

# Effects of renormalization and non-Hermiticity on nonlinear responses in strongly correlated electron systems

Yoshihiro Michishita<sup>✉\*</sup> and Robert Peters<sup>†</sup>*Department of Physics, Kyoto University, Kyoto 606-8502, Japan*
 (Received 22 December 2020; revised 2 March 2021; accepted 29 April 2021; published 17 May 2021)

Nonlinear responses in condensed matter are intensively studied because they provide rich information about the materials and hold the possibility of being applied in diodes or high-frequency optical devices. While nonlinear responses in noninteracting models have been explored widely, the effect of strong correlations on the nonlinear response is still poorly understood, even though it has been suggested that correlations can enhance the nonlinear response. In this work, we first give an analytical derivation of nonlinear responses using Green's function methods at finite temperature. Then, we discuss the difficulties of considering dissipation using conventional methods, such as the reduced density matrix method. We reveal that the relaxation time approximation leads to severe limitations when considering optical responses. Finally, we demonstrate that correlation effects, such as the renormalization of the band structure and different lifetimes in orbitals or sublattices, can significantly enhance nonlinear responses and even change the sign of the nonlinear conductivity.

DOI: [10.1103/PhysRevB.103.195133](https://doi.org/10.1103/PhysRevB.103.195133)

## I. INTRODUCTION

Nonlinear responses in condensed matter theory have attained great interest because of their rich information about the symmetries of materials and their various functionalities. For example, the breaking of the inversion symmetry in a material can be detected by measuring the second harmonic generation of the electric susceptibility [1–3]. Moreover, in noncentrosymmetric materials, the shift current and nonreciprocal (rectification) current can occur in nonlinear responses [4,5]. It was extensively studied due to its application in solar cells, photodetectors, and high-frequency rectification devices [6–10].

Although nonlinear responses in condensed matter systems have many possible applications, the magnitude of the nonlinear response, which is usually small, poses a significant obstacle for most applications. Thus, much effort has been put into enhancing the amplitude of the nonlinear response. It has been proposed that the shift current can be magnified in Dirac systems [4,7,11,12] and that superconducting fluctuations can enhance the nonreciprocity [13–15]. Another possibility to enhance nonlinear responses might be correlation effects. A strong high-harmonic generation was revealed in strongly correlated electron systems both in experiments [16,17] and numerical calculations [18–22]. A nonlinear Hall effect, which is almost  $10^3$  times as large as the *ab initio* calculation result, has been measured in the Weyl-Kondo semimetal candidate  $\text{Ce}_3\text{Bi}_4\text{Pd}_3$  [23]. Moreover, it has been suggested from a Hartree-analysis that the strong Coulomb interaction may enhance nonreciprocity [24]. Although these works show that correlation effects give large

nonlinear responses, a systematic analysis of strong correlation effects on nonlinear responses is still missing.

In this paper, we first derive a formalism based on Green's functions for calculating the nonlinear response at finite temperature and formulate a diagrammatic method to use them. We note that Parker *et al.* [25] derived a similar diagrammatic method for nonlinear responses focusing on the zero dissipation limit and João *et al.* [26] introduced a diagrammatic method based on Keldysh Green's functions. Neglecting vertex corrections, we can derive equations based on the single-particle Green's function, including correlation effects via the self-energy. Because there are many methods available to calculate the self-energy of correlated materials, the here-derived formalism makes it easy to analyze correlation effects on nonlinear responses. Next, we discuss difficulties of including the dissipation effect in conventional methods, such as the reduced density matrix (RDM) method [27–31]. In these methods, dissipation is often introduced phenomenologically by using the relaxation time approximation (RTA). We reveal that the RTA breaks the gauge invariance and is only justified in the DC limit, the high-frequency limit, and at high temperatures, while dissipation is appropriately included in the Green's function method.

Furthermore, while the RDM method for nonlinear responses mainly focuses on noninteracting systems, we demonstrate that it is possible to include correlation effects into the RDM using Green's functions. By including correlation effects into the RDM, we are able to retrieve the equations of the Green's function method in the DC limit. Finally, we use our Green's function formalism to analyze correlation effects on nonlinear responses. Notably, we look at the impact of the renormalization of the band structure and the effect of different lifetimes on the nonlinear response functions. We show that renormalization effects can enormously enhance the nonlinear response. Considering a renormalization uniform

\*michishita.yoshihiro.56e@st.kyoto-u.ac.jp

†peters@scphys.kyoto-u.ac.jp

in all orbitals, the renormalization factor  $z (< 1)$  enhances the  $n$ th-order response by a factor of  $z^{-(n-1)}$ . Furthermore, we study the effect of different lifetimes in different orbitals using the non-Hermitian band index of the effective non-Hermitian Hamiltonian describing the single-particle Green's function. We show that the occurrence of different lifetimes can not only enhance terms already existing in the Hermitian case, but also creates novel non-Hermitian terms in the nonlinear response function originating in the coalescence of several bands. Our framework can be applied to most correlated electron systems, such as heavy fermions, magnetic systems, Mott insulators, and so on. However, we note that it cannot be directly used for systems with strong spatial fluctuations because we ignore vertex corrections and the momentum dependence of the self-energy. On the other hand, by using the Nambu formalism, we can also expand our framework to superconducting systems.

The rest of the paper is organized as follows. In Sec. II, we derive the Green's function formalism for the nonlinear response at finite temperature. Next, we discuss the difficulties of including the dissipation in the RDM method in Sec. III. We reveal that the RTA under an AC electric field is a severe approximation, although it is often used in previous works. In Sec. IV, we extend the RDM method to interacting systems by using Green's functions. Finally, we analyze correlation effects, such as the renormalization of the band structure and the occurrence of different lifetimes in different orbitals, on the nonlinear response in Sec. V.

## II. NONLINEAR RESPONSE USING THE GREEN'S FUNCTION METHOD

In this section, we introduce the Matsubara formalism to express nonlinear response functions by Green's functions, which are common and easy to handle in the context of correlated systems at finite temperature. Throughout this paper, we set the Planck constant and the lattice constant to unity,  $\hbar = a = 1$ . We also set the electron charge  $e = 1$  in the numerical calculations.

We here use the velocity gauge, in which the effect of electric fields is described in the Hamiltonian as

$$\begin{aligned} \mathcal{H}(\mathbf{k}) &\rightarrow \mathcal{H}(\mathbf{k} - q\mathbf{A}(t)) \\ &= \mathcal{H}(\mathbf{k}) + \sum_{n=1}^{\infty} \frac{1}{n!} \prod_{i=1}^n (-qA^{\alpha_i}(t)\partial_{\alpha_i})\mathcal{H}(\mathbf{k}), \end{aligned} \quad (1)$$

where  $q$  is the charge of the electron and  $\alpha_i$  is a direction in the momentum space. In this paper, we suppose that there is no magnetic field and we use the Coulomb gauge  $\mathbf{A}(\mathbf{x}, t) = \mathbf{A}(t)$ . We note that there is another choice of gauge, namely the length gauge. Under the length gauge, electric fields can be described by the dipole Hamiltonian, and it is often used in the semiclassical Boltzmann equation and the RDM. It is known that both gauges give the same results for noninteracting systems when calculating exactly [29].

The action of the system in the imaginary time is given as

$$S[\mathbf{A}] = \int_0^\beta d\tau \left[ \sum_{k,a} \{ \bar{\psi}_{a,k} \partial_\tau \psi_{a,k} + \mathcal{H}(\mathbf{k} - q\mathbf{A}(-i\tau)) \} + H_{\text{int}} \right] \quad (2)$$

$$\begin{aligned} &= \int_0^\beta d\tau \left[ \sum_{k,a} \left\{ \bar{\psi}_{a,k} \partial_\tau \psi_{a,k} + \mathcal{H}(\mathbf{k}) \right. \right. \\ &\quad \left. \left. + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \prod_{i=1}^n (A^{\alpha_i}(-i\tau)) \hat{\mathcal{J}}_{\alpha_1 \dots \alpha_n}(\mathbf{k}) \right\} + H_{\text{int}} \right] \quad (3) \\ &\quad \hat{\mathcal{J}}_{\alpha_1 \dots \alpha_n}(\mathbf{k}) = q^n \partial_{\alpha_1} \dots \partial_{\alpha_n} \mathcal{H}(\mathbf{k}), \quad (4) \end{aligned}$$

where  $\bar{\psi}_a, \psi_a$  are fermionic creation and annihilation operators which construct the Hamiltonian  $\mathcal{H}$ ,  $a$  is the orbital index,  $\mathbf{A}(t)$  is the vector potential,  $\hat{\mathcal{J}}_{\alpha_1 \dots \alpha_n}(\mathbf{k}) = q^n \partial_{\alpha_1} \dots \partial_{\alpha_n} \mathcal{H}(\mathbf{k})$ , and  $H_{\text{int}}$  is the interaction part of the Hamiltonian. In this paper, we suppose that there is only a local interaction. We note that, for general nonlocal interactions, the interaction part of the Hamiltonian also depends on the vector potential.

The partition function with applied electric field is written in the path integral formalism as

$$Z[\mathbf{A}] = \int \mathcal{D}\bar{\psi} \mathcal{D}\psi \exp[-S[\mathbf{A}]]. \quad (5)$$

The expectation value of the current is

$$\langle J_\alpha(\tau) \rangle = \frac{\delta}{Z[\mathbf{A}] \delta A^\alpha(-i\tau)} Z[\mathbf{A}], \quad (6)$$

which can be written using response functions as

$$\begin{aligned} \langle J_\alpha(\tau) \rangle &= \int d\tau' \mathcal{K}_{\alpha\beta}^1(\tau, \tau') A^\beta(-i\tau') \\ &\quad + \int d\tau' \int d\tau'' \mathcal{K}_{\alpha\beta\gamma}^2(\tau, \tau', \tau'') A^\beta(-i\tau') A^\gamma(-i\tau'') \\ &\quad + \dots, \end{aligned} \quad (7)$$

where

$$\begin{aligned} &\mathcal{K}_{\alpha\alpha_1 \dots \alpha_n}^n(\tau_1, \dots, \tau_n) \\ &= \frac{1}{Z[\mathbf{A}]} \left( \prod_{i=1}^n \frac{\delta}{\delta A^{\alpha_i}(-i\tau_i)} \right) \frac{\delta}{\delta A^\alpha(-i\tau)} Z[\mathbf{A}] \Big|_{\mathbf{A}=0}. \end{aligned} \quad (8)$$

The results for the response functions in imaginary time are explicitly written in Appendix A.

After Fourier transformation to Matsubara frequencies, the current is given as

$$\begin{aligned} \langle J_\alpha(i\omega_n) \rangle &= K_{\alpha\beta}^{(1)}(i\omega_n; i\omega_n) A^\beta(i\omega_n) \\ &\quad + \sum_{\omega_m, \omega_l} K_{\alpha\beta\gamma}^{(2)}(i\omega_n; i\omega_m, i\omega_l) A^\beta(i\omega_m) A^\gamma(i\omega_l) \\ &\quad \times \delta(\omega_n - \omega_m - \omega_l) + \dots \end{aligned} \quad (9)$$

The frequency before the semicolon in the response function  $K_{\alpha\beta}^{(n)}(i\omega_n; i\omega_n, \dots)$  represents the frequency of the output response, and the frequencies after the semicolon represent the frequencies of the input forces, i.e., of the vector potentials.

Analytical continuation and using  $\mathbf{E}(\omega_i) = i\omega_i\mathbf{A}(\omega_i)$  finally yields

$$\langle J_\alpha(\omega) \rangle = K_{\alpha\beta}^{(1)}(\omega; \omega)A^\beta(\omega) + \int d\omega_1 \int d\omega_2 K_{\alpha\beta\gamma}^{(2)}(\omega; \omega_1, \omega_2)A^\beta(\omega_1)A^\gamma(\omega_2)\delta(\omega - \omega_{12}) + \dots \quad (10)$$

$$= \sigma_{\alpha\beta}^{(1)}(\omega)E^\beta(\omega) + \int d\omega_1 \sigma_{\alpha\beta\gamma}^{(2)}(\omega; \omega_1, \omega_2)E^\beta(\omega_1)E^\gamma(\omega_2)\delta(\omega - \omega_{12}) + \dots, \quad (11)$$

$$\sigma_{\alpha\beta\dots}^{(n)}(\omega; \{\omega_s\}) = K_{\alpha\beta\dots}^{(n)}(\omega; \{\omega_i\}) / \left( \prod_{s=1}^n i\omega_i \right), \quad (12)$$

where  $\omega_{12} = \omega_1 + \omega_2$ . The first- and second-order conductivities can be expressed via single-particle Green's functions as

$$\sigma_{\alpha\beta}^{(1)}(\omega_1; \omega_1) = -\frac{1}{\omega_1} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} f(\omega) \sum_{\mathbf{k}} \{ \text{Tr}[\mathcal{J}_{\alpha\beta}(\mathbf{k})(G^R(\omega, \mathbf{k}) - G^A(\omega, \mathbf{k}))] \\ + \text{Tr}[\mathcal{J}_\alpha(\mathbf{k})G^R(\omega + \omega_1, \mathbf{k})\mathcal{J}_\beta(\mathbf{k})(G^R(\omega, \mathbf{k}) - G^A(\omega, \mathbf{k})) + \mathcal{J}_\alpha(\mathbf{k})(G^R(\omega, \mathbf{k}) - G^A(\omega, \mathbf{k}))\mathcal{J}_\beta(\mathbf{k})G^A(\omega - \omega_1, \mathbf{k})] \} \quad (13)$$

$$\sigma_{\alpha\beta\gamma}^{(2)}(\omega_1 + \omega_2; \omega_1, \omega_2) = \frac{1}{\omega_1\omega_2} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} f(\omega) \sum_{\mathbf{k}} \left\{ \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha\beta\gamma}(G^R(\omega) - G^A(\omega))] \right. \\ + \text{Tr}[\mathcal{J}_{\alpha\beta}G^R(\omega + \omega_2)\mathcal{J}_\gamma(G^R(\omega) - G^A(\omega)) + \mathcal{J}_{\alpha\beta}(G^R(\omega) - G^A(\omega))\mathcal{J}_\gamma G^A(\omega - \omega_2)] \\ + \frac{1}{2} \text{Tr}[\mathcal{J}_\alpha G^R(\omega + \omega_{12})\mathcal{J}_{\beta\gamma}(G^R(\omega) - G^A(\omega)) + \mathcal{J}_\alpha(G^R(\omega) - G^A(\omega))\mathcal{J}_{\beta\gamma}G^A(\omega - \omega_{12})] \\ + \text{Tr}[\mathcal{J}_\alpha G^R(\omega + \omega_{12})\mathcal{J}_\beta G^R(\omega + \omega_2)\mathcal{J}_\gamma(G^R(\omega) - G^A(\omega)) \\ + \mathcal{J}_\alpha G^R(\omega + \omega_1)\mathcal{J}_\beta(G^R(\omega) - G^A(\omega))\mathcal{J}_\gamma G^A(\omega - \omega_2) \\ \left. + \mathcal{J}_\alpha(G^R(\omega) - G^A(\omega))\mathcal{J}_\beta G^A(\omega - \omega_1)\mathcal{J}_\gamma G^A(\omega - \omega_{12})] + [(\beta, \omega_1) \leftrightarrow (\gamma, \omega_2)] \right\}, \quad (14)$$

where  $\mathcal{J}_{\alpha\beta\dots}$  is the matrix representation of  $\hat{\mathcal{J}}_{\alpha\beta\dots}$ ,  $G^{R/A}(\omega, k)$  is the retarded/advanced Green's function, and  $f(\omega)$  is the Fermi distribution function.  $[(\beta, \omega_1) \leftrightarrow (\gamma, \omega_2)]$  means a term in which the index and the variable have been replaced by the other set. Further details of the derivation are given in the Appendixes A and B. Throughout this paper, we omit the  $\mathbf{k}$ -index of the Green's function and the velocity operator,  $\mathcal{J}_{\alpha\beta\dots}$ . Furthermore, we ignore vertex corrections in the many-particle Green's functions, which allows us to express the conductivity as a product of single-particle Green's functions. This approximation is also commonly used in the semiclassical Boltzmann equation and the RDM formalism. The results above are consistent with the results in [25,26]. Specifically, in the dissipationless limit, the results are consistent with [32] Eqs. (26) and (43) in [25]. The detail is written in Appendix E. The here-presented procedure to derive the nonlinear optical conductivity can be summarized into a diagrammatic method, which is given in Appendix D. We note that this diagrammatic method is a generalization of the diagrammatic method at zero temperature in Parker *et al.* [25] to nonlinear response functions using real-frequencies at finite temperature.

If we take the DC limit  $\omega_1, \omega_2 \rightarrow 0$ , the first- and second-order conductivities become

$$\sigma_{DC;\alpha\beta}^{(1)} = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left\{ \left( -\frac{\partial f(\omega)}{\partial \omega} \right) \text{ReTr}[\mathcal{J}_\alpha G^R(\omega)\mathcal{J}_\beta G^A(\omega)] - 2f(\omega) \text{ReTr} \left[ \mathcal{J}_\alpha \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_\beta G^R(\omega) \right] \right\}, \quad (15)$$

$$\sigma_{DC;\alpha\beta\gamma}^{(2)} = -2 \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left\{ \left( -\frac{\partial f(\omega)}{\partial \omega} \right) \text{Im} \left( \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_\beta G^R(\omega)\mathcal{J}_\gamma G^A(\omega) \right] + \frac{1}{2} \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_{\beta\gamma} G^A(\omega) \right] \right) \right. \\ \left. - f(\omega) \text{Im} \left( \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial}{\partial \omega} \left( \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_\beta G^R(\omega) \right) \mathcal{J}_\gamma G^R(\omega) \right] + \frac{1}{2} \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial^2 G^R(\omega)}{\partial \omega^2} \mathcal{J}_{\beta\gamma} G^R(\omega) \right] \right) \right\} + (\beta \leftrightarrow \gamma). \quad (16)$$

Interaction effects can be taken into account by including the retarded/advanced self-energy  $\Sigma^{R/A}(\omega)$  into the Green's function,  $G^R(\omega) = 1/[\omega - \mathcal{H} - \Sigma^{R/A}(\omega)]$ . Throughout this paper, we ignore the momentum dependence of the self-energy. Including the momentum dependence of the self-energy, we should also consider vertex corrections to satisfy the Ward-Takahashi identities. We note that the momentum dependence of the self-energy can become significant for certain

phenomena in strongly correlated materials and, in these cases, must be included in the considerations about nonlinear responses. We also note that we can recover the physical unit by substituting  $\omega \rightarrow \hbar\omega$  and multiply  $a^n$  for  $n$ th-order nonlinear conductivity.

Finally, setting  $\Sigma(\omega) = i\gamma/2$  and taking the limit  $\gamma \rightarrow 0$ , we can perform the frequency integrals and further simplify the results which are summarized in Appendix E.

### III. DIFFICULTIES DESCRIBING DISSIPATION EFFECTS IN THE REDUCED DENSITY MATRIX FORMALISM

Having introduced the Green's function technique based on a path-integral derivation to calculate nonlinear transport, we can compare with different approaches and approximations made to calculate the nonlinear response. The semiclassical Boltzmann equation and the RDM method are often used to calculate nonlinear responses. In these methods, the dissipation is usually introduced by the relaxation time approximation (RTA). In this section, we briefly introduce the RDM method. Being able to compare it with the Green's function method, we can pinpoint the problems accompanying the RTA and explain in what situation RTA is justified. We note that the results by the semiclassical Boltzmann equation can also be obtained by the RDM results [25] so that we here consider only the RDM method. We briefly introduce the Boltzmann equation approach to nonlinear transport in Appendix F.

#### A. Reduced density matrix formalism

When ignoring two-body correlations, we can write the total density matrix of the lattice system as the tensor product of the reduced density matrices  $\rho_{\text{tot}}(t) = \prod_{\mathbf{k}} \otimes \rho_{\mathbf{k}}(t)$ . We can now describe the dynamics of the density matrix for each momentum  $\mathbf{k}$  under the electric field by using the von Neumann equation, which reads

$$\frac{d}{dt}\rho_{\mathbf{k}}(t) = -i[\mathcal{H}, \rho_{\mathbf{k}}(t)] - (\rho(t) - \rho^{(0)})/\tau, \quad (17)$$

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_E, \quad (18)$$

$$\mathcal{H}_E = -q\mathbf{E} \cdot \mathbf{r}, \quad \mathbf{r} = -i\nabla_{\mathbf{k}}, \quad (19)$$

where we introduce the effect of dissipation by using the RTA,  $-(\rho(t) - \rho^{(0)})/\tau$ , and  $\rho^{(0)}$  describes the equilibrium state without the electric field. In the RDM formalism, we use the length gauge and describe the dynamics with the dipole Hamiltonian in Eq. (19). The density matrix under the velocity gauge can be obtained by using the transformation  $\rho_E = U\rho_A U^\dagger$ ,  $U = \exp[-iq\mathbf{A}(t) \cdot \mathbf{r}]$ , where  $\rho_{E/A}$  is the density matrix under the length/velocity gauge. We note again that results obtained by the length gauge are equivalent to those obtained under the velocity gauge without dissipation [29]. The recurrence equation of the  $n$ th-order density matrix  $\rho^{(n)}(t)$  about the electric field can be written as

$$\begin{aligned} \frac{d\rho^{(n)}(t)}{dt} &= -i[\mathcal{H}_0, \rho^{(n)}(t)] - i[\mathcal{H}_E, \rho^{(n-1)}(t)] - \gamma\rho^{(n)}(t) \\ &= -i\mathcal{L}\rho^{(n)}(t) + q\mathbf{E}(t) \cdot \nabla\rho^{(n-1)}(t) - \gamma\rho^{(n)}(t), \end{aligned} \quad (20)$$

$$F.T. \Leftrightarrow \rho_{\mathbf{k}}^{(n)}(\omega) = \frac{iqE^\mu(\omega_n)}{\omega - \mathcal{L} + i\gamma} \nabla_\mu \rho_{\mathbf{k}}^{(n-1)}(\omega - \omega_n), \quad (21)$$

where  $-i\mathcal{L}\rho = -i[\mathcal{H}_0, \rho]$ ,  $\gamma = 1/\tau$ ,  $\omega_n$  describes the frequency of the electric field which leads to the  $n$ th-order density matrix  $\rho^{(n)}$ ,  $\mathbf{E}(\omega_n)$  is the Fourier component of  $\mathbf{E}(t)$ , and *F.T.* means Fourier transformation. In the length gauge,

the current operator  $\mathcal{J}$  can be written as

$$\mathcal{J} = q\dot{\mathbf{r}} = -iq[\mathbf{r}, \mathcal{H}] = q\nabla_{\mathbf{k}}\mathcal{H}, \quad (22)$$

and, therefore, the  $n$ th-order conductivity can be calculated by

$$\sigma_{\alpha;\{\alpha_i\}}^{(n)}(\omega; \{\omega_i\}) = \sum_{\mathbf{k}} \text{Tr} \left[ \mathcal{J}_\alpha \rho_{\mathbf{k}}^{(n)}(\omega) / \left( \prod_i E^{\alpha_i}(\omega_i) \right) \right]. \quad (23)$$

Detailed expressions can be found in [28,33]. We note that the equations of the RDM method using RTA can be derived from the Green's function technique in the DC limit and in the dissipation-free limit for  $\omega_i \gg \epsilon_{nm}, \gamma$ . Details about this correspondence are given in Appendix E. The RDM method introduced here is exact except for the RTA, and therefore, the necessary conditions we listed above are caused by the RTA.

#### B. Velocity gauge versus length gauge under the relaxation time approximation

In an isolated system without dissipation, physical quantities calculated by the velocity and length gauge are the same, which was shown in [29]. In this subsection, we show that this correspondence between both gauges breaks down when using the RTA. The density matrix in each gauge can be written as [29]

$$\rho_E(t) = U\rho_A(t)U^\dagger, \quad U = \exp[-i\mathbf{A}(t) \cdot \mathbf{r}], \quad (24)$$

$$\begin{aligned} \rho_E^{(n)}(t) &= \rho_A^{(n)}(t) + \sum_{l=1}^n \left( \prod_{m=1}^l -i\mathbf{A}^{\alpha_m}(t) \right) \\ &\quad \times [r_{\alpha_l}, [r_{\alpha_{l-1}}, \dots [r_{\alpha_1}, \rho^{(n-l)}]]] \\ &= \rho_A^{(n)}(t) - i\mathbf{A}(t) \cdot [\mathbf{r}, \rho_A^{(n-1)}(t)] - \dots, \end{aligned} \quad (25)$$

where  $\rho_E(t)$  is the density matrix under the length gauge,  $\rho_A(t)$  is the density matrix under the velocity gauge, and  $\rho^{(n)}(t)$  represent the density matrix with the  $n$ th-order perturbation by the electric fields. By applying the RTA, the density matrices under both gauges change as  $\rho_{E/A}^{(n)}(t) \rightarrow \rho_{E/A}^{(n)}(t)e^{-\gamma t}$  when  $n \geq 1$ . The equality in Eq. (25) for the  $n = 1$ -order density matrix using the RTA becomes

$$\rho_E^1(t)e^{-\gamma t} \stackrel{?}{=} \rho_A^1(t)e^{-\gamma t} - i\mathbf{A}(t) \cdot [\mathbf{r}, \rho_A^0(t)]. \quad (26)$$

However, because  $\rho_A^{(0)}$  does not include dissipation, the equality in Eq. (26) has to break down.

One possible strategy to avoid this breakdown is to ignore the dissipation in the system and instead include photon dissipation or adiabatic switching as  $\mathbf{A}(t) \rightarrow \mathbf{A}(t)e^{-\gamma t}$ . In this case, the equality in Eq. (25) holds true. However, it gives different results from the RTA, especially in the regime  $\omega_i \ll \gamma$  [30]. When substituting  $\mathbf{A}(t) \rightarrow \mathbf{A}(t)e^{-\gamma t}$ , we do not consider the dissipation and scattering of electrons in the system. Thus, a current must not occur because there is no mechanism to change the momentum of electrons,  $\mathbf{k} \rightarrow \mathbf{k}'$ , and to induce a nonequilibrium steady-state state. Therefore, when including dissipation of electrons by applying the RTA, a breakdown of the equality between the velocity gauge and the length gauge is inevitable. We note that, in the Green's function method, this breakdown does not occur when we use  $G^R(\omega) =$

$1/(\omega - \mathcal{H} + i\gamma/2)$  and  $G^A(\omega) = 1/(\omega - \mathcal{H} - i\gamma/2)$  because it just supposes that the dissipation is constant in the absence of an electric field.

### C. Problems of the relaxation time approximation in an AC electric field

In this part, we introduce the dissipation into the RDM method without using the RTA and show under which conditions the RTA is a good approximation. This analysis reveals the problems of using the RTA in an AC electric field. Finally, we compare the RDM using the RTA with the Green's function formalism numerically.

The easiest way to introduce the dissipation microscopically is to couple the system with a dissipative bath. For the sake of simplicity, we consider the single-band case and the coupling Hamiltonian  $\mathcal{H}_c = \lambda(\psi_k^\dagger \mathcal{B} + \text{H.c.})$ , where  $\mathcal{B}^{(\dagger)}$  is the annihilation (creation) operator in the dissipative bath. In that case, the dynamics of the system can be described by the quantum master equation, which reads

$$\frac{d}{dt}\rho_k^l(t) = -\int_{t_0}^t ds \text{Tr}_B([\mathcal{H}_c^l(t), [\mathcal{H}_c^l(s), \rho_k^l(s) \otimes \rho_B]]), \quad (27)$$

where  $\rho_B$  is the density matrix of the bath and  $\text{Tr}_B$  corresponds to the trace over the bath degrees of freedom. The operators are in the interaction representation,  $\mathcal{O}^l(t) = T_{\leftarrow} \exp[i \int_{t_0}^t dt' \mathcal{H}(t')] \mathcal{O} T_{\rightarrow} \exp[-i \int_{t_0}^t dt' \mathcal{H}(t')]$ , where  $\mathcal{H}(t) = \mathcal{H}_S(t) \otimes \mathcal{H}_B$ ,  $\mathcal{H}_S(t) = \mathcal{H}_0 - q\mathbf{E}(t) \cdot \mathbf{r}$  is the system Hamiltonian,  $\mathcal{H}_B$  is the bath Hamiltonian,  $T_{\rightarrow(\leftarrow)}$  rep-

resents the (anti)time-ordering operator. Although we take here the length gauge, the correspondence between the length and the velocity gauge holds exactly in this formulation. The proof is written in Appendix G. Equation (27) includes the dissipation term, the energy shift term, and the gain and loss terms which describe the dynamics of a particle leaving or entering the system. Here, we suppose that a particle that leaves the system loses the information about the acceleration due to the electric fields, and the electric fields do not accelerate the particles in the bath. Under this assumption, the gain and loss terms do not affect the dynamics of  $\rho^{(n)}$  ( $n \neq 0$ ), and therefore, they do not affect the conductivity. Now we focus on the dissipation term and ignore the energy shift term. Then Eq. (27) can be rewritten as

$$\begin{aligned} \frac{d}{dt}(\rho_k^l(t))^{(n)} &= -\lambda^2 \int_{t_0}^t ds \text{Re}[\{iG_B^l(t-s)\psi_k^{\dagger l}(t)\psi_k^l(s), \rho_k^l(s)\}]^{(n)}, \end{aligned} \quad (28)$$

where  $G_B^l(t-s) = -i\text{Tr}_B[\mathcal{B}^l(t)\mathcal{B}^{\dagger l}(s)\rho_B]$  and  $\{\mathcal{O}, \rho\} = \mathcal{O}\rho + \rho\mathcal{O}^\dagger$ .

Now we use the Markov approximation to simplify Eq. (28), in which we take the limit  $t_0 \rightarrow -\infty$  and approximate  $\rho_k^l(s) \simeq \rho_k^l(t)$ . The Markov approximation is justified when  $\lambda\tau_B \ll 1$ , where  $\tau_B$  is the relaxation time of the bath and  $G_B(t-s) \propto \exp[-(t-s)/\tau_B]$ . Under the Markov approximation, Eq. (28) can be rewritten as

$$\frac{d}{dt}\rho_k(t) = -i[\mathcal{H}, \rho_k(t)] - \lambda^2 \left( \int_0^\infty d(t-s) \text{Re}[\{iG_B^l(t-s)\psi_k^{\dagger l} U'(t,s)\psi_k U'^{\dagger l}(t,s), \rho_k(t)\}] \right), \quad (29)$$

$$U'(t,s) = T_{\rightarrow} \exp \left[ -i \int_s^t dt' (\mathcal{H}_0 - q\mathbf{E} e^{i\omega_0 t'} \cdot \mathbf{r}) \right]. \quad (30)$$

Finally, we consider in what situation we can derive the RTA from Eq. (29). RTA should be a good approximation to describe transport when the integral in Eq. (29) becomes time-independent, thus, when  $U'(t,s)$  becomes a function of  $(t-s)$  or is constant. We see that in the DC limit  $\omega_0 \rightarrow 0$  or when the temperature of the bath is infinite and  $G_B(t-s) \propto \delta(t-s)$ , or when  $\omega_0$  is large enough so that  $q\mathbf{E} \cdot \mathbf{r}/\omega_0$  can be ignored, the integral ( $\int d(t-s) \sim$ ) becomes a constant and Eq. (17) can be derived from Eq. (29).

After having analyzed the validness of the RTA, we will now directly compare the linear and nonlinear (photogalvanic) optical conductivity calculated by the Green's function method with the RDM using the RTA for a simple model. For this purpose, we use a model describing two-dimensional transition metal dichalcogenides (TMD) in which nonlinear optical response was discussed in the literature [10,34,35]. Details about the model are given in Appendix H and the details about how to perform the numerical calculations is given in Appendix I.

The numerical results of the optical conductivity by the RDM method using the RTA and by the Green's function

method are shown in Fig. 1. For the linear optical conductivity, the results of both methods agree with each other over the full frequency range. On the other hand, for the nonlinear optical conductivity, the results only match in the DC limit, and for large frequencies  $\omega_i \gg \gamma$ , as has been discussed above.

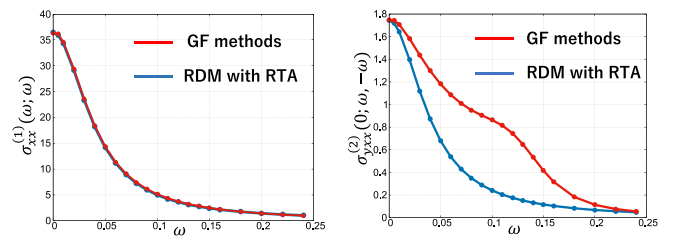


FIG. 1. Comparison between the RDM method using RTA (blue) and the Green's function method (red) for the linear (left panel) and nonlinear (right panel) optical conductivities, which is the photogalvanic effect. The parameters for the monolayer TMD model are  $t = 0.5$ ,  $\mu = 0.7$ ,  $p = 0.7$ ,  $\alpha_1 = 0.08$ ,  $\alpha_2 = 0.06$ ,  $\delta = 0.7$ ,  $\gamma = 0.05$ ,  $T = 0.02$ .

We thus find that while RTA is a good approximation for the linear optical conductivity, it leads to severe problems for the nonlinear optical conductivity except in the DC limit and for  $\omega_i \gg \gamma$ . Again, we note that the RTA supposes that all nonequilibrium states decay equally by  $\gamma$ . On the other hand, the Green's function method only assumes that the dissipation is constant in the absence of an electric field. The RTA is a more severe approximation, which affects nonlinear responses. We note that the relaxation time in most materials is usually about  $1 \sim 100$ [ps] [36]. Thus, when analyzing a Terahertz laser as input force,  $\omega_i \tau \sim 1$ , and the error of the RTA becomes large.

#### IV. EXTENSION OF THE REDUCED DENSITY MATRIX FORMALISM TO INTERACTING SYSTEM

Having derived the Green's function method for nonlinear responses, we are able to extend the RDM method to interacting systems, mainly in the DC limit, and reproduce the results of the Green's function method. For free electron systems, we use  $\rho_k^{(0)} = \sum_n f[\epsilon_n(\mathbf{k})] |n\rangle \langle n|$ . However, when we consider interacting systems, the pole of the Green's function includes the information of the quasiparticle's energy level, and therefore, the density matrix can be written as

$$\rho_k^{(0)} = \int \frac{d\omega}{2\pi i} \sum_n |n\rangle \langle n| (G_n^A(\omega) - G_n^R(\omega)) f(\omega) \quad (31)$$

$$= \int \frac{d\omega}{2\pi i} \sum_{\alpha\beta} |\alpha\rangle \langle \beta| (G^A(\omega) - G^R(\omega))_{\alpha\beta} f(\omega), \quad (32)$$

where  $|\alpha\rangle, |\beta\rangle$  are states of an arbitrary basis, and  $(\ )_{\alpha\beta}$  represent the elements of the Green's function in this basis. We note that we again omit the momentum dependence of the Green's function. Here, we can choose a momentum-independent basis  $\partial_\alpha |\alpha\rangle = 0$ . In this case, the correction of the density matrix by the electric fields only affects the Green's function matrices because  $f(\omega)$  does not depend on  $\mathbf{k}$ . Therefore, the density matrix corrected by  $n$ th-order electric fields can be written as

$$\begin{aligned} & (\rho_k^{(n)})_{\alpha\beta} \\ &= \int \frac{d\omega}{2\pi i} \sum_{l=0}^n (G^{R(l)}(\omega) (G^{R(0)}(\omega))^{-1} - (G^{A(0)}(\omega))^{-1}) \\ & \quad \times f(\omega) G^{A(n-l)}(\omega)_{\alpha\beta}. \end{aligned} \quad (33)$$

Although we need the Green's function corrected by the  $n$ th-order of the electric field, we can easily derive an equation for this using the RDM method. Here, we note that in our previous work [37], we showed that the dynamics of the matrix elements  $\rho_{\mathbf{k},\alpha}^G(0) = |\mathbf{k}, \alpha\rangle \langle 0|$  corresponds to the retarded Green's function  $G^R(\mathbf{k})$ , which reads

$$G_{\alpha\beta}^R(t) = -i\theta(t)\text{Tr}[(\psi_\alpha(t)\psi_\beta^\dagger + \psi_\beta^\dagger\psi_\alpha(t))\rho_k^{(0)}] \quad (34)$$

$$= -i\text{Tr}[\psi_\alpha \rho_\beta^G(t)], \quad (35)$$

where  $\rho_{\mathbf{k}\beta}^G(0) = \psi_\beta^\dagger \rho_k^{(0)} + \rho_k^{(0)} \psi_\beta^\dagger = |\mathbf{k}, \beta\rangle \langle 0|$  and the dynamics of  $\rho^G$  can be described [37] as

$$\frac{d}{dt} \rho_k^{GI}(t) = - \int_{t_0}^t ds i \Sigma^{RI}(t-s) \rho_k^{GI}(s), \quad (36)$$

$$\begin{aligned} \frac{d}{dt} \rho_k^G(t) &= -i[\mathcal{H}_0 + \mathcal{H}_E, \rho_k^G(t)] - \int_{t_0}^t ds i \Sigma^R(t-s) \\ & \quad \times \left( \sum_{n=0}^{\infty} \frac{(iq(t-s)\mathbf{E} \cdot \mathbf{r})^n}{n!} \right) \rho_k^G(s), \end{aligned} \quad (37)$$

$$\begin{aligned} \rho_k^{G(n)}(\omega) &= \frac{iq\mathbf{E}}{\omega - (\mathcal{H}_0 + \Sigma^R(\omega))} \left\{ \left( 1 - \frac{\partial \Sigma^R}{\partial \omega} \right) \nabla \rho_k^{G(n-1)}(\omega) \right. \\ & \quad \left. + \sum_{m=2}^{\infty} \frac{1}{m!} \frac{\partial^m \Sigma^R}{\partial \omega^m} (-iq\mathbf{E} \cdot \nabla)^m \rho_k^{G(n-m)}(\omega) \right\}. \end{aligned} \quad (38)$$

To derive Eq. (37), we approximate  $U'(t, s) \simeq \exp[-i\mathcal{H}_0(t-s)] \exp[-iq\mathbf{E} \cdot \mathbf{r}(t-s)]$  in the dissipation term, which should correspond to ignoring the vertex correction. By using this equation,  $G^{R(1)}(\omega)$  (the first-order correction of an electric field to the single-particle Green's function) can be derived as

$$\begin{aligned} & (G^{R(1)}(\omega))_{\alpha\beta} \\ &= iE^\mu \left( G^{R(0)}(\omega) \left( 1 - \frac{\partial \Sigma^R}{\partial \omega} \right) G^{R(0)}(\omega) \mathcal{J}_\mu G^{R(0)}(\omega) \right)_{\alpha\beta} \\ &= -iE^\mu \left( \frac{\partial G^{R(0)}}{\partial \omega} \mathcal{J}_\mu G^{R(0)}(\omega) \right)_{\alpha\beta}. \end{aligned} \quad (39)$$

By inserting Eq. (39) into Eq. (33), we can derive the equation for the linear conductivity as given by the path-integral method in Eq. (15). We can also calculate the higher-order DC conductivity in the same way. We note that using the RDM methods might be easier than the path-integral methods for higher-order DC conductivities. However, in the AC case, it is hard to derive an equation equivalent to Eq. (38) so that the path-integral method should be used.

#### V. CORRELATION EFFECTS ON THE NONLINEAR RESPONSE

Finally, we use the Green's function formalism and analyze the effect of renormalization and different lifetimes in different orbitals, which were not considered in previous studies. We reveal that both effects can enhance the nonlinear conductivity.

##### A. Renormalization effect

Intuitively, the renormalization effect seems to be a disadvantage for obtaining a large conductivity because it decreases the Fermi velocity. However, as the density of states might be enhanced by the renormalization at the Fermi surface, one should properly analyze how the renormalization affects the linear and the nonlinear conductivities.

First, we analyze the simple case where  $\Sigma^R(\omega) \simeq \Sigma_0^R + \alpha\omega\mathbf{1}$ . Under this approximation, the Green's function can be written as

$$\begin{aligned} G^{R-1}(\omega) &= \omega - \mathcal{H}_0 - \Sigma^R(\omega) \simeq (1-\alpha)\omega - \mathcal{H}'_0 \\ &\equiv G'^{R-1}(Z^{-1}\omega), \end{aligned} \quad (40)$$

where  $Z^{-1} = 1 - \alpha$ ,  $\mathcal{H}'_0 = \mathcal{H}_0 + \Sigma_0^R$ , and  $G'^{R-1}(\omega) = \omega - \mathcal{H}'_0$ . We can now analyze the effect of the renormalization on

the conductivities calculated by the Green's function method. By the variable transformation  $Z^{-1}\omega \rightarrow \omega'$  and  $Z^{-1}\omega_i \rightarrow \omega'_i$ , the functions which appear in the linear and nonlinear conductivities change as follows:

$$G^{R/A}(\omega) = G^{R/A}(\omega'), \quad (41)$$

$$f(\omega) \simeq \theta(-\omega) = \theta(-\omega') \simeq f(\omega'), \quad (42)$$

$$\frac{\partial f(\omega)}{\partial \omega} \simeq \delta(\omega) = Z^{-1}\delta(\omega') = \frac{\partial f(\omega')}{\partial \omega'}, \quad (43)$$

$$\frac{\partial G^R(\omega)}{\partial \omega} = Z^{-1} \frac{\partial G^R(\omega')}{\partial \omega'}, \quad (44)$$

$$d\omega = Zd\omega', \quad \frac{1}{\omega_i} = \frac{Z^{-1}}{\omega'_i}, \quad (45)$$

where the equality in Eqs. (42) and (43) are justified at zero temperature. By inserting the above equations into Eqs. (14) and (16), we can derive  $\sigma^{(2)} = Z^{-1}\sigma'^{(2)}$  in both the AC and the DC case, where  $\sigma'$  is the conductivity described by  $\omega'$ ,  $\omega'_i$ ,  $G^{R/A}(\omega')$ , which includes the energy shift by  $\Sigma_0^R$ . We note that we should compare the renormalized conductivity  $\sigma^{(n)}(\omega; \{\omega_i\})$  with  $\sigma'^{(n)}(\omega'; \{\omega'_i\}) = \sigma^{(n)}(Z^{-1}\omega; \{Z^{-1}\omega_i\})$  in the AC case. In the optical conductivity, the interband contribution becomes large when  $\omega_i \simeq \epsilon_{nm}$ . To focus on the same interband transition, we set the frequency  $Z^{-1}\omega$  for the renormalized band. We can generalize this analysis for higher-order conductivities and find

$$\sigma^{(n)}(\omega; \{\omega_i\}) \simeq Z^{-(n-1)}\sigma'^{(n)}(\omega'; \{\omega'_i\}). \quad (46)$$

By remembering that  $Z^{-1} > 1$  holds for correlated systems around the Fermi energy, we conclude that the renormalization effect enhances the higher-order nonlinear conductivity more strongly, while it does not affect the linear conductivity.

Using the Green's function technique, we can easily confirm our general discussion above by calculating the linear and the nonlinear optical conductivity for the monolayer TMD model. The results for these calculations using an unrenormalized band ( $Z = 1$ ) and a renormalized band ( $Z = 0.2$ ) are shown in Fig. 2. As we derived analytically, the numerical results confirm that the nonlinear optical response is strongly enhanced by the renormalization effect, while the linear optical response is not enhanced. We note that the renormalized nonlinear optical conductivity is not as strongly enhanced as predicted [ $Z^{-1}\sigma'^{(2)}(0; -\omega, \omega)$ ] in Fig. 2, which can be attributed to a finite temperature  $T = 0.02$ , where the Fermi function does not correspond to the step function.

## B. Different lifetimes in different orbitals

In this section, we analyze the effect of different lifetimes in different orbitals, which is not considered within the RTA. We note that there is the study by Kaplan *et al.* [38,39], where the authors analyzed the effect of different lifetimes on the nonlinear response, assuming that the conventional band-index representation is justified.

When using the RTA, the non-Hermitian (dissipation) term is described by the identity matrix. Therefore, the eigenvectors are the same as that of the Hermitian Hamiltonian. However, when different lifetimes are present in different orbitals, as in

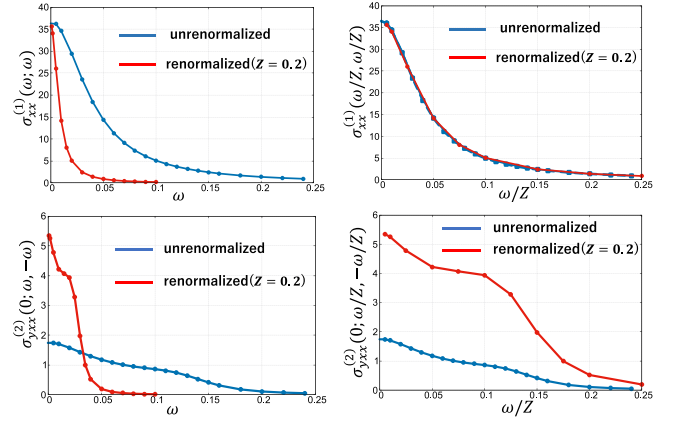


FIG. 2. Renormalization effect on the linear and nonlinear optical conductivity for the monolayer TMD model. The upper figures show the linear optical conductivity, and the lower figures show the second-order nonlinear optical conductivity (photogalvanic effect) using an unrenormalized (blue lines) and a renormalized (red lines) band. In the right figures, we use input frequencies normalized by the renormalization factor. The parameters are  $t = 0.5$ ,  $\mu = 0.7$ ,  $p = 0.7$ ,  $\alpha_1 = 0.08$ ,  $\alpha_2 = 0.06$ ,  $\delta = 0.7$ ,  $\gamma = 0.05$ ,  $T = 0.02$ ,  $Z = 1$ , or  $0.2$ . The details about how to perform the numerical calculations are written in Appendix I.

a material consisting of strongly correlated electrons coupled to weakly-correlated electrons, the eigenvectors are distinct from the Hermitian case. The eigenvectors are then determined by the effective non-Hermitian Hamiltonian, which describes the single-particle Green's function. In that case, the conventional band-index representation breaks down, and one should use a non-Hermitian band index. In this section, we first derive the non-Hermitian band index and then analyze its effect.

### 1. Band-index representation using an effective non-Hermitian Hamiltonian

In this paper, we suppose that the effective non-Hermitian Hamiltonian can be diagonalized. We note that, in general, there are situations when a non-Hermitian Hamiltonian cannot be diagonalized, which generates novel and interesting phenomena [40–45]. For a non-Hermitian Hamiltonian, its left eigenstates are different from its right eigenstates, while in the Hermitian case they correspond to each other by Hermitian conjugation. By describing the left and right eigenstates as  $\langle n_L | \mathcal{H} = \epsilon_n \langle n_L |$ ,  $\mathcal{H} | n_R \rangle = \epsilon_n | n_R \rangle$ , the following equations are satisfied:

$$\langle n_L | m_R \rangle = \langle n_R | m_L \rangle = \delta_{nm}, \quad (47)$$

$$\mathbf{1} = \sum_n | n_R \rangle \langle n_L | = \sum_n | n_L \rangle \langle n_R |, \quad (48)$$

where  $\langle n_R | = (| n_R \rangle)^\dagger$  and  $| m_L \rangle = (| m_L \rangle)^\dagger$ . We note that  $\langle n_R | \neq \langle n_L |$  and  $\langle n_R | m_R \rangle \neq \delta_{nm}$ . In Eqs. (47) and (48), we can construct the orthonormal basis by the left and right eigenstate,

and we represent the Green's functions by the band index as

$$\mathcal{H}_{nm}(\mathbf{k}) \equiv \langle n_L | \mathcal{H}(\mathbf{k}) | m_R \rangle = \delta_{nm} \epsilon_m, \quad (49)$$

$$G_{nm}^R(\omega, \mathbf{k}) \equiv \langle n_L | G^R(\omega, \mathbf{k}) | m_R \rangle = \frac{\delta_{nm}}{(\omega - \epsilon_m)}, \quad (50)$$

$$G_{nm}^A(\omega, \mathbf{k}) \equiv \langle n_R | G^R(\omega, \mathbf{k}) | m_L \rangle = \frac{\delta_{nm}}{(\omega - \epsilon_m^*)}, \quad (51)$$

where  $\mathcal{H} = \mathcal{H}_0 + \Sigma^R$  includes the lifetime of the particles and is thus a non-Hermitian operator. In the following, we consider the effect of non-Hermiticity on the conductivity through the non-Hermitian band-index representation.

## 2. Non-Hermitian effect on the conductivity

First, we consider the linear conductivity using the non-Hermitian band-index representation, which reads

$$\begin{aligned} \mathcal{K}_{\alpha\beta}^{(1)}(\omega_1) = & \sum_{n,m} \left\{ \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \text{Im}[\mathcal{J}_{LR;\alpha\beta}^{nm} G_n^R(\omega)] f(\omega) \right. \\ & + \int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} [\mathcal{J}_{LR;\alpha}^{nm} G_m^R(\omega + \omega_1) \mathcal{J}_{LR;\beta}^{mn} G_n^R(\omega) \\ & - \mathcal{J}_{RR;\alpha}^{nm} G_m^R(\omega + \omega_1) \mathcal{J}_{LL;\beta}^{mn} G_n^A(\omega) \\ & + \mathcal{J}_{RR;\alpha}^{nm} G_m^R(\omega) \mathcal{J}_{LL;\beta}^{mn} G_n^A(\omega - \omega_1) \\ & \left. - \mathcal{J}_{RL;\alpha}^{nm} G_m^A(\omega) \mathcal{J}_{RL;\beta}^{mn} G_n^A(\omega - \omega_1)] f(\omega) \right\}, \quad (52) \end{aligned}$$

where  $\mathcal{J}_{AB;i}^{mn} = \langle m_A | \mathcal{J}_i | n_B \rangle$ . In the DC limit, this becomes

$$\begin{aligned} \sigma_{DC;\alpha\beta}^{(1)} & = 2 \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left( -\frac{\partial f(\omega)}{\partial \omega} \right) \text{Re}[\mathcal{J}_{RR;\alpha}^{nm} G_m^R(\omega) \mathcal{J}_{LL;\beta}^{mn} G_n^A(\omega)] \\ & - f(\omega) \text{Re} \left[ \mathcal{J}_{LR;\alpha}^{nm} \frac{\partial G_m^R(\omega)}{\partial \omega} \mathcal{J}_{LR;\beta}^{mn} G_n^R(\omega) \right]. \quad (53) \end{aligned}$$

In the non-Hermitian band-index representation, four different types of velocity operators appear, which are  $\mathcal{J}_{LL}$ ,  $\mathcal{J}_{LR}$ ,  $\mathcal{J}_{RL}$ ,  $\mathcal{J}_{RR}$ . We note that the conventional velocity operator in the Hermitian case corresponds to  $\mathcal{J}_{LR}$  and  $\mathcal{J}_{RL}$ .  $\mathcal{J}_{LL}$  and  $\mathcal{J}_{RR}$  are unique in the Fermi surface contribution to transport in a non-Hermitian system. To compare to the conventional results, we can write  $\mathcal{J}_{LL/RR}$  by  $\mathcal{J}_{LR}$  as

$$\mathcal{J}_{LL}^{nm} = \sum_l \mathcal{J}_{LR}^{nl} \langle l_L | m_L \rangle, \quad (54)$$

$$\mathcal{J}_{RR}^{nm} = \sum_l \mathcal{J}_{LR}^{lm} \langle n_R | l_R \rangle. \quad (55)$$

By using this relation, the Fermi surface term in Eq. (53) can be rewritten as

$$\begin{aligned} & \text{Re}[\mathcal{J}_{RR;\alpha}^{nm} G_m^R(\omega) \mathcal{J}_{LL;\beta}^{mn} G_n^A(\omega)] \\ & = \text{Re}[\langle n_L | n_L \rangle \langle n_R | n_R \rangle \mathcal{J}_{LR;\alpha}^{nm} G_m^R(\omega) \mathcal{J}_{LR;\beta}^{mn} G_n^A(\omega)] \\ & + \text{Re}[\langle l_L | n_L \rangle \langle n_R | l'_R \rangle \mathcal{J}_{LR;\alpha}^{lm} G_m^R(\omega) \mathcal{J}_{LR;\beta}^{ml'} G_n^A(\omega)]. \quad (56) \end{aligned}$$

We note that the term includes  $\partial f(\omega)/\partial \omega$  is said as ‘‘the Fermi surface term.’’ The first term is the conventional term multiplied by the factor  $\gamma_{NH;n} \equiv \langle n_L | n_L \rangle \langle n_R | n_R \rangle$ . We can easily

show that  $\gamma_{NH;n} \geq 1$  is always satisfied. (See Appendix J). Therefore, we reveal that, when the system is described by a non-Hermitian Hamiltonian, with different lifetimes in different orbitals, the conventional Fermi surface term can be enhanced by the factor  $\gamma_{NH;n}$ . The second term is unique in the non-Hermitian band-index representation, which describes the mixture of eigenstates in the decay dynamics. We call this term the ‘‘band-coalescent term’’ in this paper. For the second-order conductivity, we perform the same analysis and find

$$\begin{aligned} & \sum_{l,m,n} \gamma_{NH;n} \mathcal{J}_{LR;\alpha}^{nl} \frac{\partial G_l^R(\omega)}{\partial \omega} \mathcal{J}_{LR;\beta}^{lm} G_m^R(\omega) \mathcal{J}_{LR;\gamma}^{mn} G_n^A(\omega) \\ & + \sum_{l,m,n} \sum_{k,k'(\neq n)} \langle k'_L | n_L \rangle \langle n_R | k_R \rangle \\ & \times \mathcal{J}_{LR;\alpha}^{kl} \frac{\partial G_l^R(\omega)}{\partial \omega} \mathcal{J}_{LR;\beta}^{lm} G_m^R(\omega) \mathcal{J}_{LR;\gamma}^{mk'} G_n^A(\omega), \quad (57) \end{aligned}$$

where the first term is the conventional term with non-Hermitian factor and the second term describes the band-coalescent term for nonlinear conductivity. Finally, we numerically check these results and how the non-Hermiticity changes the conventional terms and the band-coalescent terms by explicitly calculating the linear and nonlinear conductivity for two different models, including orbital (sublattice)-dependent lifetimes. First, we show the results for the one-dimensional non-Hermitian Rice-Mele model, in which the dissipation depends on the sublattice. A detailed explanation of the model is given in Appendix H. Here, we note that  $\Gamma$  denotes the difference of the dissipation strength at each sublattice. The upper panels in Fig. 3 show the  $\Gamma$  dependence of the linear and nonlinear DC conductivity in the non-Hermitian Rice-Mele model. We see that the conventional conductivity with the non-Hermitian factor is dominant for the linear conductivity, while the band-coalescent term is dominant for the nonlinear conductivity. We note that the band-coalescent term can be determined by subtracting the conventional term from the total conductivity.

Next, we analyze the monolayer TMD model with uniaxial strain and spin-dependent scattering rates, where  $\Gamma_{\uparrow/\downarrow} = \pm \Gamma$ . The lower panels in Fig. 3 show that the conventional conductivity with the non-Hermitian factor is dominant for the linear conductivity, while the band-coalescent term prevails for the nonlinear conductivity. Notably, the sign of the nonlinear conductivity changes due to the non-Hermitian effect, and the absolute value is strongly enhanced. We note that the small spike in the conventional term of the nonlinear Hall conductivity originates from numerical errors due to exceptional points. The non-Hermitian band-index is very sensitive in parameter regions, including exceptional points, where the non-Hermitian Hamiltonian cannot be diagonalized.

Although we analyzed the effect of different lifetimes in orbitals (sublattices) in two specific models, it seems to be clear that the non-Hermitian effect on nonlinear responses is highly model-dependent. Our results, however, suggest that non-Hermiticity due to a difference of lifetimes in orbitals(sublattices) can strongly enhance nonlinear transport. This enhancement of nonlinear responses should also become



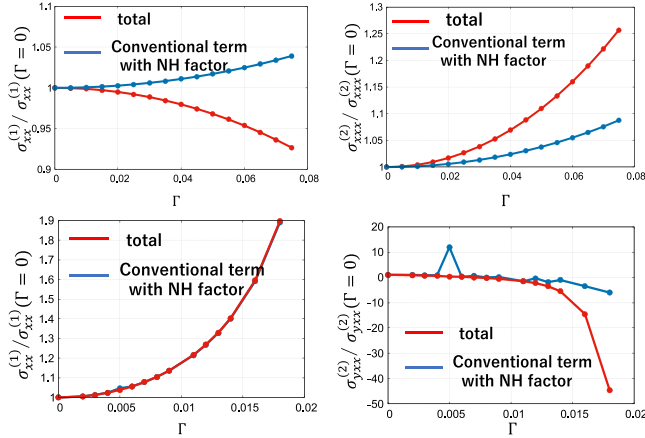


FIG. 3.  $\Gamma$ -dependence of the linear and nonlinear conductivities in the non-Hermitian Rice-Mele model and the monolayer TMD materials under uniaxial strain. The upper figures show the linear conductivity and the nonreciprocal conductivity in the 1D Rice-Mele model, and the lower figures depict the linear conductivity and the nonlinear Hall conductivity in the monolayer TMD model under uniaxial strain. The blue lines represent the original terms (also appearing in the Hermitian model) now modified by the non-Hermitian factor as in Eq. (57). The red lines describe the total conductivity, which is the sum of the conventional term with non-Hermitian factor and the band-coalescent term. The parameters are  $t = 1.0$ ,  $\delta t = 0.3$ ,  $\Delta = 0.3$ ,  $\eta = 0.05$ ,  $T = 0.02$  for the 1D Rice-Mele model and  $t = 0.5$ ,  $\mu = 0.7$ ,  $p = 0.7$ ,  $\alpha_1 = 0.08$ ,  $\alpha_2 = 0.06$ ,  $\delta = 0.7$ ,  $\eta = 0.05$ ,  $T = 0.02$  for the monolayer TMD model. The normalization coefficients are  $\sigma_{xx}^{(1)}(\Gamma = 0) = 0.0801$ ,  $\sigma_{xxx}^{(2)}(\Gamma = 0) = -0.0160$  in the Rice-Mele model and  $\sigma_{xx}^{(1)}(\Gamma = 0) = 36.21$ ,  $\sigma_{yxx}^{(2)}(\Gamma = 0) = 1.417$  in the monolayer TMD model.

important for correlated materials, where the self-energy depends on the orbital and atom.

## VI. SUMMARY AND DISCUSSION

In this paper, we constructed a formalism based on Green's functions to calculate the nonlinear response at finite temperature and generally analyze the impact of correlations on nonlinear response. By using a formalism based on Green's functions, correlations and electron scattering can be easily included via the self-energy. Previous studies on nonlinear response mainly focused on noninteracting systems using the semiclassical Boltzmann equation and the reduced density matrix formalism. In these methods, dissipation, which is necessary for the generation of a current, is introduced

phenomenologically by the RTA. We reveal that the RTA is justified for nonlinear optical response only in the DC limit and in the free limit  $\omega_i \gg \gamma$ ,  $\epsilon_{nm}$ , while the RTA seems to be a good approximation for the linear optical conductivity. We note that although Parker *et al.* [25] also derived a Green's function formalism for noninteracting systems, they considered mostly photon decay  $\omega_i \rightarrow \omega_i + i\gamma$  and neglect correlations and electron scattering.

After having established the Green's function formalism, we analyze the renormalization effect and the impact of different lifetimes in a multiorbital system as common correlation effects, which are not considered in previous studies. We demonstrate that the enhancement generated by the renormalization effect increases with the order of the nonlinear response. When considering a single-band model, the renormalization coefficient  $z (< 1)$  enhances the  $n$ th-order response by a factor of  $z^{-(n-1)}$ . Thus, the nonlinear response is more strongly increased than linear transport. Finally, we analyzed systems with different lifetimes, which commonly occur in materials where strongly correlated electrons couple to weakly interacting. The effect of different lifetimes can be analyzed by the band index of the non-Hermitian Hamiltonian. It causes the enhancement of terms that can also be derived in the Hermitian case and the emergence of a new term in which several bands coalesce. We analyzed these non-Hermitian effects on the conductivity in two specific models. In both models, the conventional term with the non-Hermitian factor is dominant for the linear conductivity, while the band-coalescent term is dominant for the nonlinear conductivity. The non-Hermitian effect can enhance the (non)linear conductivity and can even change the sign, although it depends on the model. Although the non-Hermitian band index is not well-defined at exceptional points, where the non-Hermitian factor  $\gamma_{NH}$  diverges, different lifetimes might give rise to novel transport. For example, in photonic crystals, the emergence of exceptional points induces nonreciprocal transport [46–49]. It should be possible to observe related phenomena in correlated materials. However, these questions are left for future works.

## ACKNOWLEDGMENTS

Y.M. deeply appreciates Hikaru Watanabe, Yoichi Yanase, Shun Okumura, and Yukitoshi Motome for fruitful discussions. This work is supported by the WISE program, MEXT. Y.M. is supported by a JSPS research fellowship and by JSPS KAKENHI (Grant No. 20J12265). R.P. is supported by JSPS, KAKENHI Grant No. JP18K03511. Computer simulations were done on the supercomputer of Tokyo University at the ISSP.

## APPENDIX A: DERIVATION OF THE MATSUBARA FORMALISM

In this section, we derive the conductivities using Green's function in Eqs. (13) and (14) starting from Eqs. (8) and (12). The first- and second-order response functions in the imaginary time are written as

$$\mathcal{K}_{\alpha\beta}^{(1)}(\tau, \tau_1) = \frac{1}{Z[A]} \frac{\delta}{\delta A^\beta(-i\tau_1)} \frac{\delta}{\delta A^\alpha(-i\tau)} Z[A]|_{A=0} \quad (\text{A1})$$

$$= -\langle \bar{\psi}_\mu(\tau) \mathcal{J}_\alpha^{\mu\nu} \psi_\nu(\tau) \bar{\psi}_\lambda(\tau_1) \mathcal{J}_\beta^{\lambda\eta} \psi_\eta(\tau_1) \rangle + \langle \bar{\psi}_\mu(\tau) \mathcal{J}_\alpha^{\mu\nu} \psi_\nu(\tau) \rangle \delta(\tau - \tau_1) \quad (\text{A2})$$

$$= -\delta(\tau - \tau_1) \text{Tr}[\mathcal{J}_{\alpha\beta} \mathcal{G}(0)] - \text{Tr}[\mathcal{J}_\alpha \mathcal{G}(\tau - \tau_1) \mathcal{J}_\beta \mathcal{G}(\tau_1 - \tau)], \quad (\text{A3})$$

$$\begin{aligned}
\mathcal{K}_{\alpha\beta\gamma}^{(2)}(\tau, \tau_1, \tau_2) &= \frac{1}{Z[\mathbf{A}]} \frac{\delta}{\delta A_\gamma(\tau_2)} \frac{\delta}{\delta A_\beta(\tau_1)} \frac{\delta}{\delta A_\alpha(\tau)} Z[\mathbf{A}]|_{A=0} \\
&= \delta(\tau - \tau_1)\delta(\tau - \tau_2)\text{Tr}[\mathcal{J}_{\alpha\beta\gamma}\mathcal{G}(0)] + \delta(\tau - \tau_1)\text{Tr}[\mathcal{J}_{\alpha\beta}\mathcal{G}(\tau - \tau_2)\mathcal{J}_\gamma\mathcal{G}(\tau_2 - \tau)] \\
&\quad + \delta(\tau - \tau_2)\text{Tr}[\mathcal{J}_{\alpha\gamma}\mathcal{G}(\tau - \tau_1)\mathcal{J}_\beta\mathcal{G}(\tau_1 - \tau)] \\
&\quad + \delta(\tau_1 - \tau_2)\text{Tr}[\mathcal{J}_\alpha\mathcal{G}(\tau - \tau_1)\mathcal{J}_{\beta\gamma}\mathcal{G}(\tau_1 - \tau)] + \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(\tau - \tau_2)\mathcal{J}_\gamma\mathcal{G}(\tau_2 - \tau_1)\mathcal{J}_\beta\mathcal{G}(\tau_1 - \tau)] \\
&\quad + \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(\tau - \tau_1)\mathcal{J}_\beta\mathcal{G}(\tau_1 - \tau_2)\mathcal{J}_\gamma\mathcal{G}(\tau_2 - \tau)],
\end{aligned} \tag{A4}$$

$$\begin{aligned}
&= \delta(\tau - \tau_1)\delta(\tau - \tau_2)\text{Tr}[\mathcal{J}_{\alpha\beta\gamma}\mathcal{G}(0)] + \delta(\tau - \tau_1)\text{Tr}[\mathcal{J}_{\alpha\beta}\mathcal{G}(\tau - \tau_2)\mathcal{J}_\gamma\mathcal{G}(\tau_2 - \tau)] \\
&\quad + \delta(\tau - \tau_2)\text{Tr}[\mathcal{J}_{\alpha\gamma}\mathcal{G}(\tau - \tau_1)\mathcal{J}_\beta\mathcal{G}(\tau_1 - \tau)] \\
&\quad + \delta(\tau_1 - \tau_2)\text{Tr}[\mathcal{J}_\alpha\mathcal{G}(\tau - \tau_1)\mathcal{J}_{\beta\gamma}\mathcal{G}(\tau_1 - \tau)] + \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(\tau - \tau_2)\mathcal{J}_\gamma\mathcal{G}(\tau_2 - \tau_1)\mathcal{J}_\beta\mathcal{G}(\tau_1 - \tau)] \\
&\quad + \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(\tau - \tau_1)\mathcal{J}_\beta\mathcal{G}(\tau_1 - \tau_2)\mathcal{J}_\gamma\mathcal{G}(\tau_2 - \tau)],
\end{aligned} \tag{A5}$$

where we used Wick's theorem to derive Eqs. (A3) and (A5) from Eqs. (A2) and (A4). When calculating conductivities for correlated systems, Eqs. (A3) and (A5) are exact except for vertex corrections. Correlations are included via the self-energy in the single-particle Green's function in imaginary time  $\mathcal{G}$ . We note that physical quantities obtained within the length gauge correspond to those from the velocity gauge when calculating exactly [29]. Therefore, taking the length gauge, we can derive the same results.

After Fourier transformation, we can derive the linear and second-order nonlinear response function in the Matsubara frequency as

$$\begin{aligned}
\mathcal{K}_{\alpha\beta}^{(1)}(i\omega_m; i\omega_m) &= -\frac{1}{\beta} \sum_{\omega_l} \text{Tr}[\mathcal{J}_{\alpha\beta}\mathcal{G}(i\omega_l) + \mathcal{J}_\alpha\mathcal{G}(i\omega_l + i\omega_m)\mathcal{J}_\beta\mathcal{G}(i\omega_l)], \\
\mathcal{K}_{\alpha\beta\gamma}^{(2)}(i\omega_s(= i\omega_n + i\omega_m); i\omega_n, i\omega_m) &= \frac{1}{\beta} \sum_{\omega_l} \left\{ \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha\beta\gamma}\mathcal{G}(i\omega_l)] + \left( \text{Tr}[\mathcal{J}_{\alpha\beta}\mathcal{G}(i\omega_m + i\omega_l)\mathcal{J}_\gamma\mathcal{G}(i\omega_l)] \right. \right. \\
&\quad + \frac{1}{2} \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(i\omega_n + i\omega_m + i\omega_l)\mathcal{J}_{\beta\gamma}\mathcal{G}(i\omega_l)] \\
&\quad \left. \left. + \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(i\omega_n + i\omega_m + i\omega_l)\mathcal{J}_\beta\mathcal{G}(i\omega_m + i\omega_l)\mathcal{J}_\gamma\mathcal{G}(i\omega_l)] \right) + ((\beta, i\omega_n \leftrightarrow) \gamma, i\omega_m) \right\},
\end{aligned} \tag{A6}$$

$$\begin{aligned}
&= \frac{1}{\beta} \sum_{\omega_l} \left\{ \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha\beta\gamma}\mathcal{G}(i\omega_l)] + \left( \text{Tr}[\mathcal{J}_{\alpha\beta}\mathcal{G}(i\omega_m + i\omega_l)\mathcal{J}_\gamma\mathcal{G}(i\omega_l)] \right. \right. \\
&\quad + \frac{1}{2} \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(i\omega_n + i\omega_m + i\omega_l)\mathcal{J}_{\beta\gamma}\mathcal{G}(i\omega_l)] \\
&\quad \left. \left. + \text{Tr}[\mathcal{J}_\alpha\mathcal{G}(i\omega_n + i\omega_m + i\omega_l)\mathcal{J}_\beta\mathcal{G}(i\omega_m + i\omega_l)\mathcal{J}_\gamma\mathcal{G}(i\omega_l)] \right) + ((\beta, i\omega_n \leftrightarrow) \gamma, i\omega_m) \right\},
\end{aligned} \tag{A7}$$

where  $\omega_l = (2l + 1)\pi/\beta$  are Fermionic Matsubara frequencies and  $\omega_m = 2m\pi/\beta$ ,  $\omega_n = 2n\pi/\beta$  are Bosonic Matsubara frequencies, which originate from the photons.

## APPENDIX B: ANALYTIC CONTINUATION OF THE NONLINEAR RESPONSE FUNCTION

We can calculate the (non)linear response in real frequency by using analytic continuation. By considering the paths in the complex frequency plane shown in Fig. 4, the (non)linear response functions can be written as

$$\mathcal{K}_{\alpha\beta}^{(1)}(i\omega_m; i\omega_m) = \left( \oint_{C_1+C_2} + \oint_{C_3+C_4} \right) \frac{d\omega}{2\pi i} f(\omega) \text{Tr}[\mathcal{J}_{\alpha\beta}\mathcal{G}(\omega) + \mathcal{J}_\alpha\mathcal{G}(\omega + i\omega_m)\mathcal{J}_\beta\mathcal{G}(\omega)] \tag{B1}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} f(\omega) \{ \text{Tr}[\mathcal{J}_{\alpha\beta}(G^R(\omega) - G^A(\omega))] + \text{Tr}[\mathcal{J}_\alpha G^R(\omega + i\omega_m)\mathcal{J}_\beta(G^R(\omega) - G^A(\omega))] \\
&\quad + \text{Tr}[\mathcal{J}_\alpha(G^R(\omega) - G^A(\omega))\mathcal{J}_\beta G^A(\omega - i\omega_m)] \}
\end{aligned} \tag{B2}$$

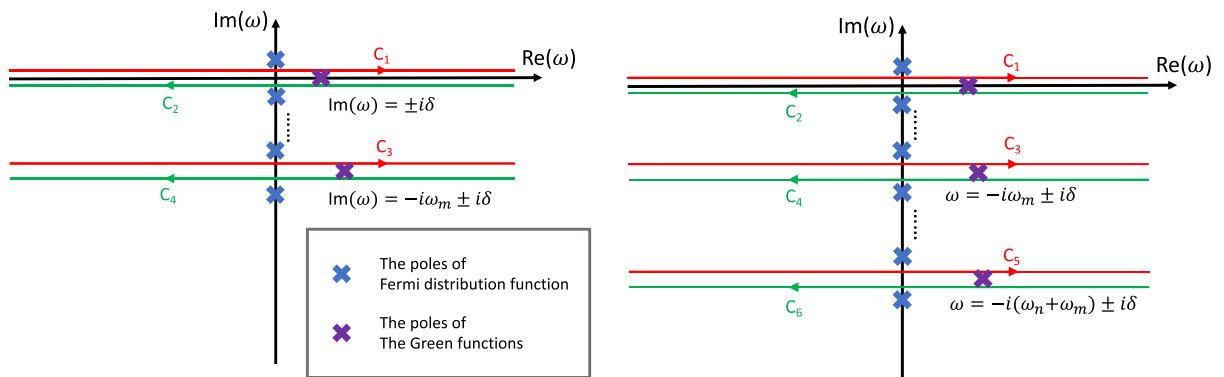


FIG. 4. Paths in the complex  $\omega$ -plane for the analytic continuation of the linear and second-order nonlinear response functions. By constructing the paths, which surround the poles of the Fermi distribution function and avoid the poles of the Green's functions, we can derive Eqs. (B3) and (B4) from Eqs. (A6) and (A7).

$$\Leftrightarrow K_{\alpha\beta}^{(1)}(\omega_1; \omega_1) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} f(\omega) \{ \text{Tr}[\mathcal{J}_{\alpha\beta}(G^R(\omega) - G^A(\omega))] + \text{Tr}[\mathcal{J}_{\alpha} G^R(\omega + \omega_1) \mathcal{J}_{\beta}(G^R(\omega) - G^A(\omega))] + \text{Tr}[\mathcal{J}_{\alpha}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\beta} G^A(\omega - \omega_1)] \}, \quad (\text{B3})$$

$$\begin{aligned} \mathcal{K}_{\alpha\beta\gamma}^{(2)}(i\omega_n + i\omega_m; i\omega_n, i\omega_m) = & - \left( \oint_{\text{up}+C_1} + \oint_{C_2+C_3} + \oint_{C_4+C_5} + \oint_{C_6+\text{low}} \right) \frac{d\omega}{2\pi i} f(\omega) \left\{ \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha\beta\gamma} \mathcal{G}(i\omega_n)] \right. \\ & + \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha} \mathcal{G}(\omega + i\omega_n + i\omega_m) \mathcal{J}_{\beta\gamma} \mathcal{G}(\omega)] + \text{Tr}[\mathcal{J}_{\alpha\beta} \mathcal{G}(\omega + i\omega_m) \mathcal{J}_{\gamma} \mathcal{G}(\omega)] \\ & \left. + \text{Tr}[\mathcal{J}_{\alpha} \mathcal{G}(\omega + i\omega_n) \mathcal{J}_{\beta} \mathcal{G}(\omega + i\omega_n + i\omega_m) \mathcal{J}_{\gamma} \mathcal{G}(\omega) + ((i\omega_n, \beta) \leftrightarrow (i\omega_m, \gamma))] \right\} \quad (\text{B4}) \end{aligned}$$

$$\begin{aligned} = & - \int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} f(\omega) \sum_k \left\{ \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha\beta\gamma}(G^R(\omega) - G^A(\omega))] \right. \\ & + \text{Tr}[\mathcal{J}_{\alpha\beta} G^R(\omega + i\omega_m) \mathcal{J}_{\gamma}(G^R(\omega) - G^A(\omega)) + \mathcal{J}_{\alpha\beta}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\gamma} G^A(\omega - i\omega_m)] \\ & + \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha} G^R(\omega + i\omega_{nm}) \mathcal{J}_{\beta\gamma}(G^R(\omega) - G^A(\omega)) + \mathcal{J}_{\alpha}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\beta\gamma} G^A(\omega - i\omega_{nm})] \\ & + \text{Tr}[\mathcal{J}_{\alpha} G^R(\omega + i\omega_{nm}) \mathcal{J}_{\beta} G^R(\omega + i\omega_m) \mathcal{J}_{\gamma}(G^R(\omega) - G^A(\omega)) \\ & + \mathcal{J}_{\alpha} G^R(\omega + i\omega_n) \mathcal{J}_{\beta}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\gamma} G^A(\omega - i\omega_m) \\ & \left. + \mathcal{J}_{\alpha}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\beta} G^A(\omega - i\omega_n) \mathcal{J}_{\gamma} G^A(\omega - i\omega_{nm})] + [(\beta, \omega_1) \leftrightarrow (\gamma, \omega_2)] \right\} \quad (\text{B5}) \end{aligned}$$

$$\begin{aligned} \Leftrightarrow K_{\alpha\beta\gamma}^{(2)}(\omega_1 + \omega_2; \omega_1, \omega_2) = & - \int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} f(\omega) \sum_k \left\{ \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha\beta\gamma}(G^R(\omega) - G^A(\omega))] \right. \\ & + \text{Tr}[\mathcal{J}_{\alpha\beta} G^R(\omega + \omega_2) \mathcal{J}_{\gamma}(G^R(\omega) - G^A(\omega)) + \mathcal{J}_{\alpha\beta}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\gamma} G^A(\omega - \omega_2)] \\ & + \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha} G^R(\omega + \omega_{12}) \mathcal{J}_{\beta\gamma}(G^R(\omega) - G^A(\omega)) + \mathcal{J}_{\alpha}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\beta\gamma} G^A(\omega - \omega_{12})] \\ & + \text{Tr}[\mathcal{J}_{\alpha} G^R(\omega + \omega_{12}) \mathcal{J}_{\beta} G^R(\omega + \omega_2) \mathcal{J}_{\gamma}(G^R(\omega) - G^A(\omega)) \\ & + \mathcal{J}_{\alpha} G^R(\omega + \omega_1) \mathcal{J}_{\beta}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\gamma} G^A(\omega - \omega_2) \\ & \left. + \mathcal{J}_{\alpha}(G^R(\omega) - G^A(\omega)) \mathcal{J}_{\beta} G^A(\omega - \omega_1) \mathcal{J}_{\gamma} G^A(\omega - \omega_{12})] + [(\beta, \omega_1) \leftrightarrow (\gamma, \omega_2)] \right\}, \quad (\text{B6}) \end{aligned}$$

where *up(low)* means the path in the complex plane surrounding the upper (lower) plane, and  $f(\omega)$  is the Fermi distribution. We use the relation  $\oint_C \frac{d\omega}{2\pi i} f(\omega) A(\omega) = -\frac{1}{\beta} \sum_n A(i\omega_n)$ , where  $\oint_C$  corresponds to the path integral only around the poles of the Fermi distribution function, while avoiding the poles of  $A(\omega)$ . Using the definitions of the response functions for real frequencies

$$\sigma_{\alpha\beta}^{(1)}(\omega_1; \omega_1) = K_{\alpha\beta}^{(1)}(\omega_1; \omega_1)/i\omega_1, \quad (\text{B7})$$

$$\sigma_{\alpha\beta\gamma}^{(2)}(\omega_1 + \omega_2; \omega_1, \omega_2) = -K_{\alpha\beta\gamma}^{(2)}(\omega_1 + \omega_2; \omega_1, \omega_2)/\omega_1\omega_2, \quad (\text{B8})$$

we can derive Eqs. (13) and (14) in the main text.

### APPENDIX C: DC LIMIT

In this section, we explicitly perform the DC-limit ( $\omega_i \rightarrow 0$ ) and derive Eqs. (15) and (16) starting from Eqs. (13) and (14). We thereby show that performing the DC limit under the velocity gauge does not yield any artificial divergence.

When  $\omega_i$  is small enough, in the sense that  $\beta\omega_i \ll 1$  and  $\tau\omega_i \ll 1$  [ $\tau$  is the inverse of the imaginary part of  $G^{R-1}(\omega)$ ], we can expand the single-particle Green's function as follows:

$$G^a(\omega + \omega_1) \simeq G^a(\omega) + \frac{\partial G^a}{\partial \omega} \omega_1, \quad (\text{C1})$$

$$G^a(\omega + \omega_1 + \omega_2) \simeq G^a(\omega) + \frac{\partial G^a}{\partial \omega}(\omega_1 + \omega_2) + \frac{\partial^2 G^a}{\partial \omega^2} \omega_1 \omega_2, \quad (\text{C2})$$

$$f(\omega + \omega_1) - f(\omega) \simeq \frac{\partial f(\omega)}{\partial \omega}(\omega_1), \quad (\text{C3})$$

where  $a = R, A$  (retarded and advanced Green's function). By using this expansion, Eq. (13) becomes

$$\sigma_{\alpha\beta}^{(1)}(\omega_1) = \frac{1}{\omega_1} \int \frac{d\omega}{2\pi} (A_0(\omega) + A_1(\omega)\omega_1) + \mathcal{O}(\omega_1^2), \quad (\text{C4})$$

$$A_0(\omega) = \int \frac{d\mathbf{k}}{(2\pi)^d} f(\omega) \text{Tr}[\mathcal{J}_{\alpha\beta}(G^R(\omega) - G^A(\omega)) + \mathcal{J}_\alpha G^R(\omega) \mathcal{J}_\beta G^R(\omega) - \mathcal{J}_\alpha G^A(\omega) \mathcal{J}_\beta G^A(\omega)], \quad (\text{C5})$$

$$A_1(\omega) = \int \frac{d\mathbf{k}}{(2\pi)^d} \left\{ \frac{\partial f(\omega)}{\partial \omega} \text{Tr}[\mathcal{J}_\alpha G^R(\omega) \mathcal{J}_\beta G^A(\omega)] + f(\omega) \left( \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_\beta G^R(\omega) + \mathcal{J}_\alpha G^A(\omega) \mathcal{J}_\beta \frac{\partial G^A(\omega)}{\partial \omega} \right] \right) \right\}. \quad (\text{C6})$$

We here used

$$- \int \frac{d\omega}{2\pi} f(\omega) (\mathcal{J}_\alpha G^R(\omega + \omega_1) \mathcal{J}_\beta G^A(\omega) - \mathcal{J}_\alpha G^R(\omega) \mathcal{J}_\beta G^A(\omega - \omega_1)) = \int \frac{d\omega}{2\pi} (f(\omega + \omega_1) - f(\omega)) \mathcal{J}_\alpha G^R(\omega + \omega_1) \mathcal{J}_\beta G^A(\omega) \quad (\text{C7})$$

to derive Eq. (C6). If  $A_0(\omega)$  would be finite after the integration, the conductivity diverges at  $\omega_1 \rightarrow 0$  even when  $1/\tau > 0$ . However, by using the identity,  $\partial_\alpha G^{R/A}(\omega) = G^{R/A}(\omega) \mathcal{J}_\alpha G^{R/A}(\omega)$ , Eq. (C5) can be rewritten as

$$A_0(\omega) = f(\omega) \int \frac{d\mathbf{k}}{(2\pi)^d} \partial_\beta \{ \mathcal{J}_\alpha (G^R(\omega) - G^A(\omega)) \} = 0. \quad (\text{C8})$$

Therefore,  $A_0(\omega)$  becomes zero at  $\omega_1 \rightarrow 0$ , an artificial divergence does not occur, and we can derive Eq. (15) using  $G^A = (G^R)^*$ .

We perform the same procedure for the second-order conductivity. By using the  $\omega_i$  expansion in Eqs. (C1) to (C3), Eq. (14) becomes

$$\sigma_{\alpha\beta\gamma}^{(2)}(\omega_1 + \omega_2; \omega_1, \omega_2) = \frac{1}{\omega_1 \omega_2} \int \frac{d\omega}{2\pi} (A_0(\omega) + (A_1(\omega)\omega_1 + A'_1(\omega)\omega_2) + A_2(\omega)\omega_1\omega_2) + \mathcal{O}(\omega_1^2, \omega_2^2, \omega_i^3), \quad (\text{C9})$$

$$\begin{aligned} A_0(\omega) &= f(\omega) \int \frac{d\mathbf{k}}{(2\pi)^d} \left\{ \frac{1}{2} \text{Tr}[\mathcal{J}_{\alpha\beta\gamma}(G^R(\omega) - G^A(\omega))] + \frac{1}{2} \text{Tr}[\mathcal{J}_\alpha G^R(\omega) \mathcal{J}_{\beta\gamma} G^R(\omega) - \mathcal{J}_\alpha G^A(\omega) \mathcal{J}_{\beta\gamma} G^A(\omega)] \right. \\ &\quad + (\text{Tr}[\mathcal{J}_{\alpha\beta} G^R(\omega) \mathcal{J}_\gamma G^R(\omega) - \mathcal{J}_{\alpha\beta} G^A(\omega) \mathcal{J}_\gamma G^A(\omega)] + \text{Tr}[\mathcal{J}_\alpha G^R(\omega) \mathcal{J}_\beta G^R(\omega) \mathcal{J}_\gamma G^R(\omega)] \\ &\quad \left. - \text{Tr}[\mathcal{J}_\alpha G^A(\omega) \mathcal{J}_\beta G^A(\omega) \mathcal{J}_\gamma G^A(\omega)]) + (\beta \leftrightarrow \gamma) \right\} \\ &= f(\omega) \int \frac{d\mathbf{k}}{(2\pi)^d} \partial_\gamma \partial_\beta \{ \mathcal{J}_\alpha (G^R(\omega) - G^A(\omega)) \} = 0, \end{aligned} \quad (\text{C10})$$

$$A_1(\omega) = f(\omega) \int \frac{d\mathbf{k}}{(2\pi)^d} \partial_\beta \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_\gamma G^R(\omega) \right] + \frac{\partial f(\omega)}{\partial \omega} \int \frac{d\mathbf{k}}{(2\pi)^d} \partial_\beta \text{Tr}[\mathcal{J}_\alpha G^R(\omega) \mathcal{J}_\gamma G^A(\omega)] + c.c. = 0, \quad (\text{C11})$$

$$A'_1(\omega) = A_1(\omega; \beta \leftrightarrow \gamma) = 0, \quad (\text{C12})$$

$$\begin{aligned} A_2(\omega) &= \int \frac{d\mathbf{k}}{(2\pi)^d} \left\{ \left( \frac{\partial f(\omega)}{\partial \omega} \right) \left( \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_\beta G^R(\omega) \mathcal{J}_\gamma G^A(\omega) \right] + \frac{1}{2} \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_{\beta\gamma} G^A(\omega) \right] \right) \right. \\ &\quad \left. - f(\omega) \text{Im} \left( \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial}{\partial \omega} \left( \frac{\partial G^R(\omega)}{\partial \omega} \mathcal{J}_\beta G^R(\omega) \right) \mathcal{J}_\gamma G^R(\omega) \right] + \frac{1}{2} \text{Tr} \left[ \mathcal{J}_\alpha \frac{\partial^2 G^R(\omega)}{\partial \omega^2} \mathcal{J}_{\beta\gamma} G^R(\omega) \right] \right) + (\beta \leftrightarrow \gamma) \right\}. \end{aligned} \quad (\text{C13})$$

In the same way as for the linear conductivity,  $A_0(\omega), A_1(\omega), A'_1(\omega)$  can be written in the form of an integration over a total derivative and thus become zero. Therefore, we can determine  $A_2(\omega)$  as the second-order DC-conductivity.

#### APPENDIX D: DIAGRAMMATIC FORMALISM FOR NONLINEAR RESPONSE AT FINITE TEMPERATURE

Parker *et al.* introduced a diagrammatic method for nonlinear responses in [25], and João *et al.* [26] introduced a diagrammatic method using Keldysh Green's functions. In this section, with the results from the previous section in mind, we construct an extension to this diagrammatic method for finite temperatures using real frequencies, which is summarized in Table I. Each diagram for the  $N$ th-order response function includes  $N$  incoming photons and one vertex for an outgoing photon. For each incoming photon a coefficient

$(i\omega_i)^{-1}$  is multiplied. The frequencies of the input vertices need to sum up to the output frequency. Furthermore, each diagram includes one object corresponding to the distribution function. Finally, retarded and advanced Green's functions are used to connect all vertices in a single loop. The difference of our results to the results by Parker *et al.* [25] is the presence of the distribution function and the distinction between the retarded and advanced Green's functions. For calculating the  $N$ th-order response, we construct all distinct diagrams using these rules. We then can easily evaluate the diagrams tracing the objects anticlockwise starting from the output vertex.

TABLE I. Objects to construct Feynman diagrams for nonlinear electromagnetic perturbations in a crystal at finite temperature. A new vertex with  $N$  incoming photons will appear in a diagram of the  $N$ th order response function. The input vertex can appear with any number of photons with a coefficient  $(-i\omega_i)^{-1}$  for each photon. The right(left)-handed arrows represent retarded(advanced) Green's functions. The direction of the arrow changes at the distribution function object and the output vertex, but not at the input vertices. We note that the input can occur at the same place as the output, such as in the first term of Eqs. (D1) and (D2). In that case, the value for the  $n$ th-order vertex becomes  $\prod_i^{n-1} \frac{1}{i\omega_i} \mathcal{J}_{\alpha_1 \dots \alpha_n}$ .

Component	Diagram	Value
(Classical) Photon Propagator		1
Electron Propagator (Retarded)		$G^R(\omega)$
Electron Propagator (Advanced)		$G^A(\omega)$
Distribution Function		$(G^R(\omega) - G^A(\omega))f(\omega)$
One-Photon Input Vertex		$\frac{1}{i\omega_1} \mathcal{J}_\alpha$
Two-Photon Input Vertex		$\frac{1}{i\omega_1} \frac{1}{i\omega_2} \mathcal{J}_{\alpha\beta}$
One-Photon Output Vertex		$\mathcal{J}_\alpha$

For example, the linear optical conductivity can be described using diagrams as

$$\begin{aligned} \sigma_{\alpha\beta}^{(1)}(\omega; \omega_1) &= \text{Diagram 1} + \text{Diagram 2} + \text{Diagram 3} \\ &= -\frac{1}{\omega_1} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} f(\omega) \sum_{\mathbf{k}} \{ \text{Tr}[\mathcal{J}_{\alpha\beta}(\mathbf{k})(G^R(\omega, \mathbf{k}) - G^A(\omega, \mathbf{k}))] \} \end{aligned}$$

$$\begin{aligned} &+ \text{Tr}[\mathcal{J}_\alpha(\mathbf{k})G^R(\omega + \omega_1, \mathbf{k})\mathcal{J}_\beta(\mathbf{k})(G^R(\omega, \mathbf{k}) - G^A(\omega, \mathbf{k})) \\ &+ \mathcal{J}_\alpha(\mathbf{k})(G^R(\omega, \mathbf{k}) - G^A(\omega, \mathbf{k}))\mathcal{J}_\beta(\mathbf{k})G^A(\omega - \omega_1, \mathbf{k})] \}. \end{aligned} \quad (\text{D1})$$

The diagrams for the second-order optical conductivity are given as

$$\begin{aligned} \sigma_{\alpha\beta\gamma}^{(2)}(\omega; \omega_1, \omega_2) &= \text{Diagram 1} \\ &+ \text{Diagram 2} + \text{Diagram 3} \\ &+ \text{Diagram 4} + \text{Diagram 5} \\ &+ \text{Diagram 6} + \text{Diagram 7} + \text{Diagram 8} \\ &+ ((\beta, \omega_1) \leftrightarrow (\gamma, \omega_2)). \end{aligned} \quad (\text{D2})$$

#### APPENDIX E: WEAK-SCATTERING LIMIT IN THE GREEN'S FUNCTION METHOD

When considering the weak-scattering limit where  $G^R(\omega) = 1/(\omega - \mathcal{H} - i\gamma/2)$  and  $\gamma \ll 1/\beta$ ,  $\omega_1, \epsilon_{nm}$ , we can perform the frequency integration by using

$$\begin{aligned} \int d\omega A(\omega, \{\omega_i\}\{\epsilon_m\})(G_n^R(\omega) - G_n^A(\omega))f(\omega) \\ \simeq -2\pi i A(\epsilon_n \pm i\gamma/2, \{\omega_i\}\{\epsilon_m\})f(\epsilon_n \pm i\gamma/2) \\ \simeq -2\pi i A(\epsilon_n \pm i\gamma/2, \{\omega_i\}\{\epsilon_m\})f(\epsilon_n), \end{aligned} \quad (\text{E1})$$

where  $A(\omega, \{\omega_i\}\{\epsilon_m\})$  is a product of Green's functions and velocities, and the sign takes  $\pm$  when  $A(\omega, \{\omega_i\}\{\epsilon_m\})$  is an analytical function in the upper/lower plane of the complex  $\omega$ -space. The plane is chosen such that  $A(\omega, \{\omega_i\}\{\epsilon_m\})$  is analytic. Other poles than  $\omega = \epsilon_n \pm i\gamma/2$  can be ignored because  $G_n^R(\epsilon) - G_n^A(\epsilon) \simeq 0$  at those due to the assumption  $\gamma \ll 1/\beta$ ,  $\omega_1, \epsilon_{nm}$ . Then we can derive the linear and nonlinear optical conductivities as

$$\sigma_{\alpha\beta}^{(1)}(\omega_1; \omega_1) \simeq \frac{i}{\omega_1} \sum_{\mathbf{k}} \left\{ \mathcal{J}_{\alpha\beta}^{nm} f(\epsilon_n) + \frac{\mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn}}{\omega_1 - \epsilon_{nm} + i\gamma} f_{nm} \right\} \quad (\text{E2})$$

$$\begin{aligned} \sigma_{\alpha\beta\gamma}^{(2)}(\omega_1 + \omega_2; \omega_1, \omega_2) \\ \simeq -\frac{1}{\omega_1 \omega_2} \sum_{\mathbf{k}} \left\{ \frac{1}{2} \left( \mathcal{J}_{\alpha\beta\gamma}^{nm} f_n + \frac{\mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn}}{\omega_{12} - \epsilon_{nm} + i\gamma} f_{nm} \right) \right. \\ \left. + \frac{\mathcal{J}_{\alpha\beta}^{nm} \mathcal{J}_\gamma^{mn}}{\omega_2 - \epsilon_{mn} + i\gamma} f_{nm} \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{\mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{ml} \mathcal{J}_\gamma^{ln} \{(\omega_1 - \epsilon_{ml} + i\gamma)f_{nl} + (\omega_2 - \epsilon_{ln} + i\gamma)f_{ml}\}}{(\omega_{12} - \epsilon_{mn} + i\gamma)(\omega_2 - \epsilon_{ln} + i\gamma)(\omega_1 - \epsilon_{ml} + i\gamma)} \\
& + ((\beta, \omega_1) \leftrightarrow (\gamma, \omega_2)), \tag{E3}
\end{aligned}$$

where  $\mathcal{J}^{nm} = \langle n | \mathcal{J} | m \rangle$  and  $G_n^R(\omega) = \langle n | G^R(\omega) | n \rangle = 1/(\omega - \epsilon_n + i/2\tau)$ ,  $\epsilon_{nm} = \epsilon_n - \epsilon_m$ , and  $f_{nm} = f(\epsilon_n) - f(\epsilon_m)$ . We also use the approximation  $(\omega_i - \epsilon_{nl})/(\omega_i - \epsilon_{nl} + i\gamma) \simeq 1$  to derive Eq. (E3). We note that these equations diverge in

the DC limit, where the assumption  $\gamma \ll \omega_i$  is not satisfied. These results correspond to the results by the RDM method with RTA under the velocity gauge. Under the assumption  $\gamma \ll \omega_i$ , we can regard  $(\omega_1 + i\gamma)/\omega_1 \simeq 1$  and derive the same results by the RDM methods under the length gauge from Eqs. (E2) and (E3).

Finally, we analyze the DC limit by first taking the limit  $\omega_i \rightarrow 0$  and assuming  $\omega \ll \gamma$ . Then we can derive the DC conductivity as

$$\sigma_{\alpha\beta;DC}^{(1)} = \sum_k \left\{ \tau \mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn} \left( -\frac{\partial f}{\partial \omega} \right) \Big|_{\epsilon_n} - \frac{\mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn}}{(\epsilon_{nm} + i\gamma)^2} f_{nm} \right\}, \tag{E4}$$

$$\begin{aligned}
\sigma_{\alpha\beta\gamma;DC}^{(2)} = & - \sum_k \left\{ \frac{\tau^2}{2} \left( \mathcal{J}_\alpha^{nn} \mathcal{J}_\beta^{nn} \mathcal{J}_\gamma^{nn} \frac{\partial^2 f_n}{\partial \epsilon_n^2} + \mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn} \left( -\frac{\partial f(\epsilon_n)}{\partial \epsilon_n} \right) \right) + \frac{i\tau \mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn}}{(\epsilon_{nm} + i\gamma)^2} \left( \mathcal{J}_\gamma^{nn} \frac{\partial f(\omega)}{\partial \omega} \Big|_{\epsilon_n} - \mathcal{J}_\gamma^{mm} \frac{\partial f(\omega)}{\partial \omega} \Big|_{\epsilon_m} \right) \right. \\
& + \mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{ml} \mathcal{J}_\gamma^{ln} \left( \frac{f_l}{(\epsilon_{lm} + i\gamma)^2 (\epsilon_{nl} + i\gamma)^2} + \frac{f_n}{(\epsilon_{nm} + i\gamma)^2 (\epsilon_{nl} + i\gamma)^2} + \frac{2f_n}{(\epsilon_{nm} + i\gamma)^3 (\epsilon_{nl} + i\gamma)} \right) \\
& \left. + \frac{1}{2} \mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn} \frac{f_n}{(\epsilon_{nm} + i\gamma)^2} + (\beta \leftrightarrow \gamma) \right\}. \tag{E5}
\end{aligned}$$

The first terms, which are proportional to  $\tau$  for  $\sigma^{(1)}$  and proportional to  $\tau^2$  for  $\sigma^{(2)}$ , represent the Drude term. The other terms for the second-order conductivity represent the Berry curvature dipole term and the Fermi sea terms.

## APPENDIX F: SEMICLASSICAL BOLTZMANN EQUATION

In the semiclassical Boltzmann treatment, transport phenomena are analyzed by calculating the distribution function for particles near equilibrium [36,50,51]. The effect of the vector potential is taken into account as

$$\mathcal{H}(\mathbf{p}) \rightarrow \mathcal{H}(\mathbf{k}(\mathbf{p}, t)) = \mathcal{H}(\mathbf{p} - q\mathbf{A}(\mathbf{x}, t)). \tag{F1}$$

$$|n(\mathbf{p})\rangle \rightarrow |\tilde{n}(\mathbf{k}(t))\rangle \simeq |n(\mathbf{p})\rangle - i \sum_{n' \neq n} \frac{|n'(\mathbf{p})\rangle \langle n'(\mathbf{p}) | \frac{\partial}{\partial t} n(\mathbf{k}(t)) \rangle}{\epsilon_n - \epsilon_{n'}}, \tag{F4}$$

$$v_{n\alpha}[\mathbf{k}(t)] = \langle \tilde{n} | \frac{\partial \mathcal{H}}{\partial k^\alpha} | \tilde{n} \rangle \simeq \frac{\partial \epsilon_n(\mathbf{p})}{\partial p^\alpha} - i \sum_{n' \neq n} \left( \frac{\langle n(\mathbf{p}) | \frac{\partial \mathcal{H}}{\partial q^\alpha} | n'(\mathbf{p}) \rangle \langle n'(\mathbf{p}) | q\mathbf{E} \cdot \nabla_{\mathbf{k}} | n(\mathbf{p}) \rangle}{\epsilon_n - \epsilon_{n'}} - \text{c.c.} \right) = v_{n\alpha}^0(\mathbf{p}) - q(\mathbf{E} \times \boldsymbol{\Omega}_n(\mathbf{p}))_\alpha, \tag{F5}$$

$$\boldsymbol{\Omega}_n(\mathbf{k}) = \nabla_{\mathbf{k}} \times \mathcal{A}_n, \quad \mathcal{A}_n = -i \langle u_{nk} | \nabla u_{nk} \rangle, \tag{F6}$$

where  $|n\rangle$  is the eigenstate of the Hamiltonian without the vector potential,  $\epsilon_n(\mathbf{p})$  is the eigenvalue, and  $\mathcal{H}(\mathbf{p}) |n(\mathbf{p})\rangle = \epsilon_n(\mathbf{p}) |n(\mathbf{p})\rangle$  holds. By taking into account the correction of the band velocity, we obtain the semiclassical equation of motion, which reads

$$\dot{\mathbf{k}}_n = q\mathbf{E}(t), \quad \dot{\mathbf{r}}_n = \frac{\partial \epsilon_n(\mathbf{p})}{\partial \mathbf{p}} - q\mathbf{E}(t) \times \boldsymbol{\Omega}_n(\mathbf{p}). \tag{F7}$$

Finally, the distribution function in the Boltzmann formalism with applied electric field using the relaxation time

By taking the Coulomb gauge  $\mathbf{A}(\mathbf{x}, t) = \mathbf{A}(t)$ , the translational symmetry is preserved, and the following equations are satisfied:

$$\dot{\mathbf{k}} = -q \frac{\partial \mathbf{A}(t)}{\partial t} = q\mathbf{E}, \tag{F2}$$

$$\frac{\partial}{\partial p^\alpha} = \frac{\partial}{\partial k^\alpha}, \quad \frac{\partial}{\partial t} = \dot{\mathbf{k}} \cdot \nabla_{\mathbf{k}} = q\mathbf{E} \cdot \nabla_{\mathbf{k}}, \tag{F3}$$

where  $p$  is the wave number of the particle without the electric field,  $E$  the electric field described by the vector potential  $A$ , and  $q$  is the wave number under the electric field. Considering the change of the eigenstates and the band velocity induced by the vector potential up to the first order of the vector potential, we find

approximation (RTA) is given by the following equation:

$$\frac{df_n(t)}{dt} = \frac{\partial f_n(t)}{\partial t} + \dot{\mathbf{k}} \cdot \nabla_{\mathbf{k}} f_n(t) = -\frac{f_n(t) - f_n^{(0)}}{\tau}, \tag{F8}$$

which can be solved as

$$\begin{aligned}
f(t) &= f^{(0)}(t) + f^{(1)}(t) + f^{(2)}(t) + \dots \\
&\Rightarrow \tau \frac{\partial f_n^{(m)}(t)}{\partial t} + f_n^{(m)}(t) = -q\tau \mathbf{E}(t) \cdot \nabla_{\mathbf{k}} f_n^{(m-1)}(t), \tag{F9}
\end{aligned}$$

where  $f_n^{(0)} = 1/(1 + \exp[\beta\epsilon_n(\mathbf{k})])$  is the Fermi distribution function,  $\beta$  is the inverse of the temperature, and  $f^{(m)}$  represent the  $m$ th-order nonequilibrium perturbative distribution function for the electric field.

The first- and second-order terms of the distribution function become

$$f_n^{(1)}(\omega, \alpha) = \frac{-q\tau}{1 - i\omega\tau} E^\alpha \partial_\alpha f_n^{(0)} \quad (\text{F10})$$

$$\begin{aligned} f_n^{(2)}((\omega_1, \beta), (\omega_2, \gamma)) &= \frac{-q\tau E^\beta \partial_\beta}{1 - i(\omega_1 + \omega_2)\tau} f_n^{(1)}(\omega_2, \gamma) + ((\omega_1, \beta) \leftrightarrow (\omega_2, \gamma)) \\ &= \frac{(q\tau)^2 E^\alpha E^\beta \partial_\alpha \partial_\beta f_n^{(0)}}{(1 - i(\omega_1 + \omega_2)\tau)(1 - i\omega_2\tau)} + ((\omega_1, \beta) \leftrightarrow (\omega_2, \gamma)). \end{aligned} \quad (\text{F11})$$

By combining the recurrence relation in Eq. (F9) with the velocity corrected by the electric field in Eq. (F5), we can derive the second-order nonlinear conductivity as

$$\begin{aligned} \sigma_{\alpha\beta\gamma}^{B(2)}(\omega_1 + \omega_2; \omega_1, \omega_2) &= q^3 \sum_{n,k} \left\{ \frac{\partial \epsilon_n}{\partial p^\alpha} \frac{\tau^2 \partial_\beta \partial_\gamma f_0}{(1 - i\omega_{12}\tau)(1 - i\omega_2\tau)} \right. \\ &\quad \left. + (\beta \leftrightarrow \gamma) \right\}. \end{aligned}$$

$$\begin{aligned} \sigma_{DC:\alpha\beta\gamma}^{B(2)} = - \sum_{n,m(\neq n),k} \int \frac{d\omega}{2\pi i} \left\{ \frac{1}{2} \mathcal{J}_\alpha^{nm} \left( \frac{\partial G_n^R(\omega)}{\partial \omega} \right) \mathcal{J}_{\beta\gamma}^{mn} G_n^A(\omega) \left( -\frac{\partial f(\omega)}{\partial \omega} \right) + \mathcal{J}_\alpha^{mn} G_n^R(\omega) \mathcal{J}_\beta^{nm} G_n^R(\omega) \mathcal{J}_\gamma^{mn} G_n^A(\omega) \frac{\partial^2 f(\omega)}{\partial \omega^2} \right. \\ \left. + \mathcal{J}_\alpha \frac{\partial G_n^R(\omega)}{\partial \omega} \mathcal{J}_\beta^{nm} G_m^R(\omega) \mathcal{J}_\gamma^{mn} G_n^A(\omega) \left( -\frac{\partial f(\omega)}{\partial \omega} \right) + \mathcal{J}_\alpha^{nm} \left( \frac{\partial G_m^R(\omega)}{\partial \omega} \right) \mathcal{J}_\beta^{mn} G_n^R(\omega) \mathcal{J}_\gamma^{mn} G_n^A(\omega) \left( -\frac{\partial f(\omega)}{\partial \omega} \right) + (\beta \leftrightarrow \gamma) \right\}, \end{aligned} \quad (\text{F14})$$

where  $\mathcal{J}^{nm} = \langle n | \mathcal{J} | m \rangle$  and  $G_n^R(\omega) = \langle n | G^R(\omega) | n \rangle = 1/(\omega - \epsilon_n + i/2\tau)$ . We use  $q\partial_\alpha \mathcal{J}_\beta^{nm} = \mathcal{J}_{\alpha\beta}^{nm} + (\mathcal{J}_\alpha^{nm} \mathcal{J}_\beta^{mn} + \mathcal{J}_\beta^{nm} \mathcal{J}_\alpha^{mn})/(\epsilon_{nm})$  to derive Eq. (F14). Here, we suppose that  $\beta\gamma$  is small and  $G_n^R(\omega)G_n^A(\omega) = 1/[(\omega - \epsilon_n)^2 + \gamma^2/4] \simeq 2\pi\delta(\omega - \epsilon_n)/\gamma$  is justified. Then, doing the frequency integration in Eq. (F14), we can obtain the original result Eq. (F13).

This Green's function representation of the Boltzmann equation Eq. (F14) can be directly derived from the original Green's function method shown in the main text, Eq. (14), by ignoring the Fermi sea terms and the interband transitions  $\mathcal{J}_{mn}(m \neq n)$  except for the second and third terms in Eq. (F14), which is justified when  $\epsilon_{nm}\tau \gg 1$ .

Next, we consider the AC case. We can recover a finite frequency  $\omega_i$  from the DC limit in Eq. (F14), which can be derived from Eq. (14) under the following assumptions:

(1) approximate  $\omega_i \simeq \omega_i + i\gamma$  which is justified in the limit

$\omega_i\tau \gg 1$ ;

(2) approximate  $f(\omega + \omega_i) - f(\omega) \simeq [\partial f(\omega)/\partial \omega]\omega_i$  and  $[\partial f(\omega)/\partial \omega] - [\partial f(\omega - \omega_i)/\partial \omega] \simeq [\partial^2 f(\omega)/\partial \omega^2]\omega_i$ , which is justified when  $\beta\omega_i \ll 1$ ;

$$+ \frac{\tau}{2(1 - i\omega_2\tau)} \epsilon_{\alpha\beta\mu} \Omega_{n\mu} \partial_\gamma f_0 + ((\omega_1, \beta) \leftrightarrow (\omega_2, \gamma)) \Big\}, \quad (\text{F12})$$

$$\begin{aligned} \sigma_{DC:\alpha\beta\gamma}^{B(2)} = q^3 \sum_{n,k} \left\{ \frac{\partial \epsilon_n}{\partial p^\alpha} \tau^2 \partial_\beta \partial_\gamma f_0 + \tau \epsilon_{\alpha\beta\mu} \Omega_{n\mu} \partial_\gamma f_0 \right. \\ \left. + (\beta \leftrightarrow \gamma) \right\}. \end{aligned} \quad (\text{F13})$$

Then, we compare our results with the semiclassical Boltzmann treatment. For the sake of comparison, we set the self-energy in the Green's function as  $G^R(\omega) = 1/(\omega - \mathcal{H} + i/2\tau) = 1/(\omega - \mathcal{H} + i\gamma/2)$ . In this case, the Green's function can be diagonalized with the eigenvalue of the free Hamiltonian, and therefore, the nonlinear conductivity calculated by the semiclassical Boltzmann treatment can be written using Green's functions. First, we focus on the Green's function representation of  $\sigma_{DC:\alpha\beta\gamma}^{B(2)}$  in the DC limit, which reads

(3) approximate  $G^R(\omega + \omega_{12}) - G^R(\omega + \omega_2) \simeq [\partial G^R(\omega + \omega_2)/\partial \omega]\omega_1$  and  $1/(\omega_2 - \epsilon_{nm}) \sim 1/(-\epsilon_{nm})$ , which is justified when  $\omega_i \ll \epsilon_{nm}$  ( $\epsilon_{nm} = \epsilon_n - \epsilon_m$ ).

Therefore, in the case of AC electric fields, there are severe approximations. Thus, the semiclassical Boltzmann equation is applicable at high temperatures or when the frequency  $\omega_i$  is very small so that the above conditions are satisfied. We note that we can also derive Eq. (F12) from Eq. (14) by supposing  $\gamma \rightarrow 0$ , which corresponds to the condition  $\omega_i \gg \gamma$  for the RDM method. We note that taking the DC limit in this situation leads to a diverging conductivity. Moreover, the relaxation time in most materials is usually about  $1 \sim 100$ [ps] [36]. Thus, when analyzing a Terahertz laser as input force,  $\omega_i\tau \sim 1$ , the conditions are not fulfilled. On the other hand, for a DC electric field in which  $\omega_i = 0$ , the only condition for the semiclassical Boltzmann treatment are  $\epsilon_{nm}\tau \gg 1$  and  $\beta\gamma \ll 1$ .

We note that, by considering higher-order corrections of the eigenstates by the electric field in Eq. (F5), we can derive a more precise semiclassical Boltzmann equation. In this way, it is possible to get rid of the approximation  $\epsilon_{nm}\tau \gg 1$  and to include the Fermi sea terms in the Boltzmann equation. The

other approximations listed above, however, remain necessary due to the relaxation time approximation.

### APPENDIX G: GAUGE INVARIANCE WITH THE DISSIPATION IN QUANTUM MASTER EQUATION FORMALISM

In this section, we analyze the correspondence between the length gauge and the velocity gauge in the quantum master equation in Eq. (29) in the main text. By using Eq. (24), we can describe Eq. (28) under the velocity gauge as

$$\begin{aligned} & \frac{d}{dt} \rho_{Ak}(t) + i[q\mathbf{E} \cdot \mathbf{r}, \rho_{Ak}(t)] \\ &= -i[\mathcal{H}_0(\mathbf{k} - q\mathbf{A}) - q\mathbf{E} \cdot \mathbf{r}, \rho_{Ak}(t)] - \lambda^2 \int_{t_0}^t ds \\ & \quad \times (\text{Re}[\{iG_B^l(t-s)\psi_k^\dagger \tilde{U}(t,s)\psi_k, \rho_{Ak}(s), \tilde{U}^\dagger(t,s)\}])^{(n)}, \end{aligned} \quad (\text{G1})$$

$$\tilde{U}(t,s) = T_{\rightarrow} U^{-1}(t) \exp\left[-i \int_s^t dt' (\mathcal{H}_0 - q\mathbf{E}(t') \cdot \mathbf{r})\right] U(s), \quad (\text{G2})$$

where  $\{\mathcal{O}_1, \rho, \mathcal{O}_2^\dagger\} = \mathcal{O}_1 \rho \mathcal{O}_2^\dagger + \mathcal{O}_2 \rho \mathcal{O}_1^\dagger$ . If we can show that  $T_{\rightarrow} \tilde{U}(t,s) = \exp[-i \int_s^t \mathcal{H}_0[\mathbf{k} - q\mathbf{A}(t)]]$ , the second term on the right side in Eq. (G1) can be written in the interaction representation in the velocity gauge Hamiltonian and the gauge invariance holds true in the open system. This can be verified by calculating the  $s$  and  $t$  derivatives of  $\tilde{U}(t,s)$  as

$$\begin{aligned} & \frac{\partial}{\partial t} \tilde{U}(t,s) \\ &= U^{-1}(t) \{-iq\mathbf{E}(t) \cdot \mathbf{r} - i(\mathcal{H}_0(\mathbf{k}) - q\mathbf{E}(t) \cdot \mathbf{r})\} \\ & \quad \times \exp\left[-i \int_s^t dt' (\mathcal{H}_0 - q\mathbf{E}(t') \cdot \mathbf{r})\right] U(s) \\ &= U^{-1}(t) (-i\mathcal{H}_0(\mathbf{k})) \exp\left[-i \int_s^t dt' (\mathcal{H}_0 - q\mathbf{E}(t') \cdot \mathbf{r})\right] U(s) \\ &= -i\mathcal{H}_0(\mathbf{k} - q\mathbf{A}(t)) \tilde{U}(t,s) \end{aligned} \quad (\text{G3})$$

$$\frac{\partial}{\partial s} \tilde{U}(t,s) = \tilde{U}(t,s) (i\mathcal{H}_0(\mathbf{k} - q\mathbf{A}(s))), \quad (\text{G4})$$

$$\tilde{U}(t,t) = 1. \quad (\text{G5})$$

We use the relation  $U^{-1}(t)\mathcal{H}_0(\mathbf{k})U(t) = \mathcal{H}(\mathbf{k} - q\mathbf{A}(t))$  to derive Eqs. (G3) and (G4). From the equality in Eqs. (G3)–(G5), we can identify  $\tilde{U}(t,s) = T_{\rightarrow} \exp[-i \int_s^t dt' \mathcal{H}_0[\mathbf{k} - q\mathbf{A}(t')]]$ . Therefore, the correspondence between the length gauge and the velocity gauge holds true in the quantum master equation, while it is broken when introducing the RTA at finite frequency.

### APPENDIX H: MODELS USED IN THE MAIN TEXT

In the main text, we use the following two models to numerically confirm our general results. In this section, we introduce the effective Hamiltonian  $\mathcal{H}_{\text{eff}} = \mathcal{H}_0 + \Sigma^R$ , which includes the dissipation effect.

### 1. One-dimensional Rice-Mele model with sublattice-dependent dissipation

We start from the Hermitian 1D Rice-Mele model, but assume that the dissipation depends on the sublattice. Such an effective non-Hermitian Hamiltonian can also be derived from the non-Hermitian matrix describing the single-particle Green's function. The effective non-Hermitian Hamiltonian reads [24]

$$\begin{aligned} \mathcal{H}_{\text{eff}}(\mathbf{k}) &= \sum_{a,b} \psi_a^\dagger (\tau^0 i\eta + \tau^x t \cos k + \tau^y \delta t \sin k + \tau^z (\Delta + i\Gamma))_{ab} \psi_b, \end{aligned} \quad (\text{H1})$$

where  $\psi_{A/B}^{(\dagger)}$  describes the annihilation (creation) operator in sublattice  $A/B$ ,  $\tau$  represents the Pauli matrices,  $\eta$  is the average of the dissipation strength at each sublattice,  $t$  is an intralattice hopping,  $\delta t$  is an interlattice hopping,  $\Delta$  is the difference of the chemical potential between the sublattices, and  $\Gamma$  is the difference of the dissipation strength at each sublattice.

### 2. Monolayer TMD materials with a spin-dependent dissipation

This model is commonly used to describe transition metal dichalcogenide (TMD) monolayers. The effective non-Hermitian Hamiltonian, which can again be understood as the non-Hermitian matrix describing the single-particle Green's function, can be written as [10,52]

$$\begin{aligned} \mathcal{H}_{\text{eff}} &= \sum_{k,s} (\epsilon(\mathbf{k}) - \mu - i\eta - i\Gamma_s) c_{k,s}^\dagger c_{k,s} \\ & \quad + \sum_{k,s,s'} \mathbf{g}(\mathbf{k}) \cdot \boldsymbol{\sigma}_{ss'} c_{k,s}^\dagger c_{k,s'}, \end{aligned} \quad (\text{H2})$$

$$\epsilon(\mathbf{k}) = 2t(p \cos(\mathbf{k} \cdot \mathbf{a}_1) + \cos(\mathbf{k} \cdot \mathbf{a}_2) + \cos(\mathbf{k} \cdot (\mathbf{a}_1 + \mathbf{a}_2))), \quad (\text{H3})$$

$$g^x(\mathbf{k}) = \frac{\alpha_1}{2} [\sin(\mathbf{k} \cdot (\mathbf{a}_1 + \mathbf{a}_2)) + \sin(\mathbf{k} \cdot \mathbf{a}_2)], \quad (\text{H4})$$

$$g^y(\mathbf{k}) = -\frac{\alpha_1}{\sqrt{3}} \left[ \sin(\mathbf{k} \cdot \mathbf{a}_1) + \frac{\sin(\mathbf{k} \cdot (\mathbf{a}_1 + \mathbf{a}_2)) - \sin(\mathbf{k} \cdot \mathbf{a}_2)}{2} \right], \quad (\text{H5})$$

$$g^z(\mathbf{k}) = \frac{2\alpha_2}{3\sqrt{3}} [\sin(\mathbf{k} \cdot \mathbf{a}_1) + \sin(\mathbf{k} \cdot \mathbf{a}_2) - \sin(\mathbf{k} \cdot \mathbf{a}_1 + \mathbf{a}_2)], \quad (\text{H6})$$

where  $c_{k,s}^{(\dagger)}$  is the annihilation (creation) operator for a conduction electron whose momentum is  $\mathbf{k}$  and spin is  $s$ .  $\mu$  is the chemical potential,  $\Gamma_{\uparrow/\downarrow} = \pm\Gamma$  is the spin-dependent dissipation,  $p$  is the effect of the strain [10] and  $\boldsymbol{\sigma}$  are the Pauli matrices, and  $\mathbf{g}$  represents the spin-orbit coupling. The lattice vectors are  $\mathbf{a}_1 = (1, 0)$  and  $\mathbf{a}_2 = (-1/2, \sqrt{3}/2)$ .

### APPENDIX I: DETAILS OF THE NUMERICAL CALCULATIONS

In this section, we write in detail how to numerically calculate the results shown in the figures of the main text. The codes



used for the numerical calculations in this paper are published in [53].

### 1. Green's function method

Here we describe the procedure of how to perform the numerical calculation using the Green's function method.

(1) A tight-binding Hamiltonian,  $\mathcal{H}(\mathbf{k})$ , describing the single-electron part of the model, such as Eq. (H1) or Eq. (H2), must be obtained.

(2) Starting from this tight-binding Hamiltonian, current operators  $\mathcal{J}_{\alpha\beta\dots}$  can be calculated by Eq. (4).

(3) For accounting for correlation effects, self-energies must be calculated. In this paper, we used the dynamical mean-field theory [54].

(4) Using  $\mathcal{H}(\mathbf{k})$  and the self-energies, the retarded and advanced Green's functions can be calculated.

(5) Having these Green's functions and current operators, one can use the Green's function formalism to calculate nonlinear response in strongly correlated systems.

To calculate the effect of renormalization of the band structure, we set  $\Sigma^R(\omega) = -(1/Z - 1)\omega - \Sigma_0^R$ , where  $\Sigma_0^R$  is the real-part of the self-energy at  $\omega = 0$ . Then, one can analyze the renormalization effects on the linear response and the nonlinear response. We note that, when calculating the optical conductivity for a small input frequency ( $\omega_i$ ), one should do the momentum integration before the frequency ( $\omega$ ) integration. Furthermore, one should use Eqs. (C6) and (C10)–(C12).

### 2. RDM methods using the RTA

When using the RDM for calculating the (non)linear conductivity, one first needs to diagonalize the free Hamiltonian  $\mathcal{H}(\mathbf{k})$ . Using the eigenvectors, one calculates the velocity operators for different bands and calculates the (non)linear conductivity by Eqs. (82) in [29].

#### APPENDIX J: PROOF THAT $\gamma_{NH} \geq 1$

The left and right eigenvectors  $\langle n_L |$ ,  $|n_R\rangle$  can be described as  $|n_R\rangle = (a_1, \dots, a_l)^T$  and  $\langle n_L | = (b_1, \dots, b_l)$ . Then, the following quantity must be larger than zero. Therefore, the non-Hermitian factor  $\gamma_{NH}$  must be larger than 1:

$$\begin{aligned} & \langle n_R | n_R \rangle \langle n_L | n_L \rangle - \langle n_L | n_R \rangle \langle n_R | n_L \rangle \\ &= \left( \sum_s |a_s|^2 \right) \left( \sum_s |b_s|^2 \right) - \left| \sum_s (a_s b_s) \right|^2 \\ &\geq \left( \sum_s |a_s|^2 \right) \left( \sum_s |b_s|^2 \right) - \left( \sum_s |a_s| |b_s| \right)^2 \\ &= \sum_{s,t} (|a_s| |b_t| - |a_t| |b_s|)^2 / 2 \geq 0 \end{aligned} \quad (J1)$$

$$\Leftrightarrow \gamma_{NH;n} = \langle n_R | n_R \rangle \langle n_L | n_L \rangle / \langle n_L | n_R \rangle \langle n_R | n_L \rangle \geq 1. \quad (J2)$$

- 
- [1] J. C. Petersen, M. D. Caswell, J. S. Dodge, I. A. Sergienko, J. He, R. Jin, and D. Mandrus, *Nat. Phys.* **2**, 605 (2006).
- [2] L. Zhao, D. H. Torchinsky, H. Chu, V. Ivanov, R. Lifshitz, R. Flint, T. Qi, G. Cao, and D. Hsieh, *Nat. Phys.* **12**, 32 (2016).
- [3] J. W. Harter, Z. Y. Zhao, J.-Q. Yan, D. G. Mandrus, and D. Hsieh, *Science* **356**, 295 (2017).
- [4] T. Morimoto and N. Nagaosa, *Sci. Adv.* **2**, e1501524 (2016).
- [5] Y. Tokura and N. Nagaosa, *Nat. Commun.* **9**, 3740 (2018).
- [6] X. Zhang, X. F. Ma, Y. Jin, T. Lu, E. P. Boden, P. D. Phelps, K. R. Stewart, and C. P. Yakymyshyn, *Appl. Phys. Lett.* **61**, 3080 (1992).
- [7] C.-K. Chan, N. H. Lindner, G. Refael, and P. A. Lee, *Phys. Rev. B* **95**, 041104(R) (2017).
- [8] F. de Juan, A. G. Grushin, T. Morimoto, and J. E. Moore, *Nat. Commun.* **8**, 15995 (2017).
- [9] H. Isobe, S.-Y. Xu, and L. Fu, *Sci. Adv.* **6**, eaay2497 (2020).
- [10] B. T. Zhou, C.-P. Zhang, and K. T. Law, *Phys. Rev. Appl.* **13**, 024053 (2020).
- [11] L. Wu, S. Patankar, T. Morimoto, N. L. Nair, E. Thewalt, A. Little, J. G. Analytis, J. E. Moore, and J. Orenstein, *Nat. Phys.* **13**, 350 (2017).
- [12] Q. Ma, S.-Y. Xu, H. Shen, D. MacNeill, V. Fatemi, T.-R. Chang, A. M. Mier Valdivia, S. Wu, Z. Du, C.-H. Hsu, S. Fang, Q. D. Gibson, K. Watanabe, T. Taniguchi, R. J. Cava, E. Kaxiras, H.-Z. Lu, H. Lin, L. Fu, N. Gedik, and P. Jarillo-Herrero, *Nature (London)* **565**, 337 (2019).
- [13] R. Wakatsuki, Y. Saito, S. Hoshino, Y. M. Itahashi, T. Ideue, M. Ezawa, Y. Iwasa, and N. Nagaosa, *Sci. Adv.* **3**, e1602390 (2017).
- [14] Y. M. Itahashi, T. Ideue, Y. Saito, S. Shimizu, T. Ouchi, T. Nojima, and Y. Iwasa, *Sci. Adv.* **6**, eaay9120 (2020).
- [15] F. Ando, Y. Miyasaka, T. Li, J. Ishizuka, T. Arakawa, Y. Shiota, T. Moriyama, Y. Yanase, and T. Ono, *Nature (London)* **584**, 373 (2020).
- [16] H. Kishida, H. Matsuzaki, H. Okamoto, T. Manabe, M. Yamashita, Y. Taguchi, and Y. Tokura, *Nature (London)* **405**, 929 (2000).
- [17] H. Liu, Y. Li, Y. S. You, S. Ghimire, T. F. Heinz, and D. A. Reis, *Nat. Phys.* **13**, 262 (2017).
- [18] H. Rostami, M. I. Katsnelson, and M. Polini, *Phys. Rev. B* **95**, 035416 (2017).
- [19] R. E. F. Silva, I. V. Blinov, A. N. Rubtsov, O. Smirnova, and M. Ivanov, *Nat. Photonics* **12**, 266 (2018).
- [20] S. Y. Kruchinin, *Phys. Rev. A* **100**, 043839 (2019).
- [21] Y. Murakami, M. Eckstein, and P. Werner, *Phys. Rev. Lett.* **121**, 057405 (2018).
- [22] N. Tancogne-Dejean, M. A. Sentef, and A. Rubio, *Phys. Rev. Lett.* **121**, 097402 (2018).
- [23] S. Dzsaber, X. Yan, M. Taupin, G. Eguchi, A. Prokofiev, T. Shiroka, P. Blaha, O. Rubel, S. E. Grefe, H.-H. Lai, Q. Si, and S. Paschen, *Proc. Natl. Acad. Sci. USA* **118**, e2013386118 (2021).
- [24] T. Morimoto and N. Nagaosa, *Sci. Rep.* **8**, 2973 (2018).
- [25] D. E. Parker, T. Morimoto, J. Orenstein, and J. E. Moore, *Phys. Rev. B* **99**, 045121 (2019).
- [26] S. M. João and J. M. V. P. Lopes, *J. Phys.: Condens. Matter* **32**, 125901 (2019).
- [27] J. E. Sipe and E. Ghahramani, *Phys. Rev. B* **48**, 11705 (1993).
- [28] J. E. Sipe and A. I. Shkrebtii, *Phys. Rev. B* **61**, 5337 (2000).

- [29] G. B. Ventura, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, *Phys. Rev. B* **96**, 035431 (2017).
- [30] D. J. Passos, G. B. Ventura, J. M. V. P. Lopes, J. M. B. Lopesdos Santos, and N. M. R. Peres, *Phys. Rev. B* **97**, 235446 (2018).
- [31] H. Watanabe and Y. Yanase, *Phys. Rev. Research* **2**, 043081 (2020).
- [32] We note that the results in [25] seem to include a typo in Eq. (B18) where  $1/(\omega - \epsilon_{ab})$  should be changed to  $1/(\omega - \epsilon_{ba})$  in the third term in Eq. (43).
- [33] H. Watanabe and Y. Yanase, *Phys. Rev. X* **11**, 011001 (2021).
- [34] Y. Zhang, J. van den Brink, C. Felser, and B. Yan, *2D Materials* **5**, 044001 (2018).
- [35] X.-Q. Yu, Z.-G. Zhu, J.-S. You, T. Low, and G. Su, *Phys. Rev. B* **99**, 201410(R) (2019).
- [36] Z. Z. Du, C. M. Wang, S. Li, H.-Z. Lu, and X. C. Xie, *Nat. Commun.* **10**, 3047 (2019).
- [37] Y. Michishita and R. Peters, *Phys. Rev. Lett.* **124**, 196401 (2020).
- [38] T. Holder, D. Kaplan, and B. Yan, *Phys. Rev. Research* **2**, 033100 (2020).
- [39] D. Kaplan, T. Holder, and B. Yan, *Phys. Rev. Lett.* **125**, 227401 (2020).
- [40] J. Wiersig, *Phys. Rev. Lett.* **112**, 203901 (2014).
- [41] H. Xu, D. Mason, L. Jiang, and J. Harris, *Nature (London)* **537**, 80 (2016).
- [42] W. Chen, Ş. K. Özdemir, G. Zhao, J. Wiersig, and L. Yang, *Nature (London)* **548**, 192 (2017).
- [43] H. Shen, B. Zhen, and L. Fu, *Phys. Rev. Lett.* **120**, 146402 (2018).
- [44] Y. Michishita, T. Yoshida, and R. Peters, *Phys. Rev. B* **101**, 085122 (2020).
- [45] R. Hanai and P. B. Littlewood, *Phys. Rev. Research* **2**, 033018 (2020).
- [46] A. Regensburger, C. Bersch, M.-A. Miri, G. Onishchukov, D. N. Christodoulides, and U. Peschel, *Nature (London)* **488**, 167 (2012).
- [47] Z. Lin, H. Ramezani, T. Eichelkraut, T. Kottos, H. Cao, and D. N. Christodoulides, *Phys. Rev. Lett.* **106**, 213901 (2011).
- [48] J. Doppler, A. A. Mailybaev, J. Böhm, U. Kuhl, A. Girschik, F. Libisch, T. J. Milburn, P. Rabl, N. Moiseyev, and S. Rotter, *Nature (London)* **537**, 76 (2016).
- [49] Y. Choi, C. Hahn, J. W. Yoon, S. H. Song, and P. Berini, *Nat. Commun.* **8**, 14154 (2017).
- [50] D. Xiao, M.-C. Chang, and Q. Niu, *Rev. Mod. Phys.* **82**, 1959 (2010).
- [51] I. Sodemann and L. Fu, *Phys. Rev. Lett.* **115**, 216806 (2015).
- [52] S. Kanasugi and Y. Yanase, *Phys. Rev. B* **102**, 094507 (2020).
- [53] [https://github.com/YoshihiroMichishita/Test\\_Codes/](https://github.com/YoshihiroMichishita/Test_Codes/).
- [54] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996).