

Regularized lattice theory for spatially dispersive nonlinear optical conductivities

Steven Gassner¹* and E. J. Mele

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA



(Received 26 April 2023; revised 19 July 2023; accepted 21 July 2023; published 3 August 2023)

Nonlinear optical responses are becoming increasingly relevant for characterizing the symmetries and quantum geometry of electronic phases in materials. Here, we develop an expanded diagrammatic scheme for calculating *spatially dispersive corrections* to nonlinear optical conductivities, which we expect to enhance or even dominate even-order responses in materials of recent interest. Building upon previous work that enforces gauge invariance of spatially uniform nonlinear optical responses, we review the cancellation of diagrams required to ensure the equivalence between velocity gauge and length gauge formulations, and provide a simple vertex rule for extending optical responses to first order in the light wave vector \mathbf{q} . We then demonstrate the method with calculations on a prototypical centrosymmetric model where spatial dispersion admits anomalous second-harmonic generation, a response that is symmetry forbidden under the dipole approximation.

DOI: [10.1103/PhysRevB.108.085403](https://doi.org/10.1103/PhysRevB.108.085403)

I. INTRODUCTION

Optical responses are an extremely valuable tool for understanding the electronic structure of quantum matter, from materials to molecules [1,2]. Not only are they vital for characterizing a host of different devices from photovoltaics to lasers but they are also immensely useful as probes for symmetry breaking and quantum geometric properties in materials [3–11]. Over the past several decades, the development of a straightforward and general means of systematically calculating these responses has been a surprisingly difficult task, revealing a number of subtleties that are still being worked out to this day. A main reason is that there are different choices of gauge for describing the light-matter coupling with various advantages and disadvantages. One option is to couple the electric field $\mathbf{E}(t)$ with the position operator $\hat{\mathbf{r}}$,

$$\hat{\mathcal{H}} \rightarrow \hat{\mathcal{H}} - q \hat{\mathbf{r}} \cdot \mathbf{E}(t), \quad (1)$$

which goes by the name length gauge (also known as position gauge or dipole gauge.) Here, $q = -|e|$ denotes the electron charge. Another option, called the velocity gauge, employs a minimal substitution in the \mathbf{k} -dependent Bloch Hamiltonian,

$$\hat{\mathcal{H}}(\mathbf{k}) \rightarrow \hat{\mathcal{H}}(\mathbf{k} - q \mathbf{A}(t)), \quad (2)$$

where $\mathbf{A}(t)$ is the magnetic vector potential satisfying $\mathbf{E}(t) = -\partial_t \mathbf{A}(t)$. (At first order in $\mathbf{A}(t)$, this takes the form of a coupling $-q \hat{\mathbf{v}} \cdot \mathbf{A}(t)$ where $\hat{\mathbf{v}} = \nabla_{\mathbf{k}} \hat{\mathcal{H}}$ is the velocity operator; hence the name velocity gauge.) These two gauges are related by a time-dependent unitary transformation, and are hence equivalent [12] in a fully microscopic theory. However, maintaining this equivalence for effective Hamiltonians that are projected into a partial sector of the full Hilbert space requires additional consideration.

Velocity gauge, while having the advantage of being diagonal in \mathbf{k} space and easier for numerical calculations, has the serious disadvantage of introducing artificial low-frequency divergences that must be regularized systematically. Historically, this has been attempted using sum rules, developed by Sipe and coworkers [13,14] and later generalized by Ventura *et al.* [15,16], which formally show that the weight of these divergent terms vanishes only after a full \mathbf{k} -space integration. In more recent work with Wannierized tight-binding models, Schüler *et al.* identified a sum rule encoding the cancellation between the paramagnetic and diamagnetic currents calculated at *linear order* in $\mathbf{A}(t)$, and showed that enforcing this sum rule improves the accuracy of velocity gauge calculations [17].

In this paper, we relate a convenient diagrammatic method [18] and the traditional reduced density matrix perturbation theory approach [15,16]. Representing terms in the perturbation theory by their associated diagrams, we can understand in a physically transparent manner the general cancellation between the paramagnetic and diamagnetic contributions in the linear response and investigate the analogs of that cancellation that occurs for nonlinear responses. We also introduce a scheme for extending these results to first order in the wave vector \mathbf{q} of the incident radiation. We call these responses *spatially dispersive*, because they take into account at lowest order the effects of the spatial gradient of the electric field on the optical response. As a prototypical example, we investigate the anomalous second-harmonic generation (SHG) response that appears in a centrosymmetric 1D model. While SHG is typically understood to vanish in centrosymmetric systems in the dipole approximation (and hence is often used experimentally as a probe for inversion breaking in crystals), this selection rule is violated when one takes into account the spatial variation of the electromagnetic field on the length scale of the electronic states being coupled in optical transitions. This becomes especially relevant for artificial lattices where the lattice constants are inflated and can greatly exceed the microscopic atomic scale or more generally for band

*Corresponding author: sgassner@sas.upenn.edu

structures in which the Wannier representation introduces coherence over large length scales.

II. CALCULATIONS OF NONLINEAR OPTICAL RESPONSES

A. Generalities

To calculate nonlinear optical conductivities, we employ the density matrix equation of motion approach detailed in Refs. [15,16] and connect it to the diagrammatic approach developed in Ref. [18]. Recall the density matrix equation of motion,

$$i\hbar \partial_t \hat{\rho}(t) = [\hat{\mathcal{H}}(t), \hat{\rho}(t)], \quad (3)$$

Assume that $\hat{\mathcal{H}}(t) = \hat{\mathcal{H}}^{(0)} + \hat{V}(t)$, with $\hat{V}(t)$ a weak perturbation. Transforming to Fourier representation,

$$\hat{\rho}(t) = \int d\omega e^{-i\omega t} \hat{\rho}(\omega), \quad (4)$$

$$\hat{V}(t) = \int d\omega e^{-i\omega t} \hat{V}(\omega), \quad (5)$$

we can reexpress the equation of motion as an iterative equation in the eigenbasis of the unperturbed Hamiltonian, $\hat{\mathcal{H}}_{ab}^{(0)} = \epsilon_a \delta_{ab}$,

$$\begin{aligned} (\hbar\omega^{(n)} - \epsilon_{ab})\rho_{ab}^{(n)}(\omega^{(n)}) &= [\hat{V}(\omega_n), \hat{\rho}^{(n-1)}(\omega^{(n-1)})]_{ab} \\ \Rightarrow \rho_{ab}^{(n)}(\omega^{(n)}) &= \frac{[\hat{V}(\omega_n), \hat{\rho}^{(n-1)}(\omega^{(n-1)})]_{ab}}{\hbar\omega^{(n)} - \epsilon_{ab}}, \end{aligned} \quad (6)$$

with $\omega^{(n)} = \sum_{j=1}^n \omega_j$ and $\epsilon_{ab} \equiv \epsilon_a - \epsilon_b$, where ϵ_a are the eigenvalues of $\hat{\mathcal{H}}^{(0)}$. Introducing a matrix $\hat{\epsilon}$ whose entries are ϵ_{ab} (and henceforth setting $\hbar = 1$), we can compactly write the solution to the equation of motion as

$$\hat{\rho}^{(n)}(\omega^{(n)}) = \frac{1}{\omega^{(n)} - \hat{\epsilon}} \circ [\hat{V}(\omega_n), \hat{\rho}^{(n-1)}(\omega^{(n-1)})], \quad (7)$$

where \circ denotes the Hadamard product, or elementwise multiplication: $(\hat{A} \circ \hat{B})_{ab} = A_{ab}B_{ab}$. As a final bit of notation, we will assume \hat{V} is a sum of terms from which we can select a different term \hat{V}_i with each iteration i of Eq. (7). We therefore write

$$\hat{\rho}_{\hat{V}_1 \dots \hat{V}_n}^{(n)} = \frac{1}{\omega^{(n)} - \hat{\epsilon}} \circ [\hat{V}_n, \hat{\rho}_{\hat{V}_1 \dots \hat{V}_{n-1}}^{(n-1)}], \quad (8)$$

dropping the implied dependence of $\hat{\rho}^{(n)}$ on $\omega^{(n)}$ as well as the dependence of \hat{V}_n on ω_n . The utility of this notation will become clear shortly.

1. Velocity gauge: Diagrammatic method

Let us first apply this density matrix perturbation theory to light-matter coupling in velocity gauge. In this case,

$$\hat{\mathcal{H}}_{\mathbf{k}}(t) = \hat{\mathcal{H}}_{\mathbf{k}+e\mathbf{A}(t)}^{(0)} \equiv \hat{\mathcal{H}}_{\mathbf{k}}^{(0)} + \hat{V}_{\mathbf{k}}(t), \quad (9)$$

which means that $\hat{V}_{\mathbf{k}}(t)$ takes the form

$$\begin{aligned} \hat{V}_{\mathbf{k}}(t) &= \frac{\partial \hat{\mathcal{H}}_{\mathbf{k}}^{(0)}}{\partial k_{\alpha_1}} e A_{\alpha_1}(t) + \frac{1}{2} \frac{\partial^2 \hat{\mathcal{H}}_{\mathbf{k}}^{(0)}}{\partial k_{\alpha_1} \partial k_{\alpha_2}} e^2 A_{\alpha_1}(t) A_{\alpha_2}(t) + \dots \\ &= \sum_{n=1}^{\infty} \frac{e^n}{n!} \hat{h}^{\alpha_1 \dots \alpha_n} A_{\alpha_1}(t) \dots A_{\alpha_n}(t), \end{aligned} \quad (10)$$

where we introduce the shorthand $\hat{h}^{\alpha_1 \dots \alpha_n} \equiv \partial^{\alpha_1} \dots \partial^{\alpha_n} \hat{\mathcal{H}}_{\mathbf{k}}^{(0)}$ with $\partial^{\alpha} \equiv \frac{\partial}{\partial k_{\alpha}}$. Here and throughout, Greek indices denote spatial directions and repeated indices are summed over. We can calculate the density matrix to any desired order in $A_{\alpha}(t)$. At first order, we have

$$\hat{\rho}^{(1)}(\omega_1) = \frac{1}{\omega_1 - \hat{\epsilon}} \circ [\hat{h}^{\alpha_1}, \hat{\rho}^{(0)}] A_{\alpha_1}(\omega_1). \quad (11)$$

At second order, we have an additional term:

$$\begin{aligned} \hat{\rho}^{(2)}(\omega^{(2)}) &= \frac{1}{\omega^{(2)} - \hat{\epsilon}} \circ [\hat{h}^{\alpha_2}, \hat{\rho}^{(1)}(\omega_1)] A_{\alpha_2}(\omega_2) \\ &+ \frac{1}{\omega^{(2)} - \hat{\epsilon}} \circ \left[\frac{1}{2} \hat{h}^{\alpha_1 \alpha_2}, \hat{\rho}^{(0)} \right] A_{\alpha_1}(\omega_1) A_{\alpha_2}(\omega_2) \\ &\equiv (\hat{\rho}_{\hat{h}^{\alpha_1}, \hat{h}^{\alpha_2}}^{(2)}(\omega^{(2)}) + \hat{\rho}_{\frac{1}{2} \hat{h}^{\alpha_1 \alpha_2}}^{(1)}(\omega^{(2)})) A_{\alpha_1}(\omega_1) A_{\alpha_2}(\omega_2). \end{aligned} \quad (12)$$

Hence, the utility of the notation in Eq. (8) is to split the density matrix at n th order in $A_{\alpha}(t)$ into a sum of density matrices $\hat{\rho}_{\hat{V}_1 \dots \hat{V}_n}^{(n)}$ that are perturbed with relatively simple operators \hat{V}_i . In fact, each of these operators \hat{V}_i is simply a j th \mathbf{k} derivative (for some j) of the bare Hamiltonian $\hat{\mathcal{H}}_{\mathbf{k}}^{(0)}$, which has the physically transparent meaning of a vertex interaction with j photons. (A word of caution: the (n) in $\hat{\rho}_{\hat{V}_1 \dots \hat{V}_n}^{(n)}$ no longer means at n th order in $A_{\alpha}(t)$ (as opposed to the more conventional notation $\hat{\rho}^{(n)}$) and now simply means perturbed with n operators \hat{V}_i , $i = 1, \dots, n$.)

This decomposition is the basis for the connection with the diagrammatic scheme in Ref. [18]. At zeroth order in $A_{\alpha}(t)$, the expectation value of the current is simply a trace of the current operator with respect to the equilibrium density matrix:

$$\langle \hat{j}^{\mu} \rangle_0 = \text{tr} \{ \hat{j}_{(0)}^{\mu} \cdot \hat{\rho}^{(0)} \}. \quad (13)$$

At nonzero order in $A_{\alpha}(t)$, both the density matrix *and* the current operator must be expanded in powers of $A_{\alpha}(t)$:

$$\begin{aligned} \hat{j}^{\mu}(t) &\equiv - \frac{\partial \hat{\mathcal{H}}_{\mathbf{k}}(t)}{\partial A_{\mu}(t)} \\ &= - \sum_{n=0}^{\infty} \frac{e^{n+1}}{n!} \hat{h}^{\mu \alpha_1 \dots \alpha_n} A_{\alpha_1}(t) \dots A_{\alpha_n}(t), \\ \hat{\rho}(t) &= \sum_{n=0}^{\infty} \hat{\rho}^{(n)}(t). \end{aligned} \quad (14)$$

Keeping only terms at some desired order in $A_{\alpha}(t)$, the result is always a sum of traces, where each trace (expressible by a diagram) is acting on a product between some term in Eq. (14) and a density matrix of the form $\hat{\rho}_{\hat{V}_1 \dots \hat{V}_n}^{(n)}$. We express this diagrammatically in terms of Feynman rules illustrated in Fig. 1. This approach can be viewed as a hybrid approach making connection between the density matrix equation of motion method of Refs. [15,16] and the Matsubara Green's function approach of Ref. [18], which lead to the same results. That is, one can alternatively view the calculation of $\text{tr} \{ \hat{\mathcal{O}} \cdot \hat{\rho}_{\hat{V}_1 \dots \hat{V}_n}^{(n)} \}$ for a particular diagram as a trace of a product of Matsubara Green's functions multiplied by vertex operators, summed over a single unconstrained fermionic Matsubara frequency.

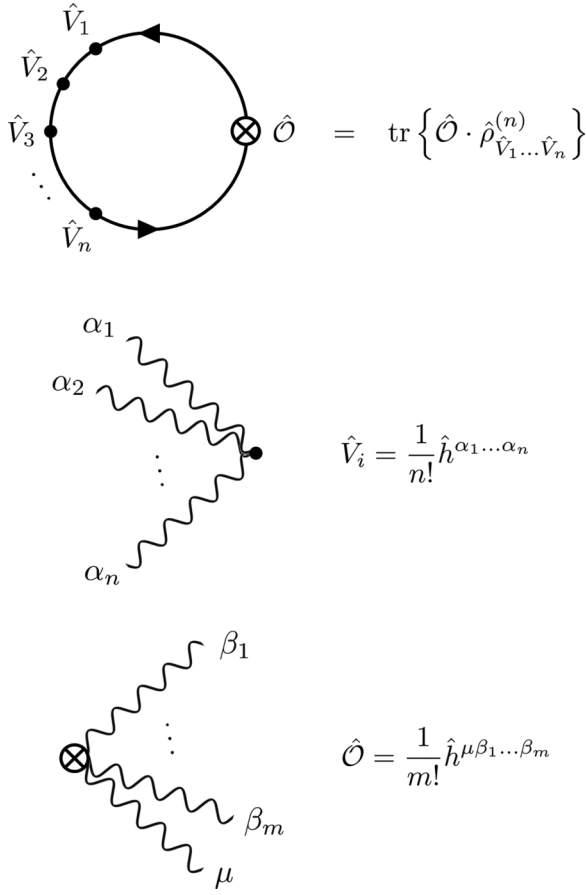


FIG. 1. Summary of Feynman rules employed for computing optical responses in velocity gauge. Following the convention of Ref. [18], the output vertex is marked with a \otimes and represents some output current operator $\hat{\mathcal{O}}$. Measuring this output current amounts to a trace of this operator with a density matrix, which is represented by a fermion loop. This density matrix is perturbed [as in Eq. (8)] with operators \hat{V}_i represented by black vertices. The vertex operators take the form of \mathbf{k} derivatives of the Hamiltonian (denoted $\hat{h}^{\alpha_1 \dots \alpha_n}$) with the directions of the derivatives given by the polarizations (denoted with Greek letters) of the photons entering the vertex. The multiphoton vertex operators are symmetrized with a factor $1/n!$, where n is the number of photons *excluding* the photon representing the output current (we reserve the index μ to indicate the polarization of the output photon).

The reduced density matrix can therefore be alternatively understood as a lesser Green's function $G^<(\mathbf{k}, \omega)$ [19].

The last step is to convert the response in terms of $A_{\alpha_j}(\omega_j)$ to a response in terms of $E_{\alpha_j}(\omega_j) \equiv i\omega_j A_{\alpha_j}(\omega_j)$. We formally define

$$\langle j^\mu(\omega^{(n)}) \rangle = \sigma^{\mu\alpha_1 \dots \alpha_n}(\omega^{(n)}; \omega_1, \dots, \omega_n) E_{\alpha_1}(\omega_1) \dots E_{\alpha_n}(\omega_n). \quad (16)$$

Recalling $\omega^{(n)} \equiv \sum_{j=1}^n \omega_j$, this is the fully general nonlinear conductivity at frequency $\omega^{(n)}$ due to input electric fields at frequencies $\{\omega_j\}_{j=1, \dots, n}$. We calculate this by first calculating

a quantity which we will call $\kappa_{\mathbf{k}}^{\mu\alpha_1 \dots \alpha_n}(\omega^{(n)}; \{\omega_j\})$:

$$\begin{aligned} \sigma^{\mu\alpha_1 \dots \alpha_n}(\omega^{(n)}; \{\omega_j\}) &= -\frac{e^{n+1}}{\hbar^n} \prod_{j=1}^n \left(\frac{1}{i\omega_j} \right) \\ &\times \int [d\mathbf{k}] \kappa_{\mathbf{k}}^{\mu\alpha_1 \dots \alpha_n}(\omega^{(n)}; \{\omega_j\}). \end{aligned} \quad (17)$$

The task is therefore reduced to calculating $\kappa_{\mathbf{k}}^{\mu\alpha_1 \dots \alpha_n}$ via Feynman diagrams. We will do this for the example cases of the linear conductivity and SHG conductivity, and note the cancellations that occur to eliminate the apparent $1/\omega$ and $1/\omega^2$ divergences, respectively.

2. Length gauge

Length gauge is another commonly used gauge in which to calculate susceptibilities to nonlinear order in the electric field [11, 14, 20–23]. In this gauge, one has the perturbation,

$$\hat{V}(t) = e \hat{r}^\alpha E_\alpha(t), \quad (18)$$

where \hat{r}^α denotes the position operator. Care needs to be taken when interpreting the \mathbf{k} -space form of the operator \hat{r}^α [24]. When calculating commutators with other operators, \hat{r}^α acts as a *covariant derivative*,

$$[\hat{r}^\alpha, \hat{\mathcal{O}}]_{ab} \equiv i[\hat{\mathcal{D}}^\alpha, \hat{\mathcal{O}}]_{ab} = i\partial^\alpha \hat{\mathcal{O}}_{ab} + [\hat{\mathcal{A}}^\alpha, \hat{\mathcal{O}}]_{ab}, \quad (19)$$

where $\mathcal{A}_{ab}^\alpha(\mathbf{k}) \equiv \langle u_a(\mathbf{k}) | \partial^\alpha u_b(\mathbf{k}) \rangle$ is the non-Abelian Berry connection. With this definition, nonlinear conductivity tensors in length gauge can be computed directly as (suppressing frequency dependence)

$$\sigma^{\mu\alpha_1 \dots \alpha_n} = -\frac{e^{n+1}}{\hbar^n} \int [d\mathbf{k}] \text{tr} \{ \hat{v}^\mu \cdot \hat{\rho}_{\hat{r}^{\alpha_1} \dots \hat{r}^{\alpha_n}}^{(n)} \}, \quad (20)$$

with

$$\hat{\rho}_{\hat{r}^{\alpha_1} \dots \hat{r}^{\alpha_n}}^{(n)} = \frac{1}{\omega^{(n)} - \hat{\epsilon}} \circ [\hat{r}^{\alpha_n}, \hat{\rho}_{\hat{r}^{\alpha_1} \dots \hat{r}^{\alpha_{n-1}}}^{(n-1)}], \quad (21)$$

without need for extra terms as in the diagrammatic scheme for velocity gauge. While the length gauge yields expressions that are analytically more compact, it is numerically less straightforward to use since \mathbf{k} derivatives act iteratively on the density matrix. In velocity gauge, all the \mathbf{k} derivatives act on the Hamiltonian, so they can be computed ahead of time to speed up calculations. We argue that the diagrammatic treatment makes velocity gauge both analytically concise and numerically expedient, so we focus primarily on velocity gauge when we make our extension to spatially dispersive responses. However, we will always make comparisons to length gauge, and numerically demonstrate the equivalence between the two gauges in each response we consider. Since our analysis makes clear the connection between the density matrix picture (favorable to length gauge) and the diagrammatic picture (favorable to velocity gauge), generating formulas in either gauge will be seamless.

3. Relaxation time approximation

We remark that, in both gauges, we adopt the relaxation time approximation that manifests in a shift of *each* frequency

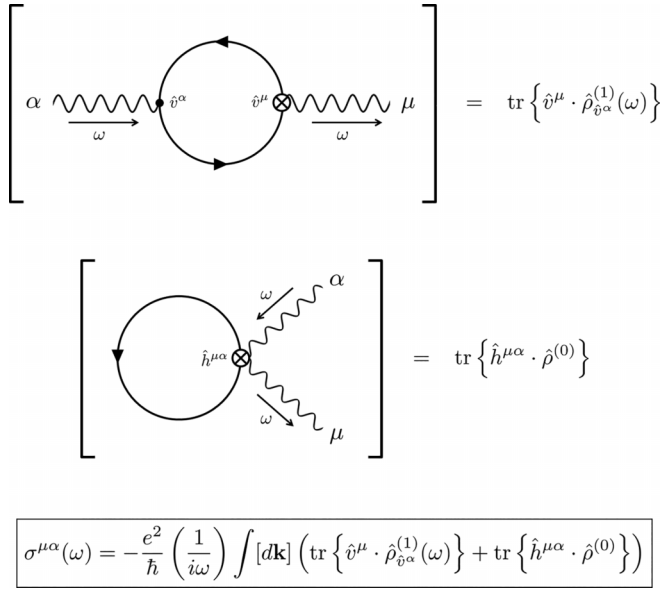


FIG. 2. Diagrams relevant to computing the linear conductivity tensor $\sigma^{\mu\alpha}(\omega)$ in velocity gauge.

$\omega_j \rightarrow \omega_j + i\eta$ for a small relaxation rate η , as discussed in Ref. [16]. Note that this is not equivalent to shifting every pole in a nonlinear conductivity by $i\eta$; for instance, it leads to shifts in second-harmonic resonances of the form $(2\omega + 2i\eta - \epsilon_{ab})^{-1}$. Treating phenomenological relaxation in this way is crucial for getting a properly regularized response at low frequency in velocity gauge.

B. Examples

1. Linear conductivity

The linear conductivity $\sigma^{\mu\alpha}(\omega; \omega) \equiv \sigma^{\mu\alpha}(\omega)$ is defined as follows:

$$\langle \hat{j}^\mu(\omega) \rangle = \sigma^{\mu\alpha}(\omega) E_\alpha(\omega). \quad (22)$$

The calculation in velocity gauge can be written as the sum of two terms, represented diagrammatically in Fig. 2,

$$\begin{aligned} \sigma^{\mu\alpha}(\omega) &= -\frac{e^2}{\hbar} \left(\frac{1}{i\omega} \right) \int [d\mathbf{k}] \kappa_{\mathbf{k}}^{\mu\alpha}(\omega), \\ \kappa_{\mathbf{k}}^{\mu\alpha}(\omega) &= \text{tr} \{ \hat{v}^\mu \cdot \hat{\rho}_{\hat{v}^\alpha}^{(1)}(\omega) \} + \text{tr} \{ \hat{h}^{\mu\alpha} \cdot \hat{\rho}^{(0)} \}, \end{aligned} \quad (23)$$

where each operator within the traces is understood to be evaluated at \mathbf{k} . Note that $\hat{v}^\mu \equiv \hat{h}^\mu$, so we will use these interchangeably. Physically, these two terms correspond to the well-known paramagnetic current and diamagnetic current, respectively. The apparent $1/\omega$ divergence in the conductivity formula is canceled when these two terms are summed and integrated over \mathbf{k} . It is interesting to note that this cancellation *does not occur* in the case of a superconductor, where the low-frequency divergence of the conductivity is physical and results in the linear-in- \mathbf{A} current in the London equation.

Assuming $\rho_{ab}^{(0)} = f_a \delta_{ab}$, we can reexpress the linear conductivity as follows:

$$\sigma^{\mu\alpha}(\omega) = \frac{ie^2}{\hbar\omega} \int [d\mathbf{k}] \left(\sum_{ab} f_{ab} \frac{v_{ab}^\mu v_{ba}^\alpha}{\omega - \epsilon_{ba}} + \sum_a f_a h_{aa}^{\mu\alpha} \right), \quad (24)$$

which matches velocity gauge results from the literature [16].

In length gauge, recalling that $[\hat{r}^\alpha, \hat{O}]_{ab} = i\partial^\alpha \mathcal{O}_{ab} + [\hat{A}^\alpha, \hat{O}]_{ab}$, where \hat{A}^α is the non-Abelian Berry connection, we write

$$\begin{aligned} \sigma_{\text{len.}}^{\mu\alpha}(\omega) &= -\frac{e^2}{\hbar} \int [d\mathbf{k}] \text{tr} \{ \hat{v}^\mu \cdot \hat{\rho}_{\hat{r}^\alpha}^{(1)}(\omega) \} \\ &= -\frac{e^2}{\hbar} \int [d\mathbf{k}] \left(\sum_{ab} f_{ab} \frac{v_{ab}^\mu \mathcal{A}_{ba}^\alpha}{\omega - \epsilon_{ba}} + \sum_a \frac{iv_{aa}^\mu \partial^\alpha f_a}{\omega} \right). \end{aligned} \quad (25)$$

Using $v_{ab}^\alpha = i\epsilon_{ab} \mathcal{A}_{ab}^\alpha$ for $a \neq b$, this allows us to write

$$\sigma_{\text{len.}}^{\mu\alpha}(\omega) = \frac{ie^2}{\hbar} \int [d\mathbf{k}] \left(\sum_{ab} f_{ab} \frac{v_{ab}^\mu v_{ba}^\alpha}{\epsilon_{ba}(\omega - \epsilon_{ba})} - \sum_a \frac{v_{aa}^\mu \partial^\alpha f_a}{\omega} \right). \quad (26)$$

2. Second-harmonic generation conductivity

The SHG conductivity $\sigma^{\mu\alpha\beta}(2\omega; \omega, \omega) \equiv \sigma_{\text{SHG}}^{\mu\alpha\beta}(\omega)$ is defined as follows:

$$\langle \hat{j}^\mu(2\omega) \rangle = \sigma_{\text{SHG}}^{\mu\alpha\beta}(\omega) E_\alpha(\omega) E_\beta(\omega). \quad (27)$$

The calculation in velocity gauge can be written as the sum of four terms, represented diagrammatically in Fig. 3:

$$\begin{aligned} \sigma_{\text{SHG}}^{\mu\alpha\beta}(\omega) &= -\frac{e^3}{\hbar^2} \left(\frac{-1}{\omega^2} \right) \int [d\mathbf{k}] \kappa_{\text{SHG},\mathbf{k}}^{\mu\alpha\beta}(\omega), \\ \kappa_{\text{SHG},\mathbf{k}}^{\mu\alpha\beta}(\omega) &= \text{tr} \{ \hat{v}^\mu \cdot \hat{\rho}_{\hat{v}^\alpha, \hat{v}^\beta}^{(2)}(2\omega) \} + \text{tr} \{ \hat{v}^\mu \cdot \hat{\rho}_{\frac{1}{2}\hat{h}^{\mu\alpha}}^{(1)}(2\omega) \} \\ &\quad + \text{tr} \{ \hat{h}^{\mu\alpha} \cdot \hat{\rho}_{\hat{v}^\beta}^{(1)}(\omega) \} + \text{tr} \left\{ \frac{1}{2} \hat{h}^{\mu\alpha\beta} \cdot \hat{\rho}^{(0)} \right\}. \end{aligned} \quad (28)$$

Assuming $\rho_{ab}^{(0)} = f_a \delta_{ab}$, this becomes

$$\begin{aligned} \kappa_{\text{SHG},\mathbf{k}}^{\mu\alpha\beta}(\omega) &= \sum_{a,b,c} v_{ab}^\mu \frac{v_{bc}^\beta \frac{v_{ca}^\alpha f_{ac}}{\omega - \epsilon_{ca}} - \frac{f_{cb} v_{bc}^\alpha}{\omega - \epsilon_{bc}} v_{ca}^\beta}{2\omega - \epsilon_{ba}} \\ &\quad + \sum_{a,b} v_{ab}^\mu \frac{\frac{1}{2} h_{ba}^{\alpha\beta} f_{ab}}{2\omega - \epsilon_{ab}} + \sum_{a,b} h_{ab}^{\mu\alpha} \frac{v_{ba}^\beta f_{ab}}{\omega - \epsilon_{ba}} \\ &\quad + \sum_a \frac{1}{2} h_{aa}^{\mu\alpha\beta} f_a. \end{aligned} \quad (29)$$

We note that, under \mathbf{k} -space symmetry considerations alone, the contributions from the first and last diagrams in Fig. 3 should vanish under \mathbf{k} -space integration over the full Brillouin zone (that is, their integrands are odd in \mathbf{k} while the integration region is symmetric in \mathbf{k}). Therefore, the content of the cancellation required to eliminate the $1/\omega^2$ prefactor in the conductivity comes solely from the second and third diagrams, just as the cancellation in the linear conductivity comes from only two diagrams.

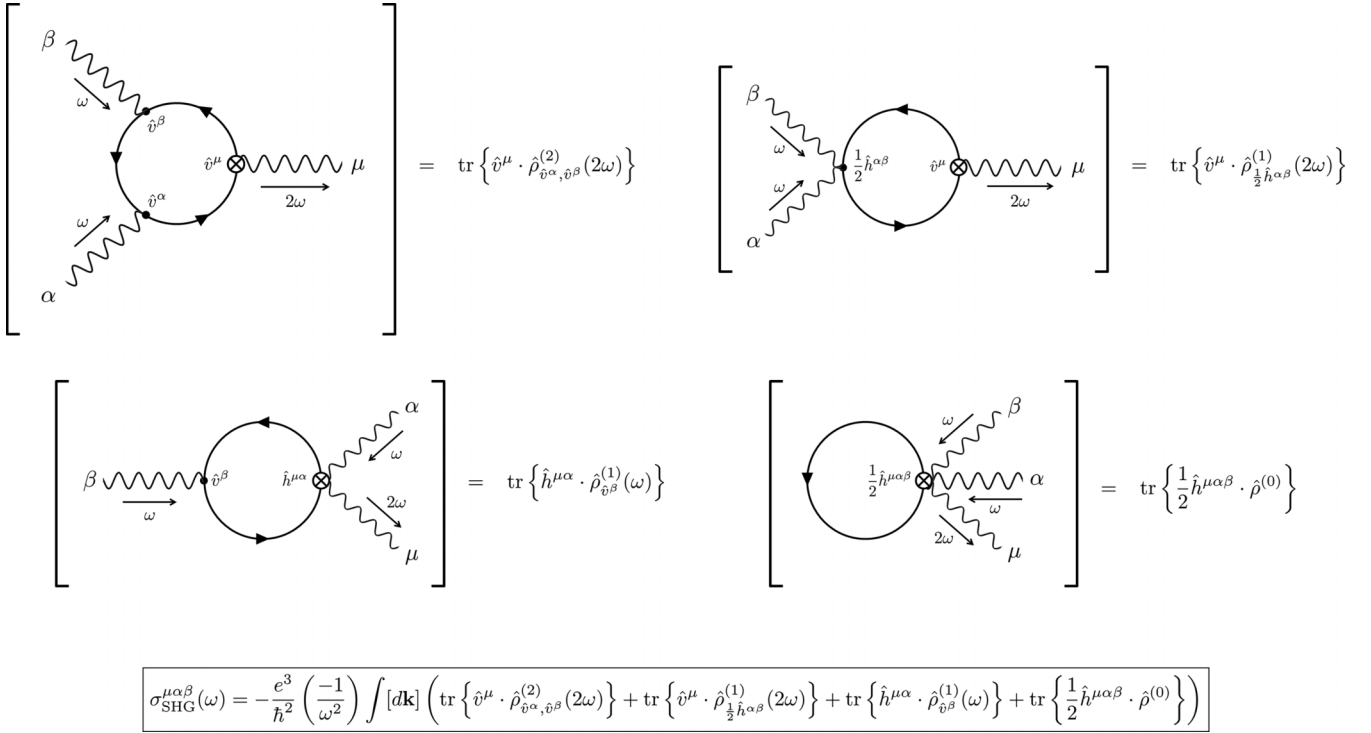


FIG. 3. Diagrams relevant to computing the SHG conductivity tensor $\sigma_{\text{SHG}}^{\mu\alpha\beta}(\omega)$ in velocity gauge.

III. CALCULATIONS OF SPATIALLY DISPERSIVE NONLINEAR OPTICAL RESPONSES

A. Generalities

A nonlinear optical conductivity tensor is in full generality a function of the wave vectors of the input/output optical fields \mathbf{q}_i , but is typically treated in the $\mathbf{q} \rightarrow 0$ limit. This is justified within the so-called dipole approximation, in which it is assumed that the length scale of the electronic states being coupled by optical transitions (for instance, the size of electronic orbitals) is very small compared to the wavelength of the field driving the transition. However, there are many cases where this approximation should be expected to fail (for instance, in band systems whose Wannier representation has poor localization). For these reasons, one may be interested in the lowest-order contribution of the spatial variation of the optical fields:

$$E_{\alpha_j}(\mathbf{r}, t) = E_{\alpha_j}(\mathbf{0}, t)(1 - i\mathbf{q}_j \cdot \hat{\mathbf{r}}) + \mathcal{O}(\mathbf{q}^2). \quad (30)$$

We can include \mathbf{q} in calculations of nonlinear optical responses by allowing photons to transfer momentum when coupling electronic states, which can be straightforwardly understood in the diagrammatic Green's function scheme written in terms of Green's functions [19]. However, it is helpful to have a physically transparent heuristic for treating just the lowest-order spatially dispersive correction, which has an interpretation in terms of beyond-electric-dipole light-matter coupling. Previous works [25,26] have identified how to expand the linear optical conductivity to first order in \mathbf{q} . Interpreting the zeroth-order contribution to the current operator as the time-derivative of the electric dipole, $\hat{\mathbf{P}}^\mu = -e\hat{\mathbf{r}}^\mu$, the first-order terms include contributions from the electric

quadrupole and magnetic dipole [27],

$$\hat{\mathbf{j}}^\mu = \partial_t \hat{\mathbf{P}}^\mu - iq_v \left(\frac{1}{2} \partial_t \hat{\mathbf{Q}}^{\nu\mu} + c\epsilon^{\nu\mu\rho} \hat{\mathbf{M}}_\rho \right) + \mathcal{O}(\mathbf{q}^2), \quad (31)$$

where $\hat{\mathbf{Q}}^{\nu\mu} = -e\hat{\mathbf{r}}^\nu \hat{\mathbf{r}}^\mu$ is the electric quadrupole operator and $\hat{\mathbf{M}}_\rho = -\frac{e}{2c} \epsilon_{\rho\alpha\beta} \hat{\mathbf{r}}^\alpha \hat{\mathbf{r}}^\beta$ is the magnetic dipole operator. Equation (31) constitutes the starting point for a velocity gauge form for light-matter coupling up to first order in \mathbf{q} . In this paper, we will focus on the electric quadrupole contribution and determine the velocity gauge and length gauge expressions for treating this contribution. In length gauge, the perturbation takes the form

$$\hat{V}(\omega) = \left(\hat{\mathbf{r}}^\alpha - iq_v \left(\frac{1}{2} \hat{\mathbf{Q}}^{\nu\alpha} + \frac{c}{\omega} \epsilon^{\nu\alpha\rho} \hat{\mathbf{M}}_\rho \right) \right) E_\alpha(\omega). \quad (32)$$

We constrain the exact form of the electric quadrupole operator by ensuring that we recover $-\frac{\partial}{\partial r^\alpha} V(\omega) = E_\alpha(\omega)(1 - iq_v r^\nu) + \mathcal{O}(\mathbf{q}^2)$. Notably, this gives a relative factor of 2 to the off-diagonal spatial components of $\hat{\mathbf{Q}}^{\nu\mu}$. For example, in $d = 2$, we have

$$\hat{\mathbf{Q}}^{\nu\mu} = \begin{pmatrix} \hat{\mathbf{Q}}^{xx} & \hat{\mathbf{Q}}^{xy} \\ \hat{\mathbf{Q}}^{yx} & \hat{\mathbf{Q}}^{yy} \end{pmatrix} = \begin{pmatrix} \hat{\mathcal{A}}^x \cdot \hat{\mathcal{A}}^x & 2\hat{\mathcal{A}}^x \cdot \hat{\mathcal{A}}^y \\ 2\hat{\mathcal{A}}^y \cdot \hat{\mathcal{A}}^x & \hat{\mathcal{A}}^y \cdot \hat{\mathcal{A}}^y \end{pmatrix}, \quad (33)$$

where we recall the non-Abelian Berry connection $\mathcal{A}_{ab}^\mu \equiv \langle u_{a\mathbf{k}} | \partial_t^\mu u_{b\mathbf{k}} \rangle$.

It has been shown [25] that the first-order-in- \mathbf{q} contribution to the current operator matrix element can be cast in terms of the following object:

$$\begin{aligned} \langle u_{n,\mathbf{k}+\frac{\mathbf{q}}{2}} | \hat{\mathbf{v}}_{\mathbf{k}}^\mu | u_{m,\mathbf{k}-\frac{\mathbf{q}}{2}} \rangle &= -\frac{i}{2} q_v \{ \hat{\mathcal{A}}_{\mathbf{k}}^v, \hat{\mathbf{v}}_{\mathbf{k}}^\mu \}_{nm} \\ &\equiv -\frac{i}{2} q_v (\partial_t \hat{\mathbf{Q}}_{\mathbf{k}}^{\nu\mu})_{nm}, \end{aligned} \quad (34)$$

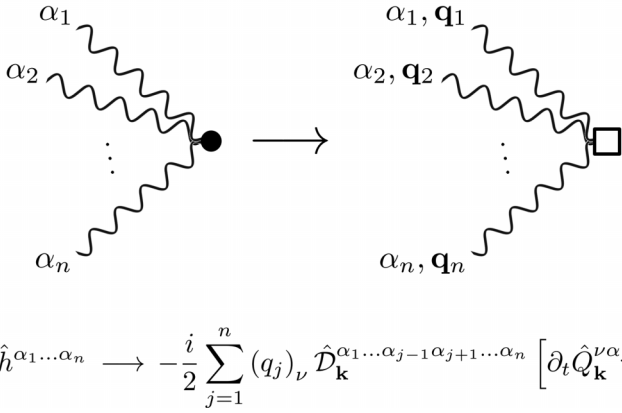


FIG. 4. Rule for determining the first-order-in- \mathbf{q} contribution from a vertex operator. We change a vertex from a dot to a square to indicate that it is being perturbed to first order in \mathbf{q} .

where in the last line we make the crucial identification that this should equal the time derivative of the electric quadrupole moment. This is the basis for the vertex rule we introduce, illustrated in Fig. 4 and summarized in the following expression for the velocity-gauge perturbation:

$$\hat{V}_{\mathbf{k}}(t) = \sum_{n=1}^{\infty} \frac{e^n}{n!} A_{\alpha_1}(t) \dots A_{\alpha_n}(t) \left(\hat{D}_{\mathbf{k}}^{\alpha_1 \dots \alpha_n} [\hat{\mathcal{H}}_{\mathbf{k}}] + \frac{1}{2} \sum_{j=1}^n (q_j)_\nu \hat{D}_{\mathbf{k}}^{\alpha_1 \dots \alpha_{j-1} \alpha_{j+1} \dots \alpha_n} [\hat{\mathcal{H}}_{\mathbf{k}}, \hat{Q}_{\mathbf{k}}^{\nu \alpha_j}] \right), \quad (35)$$

where the vectors \mathbf{q}_j are the wave vectors from the n fields $A_{\alpha_1}(t), \dots, A_{\alpha_n}(t)$. We can interpret this as introducing an additional type of vertex in the diagrammatic method, one that takes into account at first order the momentum ($\hbar \mathbf{q}_j$) of each photon connected to that vertex. We will use the terms momentum and wave vector interchangeably.

Schematically, this leads to a spatially dispersive correction to nonlinear optical conductivities,

$$\sigma^{\mu \alpha_1 \dots \alpha_n}(\mathbf{q}) = \sigma_{(0)}^{\mu \alpha_1 \dots \alpha_n} + q_\nu \sigma_{(1)}^{\nu \mu \alpha_1 \dots \alpha_n} + \mathcal{O}(\mathbf{q}^2), \quad (36)$$

where \mathbf{q} is some characteristic momentum of the input fields. To declutter notation, we have suppressed the ω dependence, which is understood to be $(\omega_{(n)}; \omega_1 \dots \omega_n)$ in general. As before, in velocity gauge, we write this as

$$\sigma_{(1)}^{\nu \mu \alpha_1 \dots \alpha_n} = -\frac{e^{n+1}}{\hbar^n} \prod_{j=1}^n \left(\frac{1}{i\omega_j} \right) \int [d\mathbf{k}] \kappa_{(1), \mathbf{k}}^{\nu \mu \alpha_1 \dots \alpha_n} \quad (37)$$

and calculate $\kappa_{(1), \mathbf{k}}^{\nu \mu \alpha_1 \dots \alpha_n}$ using a diagrammatic scheme. As a consequence of Eq. (35), this diagrammatic scheme amounts to taking each diagram for the $\mathcal{O}(\mathbf{q}^0)$ response and, for each vertex $\hat{h}^{\alpha_1 \dots \alpha_n}$, creating a diagram with the following replacement rule (dropping \mathbf{k} dependence for brevity),

$$\begin{aligned} \hat{h}^{\alpha_1 \dots \alpha_n} &\rightarrow \frac{1}{2} \sum_{j=1}^n (q_j)_\nu \hat{D}_{\mathbf{k}}^{\alpha_1 \dots \alpha_{j-1} \alpha_{j+1} \dots \alpha_n} [\hat{\mathcal{H}}, \hat{Q}_{\mathbf{k}}^{\nu \alpha_j}] \\ &\equiv -\frac{i}{2} \sum_{j=1}^n (q_j)_\nu \partial_t \hat{Q}_{\mathbf{k}}^{\nu \alpha_j; \alpha_1 \dots \alpha_{j-1} \alpha_{j+1} \dots \alpha_n}, \end{aligned} \quad (38)$$

where $(q_j)_\nu$ is the ν component of the wave vector for the photon with polarization α_j (again, summation over the repeated index ν is implied). In the second line, we introduce a compact notation, whereby indices after the semicolon indicate covariant \mathbf{k} derivatives. To make sense of Eq. (38), note that for a single-photon vertex, the replacement rule reduces to

$$\hat{h}^{\alpha_1} \rightarrow -\frac{i}{2} (q_1)_\nu \partial_t \hat{Q}_{\mathbf{k}}^{\nu \alpha_1}. \quad (39)$$

In other words, it replaces an *electric dipole* coupling $\hat{h}^{\alpha_1} \equiv \hat{v}^{\alpha_1} \equiv \partial_t \hat{r}^{\alpha_1}$ with an *electric quadrupole* coupling $\partial_t \hat{Q}_{\mathbf{k}}^{\nu \alpha_1}$. The extension to multiphoton vertices can then be understood as a sum of terms that change the coupling from dipolar to quadrupolar for *one photon at a time* while having the rest of the photons contribute covariant \mathbf{k} derivatives of the resulting quadrupolar current operator. We illustrate this vertex rule in Fig. 4.

We note that the output vertex should also be perturbed to linear-in- \mathbf{q} order in this scheme, and we denote this by a \boxtimes symbol in the diagrams. As in the spatially uniform case, the only special treatment that must be given to the output vertex is that the symmetry factor $1/n!$ should only include the n input photons (i.e., excluding the output photon, for whom we reserve the index μ to denote its polarization).

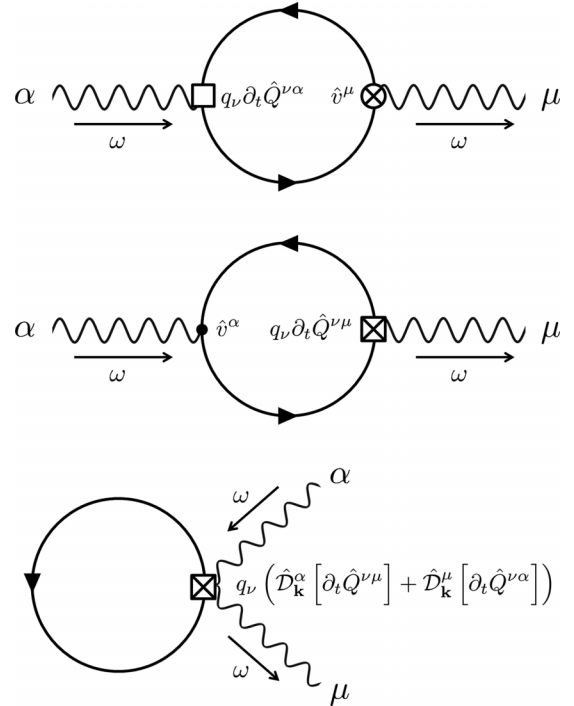


FIG. 5. Diagrams relevant to computing the *spatially dispersive* linear conductivity tensor to lowest order $[\sigma_{(1)}^{\nu \mu \alpha}(\omega)]$ in velocity gauge. The third diagram is shown for completeness, but has a vanishing contribution since it only contains a vertex with net-zero momentum flow. (We omit a common factor of $-\frac{i}{2}$ from each diagram.)

B. Examples

1. Spatially dispersive linear conductivity

Applying the Feynman rules in Fig. 4 to the linear conductivity diagrams in Fig. 2, we obtain the diagrams in Fig. 5. These diagrams represent the following two terms:

$$\begin{aligned} \kappa_{(1),\mathbf{k}}^{v\mu\alpha}(\omega) = & -\frac{i}{2}(\text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\partial_t \hat{Q}^{v\alpha}}^{(1)}(\omega)\} + \text{tr}\{\partial_t \hat{Q}^{v\mu} \cdot \hat{\rho}_{\hat{v}^\alpha}^{(1)}(\omega)\}) \\ & + \text{tr}\{(\partial_t \hat{Q}^{v\mu;\alpha} + \partial_t \hat{Q}^{v\alpha;\mu}) \cdot \hat{\rho}^{(0)}\}. \end{aligned} \quad (40)$$

The conductivity can therefore be written as

$$\begin{aligned} \sigma_{(1)}^{v\mu\alpha}(\omega) = & \frac{e^2}{2\hbar\omega} \int [d\mathbf{k}] \left\{ \sum_{ab} f_{ab} \frac{v_{ab}^\mu \dot{Q}_{ba}^{v\alpha} + \dot{Q}_{ab}^{v\mu} v_{ba}^\alpha}{\omega - \epsilon_{ba}} \right. \\ & \left. + \sum_a f_a (\dot{Q}_{aa}^{v\mu;\alpha} + \dot{Q}_{aa}^{v\alpha;\mu}) \right\}, \end{aligned} \quad (41)$$

where we abbreviate $\langle u_{a\mathbf{k}} | \partial_t \hat{Q}^{v\mu} | u_{b\mathbf{k}} \rangle \equiv \dot{Q}_{ab}^{v\mu} \equiv i[\hat{\mathcal{H}}, \hat{Q}^{v\mu}]_{ab}$.

$$\begin{aligned} \kappa_{\text{SHG},(1),\mathbf{k}}^{v\mu\alpha\beta}(\omega) = & -\frac{i}{2} \left[\text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\hat{v}^\alpha, \partial_t \hat{Q}^{v\beta}}^{(2)}(2\omega)\} + \text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\partial_t \hat{Q}^{v\alpha}, \hat{v}^\beta}^{(2)}(2\omega)\} + \text{tr}\{\partial_t \hat{Q}^{v\mu} \cdot \hat{\rho}_{\hat{v}^\alpha, \hat{v}^\beta}^{(2)}(2\omega)\} + \text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\frac{1}{2}\partial_t \hat{Q}^{v\alpha;\beta} + \frac{1}{2}\partial_t \hat{Q}^{v\beta;\alpha}}^{(1)}(2\omega)\} \right. \\ & + \text{tr}\{\partial_t \hat{Q}^{v\mu} \cdot \hat{\rho}_{\frac{1}{2}\hat{h}^{\alpha\beta}}^{(1)}(2\omega)\} + \text{tr}\{\hat{h}^{\mu\alpha} \cdot \hat{\rho}_{\partial_t \hat{Q}^{v\beta}}^{(1)}(\omega)\} + \text{tr}\{(\partial_t \hat{Q}^{v\mu;\alpha} + \partial_t \hat{Q}^{v\alpha;\mu}) \cdot \hat{\rho}_{\hat{v}^\beta}^{(1)}(\omega)\} \\ & \left. + \text{tr}\left\{\frac{1}{2}(\partial_t \hat{Q}^{v\mu;\alpha\beta} + [\text{perms.}]) \cdot \hat{\rho}^{(0)}\right\} \right], \end{aligned} \quad (43)$$

where [perms.] indicates the two other permutations of μ, α, β modulo exchanges of the indices after the semicolon. Assuming $\rho_{ab}^{(0)} = f_a \delta_{ab}$, this becomes

$$\begin{aligned} \kappa_{\text{SHG},(1),\mathbf{k}}^{v\mu\alpha\beta}(\omega) = & -\frac{i}{2} \sum_{a,b} \left(v_{ab}^\mu \frac{N_{1,ba}^{v\alpha\beta}(\omega)}{2\omega - \epsilon_{ba}} - \dot{Q}_{ab}^{v\mu} \frac{N_{2,ba}^{\alpha\beta}(\omega)}{2\omega - \epsilon_{ba}} + \frac{1}{2} v_{ab}^\mu \frac{(\dot{Q}_{ba}^{v\alpha;\beta} + \dot{Q}_{ba}^{v\beta;\alpha}) f_{ab}}{2\omega - \epsilon_{ba}} + \frac{1}{2} \dot{Q}_{ab}^{v\mu} \frac{h_{ba}^{\alpha\beta} f_{ab}}{2\omega - \epsilon_{ba}} + h_{ab}^{\mu\alpha} \frac{\dot{Q}_{ba}^{v\beta} f_{ab}}{\omega - \epsilon_{ba}} \right. \\ & \left. + (\dot{Q}_{ab}^{v\mu;\alpha} + \dot{Q}_{ab}^{v\alpha;\mu}) \frac{v_{ba}^\beta f_{ab}}{\omega - \epsilon_{ba}} \right) - \frac{i}{4} \sum_a (\dot{Q}_{aa}^{v\mu;\alpha\beta} + \dot{Q}_{aa}^{v\alpha;\beta\mu} + \dot{Q}_{aa}^{v\beta;\mu\alpha}) f_a, \end{aligned} \quad (44)$$

where

$$\begin{aligned} N_{1,ba}^{v\alpha\beta}(\omega) = & \sum_c \left(\frac{v_{bc}^{v\beta} v_{ca}^\alpha f_{ac}}{\omega - \epsilon_{ca}} - \frac{f_{cb} v_{bc}^\alpha \dot{Q}_{ca}^{v\beta}}{\omega - \epsilon_{bc}} \right. \\ & \left. + v_{bc}^\beta \frac{\dot{Q}_{ca}^{v\alpha} f_{ac}}{\omega - \epsilon_{ca}} - \frac{f_{cb} \dot{Q}_{bc}^{v\alpha} v_{ca}^\beta}{\omega - \epsilon_{bc}} \right), \\ N_{2,ba}^{\alpha\beta}(\omega) = & \sum_c \left(v_{bc}^\beta \frac{v_{ca}^\alpha f_{ac}}{\omega - \epsilon_{ca}} - \frac{f_{cb} v_{bc}^\alpha v_{ca}^\beta}{\omega - \epsilon_{bc}} \right). \end{aligned} \quad (45)$$

In our forthcoming results, we compare to the following length gauge formula, obtained by starting from the usual length gauge formulation of SHG and applying Eqs. (32)–(34):

$$\begin{aligned} \sigma_{(1)}^{v\mu\alpha\beta}(\omega) = & \frac{ie^3}{2\hbar^2} \int [d\mathbf{k}] (\text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\hat{v}^\alpha, \hat{v}^\beta}^{(2)}\} + \text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\hat{v}^\alpha, \partial_t \hat{Q}^{v\beta}}^{(2)}\}) \\ & + \text{tr}\{\partial_t \hat{Q}^{v\mu} \cdot \hat{\rho}_{\hat{v}^\alpha, \hat{v}^\beta}^{(2)}\}. \end{aligned} \quad (46)$$

Despite the apparent divergence from the prefactor $1/\omega$, this response is regularized in the low-frequency limit, as we show numerically for an example model in Sec. IV. To further justify this, we can rederive this result in length gauge, using Eqs. (32)–(34):

$$\begin{aligned} \sigma_{(1)}^{v\mu\alpha}(\omega) = & \frac{ie^2}{2\hbar} \int [d\mathbf{k}] (\text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\hat{v}^\alpha}^{(1)}(\omega)\} \\ & + \text{tr}\{\partial_t \hat{Q}^{v\mu} \cdot \hat{\rho}_{\hat{v}^\alpha}^{(1)}(\omega)\}). \end{aligned} \quad (42)$$

We show numerically in Sec. IV that Eqs. (41) and (42) are equivalent.

2. Spatially dispersive second-harmonic generation conductivity

Figure 6 displays the eight diagrams for calculating the linear-in- \mathbf{q} SHG conductivity. These yield the following seven terms:

IV. RESULTS FOR MINIMAL MODELS

In this section, we test our scheme on a minimal tight-binding model with three useful features: (1) inversion symmetry, (2) a quasi-2D structure allowing for multiple different polarizations, and (3) a clear decoupling limit in which the model reduces to the well-understood problem of multipole optical responses in a molecule [27–29]. We will confirm that, up to electric quadrupole order, a nonzero SHG response indeed exists in this centrosymmetric system, and we can therefore treat \mathbf{q} on the same footing as a static inversion-breaking parameter in this model.

A. Rectangular molecule

We first establish the response up to first order in \mathbf{q} and up to second order in the field strength for a four-site rectangular molecule with only nearest-neighbor hoppings between the sites. Here, the position operators are unambiguous up to a choice of origin, and so one can straightforwardly describe

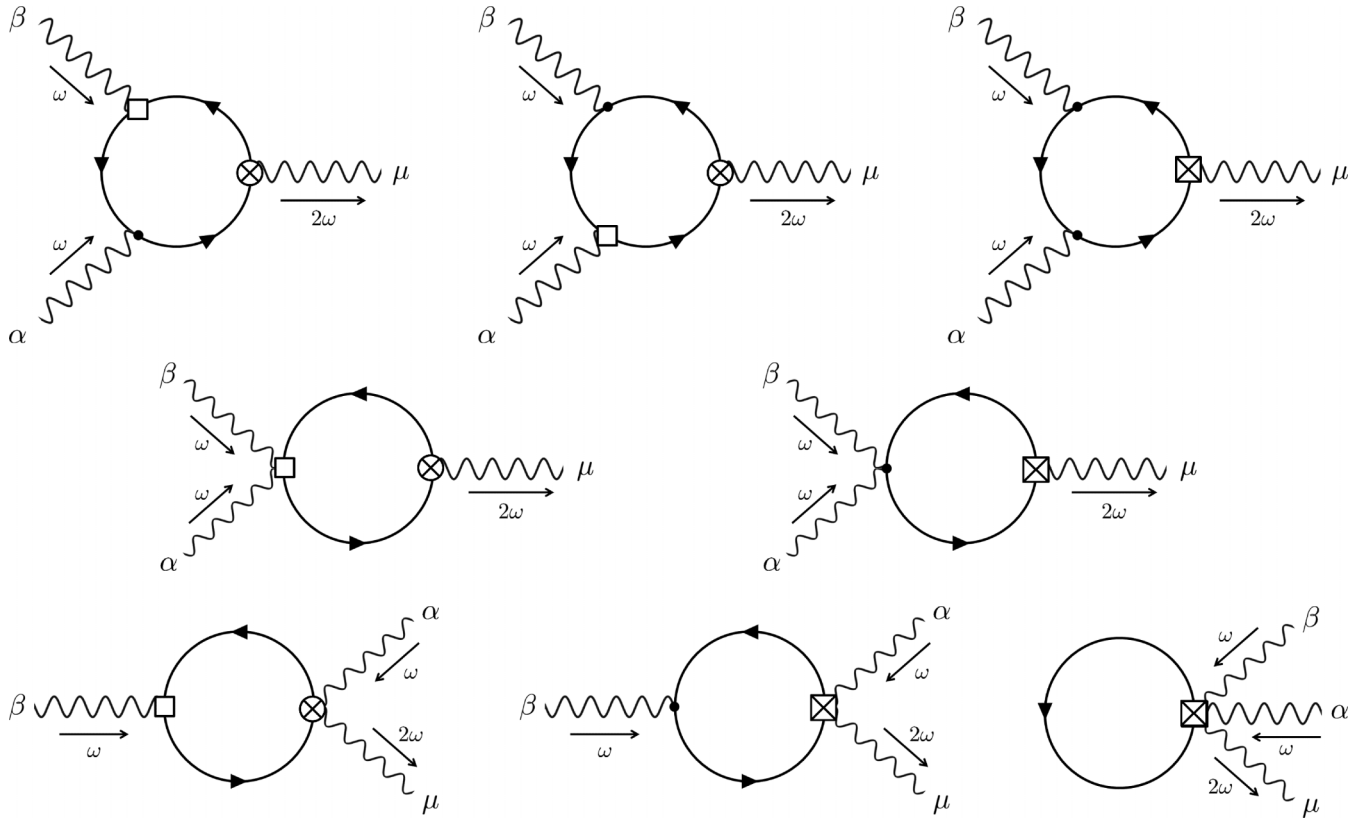


FIG. 6. Diagrams for computing lowest-order *spatially dispersive corrections* to the SHG conductivity tensor $[\sigma_{\text{SHG},(1)}^{v\mu\alpha\beta}(\omega)]$ in velocity gauge.

the problem in length gauge. We write the Hamiltonian as

$$\hat{\mathcal{H}} = - \begin{pmatrix} 0 & t_x & 0 & t_y \\ t_x & 0 & t_y & 0 \\ 0 & t_y & 0 & t_x \\ t_y & 0 & t_x & 0 \end{pmatrix}. \quad (47)$$

Assume t_x and t_y are both positive and, without loss of generality, assume $t_x > t_y$. We use a basis in which the position operators \hat{r}^x and \hat{r}^y take the form

$$\hat{r}^x = R_x \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad \hat{r}^y = R_y \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (48)$$

where R_x and R_y denote half the length of the molecule in the x and y directions, respectively. Note that the quadrupole operators $\hat{r}^x \hat{r}^x$ and $\hat{r}^y \hat{r}^y$ are proportional to the identity matrix, so they cannot mediate any nonzero responses in the density matrix equation of motion. On the other hand, $\hat{r}^x \hat{r}^y \equiv \hat{r}^y \hat{r}^x = R_x R_y \text{diag}(-1, 1, -1, 1)$, which *can* mediate a nonzero response. So one should only expect a response when the electric field is perpendicular to the incident \mathbf{q} vector, which is sensible.

This Hamiltonian is diagonalized by a unitary transformation such as

$$\hat{U} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}, \quad (49)$$

yielding levels with energies $-t_x - t_y, -t_x + t_y, t_x - t_y, t_x + t_y$, in ascending order. We calculate the electric quadrupole SHG response using

$$\begin{aligned} \sigma_{\text{SHG},(1)}^{v\mu\alpha\beta}(\omega) &= \frac{ie^3}{2\hbar^2} \left(\text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\hat{Q}^{v\alpha}, \hat{r}^\beta}^{(2)}(2\omega)\} \right. \\ &\quad + \text{tr}\{\hat{v}^\mu \cdot \hat{\rho}_{\hat{r}^\alpha, \hat{Q}^{v\beta}}^{(2)}(2\omega)\} \\ &\quad \left. + \text{tr}\{\partial_t \hat{Q}^{v\mu} \cdot \hat{\rho}_{\hat{r}^\alpha, \hat{r}^\beta}^{(2)}(\omega)\} \right), \quad (50) \end{aligned}$$

starting from an initial density matrix of $\hat{\rho}^{(0)} = \text{diag}(f_1, f_2, f_3, f_4)$. We find

$$\sigma_{\text{SHG},(1)}^{xxyy}(\omega) = -\frac{4ie^3}{\hbar^2} \frac{(f_{12} - f_{34})t_x t_y \omega R_x^2 R_y^2}{(\omega^2 - t_x^2)(\omega^2 - (2t_y)^2)}, \quad (51)$$

$$\sigma_{\text{SHG},(1)}^{yyxx}(\omega) = -\frac{4ie^3}{\hbar^2} \frac{(f_{12} - f_{34})t_x t_y \omega R_x^2 R_y^2}{(\omega^2 - (2t_x)^2)(\omega^2 - t_y^2)}. \quad (52)$$

This result reveals a number of interesting features. First, it is an example of a nonzero second-harmonic response in an inversion-symmetric system, where the inversion breaking

is allowed by the nonzero order in \mathbf{q} . Second, the response intriguingly vanishes for half-filling, owing to a symmetry between the Bloch states in the occupied and unoccupied energy levels. We note that introducing diagonal hoppings breaks this symmetry, and therefore allows a nonzero response at half filling, at the expense of the energetic factors in Eqs. (51) and (52) becoming considerably more complicated. In other words, the response takes the form

$$\sigma_{\text{SHG},(1)}^{xyy}(\omega) = -\frac{ie^3}{\hbar^2} F(\omega) R_x^2 R_y^2, \quad (53)$$

where $F(\omega)$ only depends on ω , the hopping amplitudes, and the Fermi factors.

Appealingly, $\sigma_{\text{SHG},(1)}^{v\mu\alpha\beta}(\omega)$ can be seen to depend on four factors of the length of a molecule, in the directions specified by \mathbf{q} , \mathbf{j} , and the two \mathbf{E} fields. Therefore, a larger molecule exhibits a larger magnitude for this spatially dispersive correction. The size of the molecule can be considered as a proxy for the coherence length of electronic states. Crucially, in the case of obstructed band systems forcing delocalized Wannier states, this characteristic coherence length may be on the order of many unit cells, and therefore one should expect substantial spatially dispersive corrections from these systems.

B. Chain of rectangular molecules

Transitioning to a band theory, we study a quasi-1D model in which two copies of the Rice-Mele model form the legs of a ladder. We define the Hamiltonian as

$$\hat{\mathcal{H}}_k^o = - \begin{pmatrix} \Delta & g_o(k) & 0 & t_y \\ g_o^*(k) & -\Delta & t_y & 0 \\ 0 & t_y & -\Delta & g_o^*(k) \\ t_y & 0 & g_o(k) & \Delta \end{pmatrix}, \quad (54)$$

$$g_o(k) = t_x + t'_x e^{-ik}. \quad (55)$$

This Hamiltonian reduces to the molecule Hamiltonian in Eq. (47) in the limit $t'_x \rightarrow 0$ (and $\Delta \rightarrow 0$). The purpose of the parameter Δ (which breaks inversion if $t_x \neq t'_x$) is to allow calculation of the spatially dispersive correction to the linear conductivity (which vanishes under inversion symmetry). Here, we assume translational symmetry in the x direction, with crystal momentum $k \in [0, 2\pi)$.

We have written the Hamiltonian in what we will call orbital form (hence the subscript o), in which it has the desirable property of being periodic with respect to k -space translations by a reciprocal lattice vector:

$$\hat{\mathcal{H}}_k^o = \hat{\mathcal{H}}_{k+2\pi}^o. \quad (56)$$

However, this comes at the cost of neglecting Bloch phases $e^{ik\Delta x}$ for *intracell* hoppings, and hence the normal Peierls substitution $k \rightarrow k + eA(t)$ with this Hamiltonian will neglect intracell currents. We will call this form of the Hamiltonian the orbital form, (hence the superscript o) since it may be viewed as treating all states within the unit cell as orbitals on a single generalized site in the unit cell. There are two equivalent options for dealing with this. One is to define the position operator with an intercell and intracell part,

$$\hat{r}_i \rightarrow \hat{r}_i + \hat{\tau}, \quad (57)$$

and to explicitly add the intracell part of the velocity operator $\hat{v}_{\text{intra}} \equiv i[\hat{\mathcal{H}}, \hat{\tau}]$ when calculating currents. Alternatively, one can define a form of the k -space Hamiltonian such that the Bloch phases *do* accurately reflect intracell processes,

$$\hat{\mathcal{H}}_k^s = - \begin{pmatrix} \Delta & g_s(k) & 0 & t_y \\ g_s^*(k) & -\Delta & t_y & 0 \\ 0 & t_y & -\Delta & g_s^*(k) \\ t_y & 0 & g_s(k) & \Delta \end{pmatrix}, \quad (58)$$

$$g_s(k) = t_x e^{ik/2} + t'_x e^{-ik/2}, \quad (59)$$

and find the unitary transformation $\hat{U}_k^{s \rightarrow o}$ that takes this Hamiltonian back to the orbital form

$$\hat{\mathcal{H}}_k^o = \hat{U}_k^{s \rightarrow o\dagger} \cdot \hat{\mathcal{H}}_k^s \cdot \hat{U}_k^{s \rightarrow o}. \quad (60)$$

We call this alternative form of the Hamiltonian the sublattice form since it treats states within the unit cell as being located on different sublattice sites. We now show that implementing the Peierls substitution with $\hat{\mathcal{H}}_k^o$ before the unitary transformation is equivalent to taking into account both the intercell and intracell currents (we set the electric charge $q = -1$ for simplicity):

$$\begin{aligned} \hat{j}^x &= \frac{\partial}{\partial A} (\hat{U}_k^{s \rightarrow o\dagger} \hat{\mathcal{H}}_{k+A}^s \hat{U}_k^{s \rightarrow o}) \\ &= \frac{\partial}{\partial k} (\hat{U}_k^{s \rightarrow o\dagger} \hat{\mathcal{H}}_k^s \hat{U}_k^{s \rightarrow o}) \\ &\quad - \frac{\partial}{\partial k} \hat{U}_k^{s \rightarrow o\dagger} \hat{\mathcal{H}}_k^s \hat{U}_k^{s \rightarrow o} - \hat{U}_k^{s \rightarrow o\dagger} \hat{\mathcal{H}}_k^s \frac{\partial}{\partial k} \hat{U}_k^{s \rightarrow o} \\ &= \frac{\partial \hat{\mathcal{H}}_k^o}{\partial k} - \frac{\partial}{\partial k} \hat{U}_k^{s \rightarrow o\dagger} \hat{U}_k^{s \rightarrow o} \hat{\mathcal{H}}_k^o - \hat{\mathcal{H}}_k^o \hat{U}_k^{s \rightarrow o\dagger} \frac{\partial}{\partial k} \hat{U}_k^{s \rightarrow o} \\ &= \frac{\partial \hat{\mathcal{H}}_k^o}{\partial k} + i \left[\hat{\mathcal{H}}_k^o, i \hat{U}_k^{s \rightarrow o\dagger} \frac{\partial}{\partial k} \hat{U}_k^{s \rightarrow o} \right] \\ &\equiv \hat{v}_{\text{inter}}^x + \hat{v}_{\text{intra}}^x. \end{aligned}$$

We therefore recognize the identity $i \hat{U}_k^{s \rightarrow o\dagger} \frac{\partial}{\partial k} \hat{U}_k^{s \rightarrow o} \equiv \hat{\tau}$ or, in other words, $\hat{U}_k^{s \rightarrow o} \equiv e^{-ik\hat{\tau}}$. In the case at hand, we have (up to an overall phase representing a shift in origin)

$$\hat{U}_k^{s \rightarrow o} = \begin{pmatrix} e^{ik/4} & 0 & 0 & 0 \\ 0 & e^{-ik/4} & 0 & 0 \\ 0 & 0 & e^{-ik/4} & 0 \\ 0 & 0 & 0 & e^{ik/4} \end{pmatrix} \quad (61)$$

$$\Rightarrow \hat{\tau} = \begin{pmatrix} -\frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & -\frac{1}{4} \end{pmatrix}. \quad (62)$$

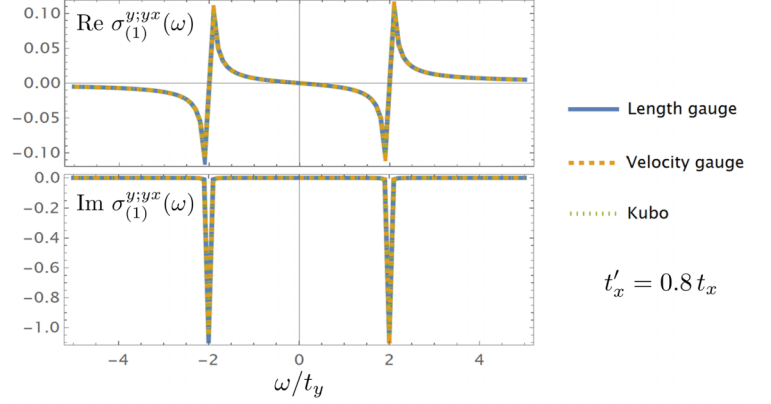
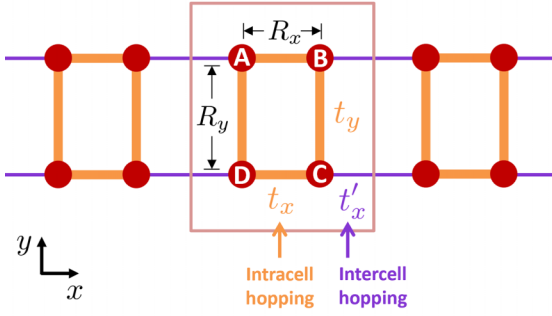
That is, the position operator acquires an intracell part encoding the fact that two of the sublattices are displaced in the x direction by half the width of the unit cell. Therefore, in what follows, we use the following form for the Peierls-substituted Hamiltonian:

$$\hat{\mathcal{H}}_k^s = - \begin{pmatrix} \Delta & g_A(k) & 0 & t_y \\ g_A^*(k) & -\Delta & t_y & 0 \\ 0 & t_y & -\Delta & g_A^*(k) \\ t_y & 0 & g_A(k) & \Delta \end{pmatrix}, \quad (63)$$

$$g_A(k) = e^{-ik/2} [(t + \delta)e^{i(k-A)/2} + (t - \delta)e^{-i(k-A)/2}], \quad (64)$$

Linear conductivity: $\sigma^{\mu\alpha}(\mathbf{q}, \omega) \approx \sigma_{(0)}^{\mu\alpha}(\omega) + q_\nu \sigma_{(1)}^{\nu\mu\alpha}(\omega)$

Molecular Chain:



Decoupling Limit: $t'_x \rightarrow 0$

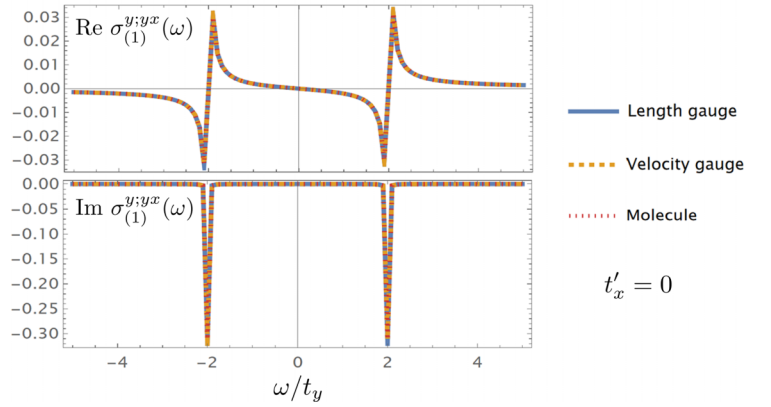
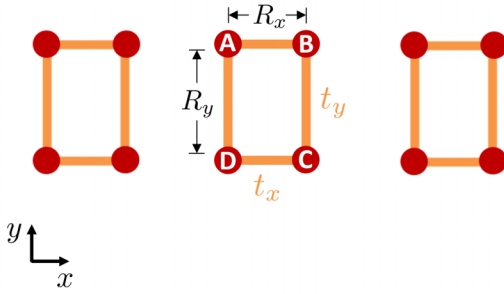


FIG. 7. Results for the spatially dispersive correction to the linear conductivity $[\sigma_{(1)}^{\nu\mu\alpha}(\omega)]$ calculated on a minimal quasi-2D model. This model takes the form of a chain of rectangular molecules with intermolecule couplings, with a sensible decoupling limit in which the problem reduces to the well-understood problem of a multipole response in a molecule. Top: Agreement is demonstrated for the molecular chain between our scheme (in both length and velocity gauge) and the expression determined in Ref. [25] from the Kubo formula expanded to first order in \mathbf{q} . Bottom: Agreement is also demonstrated in the decoupling limit between our scheme (in both length and velocity gauge) and the electric quadrupole linear response from an isolated molecule. Note that inversion breaking is needed to observe this response, so we introduce a small static on-site potential (Δ) of ± 0.1 on sublattices AD/BC to produce these results.

which has *both* the periodicity $k \rightarrow k + 2\pi$ and a dependence on A that encodes both intracell and intercell currents.

We can then define the current operators in the x direction in terms of derivatives with respect to A ,

$$\hat{v}^x \equiv \hat{h}^x = \hat{U}_k^\dagger \cdot \frac{\partial \hat{\mathcal{H}}_k^A}{\partial A} \cdot \hat{U}_k \Big|_{A=0}, \quad (65)$$

$$\hat{h}^{xx} = \hat{U}_k^\dagger \cdot \frac{\partial^2 \hat{\mathcal{H}}_k^A}{\partial A^2} \cdot \hat{U}_k \Big|_{A=0}, \quad (66)$$

$$\hat{h}^{xxx} = \hat{U}_k^\dagger \cdot \frac{\partial^3 \hat{\mathcal{H}}_k^A}{\partial A^3} \cdot \hat{U}_k \Big|_{A=0}, \quad (67)$$

where \hat{U}_k is some (nonunique) unitary transformation that diagonalizes $\hat{\mathcal{H}}_k^{A=0}$. As for the current operators in the y direction, since there is no periodicity in y , we use

$$\hat{v}^y = i \hat{U}_k^\dagger \cdot [\hat{\mathcal{H}}_k^0, \hat{p}^y] \cdot \hat{U}_k, \quad (68)$$

with \hat{p}^y defined just as in the molecule case [Eqs. (48)]. For simplicity, we set the width of the chain in the y direction

to be $1/2$, to match up with the half-unit-cell width in the x direction of a single decoupled molecule. [That is, in comparing with Eqs. (48), we will set $R_x = R_y = 1/4$.] One can then formally compute higher-order currents in the y direction using the covariant derivative, where in this case there is no partial derivative, and so the covariant derivative reduces to a commutator with the Berry connection $\hat{A}^y \equiv \hat{p}^y$,

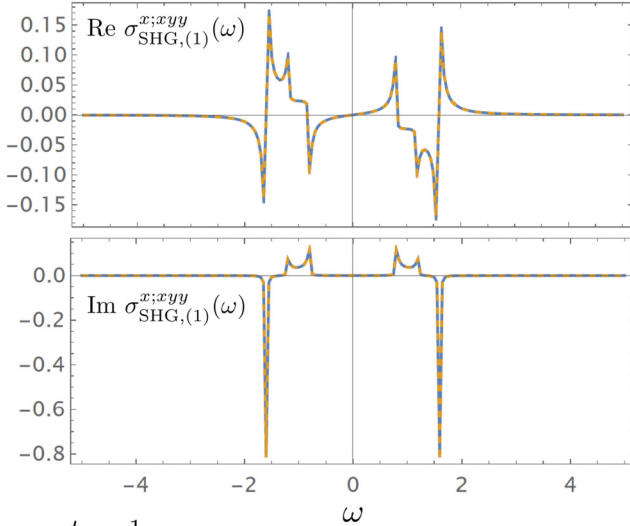
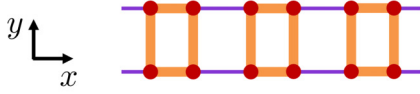
$$\hat{h}^{yy} = -i[\hat{A}^y, \hat{v}^y], \quad (69)$$

and so on for all other current operators $\hat{h}^{\alpha_1 \dots \alpha_n}$.

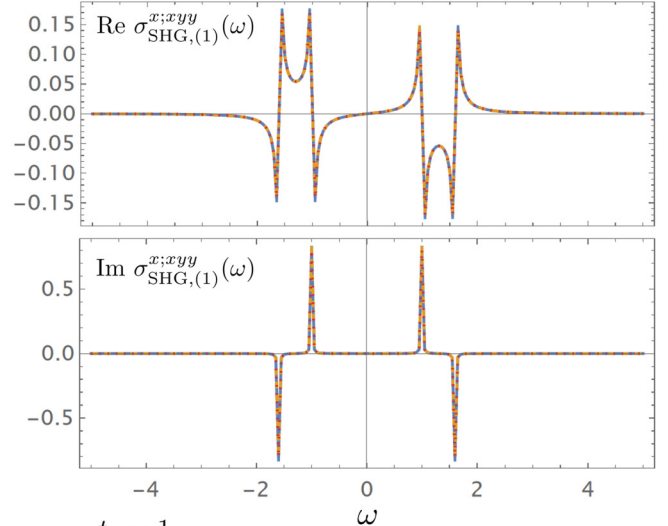
C. Numerical results

Figures 7 and 8 show numerical results for the spatially dispersive corrections to the linear conductivity and SHG conductivity, respectively. We benchmark our methodology in two ways. First, we compare the conductivities of the chain in the $t'_x \rightarrow 0$ limit to the conductivities calculated for a single rectangular molecule. Second, for the linear conductivity, we also benchmark against established results [25,26] for the spatially dispersive linear conductivity calculated by expanding

$$\text{SHG conductivity: } \sigma_{\text{SHG}}^{\mu\alpha\beta}(\mathbf{q}, \omega) \approx \sigma_{\text{SHG},(0)}^{\mu\alpha\beta}(\omega) + q_\nu \sigma_{\text{SHG},(1)}^{\nu\mu\alpha\beta}(\omega)$$

Molecular Chain:


$$\begin{aligned} t_x &= 1 \\ t'_x &= 0.2 \\ t_y &= 0.8 \end{aligned} \quad \begin{array}{l} \text{— Length gauge} \\ \text{- - - Velocity gauge} \end{array}$$

Decoupling Limit:


$$\begin{aligned} t_x &= 1 \\ t'_x &= 0 \\ t_y &= 0.8 \end{aligned} \quad \begin{array}{l} \text{— Length gauge} \\ \text{- - - Velocity gauge} \\ \cdots \cdots \cdots \text{Molecule} \end{array}$$

FIG. 8. Results for the spatially dispersive correction to the *second harmonic generation conductivity* [$\sigma_{\text{SHG},(1)}^{\nu\mu\alpha\beta}(\omega)$] calculated on the molecular chain model detailed in Fig. 7. Agreement is shown between length gauge and velocity gauge in both the molecular chain and in the decoupling limit. Agreement is also shown with the analogous single-molecule calculation. We use $t'_x = 0.8$.

the Kubo formula to first order in \mathbf{q} . Figures 7 and 8 show agreement in all cases. In our plots, we introduce a phenomenological relaxation parameter η by taking $\omega \rightarrow \omega + i\eta$, with $\eta = 0.01$. We set $t_x = 1$, without loss of generality, and in the case for the linear conductivity, we set $\Delta = 0.1$ in both the chain and the molecule to statically break inversion (and to therefore allow a nonzero spatially dispersive correction). We also always assume the *lowest three bands* in this model are filled:

$$\hat{\rho}^{(0)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (70)$$

because the spatially dispersive correction to SHG depends on a peculiar combination of the filling factors (e.g., for the molecule it is proportional to $f_1 - f_2 - f_3 + f_4$).

As a note on symmetries, the molecular chain we consider is inversion symmetric when *either* $\Delta = 0$ *or* $\delta = 0$, where we define $\delta = t_x - t'_x$. This is only evident from the sublattice form $\hat{\mathcal{H}}_k^s$ of the Hamiltonian in Eq. (58). When $\delta = 0$, the system has an inversion center on a vertical bond, represented by the following matrix transformation:

$$\hat{\mathcal{H}}_k^s \longrightarrow \hat{\mathcal{H}}_{-k}^s. \quad (71)$$

Additionally, the system has an inversion center at the *mid-point* of the unit cell whenever $\Delta = 0$, represented by

$$\hat{\mathcal{H}}_k^s \longrightarrow \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \cdot \hat{\mathcal{H}}_{-k}^s \cdot \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (72)$$

So, inversion is only broken when *both* $\delta \neq 0$ and $\Delta \neq 0$.

Given these considerations, we can directly compare SHG resulting from static inversion breaking with SHG resulting from a spatially dispersive correction. Figure 9 plots the magnitude of the *total* SHG response resulting from diagonally polarized light at a frequency of half the minimum gap in this model. We choose diagonal polarization because x -polarized light triggers a *uniform* SHG response ($\sigma_{\text{SHG},(0)}^{xxx}$) when Δ is nonzero, while y -polarized light triggers a spatially dispersive correction ($\sigma_{\text{SHG},(1)}^{xxyy}$) at first order in q_x . In that plot, we choose a window such that the contours are roughly circular in shape, and find a pair of dimensionless parameters in which the two axes have roughly the same scale. We find these relevant dimensionless parameters to be Δ/t'_x and $q_x R_x$. That is, the effects of spatial dispersion on the length scale of half the width of an individual molecule create a similar magnitude

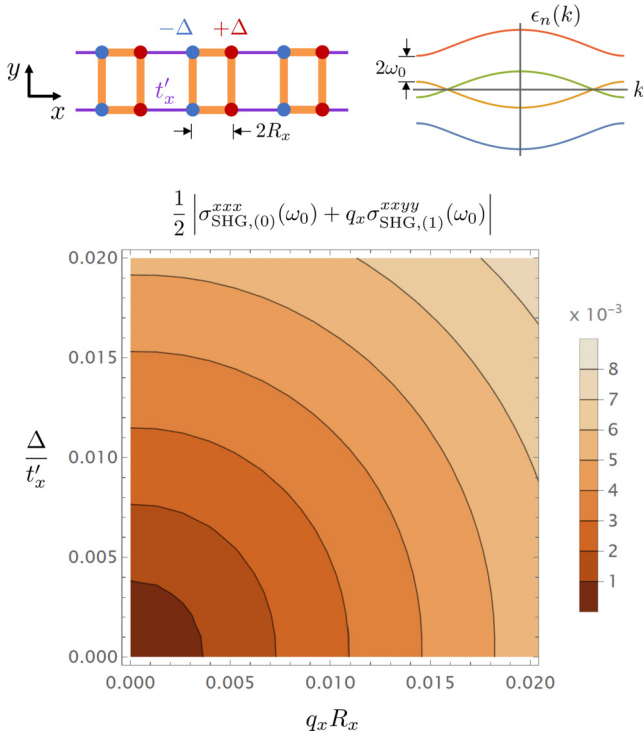


FIG. 9. Magnitude of the *total* SHG response computed up to electric quadrupole order resulting from off-axis linearly polarized light incident on the molecular chain model, evaluated at $\omega_0 = \sqrt{(t_x - t'_x)^2 + \Delta^2}$ (half the minimum gap between the filled and empty bands). Light polarized in x (chain direction) leads to a conventional SHG response when inversion is broken ($\Delta \neq 0$) while light polarized in y leads to a spatially dispersive correction ($q \neq 0$). Here, $t_x - t'_x = 0.5$ and $t_y = 0.8$.

of SHG to that of a static inversion breaking parameter on the energy scale of the intercell hopping. This heuristic has intriguing implications for real systems where spatially dispersive effects may be present. For instance, moiré systems famously create extremely inflated artificial lattice constants, in which case qR is very large (where R is interpreted generally as some coherence length for localized electronic states). So long as the hopping amplitude between superlattice sites is not too small, the spatially dispersive correction to SHG may be on the same order as a typical static inversion-breaking mechanisms for which SHG is often used as a probe. This has important ramifications for using SHG to detect spontaneous breaking of inversion symmetry, since many systems of recent interest may have sufficiently large coherence length scales R to allow spatially dispersive effects to dominate such a response.

V. OUTLOOK

We have introduced a scheme for calculating spatially dispersive corrections to nonlinear optical responses in velocity gauge. Reviewing from a diagrammatic standpoint the subtle cancellations that make up the sum rules guaranteeing equivalence between the length and velocity gauges, we compactly write velocity gauge expressions for nonlinear optical conductivities and introduce spatially dispersive corrections as an additional vertex rule. Though we only treat the electric quadrupole contribution here, we expect the formalism can be straightforwardly extended to include the magnetic dipole contribution, as well as higher-multipole corrections as desired. We expect spatially dispersive corrections to be important for understanding anomalously strong bulk second-order responses in centrosymmetric materials where the response is typically expected to vanish.

We anticipate our hybrid diagrammatic/density matrix scheme to compactly represent other extensions to traditional calculations of nonlinear susceptibilities, in addition to spatially dispersive corrections. For instance, one can treat responses beyond the relaxation time approximation $\omega_j \rightarrow \omega_j + i\eta$ by more carefully treating coupling to a bath, such as in a Keldysh formulation of dissipative systems [4,30,31]. Here, we expect the dissipation to be encoded in a lesser Green's function $G^<(\mathbf{k}, \omega)$, which is equivalent to the reduced density matrix that we associate to each diagram. Adapting this scheme to account for different models of dissipation would be an intriguing direction for future work.

Our results put forward a tool for using nonlinear optics to study quantum materials beyond electric dipole order. We expect spatially dispersive corrections to optical responses to reveal information about intrinsic nonlocality in material systems, since they should be strong when the electric field varies appreciably on the length scale of the electronic states being coupled by optical transitions. A number of materials of recent interest fall into this category, including moiré materials [32,33] and systems whose bands exhibit nontrivial quantum geometry [34–36] for which maximally localized Wannier representations still feature coherence over length scales of many unit cells. Searching for measurable spatially dispersive corrections in these systems and linking them to quantum geometry is a promising direction for future work.

ACKNOWLEDGMENTS

The authors thank M. Claassen, C. De Beule, Z. Addison, and L. Wu for insightful discussions. S.G. was supported by an NSF Graduate Research Fellowship under Grant No. DGE-1845298. This work was supported by the Department of Energy under Grant No. DE-FG02-84ER45118.

- [1] R. W. Boyd, *Nonlinear Optics* (Academic Press, London, 2003).
- [2] H. Sekino and R. J. Bartlett, Frequency dependent nonlinear optical properties of molecules, *J. Chem. Phys.* **85**, 976 (1986).
- [3] S. M. Young, F. Zheng, and A. M. Rappe, First-Principles Calculation of the Bulk Photovoltaic Effect in Bismuth Ferrite, *Phys. Rev. Lett.* **109**, 236601 (2012).

- [4] T. Morimoto and N. Nagaosa, Topological nature of nonlinear optical effects in solids, *Sci. Adv.* **2**, e1501524 (2016).
- [5] L. Wu, S. Patankar, T. Morimoto, N. L. Nair, E. Thewalt, A. Little, J. G. Analytis, J. E. Moore, and J. Orenstein, Giant anisotropic nonlinear optical response in transition metal monophenictide Weyl semimetals, *Nat. Phys.* **13**, 350 (2017).

- [6] F. de Juan, A. G. Grushin, T. Morimoto, and J. E. Moore, Quantized circular photogalvanic effect in Weyl semimetals, *Nat. Commun.* **8**, 15995 (2017).
- [7] Q. Ma, S.-Y. Xu, C.-K. Chan, C.-L. Zhang, G. Chang, Y. Lin, W. Xie, T. Palacios, H. Lin, S. Jia, P. A. Lee, P. Jarillo-Herrero, and N. Gedik, Direct optical detection of Weyl fermion chirality in a topological semimetal, *Nat. Phys.* **13**, 842 (2017).
- [8] C.-K. Chan, N. H. Lindner, G. Refael, and P. A. Lee, Photocurrents in Weyl semimetals, *Phys. Rev. B* **95**, 041104(R) (2017).
- [9] T. Holder, D. Kaplan, and B. Yan, Consequences of time-reversal-symmetry breaking in the light-matter interaction: Berry curvature, quantum metric, and diabatic motion, *Phys. Rev. Res.* **2**, 033100 (2020).
- [10] Z. Ni, N. Huang, A. V. Haglund, D. G. Mandrus, and L. Wu, Observation of giant surface second-harmonic generation coupled to nematic orders in the van der Waals antiferromagnet FePd₃, *Nano Lett.* **22**, 3283 (2022).
- [11] J. Ahn, G.-Y. Guo, N. Nagaosa, and A. Vishwanath, Riemannian geometry of resonant optical responses, *Nat. Phys.* **18**, 290 (2022).
- [12] D. H. Kobe and A. L. Smirl, Gauge invariant formulation of the interaction of electromagnetic radiation and matter, *Am. J. Phys.* **46**, 624 (1978).
- [13] J. E. Sipe and E. Ghahramani, Nonlinear optical response of semiconductors in the independent-particle approximation, *Phys. Rev. B* **48**, 11705 (1993).
- [14] C. Aversa and J. E. Sipe, Nonlinear optical susceptibilities of semiconductors: Results with a length-gauge analysis, *Phys. Rev. B* **52**, 14636 (1995).
- [15] G. B. Ventura, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, Gauge covariances and nonlinear optical responses, *Phys. Rev. B* **96**, 035431 (2017).
- [16] D. J. Passos, G. B. Ventura, J. M. Viana Parente Lopes, J. M. B. Lopes dos Santos, and N. M. R. Peres, Nonlinear optical responses of crystalline systems: Results from a velocity gauge analysis, *Phys. Rev. B* **97**, 235446 (2018).
- [17] M. Schüler, J. A. Marks, Y. Murakami, C. Jia, and T. P. Devereaux, Gauge invariance of light-matter interactions in first-principle tight-binding models, *Phys. Rev. B* **103**, 155409 (2021).
- [18] D. E. Parker, T. Morimoto, J. Orenstein, and J. E. Moore, Diagrammatic approach to nonlinear optical response with application to Weyl semimetals, *Phys. Rev. B* **99**, 045121 (2019).
- [19] H. Xu, H. Wang, and J. Li, Abnormal nonlinear optical responses on the surface of topological materials, *npj Comput. Mater.* **8**, 111 (2022).
- [20] S. A. Mikhailov, Quantum theory of the third-order nonlinear electrodynamic effects of graphene, *Phys. Rev. B* **93**, 085403 (2016).
- [21] R. K. Malla, A. Saxena, and W. J. M. Kort-Kamp, Emerging nonlinear Hall effect in Kane-Mele two-dimensional topological insulators, *Phys. Rev. B* **104**, 205422 (2021).
- [22] B. M. Fregoso, Bulk photovoltaic effects in the presence of a static electric field, *Phys. Rev. B* **100**, 064301 (2019).
- [23] B. M. Fregoso, Bulk photospin effect: Calculation of electric spin susceptibility to second order in an electric field, *Phys. Rev. B* **106**, 195108 (2022).
- [24] E. I. Blount, Bloch electrons in a magnetic field, *Phys. Rev.* **126**, 1636 (1962).
- [25] A. Malashevich and I. Souza, Band theory of spatial dispersion in magnetoelectrics, *Phys. Rev. B* **82**, 245118 (2010).
- [26] J. Ahn, S.-Y. Xu, and A. Vishwanath, Theory of optical axion electrodynamics and application to the kerr effect in topological antiferromagnets, *Nat. Commun.* **13**, 7615 (2022).
- [27] D. B. Melrose and R. C. McPhedran, *Electromagnetic Processes in Dispersive Media* (Cambridge University Press, Cambridge, England, 2009).
- [28] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, in *Photons and Atoms*, edited by C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg (John Wiley & Sons, Nashville, TN, 1997).
- [29] M. Jaszński, A. Rizzo, and K. Ruud, Molecular electric, magnetic, and optical properties, in *Handbook of Computational Chemistry*, edited by J. Leszczynski, A. Kaczmarek-Kedziera, T. Puzyn, M. G. Papadopoulos, H. Reis, and M. K. Shukla (Springer International Publishing, Cham, 2017), pp. 497–592.
- [30] A. Kamenev, Many-body theory of non-equilibrium systems, *arXiv:cond-mat/0412296*.
- [31] T. Oka and H. Aoki, Photovoltaic Hall effect in graphene, *Phys. Rev. B* **79**, 081406(R) (2009).
- [32] M. Koshino, N. F. Q. Yuan, T. Koretsune, M. Ochi, K. Kuroki, and L. Fu, Maximally Localized Wannier Orbitals and the Extended Hubbard Model for Twisted Bilayer Graphene, *Phys. Rev. X* **8**, 031087 (2018).
- [33] J. Kang and O. Vafek, Symmetry, Maximally Localized Wannier States, and a Low-Energy Model for Twisted Bilayer Graphene Narrow Bands, *Phys. Rev. X* **8**, 031088 (2018).
- [34] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, Exponential Localization of Wannier Functions in Insulators, *Phys. Rev. Lett.* **98**, 046402 (2007).
- [35] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Maximally localized Wannier functions: Theory and applications, *Rev. Mod. Phys.* **84**, 1419 (2012).
- [36] H. C. Po, H. Watanabe, and A. Vishwanath, Fragile Topology and Wannier Obstructions, *Phys. Rev. Lett.* **121**, 126402 (2018).