}^{t} dt'\ e^{i\omega_{nm}t'}\mathbf{E}(t')\cdot[\mathbf{R}_e + \mathbf{R}_i], \quad (22)$$

where

$$\mathbf{R}_e = \sum_l \left[\mathbf{r}_{nl}\rho_{lm}^{(N)} - \rho_{nl}^{(N)}\mathbf{r}_{lm}\right], \quad (23)$$

$$\mathbf{R}_i = i[\rho_{nm}^{(N)}]_{;\mathbf{k}}, \quad (24)$$

with the time labels, and indices (nm), of $\mathbf{R}_e$ and $\mathbf{R}_i$ suppressed to simplify the notation. To first order, we expect two terms to be generated from ρ_0, and more generally 2^N terms at Nth order. This is illustrated by the tree diagram in Fig. 1. As in the diagram, the two generated

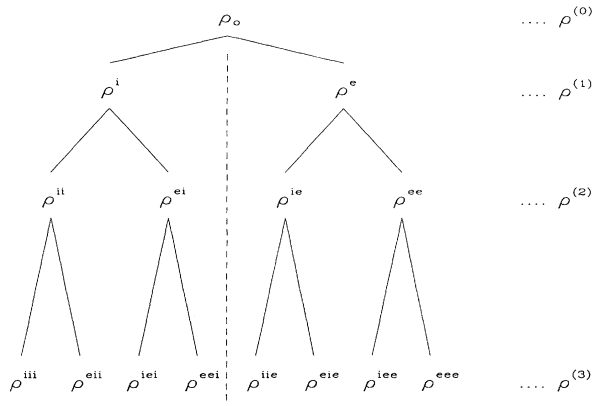

FIG. 1. Tree diagram illustrating the terms that arise in the perturbative solution for the density operator. At each node, an intraband (interband) interaction calls for progression down a left (right) branch of the tree.

$\rho^{(1)}$ terms can be labeled ρ^i and ρ^e indicating that they "originated" from $[\mathbf{r}_i, \rho_0]$ and $[\mathbf{r}_e, \rho_0]$. Similarly, four $\rho^{(2)}$ terms are found, ρ^{ii}, ρ^{ei}, ρ^{ie}, and ρ^{ee}, where for example ρ^{ie} results (roughly speaking) from $[\mathbf{r}_i, [\mathbf{r}_e, \rho_0]]$.

An important advantage of this length-gauge perturbation theory can now be appreciated: One can immediately distinguish between insulators and metals. At any given order all terms, which originate from $[\mathbf{r}_i, \rho_0]$, such as the first order ρ^i and the second order ρ^{ii} and ρ^{ei}, vanish for an insulator; such terms involve a factor

$$\langle n\mathbf{k}|[\mathbf{r}_i, \rho_0]|n\mathbf{k}'\rangle = i\frac{\partial f_n}{\partial \mathbf{k}}\delta(\mathbf{k}-\mathbf{k}'), \quad (25)$$

which vanishes for completely filled or completely empty bands. Hence, all terms on the left side of the tree in Fig. 1, half of all possible terms, will not contribute. This identification proves useful for obtaining insulator susceptibilities lacking zero-frequency divergences. In contrast, these terms which vanish for insulators survive for metals, and one can show that such terms lead to quite physical divergences in the metal susceptibilities (in the absence of scattering mechanisms) when all perturbing frequencies approach zero. The velocity gauge does not offer such a useful, transparent discrimination between metals and insulators at this stage, even though gauge invariance ensures that the derived susceptibilities will be equal. This will be discussed further in Sec. IV.

We can now identify the essential similarity between insulators and atoms: Both cases *effectively* have $[\mathbf{r}_i, \rho_0] = 0$. Our symmetric treatment of $\mathbf{r}_i$ and $\mathbf{r}_e$ suggests that $\langle n\mathbf{k}|[\mathbf{r}_i, \rho_0]|n\mathbf{k}'\rangle = 0$ is very similar to the vanishing of $\langle n\mathbf{k}|[\mathbf{r}_e, \rho_0]|m\mathbf{k}'\rangle \propto f_{nm}$ that occurs when the two bands $n \neq m$ are both filled (or both empty); that is, for $f_{nm} \equiv f_n - f_m = 0$. In this latter case, the zero matrix element is understood as a "Fermi blocking," since two occupied (or two empty) states provide no net contribution. Likewise, the vanishing of $\langle n\mathbf{k}|[\mathbf{r}_i, \rho_0]|n\mathbf{k}'\rangle$ occurs because the states involved (now two states infinitesimally "close" in the same band) are both filled or both

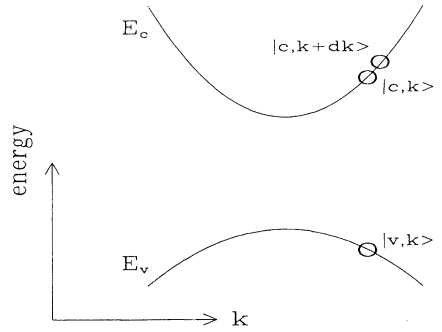

FIG. 2. A schematic that illustrates the symmetry between the interband and intraband Fermi blocking discussed in the text. The state $|c, k + dk\rangle$ is meant to represent a state infinitesimally close to $|c, k\rangle$.

empty; Fig. 2 depicts this situation.

Finally, then, the insulator $\rho^{(1)}$ is used in either of Eqs. (18,19) with a harmonic perturbation $\mathbf{E}(t) = \mathbf{E}e^{-i\omega t}$ to lead to the well known form for the first-order susceptibility tensor,

$$\chi^{(1)}_{ba}(-\omega;\omega) = \frac{e^2}{\hbar}\sum_{n,m,\mathbf{k}} \frac{r^b_{mn} r^a_{nm} f_{mn}}{\omega_{nm} - \omega}. \tag{26}$$

This first-order response has the form of a sum over crystal momenta $\mathbf{k}$ of terms, each of which resemble an atomic linear susceptibility. This often leads[23,24] to the insulator response being thought of as arising from a collection of atoms labeled by $\mathbf{k}$. However, this insulator-atom analogy strictly only follows for the linear response, since to first order only interband effects enter; the entire $\rho^{(1)}$ is given by the lone ρ^e term, and the intraband related $\mathbf{J}_\sigma$ does not contribute. These simplifications do not follow for the nonlinear susceptibilities, since both intraband and interband effects arise. For example, $\rho^{(2)}$ includes both ρ^{ee} and ρ^{ie}, and while the first will resemble an atomic expression, the latter has no analog in the atomic problem. The situation, of course gets more involved at third-order, where we expect ρ^{iie}, ρ^{iee}, ρ^{eie}, in addition to the atomic like ρ^{eee}. The nonlinear case is further complicated by the fact that, even for insulators, $\mathbf{J}_\sigma$ must be considered as well.

III. GENERAL SECOND- AND THIRD-ORDER SUSCEPTIBILITIES

A. Second-order effects

The considerations of the previous sections allow perturbative optical susceptibilities for crystals to be derived just as easily in the length gauge as in the velocity gauge. We start by considering the $\chi^{(2)}$ tensor extracted from the second-order macroscopic polarization density $P^{(2)}_c(\omega_2) = \chi^{(2)}_{cba}(-\omega_2;\omega_\beta,\omega_\alpha)E_b(\omega_\beta)E_a(\omega_\alpha)$, where $\omega_2 = \omega_\beta + \omega_\alpha$. Either Eq. (18) or Eq. (19) can be used for this, providing expressions that appear distinct; we quote both and explicitly illustrate their equivalence. Considering first Eq. (18) we employ $\rho^{(2)}$, whose two terms ρ^{ee} and ρ^{ie} spawn $\chi^{(2)}_e$ and $\chi^{(2)}_i$, respectively, of $\chi^{(2)} = \chi^{(2)}_e + \chi^{(2)}_i$, where

$$\chi^{(2)}_e = \frac{iC}{\omega_2 m}\sum_{m,n,l,\mathbf{k}} \frac{p^c_{mn}}{\omega_{nm} - \omega_2}\left(\frac{r^b_{nl} r^a_{lm} f_{ml}}{\omega_{lm} - \omega_1} - \frac{r^a_{nl} r^b_{lm} f_{ln}}{\omega_{nl} - \omega_1}\right), \tag{27}$$

$$\chi^{(2)}_i = \frac{-C}{\omega_2 m}\sum_{m,n,\mathbf{k}} \frac{p^c_{mn}}{\omega_{nm} - \omega_2}\left(\frac{r^a_{nm} f_{mn}}{\omega_{nm} - \omega_1}\right)_{;k^b}. \tag{28}$$

Here $C = e^3K/\hbar^2\mathrm{m}$ where K accounts for the usual factors which depend on whether the ω_α and ω_β are equal and/or zero.[28] Here, we have used $\omega_1 = \omega_\alpha$ and the above expressions are still not symmetrized for intrinsic permutation symmetry. The action of the generalized derivative operator, k^b, as defined in Eq. (9), implies a "chain rule,"

$$\left(\frac{r^a_{nm}}{\omega_{nm} - \omega_1}\right)_{;k^b} = \frac{(r^a_{nm})_{;k^b}}{\omega_{nm} - \omega_1} - \frac{r^a_{nm}\Delta^b_{nm}}{(\omega_{nm} - \omega_1)^2}, \tag{29}$$

using $(\omega_{nm})_{;k^b} = \Delta^b_{nm}$. Equations (27,28) could be transformed into forms more useful for numerical evaluation by using partial fraction expansions to isolate the individual resonant denominators; such manipulations, and how to handle the nonsimple poles, which the $\mathbf{k}$ operator introduces, are illustrated elsewhere.[8,11]

An alternative to the above is to start from Eq. (19) and write $\chi^{(2)} = \chi^{(2)}_\chi + \chi^{(2)}_\sigma$, where $\chi^{(2)}_\chi$ originates from substituting $\rho^{(2)}$ into the $\mathbf{P}_\chi$ expression, while both $\rho^{(2)}$ and $\rho^{(1)}$ in $\mathbf{J}_\sigma$ give $\chi^{(2)}_\sigma$,

$$\chi^{(2)}_\chi = C\sum_{m,n,l,\mathbf{k}} \frac{r^c_{mn}}{\omega_{nm} - \omega_2}\left(\frac{r^b_{nl} r^a_{lm} f_{ml}}{\omega_{lm} - \omega_1} - \frac{r^a_{nl} r^b_{lm} f_{ln}}{\omega_{nl} - \omega_1}\right) + iC\sum_{m,n,\mathbf{k}} \frac{r^c_{mn}}{\omega_{nm} - \omega_2}\left(\frac{r^a_{nm} f_{mn}}{\omega_{nm} - \omega_1}\right)_{;k^b}, \tag{30}$$

$$\chi^{(2)}_\sigma = \frac{C}{i\omega_2^2}\sum_{m,n,\mathbf{k}} \Delta^c_{mn} r^b_{mn}\left(\frac{r^a_{nm} f_{mn}}{\omega_{nm} - \omega_1}\right) + \frac{C}{i\omega_2}\sum_{m,n,\mathbf{k}} (r^b_{mn})_{;k^c}\left(\frac{r^a_{nm} f_{mn}}{\omega_{nm} - \omega_1}\right). \tag{31}$$

Equations (30,31) are equivalent to the expressions obtained with greater effort in SG. In Appendix B we confirm that the two expressions obtained here for $\chi^{(2)}$, respectively, the sum of Eqs. (27,28) and Eqs. (30,31), are equivalent, as the general considerations in Sec. II B require.

The expressions for $\chi^{(2)}_{cba}(-\omega_2;\omega_\beta,\omega_\alpha)$ involve leading factors of $1/\omega_2$ and $1/\omega_2^2$, suggesting that these expressions may diverge as $\omega_\alpha, \omega_\beta \to 0$. But a finite $\chi^{(2)}$ is expected for an insulator in this limit, and, in fact, these divergences do not exist. We illustrate the general $\chi^{(2)}$ case

here,[29] using the form $\chi_\chi^{(2)}+\chi_\sigma^{(2)}$; $\chi_\chi^{(2)}$ is already explicitly finite for vanishing frequencies and we show that $\chi_\sigma^{(2)}$ is also. Introducing intrinsic permutation symmetry, relabeling indices, and combining terms using $\omega_2=\omega_\beta+\omega_\alpha$, we find that $\chi_\sigma^{(2)}$ can be written as

$$\chi_\sigma^{(2)}=\frac{C}{2i}\sum_{m,n,\mathbf{k}}\left(\frac{r^b_{mn}}{\omega_{nm}+\omega_\beta}\right)_{;k^c}\frac{r^a_{nm}f_{mn}}{\omega_{nm}-\omega_\alpha}, \tag{32}$$

which is explicitly finite for vanishing frequencies. A (generalized) partial integration involving k^c was used to arrive at the above expression, but it is important to note that no "sum rules" such as Eqs. (10–13) were employed. As we discuss in Sec. IV, such sum rules are generally required in the velocity gauge, which is one of the characteristics that limits its usefulness.

That Eq. (32) arises so easily here can be attributed to the economy of notation afforded by the $\mathbf{k}$ operator, which naturally accounts for terms that otherwise appear as "bulky" expansions, such as Eq. (11). Similar observations were made many years ago, where the second order case was also investigated.[13] However, the full grouping and identification of the generalized derivative $\mathbf{k}$, was not recognized there, and some unattractive features remained. For example, a finite $\chi^{(2)}(0;0,0)$ could only be proven for certain crystal classes; indeed, such a finite zero-frequency limit in the general case was first proven only recently after identifying a relevant sum rule.[30]

Our well-behaved susceptibilities also allow us to reliably study cases where only some of the involved frequencies are zero. For example, for frequencies ω below the band-gap energy ("nonresonant"), we find that $\chi^{(2)}(0;\omega,-\omega)$, describing optical rectification, is also finite. Further, in the opposite resonant situation, a divergence now does appear in the optical rectification coefficient, but this divergence is not problematic and arises in order to faithfully describe the physical situation. The resonant divergences can be understood by realizing that since $\chi^{(2)}$ describes a polarization $\mathbf{P}e^{-i\omega_2 t}$, the quantities $-i\omega_2\chi^{(2)}$ and $(-i\omega_2)^2\chi^{(2)}$ describe a current $\mathbf{J}=d\mathbf{P}/dt$ and rate of change of current (which we will term current injection) $d\mathbf{J}/dt=d^2\mathbf{P}/dt^2$. Hence, a finite $\chi^{(2)}(0;\omega,-\omega)$ implies no dc current flow, while a divergence indicates currents; loosely speaking the "parts" diverging as $1/\omega_2$ and $1/\omega_2^2$ correspond to dc current and current generation, respectively.

In order to obtain expressions in the resonant case, it is convenient to return to Eqs. (30,31), satisfy intrinsic permutation symmetry, and introduce, in the standard way, a small positive imaginary part ϵ (up to now implicitly assumed) to the frequencies. Then the expression for $(-i\omega_2)^2\chi^{(2)}$ is expanded in powers of ω_2 in the limit of vanishing ϵ. This provides one term independent of ω_2, which describes the dc current injection (constant $d\mathbf{J}/dt$) and is denoted as $\nu^{(2)}(0;\omega,-\omega)$,

$$\nu^{(2)}=-\pi C\sum_{m,n,\mathbf{k}}\Delta^c_{mn}r^b_{mn}r^a_{nm}f_{mn}\delta(\omega_{nm}-\omega), \tag{33}$$

as well as a term linear in ω_2, which describes a dc current (constant $\mathbf{J}$), denoted by $\sigma^{(2)}(0;\omega,-\omega)$,

$$\sigma^{(2)}=\frac{\pi C}{2i}\sum_{m,n,\mathbf{k}}[(r^b_{mn})_{;k^c}r^a_{nm} -r^b_{mn}(r^a_{nm})_{;k^c}]f_{mn}\delta(\omega_{nm}-\omega). \tag{34}$$

As one might guess, it is only $\chi_\sigma^{(2)}$ that contributes to these expressions. The above rectification coefficients can be generally classified as photogalvanic effects, as has been previously discussed.[31–33] Here, we would mainly like to emphasize that our results show that these effects are captured within the standard framework of nonlinear optics (susceptibilities). Further, we stress that a reliable description in terms of susceptibilities strongly depends on having expressions, like ours, which do not suffer from unphysical divergences. In particular, note that the appearance of the δ functions in both expressions above explicitly shows that below the band-gap $\nu^{(2)}=\sigma^{(2)}=0$, and then the $\chi^{(2)}$ is finite, consistent with our previous conclusions. Previous incorrect predictions of nonresonant (below bandgap) photogalvanic currents have been made,[34] likely due to the presence there of unphysical divergences.

These rectification effects are very interesting in terms of applications, since one can show that the nonzero $\sigma^{(2)}$ and $\nu^{(2)}$ can imply that by controlling the relative phase of two polarizations of light at a single frequency, one could drive a current in preferred directions.[31,35] This single-beam, all optical, current-steering scheme can be considered a coherently controlled phenomena, similar to other processes that have been discussed.[37–40] Of course, a quantitative prediction of current and current generation likely requires one to go beyond our *perturbative* treatment of an *infinite* crystal without *scattering*. That is, effects such as saturation, space charge accumulation, and dephasing processes are surely important,[33] and this continuous-wave treatment would have to be refined if pulse widths become small, and/or to accurately model spatial transport.[36] Nonetheless, that such effects can be unambiguously predicted from our simple expressions illustrates the usefulness of expressions that behave as physically expected in the relevant zero-frequency limits.

B. Third-order effects

The third-order susceptibility tensor represented by $\chi^{(3)}_{dcba}(-\omega_3;\omega_\gamma,\omega_\beta,\omega_\alpha)$, where $\omega_3=\omega_\gamma+\omega_\beta+\omega_\alpha$, can also be found using either Eq. (18) or Eq. (19). The former would provide the most compact expression, with $\chi^{(3)}=\chi^{(3)}_{ee}+\chi^{(3)}_{ei}+\chi^{(3)}_{ie}+\chi^{(3)}_{ii}$, where the individual contributions arise from the four $\rho^{(3)}$ terms ρ^{eee}, ρ^{eie}, ρ^{iee}, and ρ^{iie} in Fig. 1. This form can have benefits over velocity-gauge expressions, as shown by a recent model calculation[11] of a diagonal tensor component $\chi^{(3)}_{zzzz}$. In general, however, Eq. (19) can be considered a preferred starting point because, as for $\chi^{(2)}$, unphysical divergences can then be avoided without the use of sum rules. Hence, we write $\chi^{(3)}=\chi^{(3)}_\chi+\chi^{(3)}_\sigma$; the expression for $\chi^{(3)}_\chi$ obtained by substituting $\rho^{(3)}$ into the last term of Eq. (19),

$$\frac{\chi_X^{(3)}}{C} = \sum_{l,m,n,p,\mathbf{k}} \frac{r^d_{mn}}{\omega_{nm}-\omega_3}\left[\frac{r^c_{nl}}{\omega_{lm}-\omega_2}\left(\frac{r^b_{lp}r^a_{pm}f_{mp}}{\omega_{pm}-\omega_1}-\frac{r^a_{lp}r^b_{pm}f_{pl}}{\omega_{lp}-\omega_1}\right)-\left(\frac{r^b_{nl}r^a_{lp}f_{pl}}{\omega_{lp}-\omega_1}-\frac{r^a_{nl}r^b_{lp}f_{ln}}{\omega_{nl}-\omega_1}\right)\frac{r^c_{pm}}{\omega_{np}-\omega_2}\right]$$
$$+i\sum_{l,m,n,\mathbf{k}} \frac{r^d_{mn}}{\omega_{nm}-\omega_3}\left[\frac{1}{\omega_{nm}-\omega_2}\left(\frac{r^b_{nl}r^a_{lm}f_{ml}}{\omega_{lm}-\omega_1}-\frac{r^a_{nl}r^b_{lm}f_{ln}}{\omega_{nl}-\omega_1}\right)\right]_{;k^c}$$
$$+i\sum_{l,m,n,\mathbf{k}} \frac{r^d_{mn}}{\omega_{nm}-\omega_3}\left[\frac{r^c_{nl}}{\omega_{lm}-\omega_2}\left(\frac{r^a_{lm}f_{ml}}{\omega_{lm}-\omega_1}\right)_{;k^b}-\left(\frac{r^a_{nl}f_{ln}}{\omega_{nl}-\omega_1}\right)_{;k^b}\frac{r^c_{lm}}{\omega_{nl}-\omega_2}\right]$$
$$-\sum_{m,n,\mathbf{k}} \frac{r^d_{mn}}{\omega_{nm}-\omega_3}\left[\frac{1}{\omega_{nm}-\omega_2}\left(\frac{r^a_{nm}f_{mn}}{\omega_{nm}-\omega_1}\right)_{;k^b}\right]_{;k^c}. \tag{35}$$

Both $\rho^{(3)}$ and $\rho^{(2)}$ are used in the remaining terms of Eq. (19), and after relabeling some indices, we find

$$\frac{\chi_\sigma^{(3)}}{C} = \frac{1}{i\omega_3^2}\sum_{l,m,n,\mathbf{k}} \Delta^d_{nl}\frac{r^c_{nl}}{\omega_{ln}-\omega_2}\left(\frac{r^b_{lm}r^a_{mn}f_{nm}}{\omega_{mn}-\omega_1}-\frac{r^a_{lm}r^b_{mn}f_{ml}}{\omega_{lm}-\omega_1}\right)$$
$$-\frac{1}{\omega_3^2\omega_2}\sum_{m,n,\mathbf{k}}\Delta^d_{nm}\left(\frac{r^b_{nm}r^a_{mn}f_{nm}}{\omega_{mn}-\omega_1}\right)_{;k^c}+\frac{1}{\omega_3^2}\sum_{m,n,\mathbf{k}}\Delta^d_{nm}\frac{r^c_{nm}}{\omega_{mn}-\omega_2}\left(\frac{r^a_{mn}f_{nm}}{\omega_{mn}-\omega_1}\right)_{;k^b}$$
$$-\frac{i}{\omega_3}\sum_{m,n,l,\mathbf{k}}\frac{(r^c_{nl})_{;k^d}}{\omega_{ln}-\omega_2}\left(\frac{r^b_{lm}r^a_{mn}f_{nm}}{\omega_{mn}-\omega_1}-\frac{r^a_{lm}r^b_{mn}f_{ml}}{\omega_{lm}-\omega_1}\right)+\frac{1}{\omega_3}\sum_{m,n,\mathbf{k}}\frac{(r^c_{nm})_{;k^d}}{\omega_{mn}-\omega_2}\left(\frac{r^a_{mn}f_{nm}}{\omega_{mn}-\omega_1}\right)_{;k^b}$$
$$-\frac{1}{i\omega_3\omega_2}\sum_{m,n,\mathbf{k}}\left(\frac{\partial\mathcal{E}^c_{nm}}{\partial k^d}-\frac{\partial\mathcal{E}^d_{nm}}{\partial k^c}\right)\frac{r^b_{nm}r^a_{mn}f_{nm}}{\omega_{mn}-\omega_1}, \tag{36}$$

where $\mathcal{E}_{nm}\equiv\xi_{nn}-\xi_{mm}$ and $C=e^4K/\hbar^3$, with K again accounting for factors describing the particular $\omega_\alpha,\omega_\beta$, and ω_γ of interest. Again Eqs. (35,36) remain to be symmetrized to satisfy intrinsic permutation symmetry, since, here, $\omega_1=\omega_\alpha$ and $\omega_2=\omega_\beta+\omega_\alpha$. Note that obtaining the above $\chi^{(3)}=\chi_\chi^{(3)}+\chi_\sigma^{(3)}$ expressions from the SG formalism would require substantially more effort. We further point out that the anomalous velocity in Eq. (19) contributes only at this third-order level; it did not lead to any terms in $\chi^{(1)}$ or $\chi^{(2)}$, but gives the last term in Eq. (36). This quantity, basically $\nabla\times\xi_{nn}$, is of some theoretical interest, since it is associated with the topology of the band structure,[14] and plays an important role in several recent theories, where it is related to a relevant geometric phase.[21,41] The above expressions suggest that insulator experiments sensitive to $\chi^{(3)}$ may be a valuable probe of this anomalous velocity.

The factors of $1/\omega_3$ and $1/\omega_3^2$ above suggest similar divergences to those seen in $\chi^{(2)}$. Further, factors of $1/\omega_2$ can also arise in $\chi^{(3)}$; the eventual permutation symmetry implies that $\omega_2=0$ corresponds to a zero sum for any two of the frequencies involved. However, all of these divergences are again only apparent ones, and we find a finite $\chi^{(3)}(0;0,0,0)$ as physically expected for insulators. To show this, note that only the first sum in Eq. (35) for $\chi_\chi^{(3)}$ includes terms that appear to have leading divergences (a $1/\omega_2$ factor for $l=m$ and $n=p$), but these terms can be rewritten. Using similar steps that have led to Eq. (32), the troublesome terms in $\chi_\chi^{(3)}$ reduce to

$$C\sum_{lmn\mathbf{k}}\frac{(r^d_{nm}r^c_{mn}-r^c_{nm}r^d_{mn})r^b_{ml}r^a_{lm}f_{ml}}{(\omega_{nm}-\omega_3)(\omega_{lm}-\omega_1)(\omega_{lm}+\omega_2-\omega_1)}. \tag{37}$$

Similarly, we find that the entire $\chi_\sigma^{(3)}$ can be generated from the following terms:

$$\frac{\chi_\sigma^{(3)}}{C}=\sum_{mn\mathbf{k}}\frac{1}{\omega_{mn}-\omega_2}\left(\frac{r^c_{nm}}{\omega_{mn}+\omega_3-\omega_2}\right)_{;k^d}\times\left(\frac{r^a_{mn}f_{nm}}{\omega_{mn}-\omega_1}\right)_{;k^b}+\sum_{lmn\mathbf{k}}i\frac{r^a_{mn}r^b_{lm}}{\omega_{ln}-\omega_2}\left(\frac{f_n}{\omega_{mn}-\omega_1}+\frac{f_l}{\omega_{lm}-\omega_2+\omega_1}\right)\left(\frac{r^c_{nl}}{\omega_{nl}-\omega_3+\omega_2}\right)_{;k^d}. \tag{38}$$

The above expressions must still be made to satisfy intrinsic permutation symmetry, but already they explicitly show that $\chi^{(3)}(0;0,0,0)$ is finite. Also, while obtaining the above forms involved some monotonous algebra, due to the number of terms involved, we again emphasize that no sum rules were required, and that again the properties of the generalized partial derivatives were very useful. Further, these results lead to conclusions that are analogous to our $\chi^{(2)}(0;-\omega,\omega)$ findings; that is, as long as resonances do not occur, then $\chi^{(3)}(0;-\Omega-\omega,\Omega,\omega)$ is finite and no dc current is predicted. We note also that all terms, including the contributions from the anomalous velocity in Eq. (36), play an important role in ensuring this situation. The above is to our knowledge the first ex-

plicit general proof of the finiteness of the low-frequency insulator third-order susceptibility.

Finally, we find that, analogous to the second-order situation, $\chi^{(3)}$ can have physical divergences when resonances are possible. In particular, in this case for $\omega_3 = 0$, the $\chi^{(3)}(0; -\Omega - \omega, \Omega, \omega)$ implies, for example, a dc current injection, or nonzero $\nu^{(3)}$ analogous to the second-order $\nu^{(2)}$.[35] A divergence associated with $\omega_2 = 0$ can also appear under resonant conditions; this is simply a manifestation at this third-order level of the second-order physical divergences we have already encountered. Similar to the second-order result, one can imagine interesting applications for coherent control of the current and current injection. For example, there is some experimental interest[40] in $\nu^{(3)}(0; \omega, \omega, -2\omega)$, where the phase relationship between two (now nondegenerate in frequency) beams can be used to control the direction of generated current. Although this process can be calculated by a simple Fermi's Golden rule analysis, the calculation here shows it can also be described within the usual susceptibility formalism of nonlinear optics; as with the physics associated with the physically divergent terms in $\chi^{(2)}$ discussed at the end of Sec. III A, this fact does not generally appear to have been appreciated.

IV. DISCUSSION

It is of course clear that a length-gauge calculation, such as we have advocated in this paper, must lead to the same result as the velocity-gauge calculation if both are done correctly. Yet velocity-gauge treatments appear to be more unforgiving when the inevitable approximations that allow a practical calculation are made, as we now illustrate. We begin by establishing a relationship between the length-gauge and velocity-gauge treatments. Consider the velocity-gauge analog of Eq. (21),

$$i\hbar[\tilde{\rho}_V(t) - \rho_0] = \frac{-e}{mc}\int_{-\infty}^{t} dt' \, [\tilde{\mathbf{p}}(t') \cdot \mathbf{A}(t'), \tilde{\rho}_V(t')], \quad (39)$$

where the subscript V denotes the velocity-gauge. In this interaction picture $\tilde{\mathbf{p}} = m\dot{\tilde{\mathbf{r}}}$, and since $c\mathbf{E} = -\dot{\mathbf{A}}$, a useful integration by parts is possible,

$$\begin{aligned} i\hbar[\tilde{\rho}_V(t) - \rho_0] = {} & \frac{-e}{c}[\tilde{\mathbf{r}}(t) \cdot \mathbf{A}(t), \tilde{\rho}_V(t)] \\ & -e\int_{-\infty}^{t} dt' \, [\tilde{\mathbf{r}}(t') \cdot \mathbf{E}(t'), \tilde{\rho}_V(t')] \\ & +\frac{e}{c}\int_{-\infty}^{t} dt' \, [\tilde{\mathbf{r}}(t') \cdot \mathbf{A}(t'), \dot{\tilde{\rho}}_V(t')]. \quad (40) \end{aligned}$$

To the first order this implies that

$$\tilde{\rho}_V^{(1)}(t) = \tilde{\rho}_L^{(1)}(t) - \frac{e}{i\hbar c}[\tilde{\mathbf{r}}(t) \cdot \mathbf{A}(t), \rho_0], \quad (41)$$

where L denotes the length gauge. Together with the velocity-gauge operator, $\mathbf{v} = \mathbf{p}/m - e\mathbf{A}/mc$, Eq. (14) then leads to $\chi_V^{(1)} = \chi_L^{(1)} + \mathcal{R}^{(1)}$, where for a monochromatic field at frequency ω,

$$\mathcal{R}^{(1)} = \frac{-ie^2}{\omega^2 m\hbar}\mathrm{Tr}(C_1\rho_0), \quad (42)$$

with $C_1 = [r^a, p^b] - i\hbar\delta^{ab}$. Hence, this explicitly illustrates gauge invariance, $\chi_V^{(1)} = \chi_L^{(1)}$, since, formally, $\mathcal{R}^{(1)} = 0$, due the commutator identity $[r^a, p^b] = i\hbar\delta^{ab}$. More generally, one can use Eq. (40) and show by similar steps that to any order $\chi_V^{(N)} = \chi_L^{(N)} + \mathcal{R}^{(N)}$, where again $\mathcal{R}^{(N)}$ strictly vanishes, due to commutator identities.

The above relation suggests how problems may arise in velocity-gauge treatments: In certain cases, electronic structure models implicitly contradict the commutator identities, and so $\mathcal{R}^{(N)}$ will not vanish. This can have dramatic consequences on velocity-gauge predictions. For example, we have discussed how a common two band model for semiconductors can lead to unphysical zero-frequency divergences in $\chi^{(3)}$ by exactly this scenario.[11] In this case, divergences are introduced into $\chi_V^{(3)}$ by a nonvanishing $\mathcal{R}^{(3)}$; note that leading $1/\omega$ factors appearing in $\mathcal{R}^{(1)}$ above arise generally in every $\mathcal{R}^{(N)}$. Of course, such problems could be avoided by appealing to commutator identities to "filter out" the $\mathcal{R}^{(N)}$ contribution from the exact velocity-gauge expressions prior to any modeling. However, in practice, this is not straight forward since the expressions involve sums over complete sets of states. This complicates the identification of both $\mathcal{R}^{(N)}$ and the appropriate "sum rule" — which is just a matrix element of a commutator identity — that should be employed. Further note that the $\mathcal{R}^{(N)}$ contribution can affect practical velocity-gauge treatments even if an exact electronic structure model is employed. That is, practical calculations demand a truncation of the expansions over states, and so the $\mathcal{R}^{(N)}$ term will not generally cancel from the expressions. This again may lead to differences between length-gauge and velocity-gauge predictions, and we suggest that a variety of seemingly distinct issues can be viewed from this perspective.[17,43] The fact that the length gauge leads to results that are free of unphysical zero-frequency divergences, *without* the application of sum rules that must be identified on a case-by-case basis, indicates it is more suitable for at least the calculation of frequency dependent susceptibilities.

Finally, to compare our work to SG (and that of Genkin and Mednis[9] on which it is based), transform our Hamiltonian, Eq. (2), by $U_i = e^{i(e/\hbar c)\mathbf{r}_i \cdot \mathbf{A}(t)}$, leading to the Hamiltonian

$$U_i H U_i^+ = U_i H_0 U_i^+ - eU_i \mathbf{r}_e U_i^+ \cdot \mathbf{E}(t). \quad (43)$$

Since $-e\mathbf{r}_i \cdot \mathbf{E}$ no longer appears in Eq. (43), the U_i effectively accounts exactly for this intraband part of the perturbation. The effect of U_i is easy to expose by interpreting $\mathbf{r}_i$ to be (mainly) the generator for $\mathbf{k}$ translations within a band; recall the $\nabla_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}')$ in Eq. (3). This suggests that for a general operator O, the matrix elements of interest are now of the form

$$\langle n\mathbf{k}|U_i O U_i^+|m\mathbf{k}'\rangle = e^{-i\phi}\langle n\mathcal{K}(t)|O|m\mathcal{K}'(t)\rangle, \quad (44)$$

where ϕ is a phase, and $\mathcal{K}(t) = \mathbf{k} - e/c\mathbf{A}(t)$. Comparing Eqs. (43,44) to equations occurring in SG allows us to

reinterpret the velocity-gauge based theory in SG within the framework of this paper: SG implicitly transforms to the length gauge and accounts for the intraband motion exactly (to all orders). But if perturbative results are desired, this requires more manipulations than the straight-forward perturbative approach employed here, where both intraband and interband effects are treated at the same level.

V. SUMMARY

We have offered a simple prescription to derive the electronic contributions to the optical susceptibilities of crystals. The relationship between the optical response of atoms, free electrons, and crystals is more easily exposed by our formalism. Clean, cold semiconductors (insulators) were of central interest here, but the scheme is also applicable to metallic systems, such as doped crystals or laser-excited semiconductors. We have illustrated the scheme by deriving the second- and third-order susceptibilities of insulators for arbitrary frequency mixings, and discussed their behavior for various zero-frequency limits; such expressions are given here, we believe, for the first time.

The simplicity and usefulness of our formalism is illustrated by the corrections to the standard semiconductor Bloch equations, which it easily exposes. Further, the physically expected zero-frequency limits of our nonlinear susceptibilities have allowed us to unambiguously show how the processes of optical generation and coherent control of dc current can be described within the usual susceptibility framework of nonlinear optics. In the future, the susceptibilities, in this paper, will be used with band structure models, of varying degrees of sophistication, to look at different nonlinear processes.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada, and the Ontario Laser and Lightwave Research Centre. We would like to thank Professor M. Sheik-Bahae and Professor E. W. Van Stryland for several fruitful discussions, and Professor M. Cardona for pointing out Ref. 43 to us. C.A. is also grateful to the Department of Physics and the University of Toronto for financial support.

APPENDIX A: MICROSCOPIC AND MACROSCOPIC CURRENTS

It is often useful to define a microscopic version θ_μ of an operator θ,

$$\theta_\mu \equiv \frac{1}{2}(\theta\mu + \mu\theta), \tag{A1}$$

where, for the moment explicitly indicating operators by carets, the operator $\mu \equiv |\mathbf{r}\rangle\langle\mathbf{r}| \equiv \delta(\mathbf{r} - \hat{\mathbf{r}})$; implicit in the notation θ_μ is the dependence of the quantity on the field point $\mathbf{r}$. The microscopic current operator is an example of interest here,

$$\mathbf{j}_\mu = \frac{e}{2}(\mathbf{v}\mu + \mu\mathbf{v}), \tag{A2}$$

where we have used $\mathbf{j} \equiv e\mathbf{v}$. The expected current at position $\mathbf{r}$ can then be compactly written as $\mathrm{Tr}(\rho\mathbf{j}_\mu)$. Macroscopic quantities can be obtained by performing local spatial averages of the microscopic counterparts. Such averages (which can also be effected by filtering out high spatial frequencies) eventually reduce to the macroscopic average of

$$\langle n\mathbf{k}|\mu|m\mathbf{k}\rangle = u^*_{n\mathbf{k}}(\mathbf{r})u_{m\mathbf{k}}(\mathbf{r}). \tag{A3}$$

Finally, an average of the above quantity over a unit cell reduces to $\delta_{nm}/(2\pi)^3$.

APPENDIX B: EQUIVALENCE OF $\chi^{(2)}$ EXPRESSIONS

While Sec. II B shows that the two $\chi^{(2)}$ expansions given by Eqs. (27,28) and Eqs. (30,31), respectively, must be equal, explicitly illustrating this exposes several useful points. To demonstrate the equivalence, substitute the partial fraction expansion

$$\frac{1}{\omega_2(\omega_{nm} - \omega_2)} = \frac{1}{\omega_{nm}(\omega_{nm} - \omega_2)} + \frac{1}{\omega_2\omega_{nm}} \tag{B1}$$

into $\chi^{(2)}_i$, and use Eq. (7), to find the second term in $\chi^{(2)}_\chi$ and another term which can be rewritten

$$\frac{iC}{\omega_2}\sum_{mn\mathbf{k}} r^c_{mn}\left(\frac{r^a_{nm}f_{mn}}{\omega_{nm} - \omega_1}\right)_{;k^b}$$

$$= \frac{C}{i\omega_2}\sum_{mn\mathbf{k}}(r^c_{mn})_{;k^b}\frac{r^a_{nm}f_{mn}}{\omega_{nm} - \omega_1}, \tag{B2}$$

resembling the second term of $\chi^{(2)}_\sigma$. The equality in Eq. (B2) follows from a (generalized) partial integration, or by realizing it is simply an expression of $\mathrm{Tr}(r^c_e[r^b_i, \rho^{(1)}]) = \mathrm{Tr}([r^c_e, r^b_i]\rho^{(1)})$, which follows from cyclic invariance of the trace. Next, using Eqs. (7,B1) in $\chi^{(2)}_e$ reveals the first terms of $\chi^{(2)}_\sigma$ and $\chi^{(2)}_\chi$, along with the following term:

$$\frac{C}{\omega_2}\sum_{mnl\mathbf{k}}(r^c_{ml}r^b_{ln} - r^b_{ml}r^c_{ln})\left(\frac{r^a_{nm}f_{mn}}{\omega_{nm} - \omega_1}\right). \tag{B3}$$

This combines with the term of Eq. (B2), through the sum rule in Eq. (12), to give the second term in $\chi^{(2)}_\sigma$, completing the proof that $\chi^{(2)}_e + \chi^{(2)}_i = \chi^{(2)}_\chi + \chi^{(2)}_\sigma$.

* Present address: Quantum Institute, UC Santa Barbara, Santa Barbara, CA 93106.
[1] G. Assanto *et al.*, Appl. Phys. Lett. **62**, 1323 (1993).
[2] A. Villeneuve *et al.*, Appl. Phys. Lett. **62**, 2465 (1993).
[3] K. Al-hemyari *et al.*, Appl. Phys. Lett. **63**, 3562 (1993).
[4] S. Li and Jacob B. Khurgin, Appl. Phys. Lett. **73**, 4367 (1993).
[5] D. S. Chemla, Phys. Today **38**, 57 (1985).
[6] D. J. Moss *et al.*, Phys. Rev. B **41**, 1542 (1990).
[7] M. Sheik-Bahae *et al.*, IEEE J. Quantum Electron. **QE-27**, 1296 (1991).
[8] J. E. Sipe and E. Ghahramani, Phys. Rev. B **48**, 11 705 (1993).
[9] V. M. Genkin and P. Mednis, Zh. Eksp. Teor. Fiz. **54**, 1137 (1968) [Sov. Phys. JETP **27**, 609 (1968)].
[10] C. Aversa and J. E. Sipe (unpublished).
[11] C. Aversa *et al.*, Phys. Rev. B **50**, 18 073 (1994).
[12] G. P. Agrawal, C. Cojan, and C. Flytzanis, Phys. Rev. B **17**, 776 (1978).
[13] D. E. Aspnes, Phys. Rev. B, 4648 (1972).
[14] E. I. Blount, *Solid State Physics: Advances in Research and Applications* (Academic, New York, 1962), Vol. 13.
[15] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957).
[16] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, *Photons and Atoms: Introduction to Quantum Electrodynamics* (Wiley, New York, 1989).
[17] F. Bassani, J. J. Forney, and A. Quattropani, Phys. Rev. Lett. **39**, 1070 (1977).
[18] In the case that $n \neq m$, but $\omega_{nm} = 0$, one chooses the wave functions at these degeneracies, such that the relevant components of $\mathbf{p}_{nm}$ and $\mathbf{r}_{nm}$ vanish.
[19] V. M. Genkin and P. Mednis, Fiz. Tverd. Tela (Leningrad) **10**, 1 (1968) [Sov. Phys. Solid State **10**, 1 (1968)].
[20] The similarity of the generalized derivative to a covariant derivative suggests an avenue that may establish a relationship between our work and the recent geometric phase theories of polarization; see Ref. 21 and references therein, and also E. Richter, Bull. Am. Phys. Soc. **39**, 104 (1994).
[21] R. Resta, Rev. Mod. Phys. **66**, 1 (1994).
[22] E. N. Adams and E. I. Blount, J. Phys. Chem. Solids **10**, 286 (1959).
[23] H. Haug and S. W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (World Scientific, Singapore, 1990).
[24] S. Schmitt-Rink *et al.*, Phys. Rev. B **37**, 941 (1988).
[25] See, for example, *Quantum Transport in Semiconductors* (Plenum, New York, 1992).
[26] C. Flytzanis, *Quantum Electronics* (Academic Press, New York, 1975), Vol. 1.
[27] Z. Levine *et al.*, Phys. Rev. B **49**, 4532 (1994).
[28] Butcher and Cotter, *Elements of Nonlinear Optics* (Cambridge University Press, Cambridge, 1990).
[29] In recent work for the second harmonic generation and Pockels effect, involving $\chi^{(2)}(-2\omega;\omega,\omega)$ and $\chi^{(2)}(-\omega;\omega,0)$, the susceptibilities were explicitly shown to be finite for $\omega = 0$. See, respectively, Ref. 8 and E. Ghahramani and J. E. Sipe, Appl. Phys. Lett. **64**, 2421 (1994).
[30] E. Ghahramani *et al.*, Phys. Rev. B **43**, 8990 (1991).
[31] V. M. Asnin *et al.*, Solid State Commun. **30**, 565 (1979).
[32] V. I. Belincher *et al.*, Zh. Eksp. Teor. Fiz. **83**, 649 (1982) [Sov. Phys. JETP **56**, 359 (1982)].
[33] V. I. Belincher and B. I. Sturman, Usp Fiz. Nauk. **130**, 415 (1980) [Sov. Phys. Usp. **23**, 199 (1980)].
[34] N. Kristoffel, Phys. Status Solidi B **127**, 413 (1985).
[35] C. Aversa and J. E. Sipe (unpublished).
[36] A. V. Kuznetsov and C. J. Stanton, Phys. Rev. B **48**, 10 828 (1993).
[37] V. V. Obukhovsky and A. Y. Stoyanov, Ferroelectrics **3**, 137 (1982).
[38] G. Kuriski *et al.*, Phys. Rev. B **39**, 3435 (1989).
[39] E. Dupont *et al.*, Phys. Rev. Lett. **74**, 3596 (1995).
[40] A. Hache (private communication).
[41] R. D. King-Smith and David Vanderbilt, Phys. Rev. B **47**, 1651 (1993).
[42] G. Barton, J. Phys. B **7**, 2134 (1974).
[43] R. Zeyher, H. Bilz, and M. Cardona, Solid State Commun. **19**, 57 (1976).