Diagrammatic approach to nonlinear optical response with application to Weyl semimetals

Daniel E. Parker, Takahiro Morimoto, Joseph Orenstein, and Joel E. Moore

Phys. Rev. B **99**, 045121 — Published 10 January 2019

DOI: 10.1103/PhysRevB.99.045121

Diagrammatic approach to nonlinear optical response with application to Weyl semimetals

Daniel E. Parker,[1,*] Takahiro Morimoto,[1,2,†] Joseph Orenstein,[1,2,‡] and Joel E. Moore[1,2,§]
[1] Department of Physics, University of California, Berkeley, CA 94720, USA
[2] Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
(Dated: October 26, 2018)

Nonlinear optical responses are a crucial probe of physical systems including periodic solids. In the absence of electron-electron interactions, they are calculable with standard perturbation theory starting from the band structure of Bloch electrons, but the resulting formulas are often large and unwieldy, involving many energy denominators from intermediate states. This work gives a Feynman diagram approach to calculating non-linear responses. This diagrammatic method is a systematic way to perform perturbation theory, which often offers shorter derivations and also provides a natural interpretation of nonlinear responses in terms of physical processes. Applying this method to second-order responses concisely reproduces formulas for the second-harmonic, shift current. We then apply this method to third-order responses and derive formulas for third-harmonic generation and self-focusing of light, which can be directly applied to tight-binding models. Third-order responses in the semiclasscial regime include a Berry curvature quadrupole term, whose importance is discussed including symmetry considerations and when the Berry curvature quadrupole becomes the leading contribution. The method is applied to compute third-order optical responses for a model Weyl semimetal, where we find a new topological contribution that diverges in a clean material, as well as resonances with a peculiar linear character.

CONTENTS

I. INTRODUCTION

Optical response provides a window into the quantum nature of materials. The exquisite control and precise measurements enabled by modern optical techniques frequently couple with theoretical predictions to test and confirm models of quantum materials. Nonlinear optical responses[1–4], in particular, give a wealth of information on dynamics, symmetry, and—recently—topology[5–17]. To fully reap the benefits of optical techniques, it is necessary to accurately predict optical responses in solids from theory, including both simplified tight-binding models and advanced computational approaches.

Historically, optical responses were understood first at the linear order, and then extended to nonlinear orders alongside the development of the laser in the 1960s. For molecular systems, normal quantum-mechanical perturbation theory suffices, and a convenient diagrammatic language became popular, capturing optical processes in terms of electrons changing energy levels[18]. In crystalline systems, however, there are several additional wrinkles. Simply defining the perturbation corresponding to an external electric field is a subtle task. Like in a molecule, absorbing a photon can cause an electron to jump to a different band, but can also cause the electron to move to a nearby k-point on the same band. The latter requires connecting adjacent points in k-space, and thus involves the Berry connection[19,20].

It is only relatively recently that the electromagnetic perturbation was written carefully in order to treat non-linear responses. There are two standard ways of writing an electromagnetic perturbation within the framework of independent electrons and dipole fields. First, the so-called *length gauge*

$$\widehat{H}_E = \widehat{H}_0 + e\boldsymbol{E}(t) \cdot \widehat{\boldsymbol{r}} \tag{1}$$

uses the single-particle position operator $\widehat{\boldsymbol{r}}$ whereas the second, the *velocity gauge*, uses the minimal substitution scheme

$$\widehat{H}_A = \widehat{H}_0(\boldsymbol{k} - e\boldsymbol{A}(t)) \tag{2}$$

where the vector potential $\boldsymbol{A}(t)$ is chosen so that $\boldsymbol{E}(t) = -\partial_t \boldsymbol{A}(t)$. As usual, each gauge is well-suited for a different set of tasks. The length gauge is better for analytical answers, semi-classical limits, and some questions involving topology, whereas the velocity gauge gives a cleaner resonance structure and is easier to implement numerically, especially for tight-binding models.

Over time, there has been a competition between the two approaches. The velocity gauge was initially favored in the 1980s due to easier calculation (see e.g.[21,22]). Velocity gauge calculations, however, often contain spurious divergences at zero frequency, which can be eliminated only with somewhat opaque sum rules. The position operator was defined carefully in the work of Blount in the 1960s[23], and its relation to the Berry connection was understood deeply by the early 1990s, giving rise to the modern theory of polarization[24,25]. This understanding was harnessed by Sipe and Shkrebtii to develop a widely used approach to calculate second-order responses within the length gauge[5]. The wide variety of physical effects in the second-order response—including shift current[6–9], injection current[13,14,26], and second-harmonic generation (SHG)[9,27]—are of great current interest for a wide variety of systems. These convenient formulae[5], together with putative dangers associated with the velocity gauge when truncating the number of bands, ensured the primacy of the length gauge.

Recent work[28–30] has re-examined the roots of the problem, providing careful prescriptions for both gauges and how to translate between them. It is now possible to use either gauge correctly, depending on the problem at hand. In this work we focus primarily on the relatively underappreciated velocity gauge, developing a convenient Feynman diagram prescription for calculating nonlinear responses. As noted above, diagrammatic methods have a long history in the subject. The formulation here, however, has several key differences from previous work to implement the correct form of the electromagnetic interaction and fully account for the effects of the Berry connection. Our goal is to show that any second- or third-order optical response can be calculated from diagrams in only a few lines. Two practical advantages of the resulting velocity-gauge expressions is that the resonance structure is manifest, immediately distinguishing one-, two-, and three-photon terms, and the expressions can be directly implemented in tight-binding models without the need for sum rules.

One motivation for this work is providing tools to better understand optical responses. A large body of recent work has followed the theme of studying optical responses in situations where they become particularly simple: semiclassics and Weyl semimetals. In semiclassics, the limit of a single band where $\omega \to 0$, optical responses can be understood from the semiclassical equations of motion (EOM) that describe wavepackets of Bloch electrons. Berry curvature gives an additional contribution to these equations of motion called the anomalous velocity, which leads to the Hall conductivity in linear response[31]. At second order, the anomalous velocity is responsible for the circular photogalvanic effect (CPGE), and non-linear Kerr rotation that is proportional to the dipole of Berry curvature[10–12].

The next-to-simplest situation is that of two-band models, where interband responses give rise to resonances. Perhaps the most intriguing two-band models are those for Weyl semimetals. These materials support three-dimensional gapless points called Weyl nodes that are sources and sinks of Berry curvature[32,33]. Simple tight-binding models often suffice to describe their properties. However, the fact that the Fermi surface vanishes at the Weyl nodes puts them firmly beyond the semiclassical regime. Due to their nontrivial Berry curvature structure, Weyl semimetals host a variety of non-trivial linear responses, including the chiral magnetic effect[34,35] and gyrotropic magnetic effect[36,37]. As one might expect, there is an even richer set of nonlinear optical responses due to the Berry curvature[13,14,27]. These effects can be realized, for instance, in the monopnictide TaAs, a Weyl semimetal with inversion breaking[38,39]. Recent optical experiments on TaAs revealed that TaAs shows CPGE responses closely tied to its Weyl node structure[16] and giant SHG, with the largest $\chi^{(2)}$ of any known material[15,17].

Below we connect the Feynman diagram formulation of optical response to both semiclassics and Weyl semimetals. In the semiclassical limit we show that, with particular symmetries, the largest term in the third-order response has a topological origin as the quadrupole of the Berry Curvature. We also examine the third-harmonic response of a Weyl semimetal. We find that the off-diagonal component σ^{zxxx} has large two-photon and three-photon resonances with peculiar linear profiles due to the Weyl cones.

The remainder of this paper is organized as follows. Section II introduces notation and the Feynman rules. Sections III-V derive non-linear optical responses through third order using Feynman diagrams and provide some physically interesting limits. Section VI considers the semiclassical limit, its relation to the length gauge, and some topological considerations at third order. Section VII presents a numerical example of a Weyl

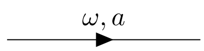
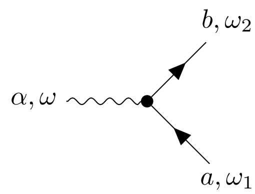
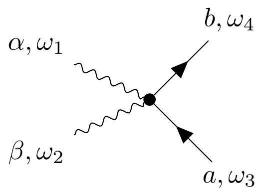
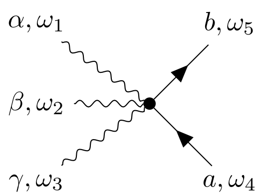
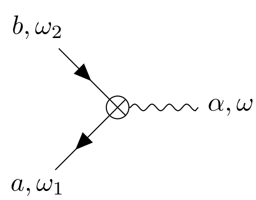
Component	Diagram	Value
(Classical) Photon Propagator		1
Electron Propagator		$G_a(\omega)$
One-Photon Input Vertex		$\frac{ie}{\hbar\omega_1} h^{\alpha}_{ab}$
Two-Photon Input Vertex		$\prod_{k=1}^{2} \left(\frac{ie}{k\hbar\omega_k}\right) h^{\alpha\beta}_{ab}$
Three-Photon Input Vertex		$\prod_{k=1}^{3} \left(\frac{ie}{k\hbar\omega_k}\right) h^{\alpha\beta\gamma}_{ab}$
One-Photon Output Vertex		eh^{μ}_{ab}

TABLE I. The Feynman rules for non-linear electromagnetic perturbations in a crystal. Following the pattern, a new vertex with N incoming photons will appear at Nth order. Energy must be integrated around each internal loop, and conserved at each vertex. The output vertex can appear with any number of photons and gains a power of $ie(k\hbar\omega_k)^{-1}$ for each additional external photon.

semimetal and, lastly, Section VIII concludes with some heuristic rules for choosing a gauge, and other comments.

II. SETUP & FEYNMAN RULES

We will work in a band theory picture of non-interacting electrons for simplicity, though most of the considerations involved carry over to Fermi liquid theory. We first recall some key definitions to set notation, then discuss perturbation theory in an external electric field for velocity gauge, and derive the Feynman rules. We also comment on the assumptions and caveats of the framework.s

A. Band Theory and Notation

Consider a crystalline material in d dimensions described by band theory. The second-quantized Hamiltonian is then

$$\widehat{H}_0 = \sum_{a\in\mathbb{Z}} \int [d\boldsymbol{k}]\ \varepsilon_a(\boldsymbol{k}) c^{\dagger}_{\boldsymbol{k}a} c_{\boldsymbol{k}a} \tag{3}$$

where $\int[d\boldsymbol{k}] = (2\pi)^{-d}\int d^d\boldsymbol{k}$ indicates the properly-normalized integral over the d-dimensional first Brillouin zone, the sum runs over all bands, and the $c^{\dagger}$ and cs's are single-particle fermion creation and annihilation operators, satisfying the usual anticommutation relations

$$\{c_{\boldsymbol{k}a}, c^{\dagger}_{\boldsymbol{k}'b}\} = (2\pi)^d \delta_{ab}\delta(\boldsymbol{k}-\boldsymbol{k}'). \tag{4}$$

(Latin indices a, b, c, d are used to label bands henceforth.) We assume that the crystal is infinite in extent, without boundary.

Because the Hamiltonian (3) involves only a single raising and lowering operator, Fermion number is a symmetry of the system. We may thus write all observables in terms of the single-particle wavefunctions. As usual, Bloch's Theorem says the single-electron wavefunctions may be written as

$$\psi_{\boldsymbol{k}a}(\boldsymbol{r}) = \langle 0|\widehat{\Psi}(\boldsymbol{r})c^{\dagger}_{\boldsymbol{k}a}|0\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{\boldsymbol{k}a}(\boldsymbol{r}), \quad a\in\mathbb{Z}, k\in \mathrm{BZ} \tag{5}$$

where $\widehat{\Psi}(\boldsymbol{r})$ annihilatesss an electron at $\boldsymbol{r}$, and the u's are periodic functions on the unit cell[40]. The u's are eigenfunctions of the k-dependent Hamiltonian $\widehat{H}_0(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\widehat{H}_0 e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$:

$$\widehat{H}_0(\boldsymbol{k})\ u_{\boldsymbol{k}a}(\boldsymbol{r}) = \varepsilon_{\boldsymbol{k}_a} u_{\boldsymbol{k}a}(\boldsymbol{r}). \tag{6}$$

Despite our assumption of independent electrons, we work in a fully many-body framework. This is necessary to implement Feynman diagrams but also, as discussed below, makes the generalization to the interacting case transparent.

B. Electromagnetic Interactions

Suppose there is an external electric field which we treat classically. We adopt the velocity gauge with the minimal coupling Hamiltonian (2). To capture nonlinear responses, we expand in powers of the vector potential in a Taylor series. This is, however, not as straightforward as it might naively seem because one must carefully consider what notion of derivative should be employed in the series. The answer is that one should use the (Berry) covariant derivative $\widehat{D}$ when working in k-space. The derivation of this fact and the defintion of the covariant derivative are reviewed in Appendix A. Appendix A shows that the covariant derivative $\widehat{\boldsymbol{D}}$ is related to the single-electron position operator via $\widehat{\boldsymbol{r}} = i\widehat{\boldsymbol{D}}$, and that it acts naturally on operators via

$$\widehat{\boldsymbol{D}}[\widehat{\mathcal{O}}] = [\widehat{\boldsymbol{D}}, \widehat{\mathcal{O}}], \tag{7}$$

where the commutator has matrix elements

$$[\widehat{D}^\mu, \widehat{\mathcal{O}}]_{ab} = \frac{\partial \mathcal{O}_{ab}}{\partial k^\mu} - i[\mathcal{A}^\alpha, \widehat{\mathcal{O}}]_{ab} \tag{8}$$

where $\boldsymbol{\mathcal{A}}$ is the Berry connection whose matrix elements are $\mathcal{A}_{ab} = i\langle u_{\boldsymbol{k}a}|\partial_{\boldsymbol{k}} u_{\boldsymbol{k}b}\rangle$. Note that the covariant derivative of an operator is *not* the derivative of its matrix elements.

In terms of the covariant derivatives, the Hamiltonian can be written as a Taylor series in terms of the electric field as

$$\widehat{H}_A = \widehat{H}_0 + \widehat{V}_E(t) = \widehat{H}_0 + \sum_{n=1}^{\infty} \frac{1}{n!}\left[\prod_{k=1}^{n} \frac{e}{\hbar} A^{\alpha_k} \widehat{D}^{\alpha_k}\right] \widehat{H}_0, \tag{9}$$

where $\alpha_k \in \{x, y, z\}$ is a spatial index with an implicit sum, and $\widehat{\boldsymbol{D}}$ is the (Berry) covariant derivative. (Greek indices $\mu, \alpha, \beta, \ldots$ will always represent spatial indices with an implicit summation henceforth.)

Equation (7) can be used to write the velocity operator of the *unperturbed* Hamiltonian as

$$\widehat{\boldsymbol{v}} = [\widehat{\boldsymbol{D}}, H_0] = -i[\widehat{\boldsymbol{r}}, \widehat{H}_0]. \tag{10}$$

For convenience, we define higher derivatives of the unperturbed Hamiltonian by

$$\widehat{h}^{\alpha_1 \ldots \alpha_N} = \widehat{D}^{\alpha_1} \cdots \widehat{D}^{\alpha_N}[\widehat{H}_0]. \tag{11}$$

The perturbation due to the external field can then be written as

$$\widehat{V}_E(t) = \frac{e}{\hbar} A^\alpha(t) \widehat{h}^\alpha + \frac{1}{2}\left(\frac{e}{\hbar}\right)^2 A^\alpha(t) A^\beta(t) \widehat{h}^{\alpha\beta} + \cdots \tag{12}$$

Fourier transforming and using $\boldsymbol{E}(\omega) = i\omega \boldsymbol{A}(\omega)$, we have

$$\widehat{V}_E(t) = \sum_{n=1}^{\infty} \frac{1}{n!} \prod_{k=1}^{n} \int d\omega_k e^{-i\omega_k t} \left(\frac{ie}{\hbar\omega_k}\right) E^{\alpha_k}(\omega_k) \widehat{h}^{\alpha_1 \ldots \alpha_n}. \tag{13}$$

It is essential that—in the velocity gauge—a seemingly new perturbation appears at each order in the electric field. Physically, the nth term corresponds to the simultaneous interaction of N photons with an electron.

The electromagnetic response of a crystal is characterized by the conductivity tensors. Incident electric fields produce a current, giving rise to a non-zero expectation of the current operator. The conductivity tensors are defined as the coefficients in an expansion of the current in powers of the external field:

$$\langle \widehat{J}^\mu \rangle(\omega) = \int d\omega_1\, \sigma^{\mu\alpha}(\omega; \omega_1) E^\alpha(\omega_1) \tag{14}$$

$$+ \int d\omega_1 d\omega_2\, \sigma^{\mu\alpha\beta}(\omega; \omega_1, \omega_2) E^\alpha(\omega_1) E^\alpha(\omega_2) + \cdots \tag{15}$$

where the first argument of the conductivity tensor σ is the "output" frequency ω and the others $(\omega_1, \omega_2, \ldots)$ are the frequencies of the incident light.

C. Feynman Rules

The task in front of us is to compute the conductivity tensors from the Hamiltonian $\widehat{H}_0$. This is an ideal task for perturbation theory, as we start with a free fermion system and have a perturbation naturally stratified in power of the external field. In the literature, the current operator has commonly been computed with a density matrix formalism in the single-particle picture[3,5]. However, we shall adopt a path-integral and Feynman diagram approach that is shorter and more physically transparent. The two approaches are, of course, equivalent.

Formally, the partition function of the perturbed system may be written as a path integral

$$\begin{aligned} Z[\boldsymbol{E}] &= \int \mathcal{D}c^\dagger \mathcal{D}c\, \exp\left(-i\int dt\, H_A(t)\right) \\ H_A(t) &= \int [d\boldsymbol{k}]\, c_k^\dagger(t) H_0 c_k(t) + c_k^\dagger(t) V_E(t) c_k(t). \end{aligned} \tag{16}$$

The expectation of the current is then

$$\begin{aligned} \langle \widehat{J}^\mu \rangle(t) &= \frac{1}{Z} \mathrm{Tr}\left[\mathcal{T} e \widehat{v}_E^\mu(t)\, e^{-i\int \widehat{H}(t')\,dt'}\right] \\ &= \frac{1}{Z} \int \mathcal{D}c^\dagger \mathcal{D}c\, \left[e v_E^\mu(t)\right] \exp\left(-i\int dt' H_A(t')\right) \end{aligned} \tag{17}$$

where $\mathcal{T}$ represents time-ordering of operators. Here $\widehat{\boldsymbol{v}}_E$ is the velocity operator in the *perturbed* system—which itself depends on the electric field:

$$\begin{aligned} \widehat{v}_E^\mu(t) &= \widehat{D}^\mu[\widehat{H}_0 + \widehat{V}_E(t)] \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \prod_{k=1}^{n} \int d\omega_k e^{i\omega_k t} \left(\frac{e}{\hbar\omega_k}\right) E^{\alpha_k}(\omega_k) \widehat{h}^{\mu\alpha_1 \cdots \alpha_n}. \end{aligned} \tag{18}$$

In terms of functional derivatives, the conductivities are then given by

$$\begin{aligned} &\sigma^{\mu\alpha_1 \ldots \alpha_n}(\omega; \omega_k) \\ &= \int \frac{dt}{2\pi} e^{i\omega t} \prod_{k=1}^{n} \int \frac{dt_k}{2\pi} e^{i\omega_k t_k} \frac{\delta}{\delta E_k^\alpha(t_k)} \langle \widehat{J}^\mu \rangle(t)\Big|_{\boldsymbol{E}=0}. \end{aligned} \tag{19}$$

As a brief technical aside, one would usually take functional derivatives in the frequency domain, but due to the explicit time-dependence in the Hamiltonian, it is neccesary to compute first in the time-domain and then Fourier transform.

Considering the form of (17), we are performing a dual expansion in $\boldsymbol{E}$, as both the exponent and the velocity operator depend on the electric field. As is usual in quantum field theory, the effect of the functional derivatives in (19) turns out to be purely combinatorial and can be entirely captured in terms of Feynman diagrams. The only wrinkle is that, since we are computing a non-standard type of current, there is an extra vertex corresponding to the "output" velocity operator.

Explicitly, the value of the Nth non-linear conductivity can be computed by drawing all connected diagrams such that:

1. There are $N+1$ external photons.
2. All electrons are internal and compose one loop.
3. Exactly one vertex is crossed to indicate the output current $\widehat{J}^{\mu}$; all other vertices are dotted.
4. Diagrams are symmetrized over all incoming photons (α_k, ω_k). The factors on the vertices in Table I are chosen to avoid double-counting.
5. The value of edges and vertices are given in Table I.

This procedure is slightly different from what is common in particle physics and thus merits some explanation. First, since $c \gg v_F$, the Fermi velocity, a negligible amount of momentum is exchanged through interactions. We thus consider only energy conservation at each vertex. Second, we assume electrons are bound to the solid, so only photons may be external. Third, since electrons must return to their equilibrium positions after a perturbation and are non-interacting, only diagrams with exactly one fermion loop are permitted. Fourth, we treat the photon as a classical background field without dynamics, whose propagator is unity. However, the electron propagator is the usual one for free fermions

$$G_{\boldsymbol{k}a}(\omega) = \frac{1}{\omega - \varepsilon_a(\boldsymbol{k})} \tag{20}$$

where the $\boldsymbol{k}$-index is suppressed below for notational convenience. We will see below that, in practice, photons may never cross the inside of a fermion loop.

This method is simple to apply in practice. Unlike many diagrammatic methods, this method involves no divergences beyond simple poles and does not require regularization. The computation of the first, second, and third-order responses are no more than a few lines. The only non-trivial part consists of one new contour integral at each order, which are performed for the reader's convenience in Appendix B.

D. Assumptions and Caveats

Though the method of Feynman diagrams outlined here is convenient, it is important to recognize the assumptions that went into it and thus determine its range of validity. The use of the velocity gauge is associated with several problems: spurious divergences and inaccurate approximations. The conductivities computed in velocity gauge are apparently divergent with $\sigma^{(N)} \sim \frac{1}{\omega^N}$. These divergences are spurious, but eliminating them requires the use of sum rules. These sum rules are now understood as identities needed to convert from velocity to length gauges (see the Appendix of[28]). However, they are still inconvenient to apply beyond first order, so when taking the $\omega \to 0$ limit, it is best to work in the length gauge. This is carried out carefully in Section VI.

The velocity gauge has often been considered badly behaved under approximations. When materials are modelled by effective Hamiltonians focusing on a few bands close to the Fermi level (such as two band models for Dirac semimetals), then effective optical responses calculated in the length gauge are generally accurate, while those in the velocity gauge can suffer from corrections the same size as the response, rendering them practically useless. It was shown in[29] that this inaccuracy only arises with models where the effective Hamiltonian is not defined on the full Brillouin zone, ruining periodicity, or from the application of incorrect sum rules. In practice, this prevents some two-band models of topological materials from being studied with velocity gauge. However, if enough care is taken in defining the model, there is no reason the velocity gauge cannot be used.

One should also take care with dynamical effects. We have taken a perturbative approach to what is actually a non-equilibrium problem. The currents described here are only the initial current created after an incident pulse of light. In practice, other dynamical effects may come into play before those currents can be observed, corrupting or distorting the current. For instance, a strong laser field could create a population of excitons whose recombination interferes with the motion of electrons. This type of issue makes it difficult to observe phenomena which manifest as electrical currents rather than optical responses, such as the shift current. We should note, however, that the perturbation theory with equilibrium Green's functions accurately describe the nonlinear conductivities, since they are obtained as finite order perturbation in the external electric field $\boldsymbol{E}(\omega)$ with respect to the equilibrium state. Namely, one could say that our diagrammatic approach generalizes the Kubo formula for linear response. Normally the Kubo formula relates the linear conductivity to the current-current correlation function $\langle \boldsymbol{J}\boldsymbol{J} \rangle$. Nonlinear conductivity generalizes this to the setting where the input and output current are different operators, so we are effective computing correlators of the form $\langle \boldsymbol{J}\,(\boldsymbol{J}\boldsymbol{J}\boldsymbol{J}) \rangle$ (at the third order).

A brief comment on the effect of scattering is also in order. In a real material, impurities, photons and other effects will perturb the free fermion band-structure. If these effects are sufficiently weak, as is often the case, one can simply replace the electron propagator with a dressed version

$$G_a(\omega) = \frac{1}{\omega - \varepsilon_a} \to \frac{1}{\omega - \varepsilon_a + i\Sigma_a(\omega)} \tag{21}$$

where Σ_a is the self-energy of the electron, and is calculable within Fermi liquid theory. In practice it is usually unnecessary to understand the full frequency dependence of the self-energy. The phenomenological approximation $i\Sigma(\omega) = i\gamma \to i0^+$ is therefore often made. All the above calculations can be carried out with this phenomenological scattering factor included by slightly moving the poles, i.e. simply substituting $\omega_1 \to \omega_1 + i\gamma$, etc. One should note that for two-photon poles, the scattering factor contributes twice, so

$$\frac{1}{\omega_1 + \omega_2 - \varepsilon} \to \frac{1}{\omega_1 + \omega_2 - \varepsilon - 2i\gamma}. \tag{22}$$

It was pointed out in[29] that this factor of two can actually have a significant effect on the shape of resonances, especially at low frequency, and is therefore crucial when making experimental predictions.

This procedure is essentially the same as incorporating interactions into the model. In principle, the technique developed here works in the fully interacting case, once the propagator and velocity operator are appropriately modified. However, performing this analytically requires either weak interactions (i.e. a Fermi liquid) or a quadratic Hamiltonian, such as in a mean-field approximation. The BCS model of superconductivity falls into the later category, and non-linear responses of superconductors will be the topic of future work.

Equipped with the Feynman rules, it is straightforward to compute the non-linear conductivity tensors at any order. At first-order there are two diagrams, four at second order, and eight at third order. Each corresponds to a unique physical process that contributes independently to the overall response.

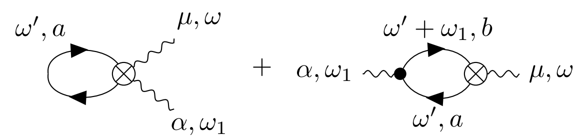
III. FIRST ORDER CONDUCTIVITY

As a pedagogical demonstration of our framework, we re-derive the first-order conductivity. Using the rules, the answer is almost immediate. As an additional confirmation, however, we offer a complementary derivation starting from the definition of the conductivity. One can see this as a derivation of the Feynman rules at first order.

A. Derivation from Diagrams

There are two Feynman diagrams at first order:

$$\sigma^{\mu\alpha}(\omega;\omega_1) = \tag{23}$$



$$= \frac{ie^2}{\hbar\omega_1} \sum_{a,b} \int [d\boldsymbol{k}] \int d\omega' \; h^{\alpha}_{ab} G_b(\omega' + \omega_1) h^{\mu}_{ba} G_a(\omega')$$
$$+ \frac{ie^2}{\hbar\omega_1} \sum_{a} \int [d\boldsymbol{k}] \int d\omega' \; h^{\mu\alpha}_{aa} G_a(\omega').$$

The frequency integrals are performed with standard techniques (see Appendix B) to find

$$I_1 = \int d\omega' G_a(\omega') \;=\; f_a \tag{24}$$

$$I_2(\omega_1) = \int d\omega' G_b(\omega' + \omega_1) G_a(\omega') \;=\; \frac{f_{ab}}{\omega_1 - \varepsilon_{ab}} \tag{25}$$

where $f_a = f(\varepsilon_{\boldsymbol{k}a})$ is the Fermi factor and $f_{ab} = f_a - f_b$, $\varepsilon_{ab} = \varepsilon_a - \varepsilon_b$ are difference of Fermi factors and energies respectively. Therefore the conductivity is

$$\sigma^{\mu\alpha}(\omega;\omega) = \frac{ie^2}{\hbar\omega} \sum_{a\neq b} \int [d\boldsymbol{k}] f_a h^{\mu\alpha}_{aa} + \frac{h^{\alpha}_{ab} h^{\mu}_{ba} f_{ab}}{\omega - \varepsilon_{ab}}. \tag{26}$$

(The sum over band indices is only performed over the indices appearing in each term; the first term is summed over a while the second is summed over both a and b. This notational abbreviation is used from now on.)

To connect this to familiar results, we convert to the length gauge and consider the $\omega \to 0$ limit, expressing all matrix elements in terms of the velocity matrix elements, $v^{\mu}_{ab} = h^{\mu}_{ab}$. Using the identity

$$h^{\mu\alpha}_{ab} = [D^{\alpha}, v^{\mu}]_{ab} = \partial^{\alpha} v^{\mu}_{ab} - i[A^{\alpha}, v^{\mu}]_{ab} \tag{27}$$

and the fact $v^{\mu}_{ab} = -i\varepsilon_{ba} A^{\mu}_{ab}$, the conductivity becomes

$$\sigma^{\mu\alpha}(\omega;\omega) = \tag{28}$$
$$\frac{ie^2}{\hbar\omega} \sum_{a,b} \int [d\boldsymbol{k}] f_a \partial^{\alpha} v^{\mu}_{aa} + f_{ab} v^{\alpha}_{ab} v^{\mu}_{ba} \left(\frac{1}{\varepsilon_{ba}} - \frac{1}{\omega - \varepsilon_{ab}} \right).$$

Combining the term in parentheses into a single fraction eliminates the spurious divergence:

$$\sigma^{\mu\alpha}(\omega,\omega) = \frac{ie^2}{\hbar} \sum_{a,b} \frac{f_a \partial^{\alpha} v^{\mu}_{aa}}{\omega} + \frac{f_{ab} v^{\alpha}_{ab} v^{\mu}_{ba}}{(\omega - \varepsilon_{ab})\varepsilon_{ba}} \tag{29}$$

This is the standard result in the length gauge[5].

In the $\omega \to 0$ limit, the second term becomes $f_{ab}v^\alpha_{ab}v^\mu_{ba}/(\varepsilon^2_{ba}) + O(\omega^2) = f_a\mathcal{F}^{\mu\alpha}_{aa} + O(\omega^2)$, the Berry curvature. We then have

$$\lim_{\omega\to 0}\sigma^{\mu\alpha}(\omega;\omega) = \frac{ie^2}{\hbar}\sum_a\int[d\boldsymbol{k}]\frac{-\partial^\alpha f_a v^\mu_{aa}}{\omega - i\gamma} + f_a\mathcal{F}^{\mu\alpha}_{aa} \tag{30}$$

The first term corresponds to the Drude weight, and the second term is responsible for the Hall conductivity. This formula matches what is derived from semiclassics in a Boltzmann equation approach, which is examined in Section VI.

B. Derivation of the Diagrams

We now give an alternative derivation of Eq. (23) from the definition of the current operator. This is essentially equivalent to a derivation of the Feynman rules and may be skipped by a reader already convinced of their validity. We start from the time-domain conductivity

$$\sigma^{\mu\alpha}(t;t_1) = \frac{\delta}{\delta E^\alpha(t_1)}\left\langle \widehat{J}^\mu\right\rangle(t)\Big|_{\boldsymbol{E}=0}. \tag{31}$$

We must therefore evaluate the expectation value of

$$\frac{\delta\widehat{v}^\mu_E(t)}{\delta E^\alpha(t_1)} - \widehat{v}^\mu_E(t)\frac{\delta}{\delta E^\alpha(t_1)}\int dt'\ \widehat{H}(t') \tag{32}$$

at $\boldsymbol{E}=0$. Writing $\widehat{H}(t') = \widehat{H}_0 + \widehat{V}_E(t')$, and using (13), the second term requires the derivative

$$\begin{aligned}&\frac{\delta}{\delta E^\alpha(t_1)}\widehat{V}_E(t)\\ &= \frac{ie}{\hbar}\int d\omega_1\frac{e^{-i\omega_1(t-t_1)}}{\omega_1}\\ &\quad\times\sum_{n=0}^\infty\frac{1}{n!}\prod_{k=1}^n\int d\omega_k e^{-i\omega_k t}E^{\alpha_k}\widehat{h}^{\alpha\alpha_1\ldots\alpha_n}\\ &= \frac{ie}{\hbar}\int d\omega_1\frac{e^{-i\omega_1(t-t_1)}}{\omega_1}\widehat{h}^\alpha + O(\boldsymbol{E}).\end{aligned} \tag{33}$$

Starting from (18) for the first term of (32), one computes

$$\begin{aligned}&\frac{\delta}{\delta E^\alpha(t_1)}\widehat{v}^\mu_E(t)\\ &= \frac{ie}{\hbar}\int d\omega_1\frac{e^{-i\omega_1(t-t_1)}}{\omega_1}\\ &\quad\times\sum_{n=0}^\infty\frac{1}{n!}\prod_{k=1}^n\int d\omega_k e^{-i\omega_k t}E^{\alpha_k}\widehat{h}^{\mu\alpha\alpha_1\ldots\alpha_n}(t)\\ &= \frac{ie}{\hbar}\int d\omega_1\frac{e^{-i\omega_1(t-t_1)}}{\omega_1}\widehat{h}^{\mu\alpha}(t) + O(\boldsymbol{E}).\end{aligned} \tag{34}$$

Hence the conductivity is

$$\begin{aligned}&\sigma^{\mu\alpha}(t;t_1)\\ &= -e\int dt'\frac{ie}{\hbar}\int d\omega_1\frac{e^{-i\omega_1(t'-t_1)}}{\omega_1}\left\langle\widehat{h}^\mu(t)\widehat{h}^\alpha(t')\right\rangle\\ &\quad+ e\frac{ie}{\hbar}\int d\omega_1\frac{e^{-i\omega_1(t-t_1)}}{\omega_1}\left\langle\widehat{h}^{\mu\alpha}(t)\right\rangle,\end{aligned} \tag{35}$$

where brackets denote expectations with respect to the *unperturbed* Hamiltonian.

To proceed, we must evaluate the expectation values in terms of the electron propagator

$$\langle c^\dagger_{\boldsymbol{k}a}(t)c_{\boldsymbol{k}b}(t')\rangle = \delta_{ab}\int d\omega\ e^{i\omega(t-t')}G_{\boldsymbol{k}a}(\omega). \tag{36}$$

Hence

$$\left\langle\widehat{h}^{\mu\alpha}(t)\right\rangle = \sum_{a,b}\int[d\boldsymbol{k}]\ \langle c^\dagger_{ka}(t)h^{\mu\alpha}_{ab}c_{kb}(t)\rangle \tag{37}$$

$$= \sum_a\int[d\boldsymbol{k}]\ h^{\mu\alpha}_{aa}\int d\omega\ G_{ka}(\omega) \tag{38}$$

and, applying Wick's theorem,

$$\begin{aligned}&\left\langle\widehat{h}^\mu(t)\widehat{h}^\alpha(t')\right\rangle\\ &= \sum_{a,b,c,d}\int[d\boldsymbol{k}]\ \langle c^\dagger_{ka}(t)h^\mu_{ab}c_{kb}(t)c^\dagger_{kc}(t')h^\alpha_{cd}c_{kd}(t')\rangle\\ &= \sum_{a,b}\int[d\boldsymbol{k}]h^\mu_{ba}G_{ka}(t-t')h^\alpha_{ab}G_{kb}(t'-t)\\ &= -\sum_{a,b}\int[d\boldsymbol{k}]\int d\omega'' e^{-i\omega''(t_1-t')}\int d\omega' e^{-i\omega'(t'-t_1)}\\ &\quad\times h^\mu_{ba}G_{ka}(\omega'')h^\alpha_{ab}G_{kb}(\omega').\end{aligned} \tag{39}$$

In the last step we dropped terms corresponding to disconnected diagrams, which contribute zero in expectation.

We have now reduced everything to the propagators and matrix elements of derivatives of the Hamiltonian—the elements present in the Feynman rules. The last step is to Fourier transform the conductivity to frequency-space to eliminate exponential factors. Thus

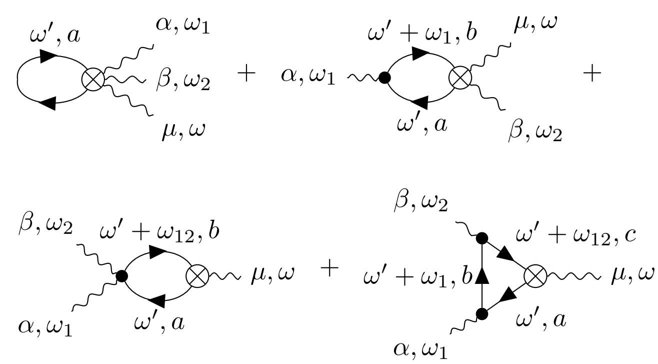
$$\begin{aligned}&\sigma^{\mu\alpha}(\omega;\omega_1)\\ &= \int\frac{dt}{2\pi}e^{i\omega t}\int\frac{dt'}{2\pi}e^{i\omega_1 t_1}\frac{ie^2}{\hbar}\left[\int d\omega_1\frac{e^{-i\omega_1(t-t_1)}}{\omega-1}\left\langle\widehat{h}^{\mu\alpha}(t)\right\rangle\right.\\ &\qquad\left.-\int dt'\int d\omega_1\frac{e^{-i\omega_1(t'-t_1)}}{\omega_1}\left\langle\widehat{h}^\mu(t)\widehat{h}^\alpha(t')\right\rangle\right]\\ &= \frac{ie^2}{\hbar\omega}\sum_{a,b}\int[d\boldsymbol{k}]\left[\int d\omega'\ h^{\mu\alpha}_{aa}G_a(\omega')\right.\\ &\qquad\left.+\int d\omega'\ h^\alpha_{ab}G_a(\omega')h^\mu_{ba}G_b(\omega'-\omega_1)\right]\delta(\omega-\omega_1)\end{aligned} \tag{40}$$

where in the first step all the t integrals have been performed to create δ-functions in frequency, eliminating t, t', ω' and ω'' and $\omega_1 \to \omega'$ in the second step. This precisely matches (23), which was obtained immediately by Feynman diagrams. The diagrams serve to eliminate the tedious steps of collapsing Fourier transforms into delta functions, thereby greatly streamlining calculations.

IV. SECOND ORDER RESPONSE

We now turn to the second-order response and demonstrate the second-order conductivity is concisely reproduced by the diagramatic formalism. There are four diagrams that contribute:

$$\sigma^{\mu\alpha\beta}(\omega;\omega_1,\omega_2) = \tag{41}$$



$$\begin{aligned}
&+ \big((\alpha,\omega_1) \leftrightarrow (\beta,\omega_2)\big) \\
&= e\left(\frac{ie}{\hbar\omega_1}\right)\left(\frac{ie}{2\hbar\omega_2}\right)\sum_a \int [d\boldsymbol{k}] \int d\omega' G_a(\omega') h^{\mu\alpha\beta}_{aa} \\
&+ \frac{-e^3}{\hbar^2\omega_1\omega_2}\sum_{a,b}\int [d\boldsymbol{k}] \int d\omega' G_a(\omega') h^{\alpha}_{ab} G_b(\omega'+\omega_1) h^{\mu\beta}_{ba} \\
&+ \frac{-e^3}{2\hbar^2\omega_1\omega_2}\sum_{a,b}\int [d\boldsymbol{k}] \int d\omega' G_a(\omega') h^{\alpha\beta}_{ab} G_b(\omega'+\omega_{12}) h^{\mu}_{ba} \\
&+ \frac{-e^3}{\hbar^2\omega_1\omega_2}\sum_{a,b,c}\int [d\boldsymbol{k}] \int d\omega' G_a(\omega') h^{\alpha}_{ab} G_b(\omega'+\omega_1) h^{\beta}_{bc} \\
&\qquad\qquad \times G_c(\omega'+\omega_{12}) h^{\mu}_{ca} + \big((\alpha,\omega_1) \leftrightarrow (\beta,\omega_2)\big).
\end{aligned}$$

There is an overall constraint $\omega = \omega_{12} \equiv \omega_1 + \omega_2$.

The frequency integrals, which are called I_1, I_2, and I_3 are performed in Appendix B. Indeed, only the triangle diagram contributes a new integral, I_3, computed in (B24); the others appeared at first order. The most convenient form of I_3 depends on the situation. For instance, one can use partial fractions to split each term into separate resonances. However, for now we adopt a more compact representation with a triple resonance:

$$I_3(\omega_1,\omega_2) = \frac{(\omega_1 - \varepsilon_{bc})f_{ab} + (\omega_2 - \varepsilon_{ba})f_{cb}}{(\omega_2 - \varepsilon_{ba})(\omega_1 - \varepsilon_{cb})(\omega - \varepsilon_{ca})} \tag{42}$$

We thus arrive at a formula for the second-order conductivity

$$\begin{aligned}
&\sigma^{\mu\alpha\beta}(\omega;\omega_1,\omega_2) = \qquad\qquad (43)\\
&\frac{-e^3}{\hbar^2\omega_1\omega_2}\sum_{a,b,c}\int [d\boldsymbol{k}]\; \frac{1}{2} f_a h^{\mu\alpha\beta}_{aa} + f_{ab}\frac{h^{\alpha}_{ab}h^{\mu\beta}_{ba}}{\omega_1 - \varepsilon_{ab}} + f_{ab}\frac{\frac{1}{2}h^{\alpha\beta}_{ab}h^{\mu}_{ba}}{\omega - \varepsilon_{ab}} \\
&\qquad + h^{\alpha}_{ab}h^{\beta}_{bc}h^{\mu}_{ca}\frac{(\omega_1 - \varepsilon_{bc})f_{ab} + (\omega_2 - \varepsilon_{ba})f_{cb}}{(\omega_1 - \varepsilon_{cb})(\omega_2 - \varepsilon_{ba})(\omega - \varepsilon_{ca})} \\
&\qquad + \big((\alpha,\omega_1) \leftrightarrow (\beta,\omega_2)\big)
\end{aligned}$$

As above, the sum over bands a, b, c should only be employed when necessary. For instance, the term $f_a h^{\mu\alpha\beta}_{aa}$ is only summed over a, and not b, c.

Let us pause for a moment to interpret the structure of this formula. Each term is a product of a matrix-element part and a resonance part from one of the integrals I_1, I_2 or I_3. This natural separation allows us to easily consider various physical limits, wherein the resonance structure simplifies but the matrix elements remain unchanged. The terms are arranged by powers of ω. The first term corresponds to the derivative of the Drude weight, the "Drude weight dipole". The second and third terms are one- and two-photon resonances respectively, which are large when two bands are separated by energies of ω_1 or $\omega_1 + \omega_2$. The last term, corresponding to the triangle diagram, is more complex. We will see below that it is still the sum of one-photon and two-photon resonances.[41] Also note that there is an overall pole $(\omega_1\omega_2)^{-1}$. Except in the first term, this second order pole is only an apparent divergence, and dissappears when the $\omega \to 0$ limit is carefully taken. However, the divergence is physical in the first term. Section VI considers this point carefully.

To provide convenient equations for important limits, as well as to gain a better understanding of the resonance structure, we next examine two limits: second-harmonic generation and the shift current. Again, this merely amounts to taking the limit of the resonance integrals I_2 and I_3, and other limits can be carried out with comparable ease.

A. Second-Harmonic Generation

The second-harmonic response is generated by both one-photon and two-photon resonances. That is, if the incident light is at frequency ω, then there will be a second-order response at both ω and 2ω. One-photon resonances come from the second and fourth diagrams in (41), while two-photon resonances are due to the third and fourth diagrams. The first diagram only contributes resonantly near $\omega = 0$. To capture these resonances carefully, we use an alternative form for the integral I_3 which makes them manifest.

Defining $\rho_1 = \omega_1/\omega, \rho_2 = \omega_2/\omega$. We may apply the partial fraction identity

$$\frac{1}{(A-\omega_1)(B-\omega_2)} = \frac{1}{(A-\rho_1 B)\,(B-\omega_2)} - \frac{\rho_1}{(A-\rho_1 B)(A-\omega_1)} \tag{44}$$

to write

$$\begin{aligned} I_3(\omega_1,\omega_2) &= \frac{1}{\omega-\varepsilon_{ca}}\left[\frac{f_a}{\varepsilon_{ab}-\rho_1\varepsilon_{ac}} + \frac{f_c}{\varepsilon_{bc}-\rho_2\varepsilon_{ac}}\right] \\ &+ \frac{\rho_1}{\omega_1-\varepsilon_{ba}}\left[\frac{f_a}{\varepsilon_{ba}+\rho_1\varepsilon_{ac}} + \frac{f_b}{\rho_2\varepsilon_{ab}+\rho_1\varepsilon_{cb}}\right] \\ &+ \frac{\rho_2}{\omega_2-\varepsilon_{cb}}\left[\frac{f_c}{\varepsilon_{cb}+\rho_2\varepsilon_{ac}} - \frac{f_b}{\rho_1\varepsilon_{cb}+\rho_2\varepsilon_{ab}}\right]. \end{aligned} \tag{45}$$

Here the first-term is a resonance due to absorbing both photons simultaneously, while the latter two are resonances in ω_1 or ω_2 only. So far, this is general, and can be used in (43) in place of (42). In the case of second harmonic generation, we take $\omega_1 = \omega_2 = \omega$, so $\rho_1 = \rho_2 = \frac{1}{2}$. After several cancellations,

$$\begin{aligned} &I_3(\omega,\omega) \\ &= \frac{1}{\varepsilon_{ab}+\varepsilon_{cb}}\left[\frac{2f_{ac}}{2\omega-\varepsilon_{ca}} + \frac{f_{bc}}{\omega-\varepsilon_{cb}} + \frac{f_{ba}}{\omega-\varepsilon_{ba}}\right]. \end{aligned} \tag{46}$$

Starting from the general equation (43), using $I_3(\omega,\omega)$, and writing out the frequency-symmetrization $(\alpha,\omega_0) \leftrightarrow (\beta,\omega_0)$ yields

$$\begin{aligned} &\sigma^{\mu\alpha\beta}(2\omega;\omega,\omega) = \\ &-\frac{e^3}{2\hbar^2\omega^2}\sum_{a,b,c}\int [d\boldsymbol{k}]\; f_a h^{\mu\alpha\beta}_{aa} + f_{ab}\frac{h^{\alpha}_{ab}h^{\mu\beta}_{ba} + h^{\beta}_{ab}h^{\mu\alpha}_{ba}}{\omega-\varepsilon_{ab}} \\ &\quad + f_{ab}\frac{h^{\alpha\beta}_{ab}h^{\mu}_{ba}}{2\omega-\varepsilon_{ab}} + \frac{\left(h^{\alpha}_{ab}h^{\beta}_{bc} + h^{\alpha}_{ab}h^{\beta}_{bc}\right)h^{\mu}_{ca}}{\varepsilon_{ab}+\varepsilon_{cb}} \\ &\quad \times \left[\frac{2f_{ac}}{2\omega-\varepsilon_{ca}} + \frac{f_{bc}}{\omega-\varepsilon_{cb}} + \frac{f_{ba}}{\omega-\varepsilon_{ba}}\right]. \end{aligned} \tag{47}$$

This result is equivalent to velocity-gauge formulas for the second-harmonic present in the literature[27,29], but did not involve any sum rules.

B. Shift Current

Another interesting limit to consider is the so-called shift current, $\sigma^{\mu\alpha\beta}(0;\omega,-\omega)$. It can be thought of as the "solar panel" response where incident light generates a DC current, and has been of recent interest in the context of two-band systems where it has a particularly simple form[9].

As with the second harmonic, the only real task is to determine what happens to the pole structure. Starting from (45), one finds

$$I_3(\omega,-\omega) = \frac{1}{\varepsilon_{ac}}\left[\frac{f_{ab}}{(\omega-\varepsilon_{ba})} - \frac{f_{cb}}{(\omega-\varepsilon_{bc})}\right]. \tag{48}$$

Then, symmetrizing explicitly,

$$\begin{aligned} &\sigma^{\mu\alpha\beta}(0;\omega,-\omega) = \\ &\quad \frac{e^3}{\hbar^2\omega^2}\sum_{a,b,c}\int [d\boldsymbol{k}]\; f_a h^{\mu\alpha\beta}_{aa} + f_{ab}\frac{h^{\alpha}_{ab}h^{\mu\beta}_{ba}}{\omega-\varepsilon_{ab}} \\ &\quad + f_{ab}\frac{h^{\beta}_{ab}h^{\mu\alpha}_{ba}}{-\omega-\varepsilon_{ab}} + f_{ab}\frac{h^{\alpha\beta}_{ab}h^{\mu}_{ba}}{\varepsilon_{ba}} \\ &\quad + \frac{h^{\alpha}_{ab}h^{\beta}_{bc}h^{\mu}_{ca}}{\varepsilon_{ac}}\left[\frac{f_{ab}}{(\omega-\varepsilon_{ba})} - \frac{f_{cb}}{(\omega-\varepsilon_{bc})}\right] \\ &\quad + \frac{h^{\beta}_{ab}h^{\alpha}_{bc}h^{\mu}_{ca}}{\varepsilon_{ac}}\left[\frac{f_{ab}}{(-\omega-\varepsilon_{ba})} - \frac{f_{cb}}{(-\omega-\varepsilon_{bc})}\right]. \end{aligned} \tag{49}$$

This result agrees with known expressions for the shift current found in the literature. One can easily check this reduces to the correct two-band limit that has been studied in previous work[9]. It is worth contrasting this result to the alternative (but equivalent) length-gauge results in, e.g. Ref.[5]. The results there involve a maximum of two bands in each term, whereas here there are three band terms. Converting between the two gauges requires the use of sum rules, which exchange some intraband matrix elements with interband ones, and visa versa. Specifically, one can convert to the shift current formula in[5] by focusing on the interband resonance from the band a to b, where we collect terms involving $\frac{f_{ab}}{(\omega-\varepsilon_{ba})}$ after switching indices $a \leftrightarrow c$ in the term $\frac{f_{cb}}{(\omega-\varepsilon_{bc})}$ and use the second-order sum rule in Eq. (13) of[8]. Moreover, since the conductivity does not depend on the choice of gauge, one may conclude that saying a particular term involves a certain number of bands is gauge-dependent information and therefore not necessarily physical. This demonstates Eq. (49) is equivalent to previously known expressions for the shift current in the literature.

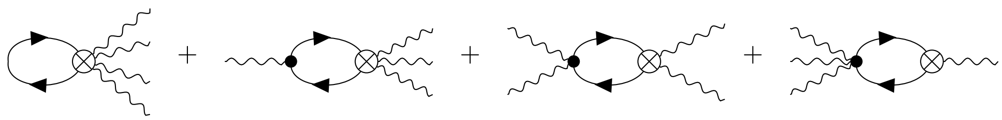
The injection current is a second-order process that describes a current whose magnitude grows linearly in time as the sample is illuminated. This process is manifest in in Ref.[5] as a term with an overall $1/\delta\omega$ divergence in the nonlinear conductivity $\sigma(\delta\omega;\omega+\delta\omega,-\omega)$. In our framework, the naive limit of Equation (49) does not contain this divergence, but it can be recovered by considering the limit $\delta\omega \to 0$ of $\sigma(\delta\omega;\omega+\delta\omega,-\omega)$ in Eq. (43). Specifically, the last two terms in Eq. (49) are related to the injection current since the factor $1/\epsilon_{ac}$ is divergent for $a = c$ which is cutoff by introducing a small $\delta\omega$.

V. THIRD ORDER RESPONSE

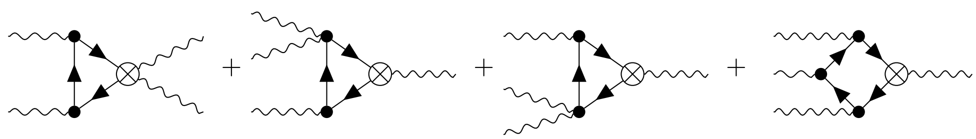
It is at third order that the diagrammatic method espoused here becomes the most useful. Unlike at second-order, the third order response is generically allowed by symmetry and expected to be present to some degree in all materials. Third order optical responses are relatively unstudied, especially in the case of non-zero Berry connection. Understanding this area is our main focus in this paper. Using our diagrammatic formalism, we derive expressions for the third-order response where each term is associated with an individual process. This allows them to be split into manifestly one-photon, two-photon, and three-photons parts, so that the origin of each resonance is clear. We then examine the limits of third harmonic generation and self-focusing of light. In subsequent sections we begin to interpret these formulas in the one band (semiclassical) and two-band (Weyl semimetal) limits.

There are eight diagrams that contribute at third order.

$$\sigma^{\mu\alpha\beta\gamma}(\omega;\omega_1,\omega_2,\omega_3) = \tag{50}$$



(51)



(52)

$$= \frac{e}{3!\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\mu\alpha\beta\gamma}_{aa} \tag{53}$$

$$+ \frac{e}{2!\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a,b} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\alpha}_{ab} G_b(\omega'+\omega_1) h^{\mu\beta\gamma}_{ba} \tag{54}$$

$$+ \frac{e}{2!\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a,b} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\alpha\beta}_{ab} G_b(\omega'+\omega_{12}) h^{\mu\gamma}_{ba} \tag{55}$$

$$+ \frac{e}{3!\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a,b} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\alpha\beta\gamma}_{ab} G_a(\omega'+\omega_{123}) h^{\mu}_{ba} \tag{56}$$

$$+ \frac{e}{\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a,b,c} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\alpha}_{ab} G_b(\omega'+\omega_1) h^{\beta}_{bc} G_c(\omega'+\omega_{12}) h^{\mu\gamma}_{bc} \tag{57}$$

$$+ \frac{e}{2!\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a,b,c} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\alpha}_{ab} G_a(\omega'+\omega_1) h^{\beta\gamma}_{bc} G_c(\omega'+\omega_{123}) h^{\mu}_{ca} \tag{58}$$

$$+ \frac{e}{2!\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a,b,c} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\alpha\beta}_{ab} G_a(\omega'+\omega_{12}) h^{\gamma}_{bc} G_c(\omega'+\omega_{123}) h^{\mu}_{ca} \tag{59}$$

$$+ \frac{e}{\omega_1\omega_2\omega_3}\left(\frac{ie}{\hbar}\right)^3 \sum_{a,b,c,d} \int [d\boldsymbol{k}] \int d\omega'\ G_a(\omega') h^{\alpha}_{ab} G_b(\omega'+\omega_1) h^{\beta}_{bc} G_c(\omega'+\omega_{12}) h^{\gamma}_{cd} G_d(\omega'+\omega_{123}) h^{\mu}_{da} \tag{60}$$

The ω' integrals are evaluated in Appendix B. For concision, however, we shall leave the expression in terms of I_3 and I_4. We must also symmetrize under all possible combinations of incoming photons, which amounts to the six permutations of (α,ω_1), (β,ω_2), and (γ,ω_3). We will denote this permutation symmetry by $\frac{1}{3!}S_3$. The full third-order non-linear response is thus

$$\begin{aligned}
&\sigma^{\mu\alpha\beta\gamma}(\omega;\omega_1,\omega_2,\omega_3) = \\
&\quad \frac{1}{3!}S_3 \frac{-ie^4}{\hbar^3\omega_1\omega_2\omega_3} \sum_{a,b,c,d} \int [d\boldsymbol{k}]\ \frac{1}{6} f_a h^{\mu\alpha\beta\gamma}_{aa} + \frac{\frac{1}{2} f_{ab} h^{\alpha}_{ab} h^{\mu\beta\gamma}_{ba}}{\omega_1 - \varepsilon_{ab}} + \frac{\frac{1}{2} f_{ab} h^{\alpha\beta}_{ab} h^{\mu\gamma}_{ba}}{\omega_{12} - \varepsilon_{ab}} + \frac{\frac{1}{6} f_{ab} h^{\alpha\beta\gamma}_{ab} h^{\mu}_{ba}}{\omega - \varepsilon_{ab}} \\
&\quad + h^{\alpha}_{ab} h^{\beta}_{bc} h^{\mu\gamma}_{bc} I_3(\omega_1,\omega_2) + \frac{1}{2} h^{\alpha\beta}_{ab} h^{\gamma}_{bc} h^{\mu}_{bc} I_3(\omega_{12},\omega_3) + \frac{1}{2} h^{\alpha}_{ab} h^{\beta\gamma}_{bc} h^{\mu}_{bc} I_3(\omega_1,\omega_{23}) + h^{\alpha}_{ab} h^{\beta}_{bc} h^{\gamma}_{cd} h^{\mu}_{da} I_4(\omega_1,\omega_2,\omega_3).
\end{aligned} \tag{61}$$

A. Third-Harmonic Generation

One physical limit of interest is third-harmonic generation, when there is a single incoming frequency. There are many simplifications in this case, giving rise to a relatively simple expression. In particular, the integral for the box diagram, I_4, can be separated into one-, two- and three-photon resonances as

$$\begin{aligned} I_4(\omega,\omega,\omega) = &\frac{f_{ab}}{(\omega-\varepsilon_{ba})(\varepsilon_{ab}+\varepsilon_{cb})(2\varepsilon_{ab}+\varepsilon_{db})} + \frac{f_{bc}}{(\omega-\varepsilon_{cb})(\varepsilon_{ab}+\varepsilon_{cb})(\varepsilon_{bc}+\varepsilon_{dc})} + \frac{f_{dc}}{(\omega-\varepsilon_{dc})(\varepsilon_{cb}+\varepsilon_{cd})(2\varepsilon_{dc}+\varepsilon_{ac})} \\ &+ \frac{4f_{db}}{(2\omega-\varepsilon_{db})(2\varepsilon_{ab}+\varepsilon_{db})(\varepsilon_{bc}+\varepsilon_{dc})} + \frac{4f_{ca}}{(2\omega-\varepsilon_{ca})(\varepsilon_{cb}+\varepsilon_{ab})(2\varepsilon_{dc}+\varepsilon_{ac})} + \frac{9f_{da}}{(3\omega-\varepsilon_{da})(2\varepsilon_{ab}+\varepsilon_{db})(2\varepsilon_{dc}+\varepsilon_{ac})}. \end{aligned} \tag{62}$$

We can similarly decompose the integrals for the triangle diagrams. The case $I_3(\omega,\omega)$ is given in Equation (47). Similarly,

$$I_3(\omega,2\omega) = \frac{1}{2\varepsilon_{ab}+\varepsilon_{cb}}\left[\frac{3f_{ac}}{3\omega-\varepsilon_{ca}} + \frac{2f_{cb}}{2\omega-\varepsilon_{cb}} + \frac{f_{ba}}{\omega-\varepsilon_{ba}}\right] \tag{63}$$

$$I_3(2\omega,\omega) = \frac{1}{2\varepsilon_{cb}+\varepsilon_{ab}}\left[\frac{3f_{ac}}{3\omega-\varepsilon_{ca}} + \frac{2f_{ba}}{2\omega-\varepsilon_{ba}} + \frac{f_{cb}}{\omega-\varepsilon_{cb}}\right]. \tag{64}$$

Combining these and applying the permutation symmetry yields

$$\sigma^{\mu\alpha\beta\gamma}(3\omega;\omega,\omega,\omega) = \tag{65}$$

$$\frac{-ie^4}{\hbar^3\omega^3}\sum_{a,b,c,d}\int[d\boldsymbol{k}]\; f_a h^{\mu\alpha\beta\gamma}_{aa} + f_{ab}\left[\frac{h^{\alpha}_{ab}h^{\mu\beta\gamma}_{ba} + h^{\beta}_{ab}h^{\mu\gamma\alpha}_{ba} + h^{\gamma}_{ab}h^{\mu\alpha\beta}_{ba}}{\omega-\varepsilon_{ab}}\right] + f_{ab}\left[\frac{h^{\alpha\beta}_{ab}h^{\mu\gamma}_{ba} + h^{\beta\gamma}_{ab}h^{\mu\beta}_{ba} + h^{\gamma\alpha}_{ab}h^{\mu\alpha}_{ba}}{2\omega-\varepsilon_{ab}}\right] \tag{66}$$

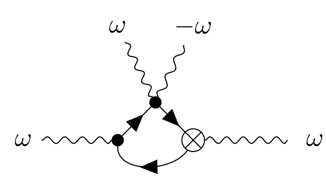
$$+\left[\left(h^{\alpha}_{ab}h^{\beta}_{bc} + h^{\beta}_{ab}h^{\alpha}_{bc}\right)h^{\mu\gamma}_{ca} + \left(h^{\beta}_{ab}h^{\gamma}_{bc} + h^{\gamma}_{ab}h^{\beta}_{bc}\right)h^{\mu\beta}_{ca} + \left(h^{\gamma}_{ab}h^{\alpha}_{bc} + h^{\alpha}_{ab}h^{\gamma}_{bc}\right)h^{\mu\alpha}_{ca}\right] I_3(\omega,\omega) \tag{67}$$

$$+\left[h^{\alpha}_{ab}h^{\beta\gamma}_{bc}h^{\mu}_{ca} + h^{\beta}_{ab}h^{\gamma\alpha}_{bc}h^{\mu}_{ca} + h^{\gamma}_{ab}h^{\alpha\beta}_{bc}h^{\mu}_{ca}\right]\Big(I_3(\omega,2\omega) + I_3(-\omega,-2\omega)\Big) \tag{68}$$

$$+\left[h^{\alpha}_{ab}h^{\beta}_{bc}h^{\gamma}_{cd} + h^{\alpha}_{ab}h^{\gamma}_{bc}h^{\beta}_{cd} + h^{\beta}_{ab}h^{\gamma}_{bc}h^{\alpha}_{cd} + h^{\beta}_{ab}h^{\alpha}_{bc}h^{\gamma}_{cd} + h^{\gamma}_{ab}h^{\alpha}_{bc}h^{\beta}_{cd} + h^{\gamma}_{ab}h^{\beta}_{bc}h^{\alpha}_{cd}\right] h^{\mu}_{da} I_4(\omega,\omega,\omega). \tag{69}$$

B. Self-Focusing

Another common third-order response is the self-focusing of light, which is the modification to the linear conductivitiy due to nonlinear effects. For instance, the process wherein an excited electron absorbs photons of energy ω and then $-\omega$,



(70)

can masquarade as the diagram for first order conductivity from Equation (23). To describe this effect, one can define the effective conductivity, via $\langle J^\mu\rangle(\omega) = \sigma^{\mu\alpha}_{\text{eff}}(\omega)E^\alpha(\omega)$ where

$$\sigma^{\mu\alpha}_{\text{eff}}(\omega) = \sigma^{\mu\alpha}(\omega) + \sigma^{\mu\alpha\beta\gamma}(\omega;\omega,-\omega,\omega)E^\beta(-\omega)E^\gamma(\omega) + O(E^4). \tag{71}$$

The third-order correction term, $\sigma^{\mu\alpha\beta\gamma}(\omega;\omega,-\omega,\omega)$, is also called the self-focusing effect.

In the self-focusing limit, the conductivity is a sum of resonances at $0\omega, 1\omega$ and 2ω, corresponding to the sums and differences of the incident frequencies. Unfortunately, the minus sign from the $-\omega$ photons lifts the permutation symmetry between the various incident photons, creating a more complex resonance structure than in the third-harmonic case. It is convenient to express the conductivity in terms of the following expressions:

$$I_3(\omega,-\omega) = \frac{1}{\varepsilon_{ac}}\left[\frac{f_{ab}}{\omega-\varepsilon_{ba}} + \frac{f_{bc}}{\omega-\varepsilon_{bc}}\right],\quad I_3(0,\omega) = \frac{1}{\varepsilon_{ab}}\left[\frac{f_{ac}}{\omega-\varepsilon_{ca}} + \frac{f_{cb}}{\omega-\varepsilon_{cb}}\right],\quad I_3(\omega,0) = \frac{1}{\varepsilon_{bc}}\left[\frac{f_{ab}}{\omega-\varepsilon_{ba}} + \frac{f_{ca}}{\omega-\varepsilon_{ca}}\right], \tag{72}$$

and $I_3(\omega,\omega)$, which is given by (47). For the box diagram, one must also consider

$$\begin{aligned}
&I_4(-\omega,\omega,\omega) &&(73)\\
&= \frac{1}{\varepsilon_{ac}(\varepsilon_{ab}+\varepsilon_{ad})}\left[\frac{f_{ba}}{\omega-\varepsilon_{ab}}+\frac{f_{ad}}{\omega-\varepsilon_{da}}\right]+\frac{1}{\varepsilon_{ac}(\varepsilon_{cb}+\varepsilon_{cd})}\left[\frac{f_{cb}}{\omega-\varepsilon_{dc}}+\frac{f_{cd}}{\omega-\varepsilon_{dc}}\right]+\frac{4f_{db}}{(2\omega-\varepsilon_{db})\,(\varepsilon_{bc}+\varepsilon_{dc})(\varepsilon_{ab}+\varepsilon_{ad})},\\
&I_4(\omega,-\omega,\omega) = \frac{1}{\varepsilon_{ac}\varepsilon_{bd}}\left[\frac{f_{ab}}{\omega-\varepsilon_{ba}}+\frac{f_{bc}}{\omega-\varepsilon_{bc}}+\frac{f_{da}}{\omega-\varepsilon_{da}}+\frac{f_{cd}}{\omega-\varepsilon_{dc}}\right] &&(74)\\
&I_4(\omega,\omega,-\omega) &&(75)\\
&= \frac{1}{\varepsilon_{db}(\varepsilon_{ad}+\varepsilon_{ac})}\left[\frac{f_{cd}}{\omega-\varepsilon_{cd}}+\frac{f_{ad}}{\omega-\varepsilon_{da}}\right]+\frac{1}{\varepsilon_{bd}\,(\varepsilon_{ab}+\varepsilon_{ac})}\left[\frac{f_{ba}}{\omega-\varepsilon_{ba}}+\frac{f_{cb}}{\omega-\varepsilon_{cb}}\right]+\frac{4f_{ac}}{(2\omega-\varepsilon_{ca})\,(\varepsilon_{ab}+\varepsilon_{cb})\,(\varepsilon_{ad}+\varepsilon_{cd})}.
\end{aligned}$$

Then, applying the symmetrization over $(\alpha,\omega),(\beta,-\omega),(\gamma,\omega)$, the self-focusing is

$$\begin{aligned}
&\sigma^{\mu\alpha\beta\gamma}(\omega;\omega,-\omega,\omega) = &&(76)\\
&\frac{ie^4}{3!\hbar^3\omega^3}\sum_{a,b,c,d}\int[d\boldsymbol{k}]f_a h_{aa}^{\mu\alpha\beta\gamma}+f_{ab}\left[\frac{h_{ab}^{\alpha}h_{ba}^{\mu\beta\gamma}}{\omega-\varepsilon_{ab}}+\frac{h_{ab}^{\beta}h_{ba}^{\mu\gamma\alpha}}{(-\omega)+\varepsilon_{ab}}+\frac{h_{ab}^{\gamma}h_{ba}^{\mu\alpha\beta}}{\omega-\varepsilon_{ab}}\right]\\
&+f_{ab}\left[\frac{h_{ab}^{\alpha\beta}h^{\mu\gamma}}{0-\varepsilon_{ab}}+\frac{h_{ab}^{\beta\gamma}h_{ba}^{\mu\alpha}}{0-\varepsilon_{ab}}+\frac{h_{ab}^{\gamma\alpha}h_{ba}^{\mu\beta}}{2\omega-\varepsilon_{ab}}\right]+f_{ab}\frac{h_{ab}^{\alpha\beta\gamma}h_{ba}^{\mu}}{\omega-\varepsilon_{ab}}\\
&+\Big[\left(h_{ab}^{\alpha}h_{bc}^{\beta}h_{ca}^{\mu\gamma}+h_{ab}^{\gamma}h_{bc}^{\beta}h_{ca}^{\mu\alpha}\right)I_3(\omega,-\omega)+\left(h_{ab}^{\beta}h_{bc}^{\alpha}h_{ca}^{\mu\gamma}+h_{ab}^{\beta}h_{bc}^{\gamma}h_{ca}^{\mu\alpha}\right)I_3(-\omega,\omega)+\left(h_{ab}^{\gamma}h_{bc}^{\alpha}h_{ca}^{\mu\beta}+h_{ab}^{\alpha}h_{bc}^{\gamma}h_{ca}^{\mu\beta}\right)I_3(\omega,\omega)\Big]\\
&+\Big[\left(h_{ab}^{\alpha\beta}h_{bc}^{\gamma}h_{ca}^{\mu}+h_{ab}^{\beta\gamma}h_{bc}^{\alpha}h_{ca}^{\mu}\right)I_3(0,\omega)+h_{ab}^{\gamma\alpha}h_{bc}^{\beta}h_{ca}^{\mu}I(2\omega,-\omega)\Big]\\
&+\Big[\left(h_{ab}^{\alpha}h_{bc}^{\beta\gamma}h_{ca}^{\mu}+h_{ab}^{\gamma}h_{bc}^{\alpha\beta}h_{ca}^{\mu}\right)I(\omega,0)+h_{ab}^{\beta}h_{bc}^{\gamma\alpha}h_{ca}^{\mu}I(-\omega,2\omega)\Big]\\
&+\Big[\left(h_{ab}^{\alpha}h_{bc}^{\beta}h_{cd}^{\gamma}+h_{ab}^{\gamma}h_{bc}^{\beta}h_{cd}^{\alpha}\right)h_{da}^{\mu}I_4(\omega,-\omega,\omega)+\left(h_{ab}^{\beta}h_{bc}^{\gamma}h_{cd}^{\alpha}+h_{ab}^{\beta}h_{bc}^{\alpha}h_{cd}^{\gamma}\right)h_{da}^{\mu}I_4(-\omega,\omega,\omega)\\
&\quad+\left(h_{ab}^{\gamma}h_{bc}^{\alpha}h_{cd}^{\beta}+h_{ab}^{\alpha}h_{bc}^{\gamma}h_{cd}^{\beta}\right)h_{da}^{\mu}I_4(-\omega,\omega,\omega)\Big]
\end{aligned}$$

Note that there is an exact permutation symmetry $\alpha\leftrightarrow\gamma$ since the second and forth frequencies are both ω.

VI. SEMICLASSICAL LIMIT

This section carefully examines the semiclassical limit of non-linear optical responses. This crucial physical limit, where on focusing on independent bands in the limit $\omega\to 0$, has been the subject of much recent work, as described in the introduction. The goal of this section is to carefully take this limit. Per the discussion in Section II D, this is most easily carried out in the length gauge. The alternative is to start from the velocity gauge and apply many sum rules. However, the source of these sum rules is expanding the change of gauge which converts from velocity to length gauge[28], so it clear that the length gauge is the natural physical setting for this limit. We will start with a purely semiclassical derivation, then show that this matches the results from the length gauge, and lastly comment on the topological properties of the third-order semiclassical conductivity.

A. Semiclassical Derivation

We work with a single band and ignore interband contributions. Recall that the equations for semi-classical electron dynamics in an electric field (but no magnetic field) are given by

$$\begin{aligned}
\hbar\frac{d}{dt}\boldsymbol{r} &= \boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_{\boldsymbol{k}}+e\boldsymbol{E}\times\boldsymbol{\Omega}(\boldsymbol{k}),\\
\hbar\frac{d}{dt}\boldsymbol{k} &= -e\boldsymbol{E}
\end{aligned}$$

where $\boldsymbol{E}=\boldsymbol{E}(t)$ is the applied electric field and $\boldsymbol{\Omega}$ is the standard vector representation for the Berry curvature in three dimensions. In the notation of this paper, for a single band a, $\Omega_a^{\alpha}=\varepsilon^{\alpha\beta\gamma}\mathcal{F}_{aa}^{\beta\gamma}$ where ε is the Levi-Civita symbol, so $(\boldsymbol{E}\times\boldsymbol{\Omega})^{\mu}=\mathcal{F}^{\mu\alpha}E_{\alpha}$.

We take a Boltzmann equation approach, writing the charge and current density as, respectively,

$$\rho(t)=-i\int[d\boldsymbol{k}]f(t),\ \text{and}\ \boldsymbol{J}(t)=-e\int[d\boldsymbol{k}]\frac{d\boldsymbol{r}}{dt}f(t) \quad (77)$$

where $f=f(t,\boldsymbol{k})$ is the distribution function of electrons, and is taken to be Fermi-Dirac distribution f_{FD}

in equilibrium. The time-evolution of f is given by the Boltzmann Equation

$$\frac{d\boldsymbol{k}}{dt} \cdot \boldsymbol{\nabla}_{\boldsymbol{k}} f + \partial_t f = \frac{f_{\mathrm{FD}} - f}{\tau} \tag{78}$$

for some relaxation time τ.

We take a monochromatic perturbation $\boldsymbol{E}(t) = E^{\alpha} \boldsymbol{e}_{\alpha} e^{i\omega t}$. Expanding $f(t) = \sum_{K\in\mathbb{Z}} f^{(K)} e^{-i\omega K t}$ and equating terms of the same order in (78), we have

$$-e\boldsymbol{E} \cdot \boldsymbol{\nabla}_{\boldsymbol{k}} f^{(K)} + (-iK\omega) f^{(K+1)} = -\frac{1}{\tau} f^{(K+1)}. \tag{79}$$

With the initial condition $f^{(0)} = f_{\mathrm{FD}}$, this gives an order-by-order solution as

$$f^{(K+1)} = \frac{-i}{K\omega + i\gamma} e\boldsymbol{E} \cdot \boldsymbol{\nabla} f^{(K)} \tag{80}$$

where $\gamma = 1/\tau$ is a dissipation rate. The first-order current is thus

$$\boldsymbol{J}^{\mu}(\omega) = -e \int [d\boldsymbol{k}]\ v^{\mu} f^{(1)} + F^{\mu\alpha} E_{\alpha} f^{(0)} \tag{81}$$

so, integrating by parts, the linear conductivity is

$$\sigma^{\mu\alpha}(\omega;\omega) = \frac{e^2}{\hbar} \int [d\boldsymbol{k}]\ f_{\mathrm{FD}} \left(-i \frac{\partial^{\alpha} v^{\mu}}{\omega + i\gamma} - F^{\mu\alpha} \right) \tag{82}$$

One can check this exactly reproduces the fully quantum equation for $\sigma^{\mu\alpha}$, Equation (30), for the case of a single band.

At higher orders, the semiclassical conductivities are essentially the same, comprised of a “Drude-like” and “Berry-curvature”-like term:

$$\sigma^{\mu\alpha\beta} \ = \ e^3 \int [d\boldsymbol{k}] f_{\mathrm{FD}} \left(\frac{\partial^{\beta}\partial^{\alpha} v^{\mu}}{(2\widetilde{\omega})\widetilde{\omega}} - i \frac{\partial^{\beta} F^{\mu\alpha}}{\widetilde{\omega}} \right) \tag{83}$$

$$\sigma^{\mu\alpha\beta\gamma} \ = \ e^4 \int [d\boldsymbol{k}] f_{\mathrm{FD}} \left(i \frac{\partial^{\gamma}\partial^{\beta}\partial^{\alpha} v^{\mu}}{(3\widetilde{\omega})(2\widetilde{\omega})\widetilde{\omega}} + \frac{\partial^{\gamma}\partial^{\beta} F^{\mu\alpha}}{(2\widetilde{\omega})\widetilde{\omega}} \right) \tag{84}$$

where $\widetilde{\omega}$ should be read as $\omega + i\gamma$.

We will show that these equations reproduce the leading order divergences at $\omega \to 0$ for the quantum calculations of the second- and third-order conductivities. However, the numerous other terms in the quantum formulas are not captured here due to their essential interband nature. It would be interesting to examine a modified semiclassical picture involving interband corrections, which should be able to reproduce more of the full response.

Under time-reversal symmetry, $\boldsymbol{\nabla}_{\boldsymbol{k}}$, $\boldsymbol{v}$, F, and $\boldsymbol{k}$ all change sign, so at second order only the derivative of the Berry curvature survives, while at third order only the velocity term remains. One can extrapolate the pattern in (84) to all orders in semiclassics.

B. Length Gauge

Let us now derive the semiclassical limit starting in the length gauge formulation, (1). We adopt the standard density-matrix approach pioneered by Sipe and Shkrebtii[5], defining the single-particle reduced density matrix

$$\rho_{\boldsymbol{k}ab}(t) = \langle c^{\dagger}_{\boldsymbol{k}a}(t) c_{\boldsymbol{k}b}(t) \rangle\,. \tag{85}$$

Then the current is given by $J^{\mu}(t) = e\,\mathrm{Tr}\,[\widehat{v}^{\mu}\rho(t)]$, where the trace is taken over the single-particle Hilbert space, i.e. it stands for the integral of the Brillouin zone and sum over bands.

The time-dependence of the density matrix in the interaction picture is given by the Schwinger-Tomonaga Equation

$$i\frac{d}{dt}\widehat{\rho}_I(t) = \left[\widehat{H}_{E,I}(t), \widehat{\rho}_I(t)\right], \tag{86}$$

where the subscript I indicates the interaction picture: $\widehat{\mathcal{O}}_I(t) = U(t)^{\dagger}\widehat{\mathcal{O}}U(t)$ for $U(t) = e^{-it\widehat{H}_0}$. We can solve (86) within the framework of perturbation theory by employing Dyson series. Expand $\widehat{\rho} = \sum_K \widehat{\rho}^{(K)}$ as a power series in power of the electric field. We can then integrate (86) to find an order-by-order solution

$$\widehat{\rho}^{(K+1)}(t) = -i \int_{-\infty}^{t} d\tau\ \left[\widehat{H}_{E,I}(\tau), \widehat{\rho}^{(K)}_I(\tau)\right]. \tag{87}$$

This also requires an initial condition $\widehat{\rho}^{(0)} = \delta_{ab} f_a$, taken to be the Fermi-Dirac distribution. We can now write a computable expression for the current. At nth order, the current can be written as a nested commutator

$$\begin{aligned} J^{\mu}(t) = e \prod_{k=1}^{n} \int_{-\infty}^{\tau_{k-1}} d\tau_k (ieE^{\alpha_k}) \\ \times\ \mathrm{Tr}\left\{ [\cdots[\widehat{v}^{\mu}, \widehat{r}^{\alpha_1}(\tau_1)]\cdots, \widehat{r}^{\alpha_n}(\tau_n)]\,\widehat{\rho}_0 \right\} \end{aligned} \tag{88}$$

where $\tau_{-1} \equiv t$. Rearranging commutators and Fourier transforming, the nth nonlinear conductivity can then be written

$$\begin{aligned} \sigma^{\mu\alpha_1\ldots\alpha_n}(\omega;\omega_1,\ldots,\omega_n) = \frac{1}{n!}\mathcal{S}_n e \prod_{k=1}^{n} \int_{-\infty}^{\tau_{k-1}} d\tau_k e^{-i\omega_k\tau_k} \\ \times (ie)\,\mathrm{Tr}\left\{ \widehat{\rho}_0\,[\widehat{r}^{\alpha_n}(\tau_n), \cdots, [\widehat{r}^{\alpha_1}(\tau_1), \widehat{v}^{\mu}]\cdots] \right\} \end{aligned} \tag{89}$$

where S_n symmetrizes over all incoming frequencies $\{(\alpha_k,\omega_k)\ :\ 1 \le k \le n\}$.

Our task is now to evaluate this commutator at leading order in ω. This is done most expediently by using the relation between the position operator and the covariant derivative $\widehat{\boldsymbol{r}} = i\widehat{\boldsymbol{D}}$, which is described in Appendix A. The form of the commutators in (89) is almost the same as the covariant derivative repeatedly acting on the velocity operator—but we must account for the time dependence. The time-evolved operator $\widehat{r}(t) = U(t)^\dagger \,\widehat{r}\, U(t)$ is easily computed by noting that, in the energy basis, $U(t)_{ab} = e^{-i\varepsilon_{\boldsymbol{k}a}(t)}\delta_{ab}$ is a one-parameter family of gauge transformations. Equation (A7) implies

$$i\widehat{\boldsymbol{D}}(t) = i\boldsymbol{\nabla} + \boldsymbol{A}'(t), \boldsymbol{A}'(t) = e^{iH_0 t}\boldsymbol{A}e^{-iH_0 t} + t\nabla H_0 \quad (90)$$

where $(\boldsymbol{\nabla}\widehat{H}_0)_{ab} = \delta_{ab}\boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_a(\boldsymbol{k})$ is the regular gradient of the matrix elements. In components, this implies the identity

$$\begin{aligned}[\widehat{r}^\alpha(\tau), \mathcal{O}]_{ab} \;&=\; (i\partial^\alpha + \tau\Delta^\alpha_{ab})\,\mathcal{O}_{ab} \\ &\quad + \sum_c e^{i\varepsilon_{ac}t}A^\alpha_{ac}\mathcal{O}_{cb} - \mathcal{O}_{ac}A^\alpha_{cb}e^{i\varepsilon_{cb}t},\end{aligned} \quad (91)$$

where we have defined $\Delta^\alpha_{ab} \equiv h^\alpha_{aa} - h^\alpha_{bb} = \partial^\alpha\varepsilon_{ab}$.

Employing (91) with $\widehat{\mathcal{O}} = \widehat{v}^\mu$,

$$\begin{aligned}\sigma^{\mu\alpha}(\omega;\omega) = \frac{ie^2}{\hbar}\sum_{a,b}\int[d\boldsymbol{k}]f_a\int_{-\infty}^{0} e^{-i\omega\tau} \\ \times\left(i\partial^\alpha v^\mu + e^{i\varepsilon_{ab}\tau}A^\alpha_{ab}v^\mu_{ba} - e^{i\varepsilon_{ba}\tau}v^\mu_{ab}A^\alpha_{ba}\right).\end{aligned} \quad (92)$$

As is customary, when performing the time integral, a phenomenological relaxation rate $\omega \to \omega + i\gamma$ is added so that $\int_{-\infty}^{0} d\tau\; e^{i(\zeta-\omega-i\gamma)\tau} = \frac{i}{\omega+i\gamma-\zeta}$. Therefore the linear conductivity is

$$\sigma^{\mu\alpha}(\omega;\omega) = \frac{ie^2}{\hbar}\sum_{a\neq b}\int[d\boldsymbol{k}]f_a\frac{\partial^\alpha v^\mu}{\omega+i\gamma} + f_{ab}\frac{A^\alpha_{ab}v^\mu_{ba}}{\varepsilon_{ab}-\omega-i\gamma}. \quad (93)$$

The first term reproduce the Drude formula, and the second term is almost $[A^\alpha, A^\mu]$. This is equivalent to the expression derived in the velocity gauge, Equation (29). Under the limit $\omega \ll \varepsilon_{ab}$, one arrives at (28), which matches the semiclassical result (82) at linear order in $\boldsymbol{E}$.

At nonlinear order, one must evaluate further nested commutators. Since we are only interested in the $\omega \to 0$ limit, we will limit ourselves to the leading order terms. However, this procedure can be easily continued to give expressions for the general conductivity tensors in the length gauge; such a calculation is carried out in[28]. At second order, we consider the expression $[\widehat{r}^\beta(\tau_2), [\widehat{r}^\alpha(\tau_1), \widehat{v}^\mu]]_{aa}$. Expanding,

$$\begin{aligned}[\widehat{r}^\beta(\tau_2), [\widehat{r}^\alpha(\tau_1), \widehat{v}^\mu]]_{aa} \;&=\; i\partial^\beta i\partial^\alpha v^\mu_{aa} \\ &\quad + i\partial^\beta\left[A^\alpha(\tau_1), v^\mu\right]_{aa} \\ &\quad + \left[A^\beta(\tau_2), (r^\alpha(\tau_1)v^\mu)\right]_{aa} \\ &\quad + \left[A^\beta(\tau_2), [A^\alpha(\tau_1), v^\mu]\right]_{aa}\end{aligned} \quad (94)$$

where $A^\alpha_{ab}(\tau) \equiv e^{i\varepsilon_{ab}\tau}A^\alpha_{ab}$ is the time-evolved operator. Each factor of $e^{i\varepsilon_{ab}\tau}$ is Fourier transformed to a denominator of the form $\frac{1}{\omega-\varepsilon_{ab}}$. However, the number of such exponential factors is different in each term. The first term has none, the second term has one, and the latter terms generically have two. The Fourier transform of the first two terms is therefore

$$\frac{\partial^\beta\partial^\alpha v^\mu_{aa}}{\omega\omega_2} + \sum_{b\neq a}\frac{-i\partial^\beta}{\omega_2}\left[\frac{A^\alpha_{ab}v^\mu_{ba}}{\omega-\varepsilon_{ba}} - \frac{v^\mu_{ab}A^\alpha_{ba}}{\omega-\varepsilon_{ab}}\right]. \quad (95)$$

After Fourier transforming the third and fourth terms, either there are factors $\frac{1}{\varepsilon_{cd}-\omega}$, which are $O(\omega^0)$ and hence subleading or, when $c = d$ there are poles $\frac{1}{\omega}$, which cancel out due to the commutator in the $\omega \to 0$ limit. Hence only the terms (95) survive in the semiclassical limit, so

$$\begin{aligned}\lim_{\omega,\omega_i\to 0}\sigma^{\mu\alpha\beta}(\omega,\omega_1,\omega_2) = \frac{-e^3}{\hbar^2}\sum_{a,b}\int[d\boldsymbol{k}]\frac{f_a\partial^\beta\partial^\alpha v^\mu_{aa}}{\omega\omega_2} \\ + f_a\frac{-i\partial^\beta}{\omega_2}\left[\frac{A^\alpha_{ab}v^\mu_{ba}}{\omega-\varepsilon_{ba}} - \frac{v^\mu_{ab}A^\alpha_{ba}}{\omega-\varepsilon_{ab}}\right] + O(\omega^0)\end{aligned} \quad (96)$$

or

$$\begin{aligned}\lim_{\omega,\omega_i\to 0}\sigma^{\mu\alpha\beta}(\omega,\omega_1,\omega_2) = \frac{-e^3}{\hbar^2}\sum_{a}\int[d\boldsymbol{k}]\frac{f_a\partial^\beta\partial^\alpha v^\mu_{aa}}{\omega\omega_2} \\ + f_a\frac{-i\partial^\beta}{\omega_2}\mathcal{F}^{\alpha\mu}_{aa} + O(\omega^0).\end{aligned} \quad (97)$$

The two terms are clearly just $\frac{ie}{\hbar}\frac{\partial^\beta}{\omega_2}$ acting on the first order expression — exactly in line with the semiclassical prediction (83). The first term is the derivative of the Drude weight, while the second is the Berry curvature dipole, which was studied in semiclassics[10,11] and with a Floquet formalism[12]. As mentioned above, this is the only term that survive in the presence of time-reversal symmetry.

At third order, one must consider

$$\begin{aligned}&\left[\widehat{r}^\gamma(\tau_3), [\widehat{r}^\beta(\tau_2), [\widehat{r}^\alpha(\tau_1), \widehat{v}^\mu]]\right]_{aa} \\ &\quad = i\partial^\gamma i\partial^\beta i\partial^\alpha v^\mu_{aa} + i\partial^\gamma i\partial^\beta[A^\alpha(\tau), v^\mu]_{aa} + \cdots.\end{aligned} \quad (98)$$

Due to the same logic that applied at second order, only these first few times survive at lowest order in ω. Hence

$$\begin{aligned}\sigma^{\mu\alpha\beta\gamma}(\omega;\omega_1,\omega_2,\omega_3) = \frac{e^4}{\hbar^3}\sum_{a,b}\int[d\boldsymbol{k}]f_a\frac{\partial^\gamma\partial^\beta\partial^\alpha v^\mu_{aa}}{\omega\omega_{23}\omega_3} \\ - if_a\frac{\partial^\gamma\partial^\beta}{\omega_{23}\omega_3}\left[\frac{A^\alpha_{ab}v^\mu_{ba}}{\omega-\varepsilon_{ba}} - \frac{v^\mu_{ab}A^\alpha_{ba}}{\omega-\varepsilon_{ab}}\right] + O(\omega^{-1})\end{aligned} \quad (99)$$

so

$$\begin{aligned}&\lim_{\omega,\omega_i\to 0}\sigma^{\mu\alpha\beta\gamma}(\omega;\omega_1,\omega_2,\omega_3) = \\ &\frac{e^4}{\hbar^3}\sum_{a,b}\int[d\boldsymbol{k}]f_a\frac{\partial^\gamma\partial^\beta\partial^\alpha v^\mu_{aa}}{\omega_{23}\omega_3\omega} - if_a\frac{\partial^\gamma\partial^\beta}{\omega_{23}\omega_3}\mathcal{F}^{\alpha\mu}_{aa} + O(\omega^{-1})\end{aligned} \quad (100)$$

Here the first term is the third derivative of the Drude weight while the second is the Berry curvature quadrapole. Of course, this matches the semiclassical result (84).

C. Symmetry Considerations

This subsection considers the effect of symmetry on the two terms of the semiclassical third order response. We focus on the effect of inversion $\mathcal{I}$, time-reveral $\mathcal{T}$, and reflection in the b direction $\mathcal{R}^b$. The semiclassical response involves the group velocity v^m, Berry curvature $\mathcal{F}^{\alpha\mu}$, and k-derivatives thereof, so we start by looking at their transformations under symmetry. Their transformation laws can be deduced from the fact that v^μ and ∂^α are vectors, while the Berry curvature behaves as a psuedovector defined by $\mathcal{F}^\beta \equiv \varepsilon_{\beta\alpha\mu}\mathcal{F}^{\alpha\mu}/2$.

The effect of inversion $\mathcal{I}$ is

$$v^\mu \to -v^\mu \tag{101}$$

$$\mathcal{F}^\beta \to \mathcal{F}^\beta \tag{102}$$

$$\partial^\alpha \to -\partial^\alpha. \tag{103}$$

Applying time reversal $\mathcal{T}$ gives

$$v^\mu \to -v^\mu \tag{104}$$

$$\mathcal{F}^\beta \to -\mathcal{F}^\beta \tag{105}$$

$$\partial^\alpha \to -\partial^\alpha. \tag{106}$$

Lastly, the reflection $\mathcal{R}^b$ leads to

$$v^\mu \to (-1)^{\delta_{b\mu}} v^\mu \tag{107}$$

$$\mathcal{F}^\beta \to -(-1)^{\delta_{b\beta}} \mathcal{F}^\beta \tag{108}$$

$$\partial^\alpha \to (-1)^{\delta_{b\alpha}} \partial^\alpha. \tag{109}$$

These constraints indicate that the group velocity term and the Berry curvature term (the first and the second terms in Eq. (100), respectively) are both even under $\mathcal{I}$, and even and odd under $\mathcal{T}$, respectively. Under $\mathcal{R}^b$, either both terms are even or both terms are odd, depending on the component of nonlinear conductivity and the direction of mirror plane. For example, in σ^{zxxx}, both are odd under $\mathcal{R}^z$, and both are even under $\mathcal{R}^y$.

An interesting question is when the group velocity term vanishes, whereupon the Berry curvature contribution dominates, if it is non-zero. First, this requires $\mathcal{T}$ breaking. Next we need a symmetry such that the group velocity term is odd and the Berry curvature is even. Such a symmetry is obtained by combining $\mathcal{R}^b$ in which both terms are odd and $\mathcal{T}$. For example, σ^{zxxx} has a nonzero contribution only from the Berry curvature term when $\mathcal{T}\mathcal{R}^z$ symmetry is preserved and both $\mathcal{T}$ and $\mathcal{R}^z$ symmetries are broken. This situation can be realized in materials with antiferomagnetic order in z direction. Since σ^{zxxx} is measured as intensity dependent Hall conductivity or intensity dependent transmission of circular polarized light, measuring these quantities in suitable antiferromagnetic materials will allow us to access the Berry curvature effect in third order responses. This Berry curvature effect might be measured in the magnetic Weyl semimetal Mn_3Sn[42–44] since it breaks $\mathcal{T}$ and some candidate AFM structures breaks $\mathcal{R}^z$ while preserving $\mathcal{T}\mathcal{R}^z$[45].

D. Length versus Velocity Gauges

Overall, we have shown that the semiclassical limit is straightforwardly accomplished in the length gauge and matches the answer from the simple Boltzmann equation approach. A few comments on the relation between the length and velocity gauge are in order. It was shown in[29] that on can convert between the two gauges with the time-dependent unitary transformation $S(t) = e^{-\frac{e}{\hbar}\boldsymbol{A}(t)\cdot\boldsymbol{D}}$. The equivalence of expectations of any physical observable $\mathcal{O}$ in the two gauges leads to sum rules of the form[28]

$$\int [d\boldsymbol{k}]\,\mathrm{Tr}\left\{D^{\alpha_1}\cdots D^{\alpha_n}\left[\mathcal{O}\widehat{\rho}_{\boldsymbol{k}}(t)\right]\right\} = 0 \tag{110}$$

where $\rho_{\boldsymbol{k}}(t)$ is the single-particle density matrix defined above. Expanding this with $\mathcal{D} = i\nabla + \mathcal{A}$ leads to the sum rules of Aversa and Sipe[46]. In particular, one can use $\mathcal{O} = \widehat{v}^m$ to convert from velocity to length gauge at order n. This will eliminate terms like $h^{\mu\alpha}$ in favor of $\partial^\alpha v^\mu + \cdots$. However, this algebra is quite involved in practice, so it is usually better to choose the correct gauge from the outset rather than painstakingly changing gauge after writing the answer to a computation.

VII. NUMERICAL EXAMPLE

This section applies the techniques developed in this paper to a model of Weyl semimetals. Numerical calculations of nonlinear optics are usually done within the frameworks of either tight-binding models or Density Functional Theory (DFT). Tight-binding models are usually simple enough to perform analytical calculations and, when chosen wisely, will reproduce the main qualitative features of a material, such as the frequencies of resonances. For more quantitative predictions in specific materials, DFT is the favored technique. Velocity gauge formulas are particularly well-suited for tight-binding models, where operators such as $h^{\mu\alpha}$ may be computed analytically. We will therefore present the example of a simple tight-binding model of a Weyl semimetal where the leading contribution is topological in origin.

The primary feature of Weyl semimetals are their paired Weyl- and anti-Weyl cones, whose linear dispersion acts as a sources and sinks of Berry curvature. As mentioned in the introduction, a wide variety of linear and second order optical responses have been studied in Weyl semimetals, many with a topological origin. Here we study the third-order response σ^{zxxx}, for which the leading contribution comes from the (topological) Berry curvature. To our knowledge, this is the first prediction for a third-order response in Weyl semimetals.

Consider the following two-band Hamiltonian for a Weyl semimetal with a Wilson mass:

$$H(\boldsymbol{k}) = d_0 I + \boldsymbol{d}(\boldsymbol{k})\cdot\boldsymbol{\sigma} \tag{111}$$

where $\boldsymbol{\sigma} = \{\sigma_x, \sigma_y, \sigma_z\}$ is the vector of Pauli matrices, $\boldsymbol{d} = \{d_x, d_y, d_z\}$, and

$$\begin{aligned} d_0(\boldsymbol{k}) &= t \sin ak_y \\ d_x(\boldsymbol{k}) &= \sin ak_x \\ d_y(\boldsymbol{k}) &= \sin ak_y \\ d_z(\boldsymbol{k}) &= \cos ak_z + m\left(2 - \cos ak_x - \cos ak_y\right), \end{aligned}$$

where a is the lattice spacing. Generically, this model supports four Weyl–anti-Weyl pairs, but we gap out three of them by adding Wilson mass term where we set $m = 1$[47]. The remaining Weyl nodes are at $\boldsymbol{k} = (0, 0, \pm\pi/2)$. The parameter t controls the tilting of the Weyl nodes.

In Section VI C, we showed that materials where time-reversal symmetry $\mathcal{T}$ and a mirror symmetry $\mathcal{R}^z$ are broken, but their product $\mathcal{T}\mathcal{R}^z$ is preserved, then the leading order contribution at third order in the $\omega \to 0$ limit is

$$\begin{aligned} &\sigma^{\mu\alpha\beta\gamma}(\omega; \omega_1, \omega_2, \omega_3) \\ &= -i\frac{e^4}{\hbar^3}\sum_a \int [d\boldsymbol{k}] \frac{f_a \partial^\gamma \partial^\beta \mathcal{F}^{\alpha\mu}_{aa}}{(\omega_3 + i\gamma)(\omega_2 + \omega_3 + 2i\gamma)} + O(\omega^{-1}). \end{aligned} \tag{112}$$

Whenever the tilting parameters t is nonzero, the model satisfies these considerations and thus we expect a topological leading response in the off-diagonal component of the third-harmonic response $\sigma^{zxxx}(3\omega; \omega, \omega, \omega)$. (Here z is the direction of the emitted light.) The tilting is selected to be in the y-direction so that both nodes are tilted the same way, making the resonances symmetry-allowed.

We compute the response via numerically integrating Equation (65) for the third-harmonic on a mesh of k-points until convergence is achieved. This involves evaulating the band energies, wavefunctions, and higher derivatives of the Hamiltonian, which may all be computed analytically. As usual for a two-band model, the energies and wavefunctions are, up to normalization,

$$\varepsilon_\pm(\boldsymbol{k}) = d_0 \pm |\boldsymbol{d}|, \quad |u_\pm\rangle = \begin{pmatrix} \frac{d_3 \pm |\boldsymbol{d}|}{d_1 + i d_2} \\ 1 \end{pmatrix}. \tag{113}$$

The velocity operators are easily found by differentiating:

$$h^x(\boldsymbol{k}) = ta\cos ak_y I + a\cos ak_x \sigma_x + ma \sin ak_x \sigma_z, \tag{114}$$

$$h^y(\boldsymbol{k}) = a\cos ak_y \sigma_y + ma\sin ak_y \sigma_z, \tag{115}$$

$$h^z(\boldsymbol{k}) = -a \sin k_z \sigma_z. \tag{116}$$

Higher derivatives are similarly straightforward. For example,

$$h^{xx}(\boldsymbol{k}) = -ta^2 \sin ak_y I - a^2 \sin ak_x \sigma_x \tag{117}$$

$$+ ma^2 \cos ak_x \sigma_z, \tag{118}$$

$$h^{xxx}(\boldsymbol{k}) = -a^2 h^x(\boldsymbol{k}), \tag{119}$$

$$h^{xxxx}(\boldsymbol{k}) = -a^2 h^{xx}(\boldsymbol{k}). \tag{120}$$

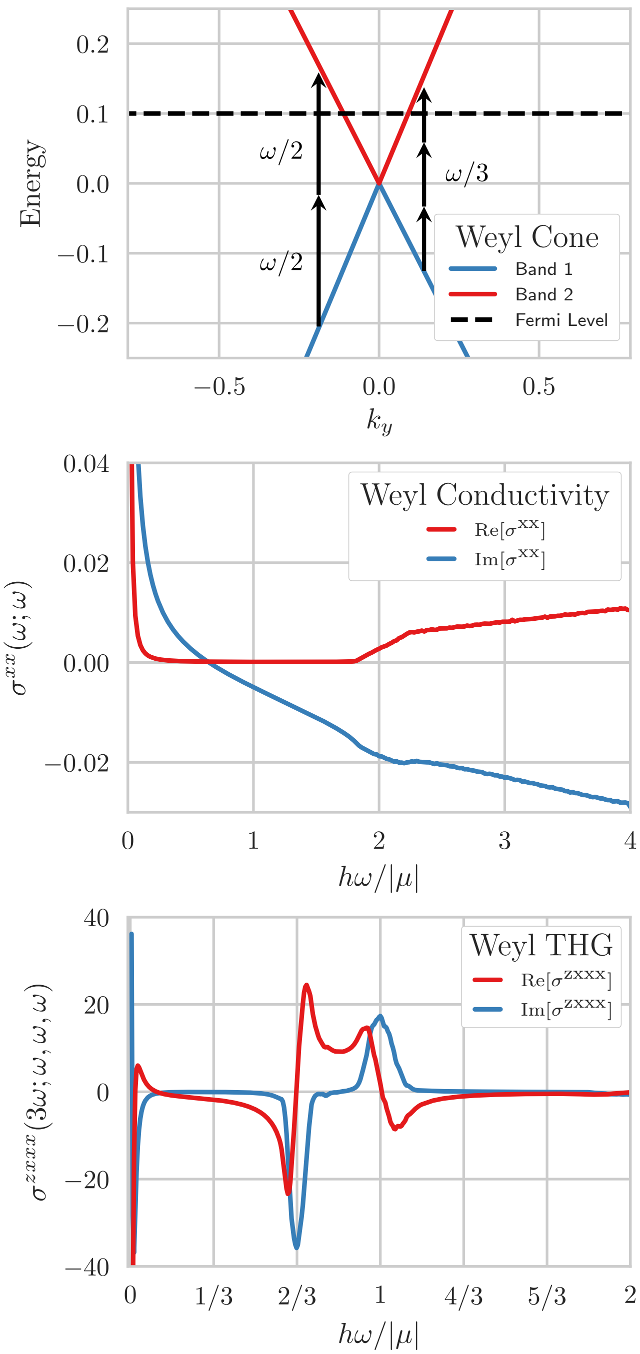


FIG. 1. *(Top)* Band structure of Equation (111) along the $k_x = 0$, $k_z = \pi/2$ plane. A (tilted) Weyl node is visible at $k_y = 0$. The dispersion is approximately linear for $|\varepsilon| \leq 8\mu$. *(Middle)* Linear conductivity in the x-direction. The Drude peak is visible at low frequencies, the conductivity increases linearly for $\omega \gg 2\mu$. *(Bottom)* Off-diagonal component of the third harmonic response. The $O(\omega^{-2})$ divergence due to the Berry curvature is visible at low frequencies, and wide resonances are visible at $\omega \sim 2\mu/3$ and $\omega \sim \mu$. The parameters used for all data are $\mu = 0.1$, $\gamma = 0.001$, $t = 0.1$, $m = a = 1$.

Figure 1 shows the well-understood linear reponse of the material and the third-harmonic response σ^{zxxx}.

The linear response of Weyl semimetals is shown in the middle panel of FIG 1. This response is already well-understood[48]. The conductivity exhibits a typical Drude peak at low frequencies, and $\Re[\sigma^{xx}] \sim \omega$ at higher frequencies. This is due to the interband resonance that becomes possible once the frequency exceeeds twice the chemical potential. Due to the tilting, the Fermi surface with $\mu = 0.1$ is an ellipse with a smaller bandgap on one side than another. This causes the linear conducitivity to onset in the range $2\mu - t \leq \omega \leq 2\mu + t$.

The third harmonic response is shown in the bottom panel of FIG. 1, and displays several features of interest. At low frequency, where $\omega \sim \gamma$, we observe the predicted divergence (112), in accordance with the semiclassical considerations of Section VI. The divergence is visible in both the real and imaginary parts due to the phenomenological broadening. Resonances are visible near $\omega \sim \mu$ and $\omega \sim 2\mu/3$, due to two- and three-photon processes respectively. The top panel of FIG. 1 indicates these processes schematically.

One can see that the two-photon process becomes resonant around $\omega \sim 2\mu/3 - t$, corresponding to the side of the Weyl cone with the smaller bandgap, and continues up to $\omega \sim 2\mu/3 + t$, when the resonance is on the other side of the cone. This causes a peculiar linear increase in $\mathrm{Re}\,\sigma^{zxxx}$ in the range $2\mu/3 - t \leq \omega \leq 2\mu/3 + t$, and similar considerations apply to the third-order response in the range $\mu - t \leq \omega \leq \mu + t$. As the tilting is increased, the range of this linear regime grows. This not too surprising, since a similar linear onset due to the tilt is present at first order. To our knowledge, this is the first prediction of a third-order response in Weyl semimetals. Again, the key topological feature is the divergence at low frequency, which is proportional to the quadrupole moment of the Berry curvature.

One should note that, although we have focused on the third-harmonic contribution here, the equations for the third-order response from Section V are generic. One can just as easily evaluate the self-focusing correction, totally off-diagonal components such as σ^{zyxz}, or other effects such as the AC Kerr effect. Similarly, any other tight-binding model can be used instead of (111). The only restriction is that it must be defined on the entire Brillouin zone, so that the equivalence with length-gauge is maintained.

Let us comment briefly on the use of DFT. The optical response formulas in previous sections require the matrix elements of derivatives of the Hamiltonian operator (11). Using the covariant derivative (8), these can be written in terms of the matrix elements of the velocity operator and the Berry connection. There are well-established techniques for calculating linear responses within DFT[49]—which already involves computing the Berry connection and matrix elements of the velocity operators—and frequently achieves predictions within $1\% - 10\%$ of experimental values. Naively, this is somewhat surprising, as DFT does not necessarily give good wavefunctions, but only energies. Nevertheless, tools such as the "GW" approximation or the use of specific functionals permit accurate determination of the wavefunctions in many cases. With sufficiently fine k-space meshes, one can in principle converge the numerical derivatives required and make accurate predictions for non-linear optical responses within DFT[50]. Another option is to use the technique of "Wannierization" to produce accurate tight-binding models by Fourier transforming Wannier functions derived *ab initio*[51]. In sum, the nonlinear optical responses presented here are may, in principle, be accurately computed within DFT.

VIII. DISCUSSION AND CONCLUSIONS

This work has elucidated a diagrammatic approach to nonlinear optical responses and applied it to predict the third order optical response of Weyl semimetals. In this final section we will reiterate the main results of the paper and discuss the choice of gauge.

As mentioned in the introduction, the choice of the length or velocity gauge in optical response calculations is a longstanding issue. The modern definitions of the gauges—which depend crucially on the Berry connection—permit the use of either gauge to compute optical responses. Therefore one is now free to choose the best gauge for the problem at hand. The computations in this work suggest a few rules of thumb for when each gauge should be applied. Equations in the velocity gauge have a natural separation between matrix elements and resonances, and contain only simple poles, making them preferable whenever it is necessary to separately examine one-, two-, and three-photon resonances. Since the only matrix elements that appear are derivatives of the Hamiltonian, the velocity gauge is particularly well-suited for tight-binding calculations. However, in the $\omega \to 0$ limit, the velocity gauge suffers from (cancelling) apparent divergences. Hence, for analytical work in this limit the length gauge is often preferable.

Let us comment on why our diagrammatic approach necessarily employs the velocity gauge. The key issue is the presence of the position operator $\hat{\boldsymbol{r}}$, which acts on *all operators to the right* by differentiation. The vertices needed in the length gauge become complicated quite quickly, as they involve not only the position and velocity operators, but objects such as the derivatives of the position operator and a resonance; virtually every term uses its own, unique, vertex. A naive diagrammatic approach to nonlinear response in the length gauge is therefore impractical. One should note that, historically, diagrammatic methods have indeed employed the length gauge[18]. However, these techniques do not account for the Berry connection, but only the fully interband parts of the position operator. In any material with non-vanishing Berry connection, these old-style diagrams will miss important contributions to nonlinear responses, including some resonances.

The diagrammatic method of this work provides an efficient computational framework to calculate nonlinear responses in the velocity gauge. The results are general for any component and frequency, without unphysical diverjoao2018nonjoao2018nongences. We have provided convenient formulae for the general second and third order responses, as well as the particular cases of second harmonic, shift current, third harmonic and self-focusing. To interpret these equations, we examined the semiclassical limit and linked it to the length gauge. On a technical level, the method of this work should often be the shortest way to compute nonlinear optical responses.

The expressions for nonlinear optical responses given here are equivalent to those previously given in the literature in all cases we are aware of (so long as the correct definitions for the length and velocity gauges are employed). We have checked that our formalism explicitly reproduces the results of Refs[5,10,11,13,52,53], as well as the equivalence of our equations for the first-order conductivity, shift current, and second-harmonic generation with those present in the literature. This is exactly what is expected. After all, one can recover many other schemes for computing non-linear responses as limits of ours, including (i) Boltzmann/semiclassical transport theory, (ii) quantum mechanical perturbation theory in the length or velocity gauge, (iii) Floquet formalism. Recent work[54] develops a diagrammatic expansion for non-linear optical responses in the Keldysh formalism which reduces to our formalism when the applied electric fields are periodic in time (i.e. plane-waves). We expect, however, that our results hold for general wavepackets $\boldsymbol{E}(t)$ so long as the duration of the wavepacket and measurement are much less than the timescale associated with dissipation.

Optical responses are most useful when connected to experiment. To this end, we have predicted the third harmonic response of a Weyl semimetal. At small frequencies, the third harmonic response is dominated by a divergent term due to the quadrupole of the Berry curvature, and hence of topological origin. There are also large resonant contributions from both two- and three-photon processes, with a peculiar linear character.

The results of this work can be expanded in both technical and practical directions. Technically, the diagrammatic formalism enables interacting electrons to be treated on the same level as free ones; we are currently expanding these results to the case of Fermi liquids and possibly even magnetic fields. On a practical level, third order responses are somewhat understudied at present, despite being present in most materials and technologically important. The formulae and techniques of this work should enable or simplify prediction of the third order optical response in a wide variety of materials.

ACKNOWLEDGMENTS

The authors thank Adolfo Grushin, Daniel Passos, and Jonah Haber for useful discussions. This work was primarily funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231 (Quantum materials program KC2202). T.M was supported by the Gordon and Betty Moore Foundations EPiQS Initiative Theory Center Grant to UC Berkeley. J.O. received support from the Gordon and Betty Moore Foundations EPiQS Initiative through Grant GBMF4537. D.E.P. received support from the NSF GRFP, DGE 1752814.

Appendix A: The Position Operator and The Berry Connection

This Appendix discusses the position operator and its close relation to the Berry connection, giving some mathematical details thereof.

1. The Position Operator as Covariant Derivative

Suppose we have a crystal with a finite number of bands, N, which are all close to the Fermi level and separated from all other bands by a large energy gap. We can then consider those N bands as an effective model for the material. What form does the single-particle position operator take in this situation? The correct answer to this question was known at least as early as 1962, where it is discussed in the classic paper of Blount[23]. Morally, just as derivatives and polynomials are exchanged by Fourier transforms, the real-space position operator $\widehat{\boldsymbol{r}}$ should become a $\boldsymbol{k}$-derivative. We briefly recall Blount's derivation, adapted to modern notation.

Any wavefunction $|f\rangle$ can be written in terms of the Bloch functions $\psi_{\boldsymbol{k}a}$ as

$$\langle \boldsymbol{r}|f\rangle = f(\boldsymbol{r}) = \sum_a \int [d\boldsymbol{k}] \psi_{\boldsymbol{k}a} f_a(\boldsymbol{k}) \tag{A1}$$

Then

$$\begin{aligned} \langle \boldsymbol{r}|\widehat{\boldsymbol{r}}|f\rangle &= \sum_a \int [d\boldsymbol{k}]\ \psi_{\boldsymbol{k}a}(\boldsymbol{r}) \boldsymbol{r} f_a(\boldsymbol{k}) \\ &= \sum_a \int [d\boldsymbol{k}]\ \left[-i\partial_{\boldsymbol{k}} \left(e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\right)\right] u_{\boldsymbol{k}a}(\boldsymbol{r}) f_a(\boldsymbol{k}) \end{aligned}$$

Integrating by parts (the surface term vanishes because the Brillouin zone is a closed manifold)

$$\begin{aligned} \langle \boldsymbol{r}|\widehat{\boldsymbol{r}}|f\rangle &= \sum_a \int [d\boldsymbol{k}]\ e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \left[i\partial_{\boldsymbol{k}} u_{\boldsymbol{k}a}(\boldsymbol{r}) + u_{\boldsymbol{k}a}(\boldsymbol{r}) i\partial_{\boldsymbol{k}} f\right] \\ &= \sum_{a,b} \int [d\boldsymbol{k}]\ \psi_{\boldsymbol{k}b}(\boldsymbol{r}) \left[\delta_{ab} i\partial_{\boldsymbol{k}} + u_{\boldsymbol{k}b} i\partial_{\boldsymbol{k}} u_{\boldsymbol{k}a}\right] f_a. \end{aligned}$$

We can therefore identify

$$\widehat{\boldsymbol{r}} = i\widehat{\boldsymbol{\mathcal{D}}} = i\left[\boldsymbol{\nabla}_{\boldsymbol{k}} - i\boldsymbol{\mathcal{A}}\right] \tag{A2}$$

where

$$\boldsymbol{\mathcal{A}}_{ab} = i\left\langle u_{\boldsymbol{k}a} | \partial_{\boldsymbol{k}} u_{\boldsymbol{k}b}\right\rangle . \tag{A3}$$

To be clear, in (A2), $\boldsymbol{\nabla}_{\boldsymbol{k}} = \delta_{ab}\boldsymbol{\nabla}_{\boldsymbol{k}}\delta(\boldsymbol{k}' - \boldsymbol{k})$ is the gradient operator which acts on *all matrix elements to the right*. Here we have used the standard notation $\psi_{\boldsymbol{k}a}$ for the Bloch functions, but nowhere was the fact that they are eigenvectors of the Hamiltonian necessary. Indeed, nothing about the Hamiltonian was needed! The connection we have defined is a generalization of the Berry connection to the case of multiple bands; $\boldsymbol{\mathcal{D}}$ is a $U(N)$ connection. It depends only on the choice of which N bands are involved and not on any details of the dynamics.

This is particularly clear once we consider a change of basis. Suppose U is a general change of basis, i.e. a $U(N)$ gauge transformation: $\psi^{'}_{\boldsymbol{k}a'} = U_{a'a}(\boldsymbol{k})\psi_{\boldsymbol{k}a}$ where the $U_{a'a}$'s vary smoothly with $\boldsymbol{k}$. Gauge transforms act naturally on basis vectors, and therefore act through the dual representation on wavefunctions, which are coefficients. Concretely, $\langle \boldsymbol{r}|f\rangle = \sum_a \int [d\boldsymbol{k}]\psi_{\boldsymbol{k}a} f_a$ transforms to

$$\sum_{a'} \int [d\boldsymbol{k}]\psi^{'}_{\boldsymbol{k}a'} f_{a'} = \sum_{a,a'} \int [d\boldsymbol{k}]\psi_{\boldsymbol{k}a} U_{aa'}(\boldsymbol{k}) f_{a'} . \tag{A4}$$

Hence wavefunctions transform as $f \to U^{\dagger} f$. We therefore mandate that $\widehat{\boldsymbol{\mathcal{D}}}f$, which is itself a wavefunction, must transform as

$$\widehat{\boldsymbol{\mathcal{D}}}f \to U^{\dagger}\widehat{\boldsymbol{\mathcal{D}}}f = \left(U^{\dagger}\widehat{\boldsymbol{\mathcal{D}}}U\right)\left(U^{\dagger}f\right) \tag{A5}$$

under a gauge transformation. The action of $\boldsymbol{D}$ gives

$$\left[U^{\dagger}U\boldsymbol{\nabla}_{\boldsymbol{k}} + U^{\dagger}\left(\boldsymbol{\nabla}_{\boldsymbol{k}}U\right) - iU^{\dagger}\boldsymbol{A}U\right]\left(U^{\dagger}f\right) . \tag{A6}$$

Comparing with $\left[\boldsymbol{\nabla}_{\boldsymbol{k}} - i\boldsymbol{A}'\right]\left(U^{\dagger}f\right)$ we find

$$\boldsymbol{A}' = U^{\dagger}\boldsymbol{A}U + iU^{\dagger}\boldsymbol{\nabla}_{\boldsymbol{k}}U \tag{A7}$$

This confirms that $\widehat{\boldsymbol{\mathcal{D}}}$ is a $U(N)$ non-Abelian connection.

Beyond being necessary to define the position operator correctly, this connection allows us to define the $\boldsymbol{k}$-derivatives of operators. The connection acts on operators naturally via

$$\widehat{\boldsymbol{D}}[\widehat{\mathcal{O}}] = [\widehat{\boldsymbol{D}}, \mathcal{O}]. \tag{A8}$$

This is used extensively in the main text.

2. Generalized Berry Connections

Let us give a few more mathematical comments. Readers curious for a more formal treatment are recommended to consult Chapter 7 of[55] or Appendix D of[56]. The normal Berry connection[20] is a $U(1)$ connection defined for a single band. For our setting of N bands, there are two possible generalizations to consider: a $U(1)^N$ connection or a $U(N)$ connection, the latter of which we have described above. Let us see how each of these arise and what role they play physically.

From the perspective of differential geometry, we are working with an infinite dimensional Hilbert bundle over the Brillouin torus. The exterior derivative d is a provides a (curvature-free) connection on the Hilbert bundle. When we select an N-dimensional effective Hilbert space, there is a projection map

$$P = \frac{1}{N}\sum_{a=1}^{N}\int [d\boldsymbol{k}]\ \left|u_{\boldsymbol{k}a}\right\rangle\left\langle u_{\boldsymbol{k}a}\right| \tag{A9}$$

from the Hilbert bundle to the $\mathbb{C}^N$ bundle of interest, and this projection naturally induces a connection on the $\mathbb{C}^N$ bundle which acts on $\mathbb{C}^N$-valued differential forms ω as[57]

$$D\,\omega = Pd\,\omega = (d + i\mathcal{A})\,\omega \tag{A10}$$

An important, yet subtle, point is that the considerations above do not uniquely define a connection. There is still a residual freedom corresponding to the choice of origin in the (real space) unit cell. This is intimately related to the modern theory of polarization and is carefully considered from a mathematical point of view in[58].

Due to the non-Abelian nature of the $U(N)$ connection, its gauge-invariant quantities are Wilson loops, which cannot be computed directly from the curvature. It would be interesting to compute these and determine if they have any physical meaning or utility. However, it seems unlikely that expressions involving Wilson loops are buried inside nonlinear conductivities. In the special case of degenerate bands, the Wilson loops have been used, for instance, to classify topological parts of Fermi-surface oscillations under magnetic fields[59].

Now let us identify the second, Abelian, connection. In practice, one virtually always chooses to work in the energy basis with Bloch functions $u_{\boldsymbol{k}a}$. However, as is well-known, these are only defined up to a phase. The functions $u'_{\boldsymbol{k}a} = e^{i\theta_a(\boldsymbol{k})}u_{\boldsymbol{k}a}$ also satisfy (6). In other words, the choice of the energy basis does not completely fix the gauge, but only up to a change of phase in each band; there is a residual $U(1)^N$ gauge freedom. This allows us to define the second connection, which is Abelian and is denoted by a non-calligraphic letter:

$$\widehat{\boldsymbol{D}}(\boldsymbol{k})_{ab} = \delta_{ab}\left[\boldsymbol{\nabla}_{\boldsymbol{k}} - i\boldsymbol{A}_{aa}\right] \tag{A11}$$

where $\boldsymbol{A}$ is the same as above, but this is now *diagonal* in the band indices. Under a $U(1)^N$ gauge transformation $U(\boldsymbol{k}) = \delta_{ab}e^{i\theta_a(\boldsymbol{k})}$, Equation (A7) reduces to

$$\boldsymbol{A}_{aa} = \boldsymbol{A}_{aa} - \boldsymbol{\nabla}_{\boldsymbol{k}}\theta_a(\boldsymbol{k}). \tag{A12}$$

So the Abelian connection transforms as $\boldsymbol{D} \to \boldsymbol{D}' = e^{-i\theta_a}\boldsymbol{D}_{aa}e^{i\theta_a} = \boldsymbol{D}$ and is thus gauge invariant. This

$U(1)^N$ connection is nothing more than one copy of the normal Berry connection for each band. As above, we get an associated connection on operators given by $\boldsymbol{D}[\widehat{\mathcal{O}}] = [\boldsymbol{D}, \widehat{\mathcal{O}}]$, and $\boldsymbol{D}[\widehat{\mathcal{O}}]$ will be gauge-invariant whenever $\mathcal{O}$ is.

Let us briefly contrast the $U(N)$ and $U(1)^N$ connections and identify when each should be used. The non-Abelian connection $\boldsymbol{\mathcal{D}}$ is a strictly more complicated object than the Abelian connection $\boldsymbol{D}$. In general, objects involving $\boldsymbol{\mathcal{D}}$ will be gauge-covariant after choosing the energy basis, but objects with $\boldsymbol{D}$ may be gauge-*invariant*. For example, the curvature

$$\mathcal{F}^{\boldsymbol{\mathcal{D}}} = i[\boldsymbol{\mathcal{D}}, \boldsymbol{\mathcal{D}}] \to \left(\mathcal{F}^{\boldsymbol{\mathcal{D}}}\right)' = U^\dagger \mathcal{F}^{\boldsymbol{\mathcal{D}}} U, \tag{A13}$$

is gauge-covariant, whereas in the Abelian case

$$F^{\boldsymbol{D}} = i[\boldsymbol{D}, \boldsymbol{D}] \to \left(F^{\boldsymbol{D}}\right)' = e^{-i\theta_a} F^{\boldsymbol{D}}_{ab} \delta_{ab} e^{i\theta_b} = F^{\boldsymbol{D}} \tag{A14}$$

is gauge-invariant. (This is a standard fact for non-Abelian versus Abelian connections.) As any observable must be strictly gauge-invariant, so it is necessarily much easier to produce observables out of the second connection. In an ideal world, we would be able to work only with $\boldsymbol{D}$ and not $\boldsymbol{\mathcal{D}}$. Indeed, for a single band when $N = 1$, this is the case. There is some hope of eliminating $\boldsymbol{\mathcal{D}}$, because for all operators that act diagonally in band space, with $\widehat{\mathcal{O}} = \delta_{ab}\mathcal{O}_{aa}$, the induced connection $\boldsymbol{\mathcal{D}}$ reduces to $\boldsymbol{D}$. However, this is a vain hope: measuring electromagnetic responses inevitably involves the off-diagonal components $\boldsymbol{\mathcal{A}}_{ab}$, and we must use the full generality of the non-Abelian connection. Moreover, when bands are degenerate or cross, such as at a Dirac point, there is no unique way to define the Bloch wavefunctions of each band. These points, which play a crucial role in topological band structures, therefore cannot be fully described via this $U(1)^N$ connection.

In a philosophical sense, the presence of the non-Abelian connection helps to explain why non-linear conductivity responses are often devoid of simple forms: they must be gauge-invariant, but their "building blocks" are only gauge-covariant, and so much be composed of tricky combinations that cancel out among themselves. More optimistically, however, one can *harness* this gauge invariance. We will use it to conceptually simplify our perturbation theory approach to non-linear conductivities in the length gauge. A theme from recent years is that the converse is also true: once a new gauge invariant combination has been isolated, it is usually physically measurable, perhaps in a limit. To search for new and interesting quantities to measure, one need only consider what combinations are gauge invariant.

Appendix B: Useful Integrals

In this section we will evaluate the loop integrals in the Feynman diagrams. Following Chapter 3 of Mahan[60], we work with Matsubara frequencies, which allows the evaluation of the integrals with straightforward contour integral techniques.

We wish to evaluate integrals such as

$$I_1 = \int d\omega\, G_a(\omega) = \int d\omega\, \frac{1}{\omega - \varepsilon_a} \tag{B1}$$

$$I_2(\omega_1) = \int d\omega\, G_a(\omega) G_b(\omega + \omega_1) \tag{B2}$$

$$I_3(\omega_1, \omega_2) = \int d\omega\, G_a(\omega) G_b(\omega + \omega_1) G_c(\omega + \omega_1 + \omega_2) \tag{B3}$$

$$I_4(\omega_1, \omega_2, \omega_3) = \int d\omega\, G_a(\omega) G_b(\omega + \omega_1) G_c(\omega + \omega_1 + \omega_2) \tag{B4}$$

$$\times\, G_d(\omega + \omega_1 + \omega_2 + \omega_3) \tag{B5}$$

In imaginary time, fermions only have frequencies at odd imaginary integers: $i\omega_n = i(2n+1)\pi/\beta$ for $n \in \mathbb{Z}$. The integral is then analytically continued to a sum

$$I_1 \to S_1 = \frac{1}{\beta} \sum_{n \in \mathbb{Z}} \frac{1}{i\omega_n - \varepsilon_a}. \tag{B6}$$

To evaluate this sum, note that the Fermi-Dirac Distribution $f(z) = \frac{1}{e^{\beta z} - 1}$ has poles at exactly these complex frequencies $i\omega_n$, each with residue $-1/\beta$. We can therefore use the following trick of trading the sum for a contour integral. Consider

$$0 = J_1 = \lim_{R \to \infty} \oint_{C_R} \frac{dz}{2\pi i} f(z) F_1(z) \tag{B7}$$

where the contour is the circle of radius R and

$$F_1(z) = \frac{1}{z - \varepsilon_a}. \tag{B8}$$

The integral on the right-hand side is easy to evalute. The poles of $f(z)F_1(z)$, shown in Figure 2 are at $z_n = i\omega_n$ with residue $R_n = -\frac{1}{\beta} F_1(i\omega_n)$, coming from the Fermi-Dirac distribution, and then $z_1 = \varepsilon_a$ with residue $R_1 = f(\varepsilon_a)$. So

$$0 = J_1 = -\frac{1}{\beta} \sum_{n \in \mathbb{Z}} F_1(i\omega_n) + f(\varepsilon_a). \tag{B9}$$

Rearranging,

$$I_1 = S_1 = f(\varepsilon_a), \tag{B10}$$

where the first equality is true since the analytic continuation back is trivial here.

Precisely the same technique will work for the more complex integrals with one extra residue for each Green's function. For I_2 we analytically continue to

$$S_2(i\omega_1) = \frac{1}{\beta} \sum_{n \in \mathbb{Z}} \frac{1}{z - \varepsilon_a} \frac{1}{z + i\omega_1 - \varepsilon_b}. \tag{B11}$$

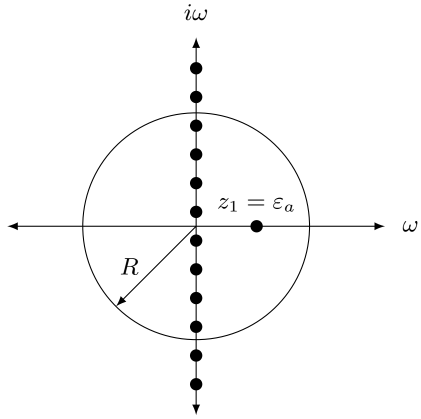


FIG. 2. Depiction of the poles of the function $f(z)F_1(z)$ and the integration contour. The poles z_n are on the $i\omega$ axis and the pole z_1 is on the ω axis.

Since $i\omega_1$ is due to an incoming photon, it is a bosonic Matsubara frequency and thus an *even* integer instead of odd: $i\omega_1 = i(2M)\pi/\beta$ for some integer M. Now consider

$$0 = J_2 = \lim_{R\to\infty} \oint_{C_R} \frac{dz}{2\pi i} f(z) F_2(z) \tag{B12}$$

$$F_2(z) = \frac{1}{z-\varepsilon_a}\frac{1}{z+i\omega_1-\varepsilon_b}. \tag{B13}$$

The function $f(z)F_2(z)$ has poles and residues

$$z_n = i\omega_n; \qquad R_n = -\frac{1}{\beta}F_2(i\omega_n) \tag{B14}$$

$$z_1 = \varepsilon_a; \qquad R_1 = \frac{f(\varepsilon_a)}{\varepsilon_a + i\omega_1 - \varepsilon_b} = \frac{-f(\varepsilon_a)}{\varepsilon_{ba} - i\omega_1} \tag{B15}$$

$$z_2 = \varepsilon_a - i\omega_1; \qquad R_2 = \frac{f(\varepsilon_b - i\omega_1)}{\varepsilon_b - i\omega_1 - \varepsilon_a} = \frac{f(\varepsilon_b)}{\varepsilon_{ba} - i\omega_1}. \tag{B16}$$

In the last equality for R_2, the fact $e^{\beta(i\omega_1)} = 1$ implies $f_a(\varepsilon_a - i\omega_1) = f(\varepsilon_b)$. Therefore

$$S_2(i\omega_1) = R_1 + R_2 = \frac{f_{ab}}{i\omega_1 - \varepsilon_{ab}}. \tag{B17}$$

Analytically continuing back we then have

$$I_2(\omega_1) = \frac{f_{ab}}{\omega_1 - \varepsilon_{ab}}. \tag{B18}$$

The generalization to I_3 and I_4 follows the same pattern. For I_3 we consider the contour integral against

$$F_3(z) = \frac{1}{z-\varepsilon_a}\frac{1}{z+i\omega_1-\varepsilon_b}\frac{1}{z+i\omega_{12}-\varepsilon_c} \tag{B19}$$

where $\omega_{12} = \omega_1 + \omega_2$. Then $f(z)F_3(z)$ has poles and residues

$$z_n = i\omega_n; \qquad R_n = -\frac{1}{\beta}F_3(i\omega_n) \tag{B20}$$

$$z_1 = \varepsilon_a; \qquad R_1 = \frac{f(\varepsilon_a)}{(\varepsilon_{ab}+i\omega_1)(\varepsilon_{ac}+i\omega_{12})} \tag{B21}$$

$$z_2 = \varepsilon_b - i\omega_1; \qquad R_2 = \frac{f(\varepsilon_b)}{(\varepsilon_{ba}-i\omega_1)(\varepsilon_{bc}+i\omega_2)} \tag{B22}$$

$$z_3 = \varepsilon_c - i\omega_{12}; \qquad R_3 = \frac{f(\varepsilon_c)}{(\varepsilon_{ca}-i\omega_{12})(\varepsilon_{cb}-i\omega_2)}. \tag{B23}$$

Then $S_3(i\omega_1, i\omega_2) = R_1 + R_2 + R_3$. Analytically continuing back to real frequency,

$$I_3(\omega_1,\omega_2) = \frac{f(\varepsilon_a)}{(\varepsilon_{ab}+\omega_1)(\varepsilon_{ac}+\omega_{12})} \tag{B24}$$

$$- \frac{f(\varepsilon_b)}{(\varepsilon_{ab}+\omega_1)(\varepsilon_{bc}+\omega_2)} + \frac{f(\varepsilon_c)}{(\varepsilon_{ac}+\omega_{12})(\varepsilon_{bc}+\omega_2)}. \tag{B25}$$

Employing the same procedure, S_4 is made up of 4 poles, which sum to give

$$\begin{aligned} &I_4(\omega_1,\omega_2,\omega_3) \\ &= \frac{f(\varepsilon_a)}{(\varepsilon_{ab}+\omega_1)(\varepsilon_{ac}+\omega_{12})(\varepsilon_{ad}+\omega_{123})} \\ &+ \frac{f(\varepsilon_b)}{(\varepsilon_{ba}-\omega_1)(\varepsilon_{bc}+\omega_2)(\varepsilon_{bd}+\omega_{23})} \\ &+ \frac{f(\varepsilon_c)}{(\varepsilon_{ca}-\omega_{12})(\varepsilon_{cb}-\omega_2)(\varepsilon_{cd}+\omega_3)} \\ &+ \frac{f(\varepsilon_d)}{(\varepsilon_{da}-\omega_{123})(\varepsilon_{db}-\omega_{23})(\varepsilon_{dc}-\omega_3)}. \end{aligned} \tag{B26}$$

[*] daniel_parker@berkeley.edu
[†] tmorimoto@berkeley.edu
[‡] jworenstein@lbl.gov
[§] jemoore@berkeley.edu

[1] R. W. Boyd, *Nonlinear optics* (Academic press, London, 2003).
[2] N. Bloembergen, *Nonlinear optics* (World Scientific, Singapore, 1996).
[3] P. N. Butcher and D. Cotter, *The elements of nonlinear optics*, Vol. 9 (Cambridge university press, 1991).
[4] P. J. Sturman and V. M. Fridkin, *Photovoltaic and Photo-refractive Effects in Noncentrosymmetric Materials*, Vol. 8 (CRC Press, Philadelphia, 1992).
[5] J. Sipe and A. Shkrebtii, Physical Review B **61**, 5337 (2000).
[6] S. M. Young and A. M. Rappe, Phys. Rev. Lett. **109**, 116601 (2012).
[7] S. M. Young, F. Zheng, and A. M. Rappe, Phys. Rev.

Lett. **109**, 236601 (2012).

[8] A. M. Cook, B. M. Fregoso, F. De Juan, S. Coh, and J. E. Moore, Nature communications **8**, 14176 (2017).

[9] T. Morimoto and N. Nagaosa, Science Advances **2**, e1501524 (2016).

[10] J. Orenstein and J. E. Moore, Phys. Rev. B **87**, 165110 (2013).

[11] I. Sodemann and L. Fu, Phys. Rev. Lett. **115**, 216806 (2015).

[12] T. Morimoto, S. Zhong, J. Orenstein, and J. E. Moore, Physical Review B **94**, 245121 (2016).

[13] C.-K. Chan, N. H. Lindner, G. Refael, and P. A. Lee, Phys. Rev. B **95**, 041104 (2017).

[14] F. de Juan, A. G. Grushin, T. Morimoto, and J. E. Moore, Nature Communications **8** (2017).

[15] L. Wu, S. Patankar, T. Morimoto, N. L. Nair, E. Thewalt, A. Little, J. G. Analytis, J. E. Moore, and J. Orenstein, Nature Physics **13**, 350 (2017).

[16] Q. Ma, S.-Y. Xu, C.-K. Chan, C.-L. Zhang, G. Chang, Y. Lin, W. Xie, T. Palacios, H. Lin, S. Jia, *et al.*, Nature Physics **13**, 842 (2017).

[17] S. Patankar, L. Wu, B. Lu, M. Rai, J. D. Tran, T. Morimoto, D. Parker, A. Grushin, N. L. Nair, J. G. Analytis, J. E. Moore, J. Orenstein, and D. H. Torchinsky, ArXiv e-prints (2018), arXiv:1804.06973 [cond-mat.mtrl-sci].

[18] J. F. Ward, Rev. Mod. Phys. **37**, 1 (1965).

[19] M. V. Berry, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **392**, 45 (1984).

[20] B. Simon, Physical Review Letters **51**, 2167 (1983).

[21] W. Kraut and R. von Baltz, Physical Review B **19**, 1548 (1979).

[22] R. von Baltz and W. Kraut, Physical Review B **23**, 5590 (1981).

[23] E. Blount, in *Solid state physics*, Vol. 13 (Elsevier, 1962) pp. 305–373.

[24] R. Resta, Rev. Mod. Phys. **66**, 899 (1994).

[25] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[26] K. Taguchi, T. Imaeda, M. Sato, and Y. Tanaka, Phys. Rev. B **93**, 201202 (2016).

[27] X. Yang, K. Burch, and Y. Ran, arXiv:1712.09363 (2017).

[28] G. Ventura, D. Passos, J. L. dos Santos, J. V. P. Lopes, and N. Peres, Physical Review B **96**, 035431 (2017).

[29] D. Passos, G. Ventura, J. Lopes, J. Santos, and N. Peres, arXiv preprint arXiv:1712.04924 (2017).

[30] A. Taghizadeh and T. G. Pedersen, Physical Review B **97**, 205432 (2018).

[31] N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, and N. P. Ong, Rev. Mod. Phys. **82**, 1539 (2010).

[32] S. Murakami, New J. Phys. **9**, 356 (2007).

[33] X. Wan, A. M. Turner, A. Vishwanath, and S. Y. Savrasov, Phys. Rev. B **83**, 205101 (2011).

[34] K. Fukushima, D. E. Kharzeev, and H. J. Warringa, Phys. Rev. D **78**, 074033 (2008).

[35] D. T. Son and B. Z. Spivak, Phys. Rev. B **88**, 104412 (2013).

[36] S. Zhong, J. E. Moore, and I. Souza, Physical review letters **116**, 077201 (2016).

[37] J. Ma and D. Pesin, Physical Review B **92**, 235205 (2015).

[38] H. Weng, C. Fang, Z. Fang, B. A. Bernevig, and X. Dai, Phys. Rev. X **5**, 011029 (2015).

[39] X. Huang, L. Zhao, Y. Long, P. Wang, D. Chen, Z. Yang, H. Liang, M. Xue, H. Weng, Z. Fang, X. Dai, and G. Chen, Phys. Rev. X **5**, 031023 (2015).

[40] We note that the Bloch wavefunctions are not necessarily periodic in $\boldsymbol{k}$. In fact, the natural gauge choice is given by the convention $u_{\boldsymbol{k}+\boldsymbol{G}a}(\boldsymbol{r}) = e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}u_{\boldsymbol{k}a}(\boldsymbol{r})$. This gauge choice should be adopted when trying to compute polarization and related quantities in tight-binding models.

[41] However, this term is familiar from atomic physics: the second-order response of a molecular system has the same form as this last term, but without the Brillouin zone integral. Of course, the meaning of the matrix elements is different in that situation.

[42] S. Nakatsuji, N. Kiyohara, and T. Higo, Nature **527**, 212 (2015).

[43] N. Kiyohara, T. Tomita, and S. Nakatsuji, Physical Review Applied **5**, 064009 (2016).

[44] A. K. Nayak, J. E. Fischer, Y. Sun, B. Yan, J. Karel, A. C. Komarek, C. Shekhar, N. Kumar, W. Schnelle, J. Kübler, *et al.*, Science advances **2**, e1501870 (2016).

[45] M.-T. Suzuki, T. Koretsune, M. Ochi, and R. Arita, Phys. Rev. B **95**, 094406 (2017).

[46] C. Aversa and J. Sipe, Physical Review B **52**, 14636 (1995).

[47] O. Vafek and A. Vishwanath, Annual Review of Condensed Matter Physics **5**, 83 (2014).

[48] P. Hosur, S. A. Parameswaran, and A. Vishwanath, Phys. Rev. Lett. **108**, 046602 (2012).

[49] M. Rohlfing and S. G. Louie, Physical Review B **62**, 4927 (2000).

[50] T. Rangel, B. M. Fregoso, B. S. Mendoza, T. Morimoto, J. E. Moore, and J. B. Neaton, Physical review letters **119**, 067402 (2017).

[51] J. Ibañez-Azpiroz, S. S. Tsirkin, and I. Souza, arXiv preprint arXiv:1804.04030 (2018).

[52] K. Taguchi, T. Imaeda, M. Sato, and Y. Tanaka, Phys. Rev. B **93**, 201202 (2016).

[53] H. Ishizuka, T. Hayata, M. Ueda, and N. Nagaosa, Phys. Rev. Lett. **117**, 216601 (2016).

[54] S. João and J. Lopes, arXiv preprint arXiv:1810.03732 (2018).

[55] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanziger, *The Geometric Phase in Quantum Systems: Foundations, Mathematical Concepts, and Applications in Molecular and Condensed Matter Physics* (Springer Science & Business Media, 2013).

[56] D. S. Freed and G. W. Moore, in *Annales Henri Poincaré*, Vol. 14 (Springer, 2013) pp. 1927–2023.

[57] J. Avron, L. Sadun, J. Segert, and B. Simon, Communications in mathematical physics **124**, 595 (1989).

[58] G. W. Moore, arXiv preprint arXiv:1706.01149 (2017).

[59] A. Alexandradinata, C. Wang, W. Duan, and L. Glazman, arXiv preprint arXiv:1707.08586 (2017).

[60] G. D. Mahan, *Many-particle physics* (Springer Science & Business Media, 2013).