


**Diagrammatic approach to excitonic effects on nonlinear optical response**Yu-Tzu Chang<sup>1</sup> and Yang-Hao Chan<sup>1,2,\*</sup><sup>1</sup>*Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei 10617, Taiwan*<sup>2</sup>*Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan* (Received 6 November 2023; revised 20 March 2024; accepted 4 April 2024; published 30 April 2024)

Optical responses of atomically thin 2D materials are greatly influenced by electron-hole interactions. It is by far established that exciton signatures can be well identified in the optical absorption spectrum of quasi-2D materials. However, the same level of understanding of excitonic effects on nonlinear optical responses and the ability to compute them accurately is still much desired. Based on the functional integral formalisms and working in the velocity gauge, we introduce a convenient Feynman diagram approach for calculating nonlinear responses including excitonic effects. By dressing electron-photon interactions with electron-hole ladder diagrams, we derive an expression for second-order optical responses and provide a comprehensive description of excitonic effects. We apply our approach to a monolayer h-BN model and show qualitative changes in the second harmonic generation spectrum when comparing with results assuming independent particles. Our approach can be readily extended to higher-order optical responses and is feasible for first-principles calculations.

DOI: [10.1103/PhysRevB.109.155437](https://doi.org/10.1103/PhysRevB.109.155437)**I. INTRODUCTION**

Modern experimental techniques integrated with theoretical simulations enable precise measurements and control of optical responses, facilitating the validation of quantum material models. To accurately describe optical responses in solid-state systems, significant progress has been made in the past few decades. Theoretically, light-matter interactions can be written in terms of the vector potential and electron momentum in a minimal-coupling scheme, which is often called the velocity gauge. Alternatively, the coupling can also be cast as the product of the electron position operator and the electric field, which is known as the length gauge. A well-defined light-matter interaction in periodic systems was established by Blount [1], which later led to the development of modern polarization theory in terms of Berry connection [2,3]. Based on Blount's treatment, Sipe *et al.*, derived nonlinear optical responses in terms of density matrix formalism through perturbation expansion of light-matter couplings [4]. A systematic comparison of these two gauges is made and it has been shown that the two are formally equivalent [5,6]. Applications of both treatments to study optical responses of real materials within the independent particle approximation (IPA) have been widely conducted.

Nonlinear optical responses describe a plethora of phenomena in which the light-induced polarization or current does not scale linearly with the electric field strength of the incident light. Typical examples of nonlinear responses include harmonic generation, optical rectification, shift current and frequency mixing effects, etc., [7] many of which have practical applications. Recent investigations on nonlinear optical spin Hall conductivity and nonlinear anomalous Hall effect

[8,9] further reveal the potential applications in spintronics devices. Aside from applications in optoelectronic devices, studies of nonlinear optical response have also advanced our fundamental understanding of light-matter interactions. Recent theoretical study on nonlinear optical responses culminates with the establishment of its connection to quantum geometry [10–12] and topology [13]. Notably, the role of electron-hole correlations hence excitonic effects is largely ignored owing to the typical small exciton binding energy in 3D bulk semiconductors.

The ability to prepare atomically thin 2D materials has brought opportunities to study excitonic effects on optical responses as it is well known by now that excitonic effects become particularly strong due to reduced screening and quantum confinement effects in quasi-2D materials. Combining efforts of experimental and theoretical first-principles investigations have shown that bound excitons with binding energies of hundreds of meV can be clearly identified in optical absorption or photoluminescence spectrum [14–16]. In recent years, excitons have emerged as a highly effective sensor for probing charge ordering and quantum phases in transition metal dichalcogenide (TMD) systems, including Wigner crystal states in TMD monolayers [17], excitonic insulators, charge transfer and Mott insulators in twisted bilayer moiré superlattices [18], and fractional quantum Hall states in twisted bilayer MoTe<sub>2</sub> [19], which makes the investigation of their response spectra an urgent and topical issue. However, our understanding of excitonic effects on optical responses is far from complete as the role of excitons in the nonlinear optical spectrum is under debate [20–22] and accurate *ab initio* computational tools are still under development [23–31].

Excitonic effects on optical responses have been studied by treating electron-hole interactions on the mean-field type of approximation with either Hartree-Fock, screened exchange self-energy or range-separated hybrid exchange-correlation

\*yanghao@gate.sinica.edu.tw

potentials [25–28]. In the wave function-based method, real-time propagation approaches have been conducted to study correlation effects on nonlinear optical responses. In a recent study, excitonic effects were considered in the framework of dynamical Berry phase polarization where the light-matter coupling is written in the length gauge. Expressions of general second-order optical responses including excitonic effects in both length and velocity gauge [30] were derived with the density matrix formalism. In particular, velocity and position operators are defined in the exciton basis to ensure the gauge invariance [32].

Although the perturbative derivation within the density matrix formalism is conceptually clean, the derivation involves tedious bookkeeping of various orders of perturbations even in the IPA. To provide a concise physical interpretation of high-order optical responses, a recent paper demonstrates a Feynman diagram approach to calculate nonlinear optical responses in the velocity gauge within IPA [33]. An extension of this method to spatially dispersive nonlinear responses was reported [34]. This motivates us to pursue a diagrammatic approach for the derivation of nonlinear optical responses with excitonic effects.

In the practical calculation, excitons are solutions of the Bethe-Salpeter equation (BSE) within a static approximation to screened Coulomb interactions. Diagrammatically, BSE can be derived from the equation of motion of electron-hole correlations, in which a series of interactions between electrons and holes is written as the so-called ladder diagrams. In Ref. [35], incorporations of ladder diagrams into Raman scattering intensity were considered to derive the excitonic effects on resonant Raman spectroscopy of layered materials. The derived expression can be naturally written in terms of exciton-phonon coupling matrix elements introduced in other contexts [36,37]. However, ladder diagrams' incorporation into nonlinear optical responses has not yet been reported.

In this paper, we develop a convenient Feynman diagram approach for nonlinear optical responses including excitonic effects. Our goal is to derive a general expression for second-order optical response with excitonic effects using the diagrammatic approach in the velocity gauge. The resulting expression offers practical advantages, revealing the physical interpretations of seemingly complicated summations over matrix elements and readily distinguishing one-, two-, and three-photon processes. Moreover, the expression can be straightforwardly implemented for tight-binding models and for first-principles calculations of real materials. The rest of the paper is structured as follows: In Sec. II, we revisit the functional derivative formalism and Feynman rules for the derivation of the linear and second-order optical conductivity in the IPA. We introduce the electron-hole correlation function, ladder diagrams, and BSE for electron-photon coupling vertex in Sec. III. By employing the vertex correction method, we derive the expression of the first-order and the second-order optical responses with excitonic effects in Sec. IV. In Sec. V, we apply our approach to an effective two-band tight-binding model for a monolayer h-BN, along with a comparison with other derivations. We conclude and discuss the outlook of our approach in Sec. VI.

## II. FUNCTIONAL INTEGRAL SETUP

### A. Independent-particle level

A detailed introduction of the functional integral approach in independent-particle level can be found in Refs. [33,38] and references therein. In this section, we revisit the key equations and introduce our notation to set the stage for the derivations in the next section.

To determine the optical conductivity using the functional integral formalism, we start by writing the partition function in the form of a path integral. The partition function in the imaginary time  $\tau$  can be written down as

$$Z = \int D[c(\tau), \bar{c}(\tau)] e^{-S[c(\tau), \bar{c}(\tau)]}, \quad (1)$$

where the action  $S$  is

$$S[c(\tau), \bar{c}(\tau)] = \int_0^\beta d\tau \{ \bar{c}(\tau) \partial_\tau c(\tau) + H[c(\tau), \bar{c}(\tau)] \}. \quad (2)$$

The first term is known as the Berry phase, and the normal-ordered Hamiltonian operator in the functional integral formalism is rewritten in terms of the Grassmann numbers [39,40]. The Hamiltonian of the system subjected to a time-dependent optical field is represented in the second quantization basis within the real-time domain as

$$\hat{H}(t) = \hat{H}_0 + \hat{H}_A(t), \quad (3)$$

which comprises of the independent particle term,

$$\hat{H}_0 = \sum_{ab} \int d\mathbf{k} \varepsilon_{ab\mathbf{k}} c_{\mathbf{k}a}^\dagger c_{\mathbf{k}b}. \quad (4)$$

Following Refs. [33,41], the light-matter couplings can be written as an expansion of the unperturbed  $H_0$  in powers of the vector potential field  $A^\alpha$ ,

$$\begin{aligned} \hat{H}_A(t) &= eA^{\alpha_1}(t)\hat{h}^{\alpha_1} + \frac{e^2}{2}A^{\alpha_1}(t)A^{\alpha_2}(t)\hat{h}^{\alpha_1\alpha_2} + \dots \\ &= \sum_{n=1}^{\infty} \frac{e^n}{n!} A^{\alpha_1}(t)A^{\alpha_2}(t)\dots A^{\alpha_n}(t)\hat{h}^{\alpha_1\alpha_2\dots\alpha_n}, \end{aligned} \quad (5)$$

where  $e$  is the electron charge and  $\alpha$  is the Cartesian index of the polarization direction of the external field. The  $n$ th-order derivative of  $\hat{h}$  is defined as

$$\hat{h}^{\alpha_1\alpha_2\dots\alpha_n} = \hbar^{-n} D^{\alpha_n} \dots D^{\alpha_2} D^{\alpha_1} \hat{H}_0, \quad (6)$$

where  $D^\alpha$  is the covariant derivative operator and its operation is defined through the commutator with another operator,  $[D^\alpha, \hat{O}]_{ab} = \frac{\partial O_{ab}}{\partial k^\alpha} - i[\xi^\alpha, \hat{O}]_{ab}$ . At the lowest order, the velocity operator in the eigenbasis of the unperturbed Hamiltonian reads

$$h_{ab}^\alpha = \frac{1}{\hbar} [D^\alpha, \hat{H}_0]_{ab} = \frac{1}{\hbar} \frac{\partial \varepsilon_{ab\mathbf{k}}}{\partial k^\alpha} \delta_{ab} - i \frac{\xi_{ab}^\alpha}{\hbar} (\varepsilon_a - \varepsilon_b), \quad (7)$$

where  $\xi_{ab}^\alpha$  is the matrix element of the interband Berry connection and we hide the crystal momentum  $\mathbf{k}$  to make the notation clean when there is no ambiguity. The lowest-order light-matter coupling term  $H_A(t)$  is the familiar form,  $e \int \hat{\mathbf{v}}_0 \cdot \mathbf{A}(t)$ .

To work with the electric field directly, we employ  $E^\alpha(\omega) = i\omega A^\alpha(\omega)$  to convert  $\hat{H}_A$  into  $\hat{H}_E$  using the Fourier

transformation

$$\hat{H}_E(t) = \sum_{n=1}^{\infty} \frac{e^n}{n!} \prod_{l=1}^n \int d\omega_l e^{-i\omega_l t} \frac{E^{\alpha_l}(\omega_l)}{i\omega_l} \hat{h}^{\alpha_1 \dots \alpha_n}. \quad (8)$$

The current density is calculated as the expectation value of the current density operator  $\hat{J}$ . In terms of the partition function, we have

$$\langle \hat{J}^\mu(t) \rangle = \frac{1}{Z} \int D[c, \bar{c}] e^{\hat{v}^\mu(t)} e^{-i \int dt' (H_0 + H_E(t'))}, \quad (9)$$

where  $\hat{v}^\mu(t) = D^\mu[\hat{H}_0 + \hat{H}_E(t)]$ , denotes the time-dependent velocity operator obtained by taking the derivative of the total Hamiltonian. Explicitly, it reads

$$\hat{v}^\mu(t) = \sum_{n=0}^{\infty} \frac{e^n}{n!} \prod_{l=1}^n \int d\omega_l e^{-i\omega_l t} \frac{E^{\alpha_l}(\omega_l)}{i\omega_l} \hat{h}^{\mu\alpha_1 \dots \alpha_n}. \quad (10)$$

Therefore, both  $\hat{v}(t)$  and  $\hat{H}_E(t)$  are functionals of the electric field  $E(t)$ . In the frequency domain, we define conductivity tensors, which are related to the current density as [41]

$$\begin{aligned} J^\mu(\omega) &= \int \frac{d\omega_1}{2\pi} \sigma^{\mu\alpha}(\omega_1) E^\alpha(\omega_1) (2\pi) \delta(\omega - \omega_1) \\ &+ \int \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} \sigma^{\mu\alpha\beta}(\omega_1, \omega_2) E^\alpha(\omega_1) E^\beta(\omega_2) \\ &\times (2\pi) \delta(\omega - \omega_1 - \omega_2) + \dots \end{aligned} \quad (11)$$

As a demonstration of the functional derivative approach and an introduction of notations and diagrams, we reproduce the derivation of the linear optical conductivity tensor given in Ref. [33] below. The derivation of the second-order conductivity tensor will be given in Appendix A.

At the linear order, we compute the conductivity tensor by taking the functional derivative of  $J(t)$  with respect to  $E(t)$  and performing a Fourier transformation

$$\begin{aligned} \sigma^{\mu\alpha}(\omega, \omega_1) \delta(\omega - \omega_1) &= \frac{\delta J^\mu(\omega)}{\delta E^\alpha(\omega_1)} \\ &= \int \frac{dt_1}{2\pi} e^{-i\omega_1 t_1} \frac{\delta}{\delta E^\alpha(t_1)} \int e^{i\omega t} dt J^\mu(t) \\ &= \int \frac{dt_1}{2\pi} e^{-i\omega_1 t_1} \int e^{i\omega t} dt \sigma^{\mu\alpha}(t, t_1), \end{aligned} \quad (12)$$

where in the second line we convert the functional derivative with respect to the external field in the frequency domain to the time domain.

To evaluate  $\sigma^{\mu\alpha}(t, t_1)$ , we note that the velocity operator in the perturbed system also depends on the external field, cf. Eqs. (9) and (10), so the functional derivatives can be taken on the observable  $v^\mu(t)$  or on the exponent in  $H_E(t)$ . They are

$$\frac{\delta \hat{v}^\mu(t)}{\delta E^\alpha(t_1)} - i \hat{v}^\mu(t) \frac{\delta}{\delta E^\alpha(t_1)} \int dt' H_E(t'). \quad (13)$$

We observe that velocity operators, hence the electron-photon coupling vertex can come from either the current density operator or from  $H_E$  in Eq. (13), which motivates the authors in Ref. [33] to define the outgoing vertex for the former and the incoming vertex for the latter, respectively. After performing

the functional derivative on Eqs. (10) and (8), we have

$$\begin{aligned} \sigma^{\mu\alpha}(t, t_1) &= i e^2 \int \frac{d\omega_1}{2\pi} \frac{e^{-i\omega_1(t-t_1)}}{\omega_1} \langle \hat{h}^{\mu\alpha}(t) \rangle \\ &- i e^2 \int dt' \int \frac{d\omega_1}{2\pi} \frac{e^{-i\omega_1(t'-t_1)}}{\omega_1} \langle \hat{h}^\mu(t) \hat{h}^\alpha(t') \rangle. \end{aligned} \quad (14)$$

We proceed by writing the expectation values explicitly in terms of the matrix element  $h_{ab}^\alpha$  and the two-particle correlation function. We focus on the second term, which dominates the finite frequency response for a semiconductor when including excitonic effects. We have within the IPA, reads

$$\begin{aligned} \langle \hat{h}^\mu(t) \hat{h}^\alpha(t') \rangle &= -h_{ab}^\mu h_{ba}^\alpha G_b(t, t') G_a(t', t) \\ &= -h_{ab}^\mu h_{ba}^\alpha \int \frac{d\omega''}{2\pi} e^{-i\omega''(t-t')} \int \frac{d\omega'}{2\pi} e^{-i\omega'(t'-t)} \\ &\times G_b(\omega'') G_a(\omega'), \end{aligned} \quad (15)$$

where the single-particle Green's function is defined as

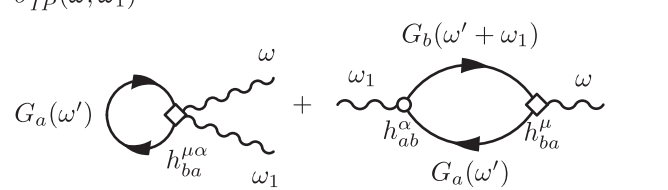
$$G_a(\mathbf{k}; t, t') \equiv -\delta_{ab} \langle c_{\mathbf{k}a}(t) c_{\mathbf{k}b}^\dagger(t') \rangle. \quad (16)$$

It is important to note that the expectation value is taken at the unperturbed state. We adopt the convention that the repeated indices that do not show up on the left-hand side of the equation are summed. Inserting the frequency integral of Green's function given in Appendix C, and using Eq. (12), we obtain the first-order conductivity from the second term in Eq. (13), which describes the interband transition,

$$\sigma_{IP}^{\mu\alpha}(\omega; \omega_1) = -\frac{iC_1}{\hbar\omega} \frac{f_{ab} h_{ba}^\alpha h_{ab}^\mu}{\hbar\omega_1 - \varepsilon_{ba} + i\eta}, \quad (17)$$

where  $f_{ab} = f_a - f_b$  is defined as the difference of electron occupations between  $a$  and  $b$  bands, and  $\varepsilon_{ba} = \varepsilon_b - \varepsilon_a$  is the difference of their energy. We define  $C_1 = ge^2 \hbar / V_{tot}$ , where the factor  $g$  accounts for the spin degeneracy factor, and  $V_{tot} = N_k V_u$  is the total volume with  $N_k$  being the total number of  $k$  points and  $V_u$  being the unit cell volume and  $\eta$  is a small positive number. This result agrees with the derivation from density matrix formulation [4,33].





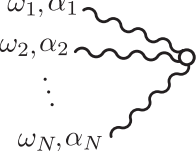
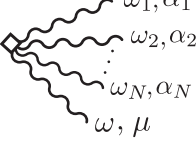

Within the IPA, the first-order conductivity is given by the two diagrams,

$$\sigma_{IP}^{\mu\alpha}(\omega; \omega_1) =$$


$$\quad (18)$$

where  $\omega_1, \omega$  indicate absorbed and emitted photons respectively. The first term corresponds to the Drude weight [33] and the second term describes the interband optical transitions. The bubble diagram in the second term represents the free electron-hole correlation function. Physically, the second diagram can be read as follows: An incoming photon generates a free electron-hole pair, which later recombines and emits a photon. Expressions of components in the diagrams are given

TABLE I. Diagram components for the conductivity within IPA.

Physical description	Diagram	Mathematical expression
Incoming vertex		$h$
Outgoing vertex		$h$
Incoming photon with polarization $\alpha_n$		$\omega_n, \alpha_n$
Outgoing photon with polarization $\mu$		$\omega, \mu$
Multi-photon absorption		$h^{\mu\alpha_1\alpha_2\cdots\alpha_N}$
Multi-photon absorption and single photon emission		$h^{\mu\alpha_1\alpha_2\cdots\alpha_N}$
Electron propagator		$G$

in Table I. We note that different symbols are used to distinguish incoming and outgoing photon vertices. In particular, the outgoing vertex associated with the current observable is represented by an empty diamond while the incoming vertex, which is associated with the perturbation expansion of  $H_E(t)$ , is represented by an empty circle. For the second-order response, the conductivity tensor  $\sigma^{\mu\alpha\beta}(\omega; \omega_1, \omega_2)$  can be computed from the second derivative of current expectation value,  $\frac{\delta J(t)}{\delta E(t_1)\delta E(t_2)}$ . A detailed derivation is given in Appendix A.

### B. Coulomb interaction

Going beyond the independent particle approximation, our main focus is to derive the response with excitonic effects. We start by considering a Hamiltonian with Coulomb interactions. The full Hamiltonian operator reads

$$\hat{H}(t) = \hat{H}_0 + \hat{H}_{\text{int}} + \hat{H}_A(t), \quad (19)$$

with the four-point interaction term,

$$\hat{H}_{\text{int}} = \sum_{abcd, \mathbf{k}_1, \mathbf{k}_2, \mathbf{q}} V_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}}^{abcd} c_{\mathbf{k}_1 + \mathbf{q}a}^\dagger c_{\mathbf{k}_2 - \mathbf{q}b}^\dagger c_{\mathbf{k}_2c} c_{\mathbf{k}_1d}, \quad (20)$$

where  $V_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}}^{abcd}$  represents matrix elements of the Coulomb potential and the coupling to the external field part,  $H_A(t)$  was given in Eq. (5).

Following the same procedure in the last section, we derive the first-order conductivity tensor by taking the functional derivative of the current expectation value with respect to the external field. We notice that as the interaction Hamiltonian does not depend on the external field, we obtain a seemingly identical expression of the first-order conductivity tensor in Eq. (15). However, the expectation value is now evaluated over the interacting ground state. The velocity-velocity

correlation function reads

$$\begin{aligned} \langle \hat{h}^\mu(t) \hat{h}^\alpha(t') \rangle_{\text{int}} &= \frac{1}{Z_{\text{int}}} \int D[c_{\mathbf{k}}^\dagger(t''), c_{\mathbf{k}}(t'')] \\ &\times [c_{\mathbf{k}a}^\dagger(t) h_{ab}^\mu c_{\mathbf{k}b}(t) c_{\mathbf{k}c}^\dagger(t') h_{cd}^\alpha c_{\mathbf{k}d}(t')] \\ &\times e^{-i \int dt'' [H_0(t'') + H_{\text{int}}(t'')]}, \end{aligned} \quad (21)$$

where  $\langle \dots \rangle_{\text{int}}$  indicates that the expectation value is taken over the interacting ground state and  $Z_{\text{int}}$  is the corresponding partition function. The first-order perturbative expansion over the interaction term,  $H_{\text{int}}$  of Eq. (21) is shown in Appendix B 1, which motivates a resummation over ladder diagrams to describe excitonic effects [42]. A more systematic derivation of ladder diagrams, which aligns with the *ab initio* GW-BSE method for treating excitonic effects can be done by using the functional derivative approach [43,44], which will be introduced in Sec. III A.

### III. DIAGRAMMATIC APPROACH TO EXCITONIC EFFECTS

To incorporate excitonic effects into our derivation, we start by solving for the interacting electron-hole correlation function, the equation of motion of which is described by the BSE [44–46]. By taking the static approximation on the screened Coulomb interactions one can construct the correlation function from the eigensolutions of the BSE. For linear optical responses, excitonic effects can be included by replacing the bubble diagram with the interacting correlation function, which involves an infinite sum of ladder diagrams. Such replacement is equivalent to dressing the electron-photon vertex in the IP conductivity diagrams with the ladder diagrams [47]. In light of this observation, we apply the procedure to derive nonlinear optical responses. The same procedure is recently used in the derivation of Raman cross section [35].

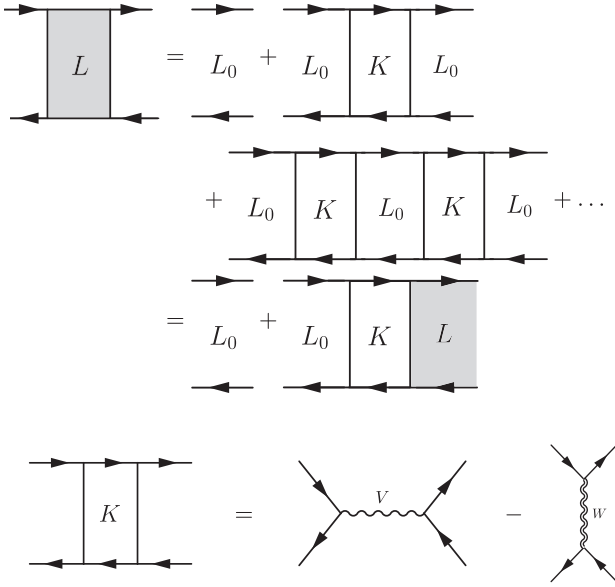


FIG. 1. (Top) Diagrammatic representation of the Bethe-Salpeter equation for the electron-hole correlation function. The symbols  $L$  and  $L_0$  represent the electron-hole correlation function for interacting and independent particles, respectively. (Bottom)  $K$  denotes the interaction kernel, which includes an attractive screened electron-hole interaction,  $W$  (double-wiggly line), and a repulsive Coulomb exchange term,  $V$  (single-wiggly line).

### A. Electron-hole interaction

Excitons are bound electron-hole pairs, where electrons and holes interact with each other via Coulomb interactions. The equation of motion of a general two-particle correlation function,  $L$  defined as  $L \equiv \frac{\delta G}{\delta \bar{h}}$ , can be derived from the Dyson's equation of a dressed single-particle Green's function  $G$ . Symbolically, we have  $G = G_0 + G_0 h G + G_0 \Sigma G$ , where  $h$  is the electron-photon coupling vertex and  $\Sigma$  is the electron self-energy due to Coulomb interactions. We delegate the details of the derivation in Appendix B 2. The resulting equation of motion in the frequency domain reads

$$L_{\mathbf{k}ab, \mathbf{k}'cd}(\omega) = L_{0; \mathbf{k}, ab}(\omega) \delta_{ac} \delta_{bd} + L_{0; \mathbf{k}, ab}(\omega) \times \sum_{\mathbf{k}'', l, l'} K_{\mathbf{k}ab, \mathbf{k}''l'} L_{\mathbf{k}''l', \mathbf{k}'cd}(\omega), \quad (22)$$

where we denote the noninteracting electron-hole correlation function as  $L_{0; \mathbf{k}, ab}$  with band indices in alphabet letters and the crystal momentum  $\mathbf{k}$ . The BSE kernel  $K$  describes the interaction between electron-hole pairs. In the approximation consistent with the GW quasiparticle self-energy, the kernel  $K$  includes an attractive screened Coulomb potential  $W(\omega)$  and a repulsive bare exchange term  $V$ . Equation (22) is shown in terms of diagrams in Fig. 1.

For a general BSE, the screened Coulomb interaction in the kernel  $K$  is frequency dependent and the band indices run over all bands. In practice, a static approximation to the screened Coulomb interaction  $W(\omega) = W(0)$  is often assumed so that the BSE can be transformed into an eigenvalue problem [45,46]. Moreover, we focus on semiconductors and use the Tamm-Dancoff approximation. With these

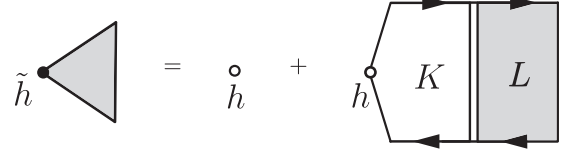


FIG. 2. The BSE for the dressed electron-photon vertex.

assumptions, we can write the eigenvalue equation for excitons with zero center-of-mass momentum (see Ref. [43–46] and Appendix C 1),

$$H_{c\mathbf{k}, c'v'\mathbf{k}'}^{BSE} Y_{c'v'\mathbf{k}'}^s = E_s Y_{c\mathbf{k}}^s, \quad (23)$$

where the index  $s$  labels exciton states,  $E_s$  and  $Y_{c\mathbf{k}}^s$  are the exciton energy and the exciton envelope function, respectively, and the effective Hamiltonian  $H_{c\mathbf{k}, c'v'\mathbf{k}'}^{BSE}$  is

$$H_{c\mathbf{k}, c'v'\mathbf{k}'}^{BSE} = (\varepsilon_{c\mathbf{k}} - \varepsilon_{v'\mathbf{k}}) \delta_{cc'} \delta_{vv'} \delta_{\mathbf{k}\mathbf{k}'} + K_{c\mathbf{k}, c'v'\mathbf{k}'}, \quad (24)$$

where  $\varepsilon_{v'\mathbf{k}}$  and  $\varepsilon_{c\mathbf{k}}$  are the energy of the conduction and the valence band electrons, respectively. The interacting electron-hole correlation function  $L_{ij\mathbf{k}, nm\mathbf{k}'}$  can be written in terms of exciton solutions as

$$L_{ij\mathbf{k}, nm\mathbf{k}'}(\omega) = i \sum_{\lambda} \left[ \bar{f}_i \bar{f}_j \bar{f}_n f_m \frac{Y_{ij\mathbf{k}}^{\lambda} Y_{nm\mathbf{k}'}^{\lambda*}}{\omega - E_{\lambda} + i\eta} - f_i \bar{f}_j \bar{f}_m f_n \frac{Y_{jik'}^{\lambda*} Y_{mn\mathbf{k}}^{\lambda}}{\omega + E_{\lambda} + i\eta} \right], \quad (25)$$

where  $\bar{f}_n = 1 - f_n$ . The first term in the parenthesis is the resonant part while the second term is the antiresonant contribution [46].

### B. Vertex correction

In Eq. (17), we derive the interband IP conductivity by connecting the edges of electron and hole propagators of a free electron-hole correlation function with the velocity operators as shown in the second diagram in Eq. (18). To derive the optical conductivity tensor with excitonic effects, we are motivated to replace free electron-hole propagators with their interacting counterparts following the discussion in Appendix B. The derivation of the first-order optical conductivity can be done straightforwardly as shown in the next section. For higher-order responses, diagrams of three- or multi-particle correlation functions appear. In particular, the three-particle triangle diagram, which represents the three-particle correlation function appears in the second-order optical response.

Since a direct calculation of the interacting triangle diagram is challenging, inspired by Ref. [35] we approximate it with a series of ladder diagrams. The essential idea is to add all possible noncrossing interaction lines for each pair of electron and hole legs. We illustrate the procedure for the electron-photon vertex in Fig. 2. The shaded triangle in Fig. 2 stands for a dressed vertex, which can be expanded with respect to the order of interactions. At the zeroth order, we have a bare noninteracting vertex  $\hat{h}$ . In the first order, a single-interaction kernel together with a free electron-hole correlation is included in the diagram. At higher order, we

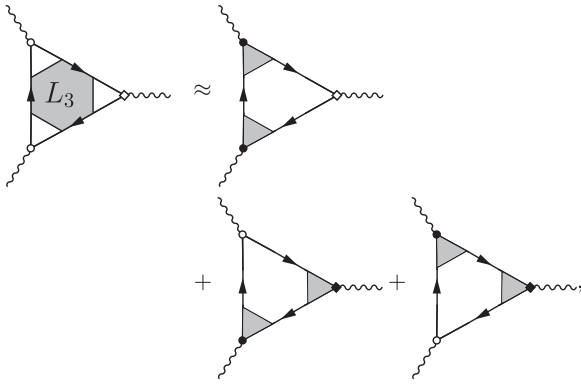


FIG. 3. The interacting three-particle correlation function diagram is approximated as a sum of three bare ones with two dressed vertices in each diagram.

insert more interaction kernels and  $L$ . The infinite-order sum leads to the right-hand side of Fig. 2, where the infinite sum of ladder diagrams is replaced with the interacting electron-hole correlation. The BSE for the dressed incoming electron-photon vertex  $\tilde{h}$  satisfies [35]

$$\tilde{h}(\omega) = h + KL(\omega)h, \quad (26)$$

where  $h^\mu$  is the bare vertex. Formally, in the matrix notation, the dressed incoming vertex can be solved by the inverse of the free electron-hole correlation,

$$\tilde{h}_{ab\mathbf{k}}^\alpha(\omega) = L_{0,ab\mathbf{k}}^{-1}(\omega)L(\omega)_{ab\mathbf{k}cd\mathbf{k}'}h_{cd\mathbf{k}'}^\alpha. \quad (27)$$

The dressed outgoing vertex is solved by

$$\tilde{h}_{ab\mathbf{k}}^{\mu*}(\omega) = h_{cd\mathbf{k}'}^{\mu*}L(\omega)_{cd\mathbf{k}'ab\mathbf{k}}L_{0,ab\mathbf{k}}^{-1}(\omega). \quad (28)$$

The dressed vertex provides a systematical way to include excitonic effects in diagrams. For second-order optical

conductivity, within the ladder approximation, the interacting triangle diagram can be decomposed into a sum of three diagrams, each of which has two dressed vertices as shown in Fig. 3. We note that the diagram with all three vertices dressed simultaneously is not included since it would involve an electron-electron or hole-hole correlation function, which is beyond the Tamm-Dancoff approximation.

We conclude that excitonic effects can be included in the  $n$ th-order conductivity within the Tamm-Dancoff approximation by simply dressing all vertices in a way that

(i) Electron and hole propagators of any two e-h correlation functions associated with the dressed vertices can not directly connect to each other.

(ii) By associating one of the two legs of an e-h correlation function with an electron and the other with a hole for every propagator in a diagram, all legs can not have mixed characters of electron and hole.

The first rule is to avoid the double counting while the second follows from the Tamm-Dancoff approximation. New components with dressed vertices up to the second-order response are listed in Table II.

### C. Assumptions and caveats

It is crucial to acknowledge the underlying assumptions made to simplify the calculation process. We collect all assumptions throughout the entire derivation and discuss the possible errors from them. We first discuss the approximations similar to those in Ref. [33] and the caveats in the following.

Concerning the choice of the gauge, it has been shown that velocity gauge could lead to incorrect results for two cases: (a) an effective Hamiltonian, which is limited to a subset of Brillouin zone and (b) few band model systems [5,6,48]. In the example we show below, we deal with an effective Hamiltonian in the full Brillouin zone so point (a) is irrelevant. For

TABLE II. Diagram components for the conductivity tensor including excitonic effects within the ladder approximation.

Physical description	Diagram	Mathematical expression
Dressed incoming/outgoing vertex		$\tilde{h}$
Interacting two-particle correlation with two bare vertices		$hLh$
Bare two-particle correlation with the left vertex dressed		$\tilde{h}L_0h$
Bare two-particle correlation with the right vertex dressed		$hL_0\tilde{h}$
One term of the approximation to the interacting three-particle correlation		$\tilde{h}G\tilde{h}GhG$

few band systems, indeed a convergence check with respect to the number of bands needs to be checked carefully. Our derivations in the velocity gauge should be valid in general and can be applied for different systems if cares are taken in the numerical calculations.

One should also be aware of the limit of the perturbative expansion over the equilibrium Green's function. Both our derivation and the derivation in Ref. [33] are based on a perturbative expansion with respect to light-matter couplings so the applicability is typically limited to the weak field case. For strong field cases, dynamical effects beyond the perturbation theory should be treated with the nonequilibrium Green's function approach on the Keldysh contour. For instance, light-excited carriers from strong pulses can change the screening environment and lead to both band and exciton energy renormalizations. Such effects are not considered from the perturbation of the equilibrium Green's function. Nevertheless, the derivation in Ref. [33] and in the present paper can be considered as a generalization of the Kubo formula to higher-order responses.

Many-body effects on the single-particle Green's function can be approximated with a quasiparticle form

$$G_a(\omega) = \frac{1}{\omega - \epsilon_a - \Sigma_a(\omega = \epsilon_a) + i\eta}. \quad (29)$$

Within the quasiparticle approximation, the main effect of the self-energy correction is a shift of energy level, a state-dependent broadening and a correction to the velocity operator from the nonlocal self-energy. The former two corrections can be straightforwardly incorporated in the derived expression by replacing the bare electron energy with the quasiparticle energy. It is known that corrections from the nonlocal potential to the velocity operators is necessary for accurate *ab initio* calculations [49]. However, it is less clear how the velocity operator is affected by the nonlocal self-energy. Since these corrections only change the velocity operator quantitatively, we ignored them in our derivation for simplicity. The possible errors will be discussed further when comparing with the derivation in Ref. [32] later. Therefore, as in Ref. [33], we use the bare Green's function and the bare velocity operator from Eq. (7) in all our calculations.

Additional approximations concerning the two-particle correlations introduced in the present work are the following. We include only ladder diagrams to approximate the electron-hole correlation function. As we show in Appendix B 2, this approximation is equivalent to the choice of GW electron self-energy in the single-particle level and leads to the standard BSE for excitons. In the linear response regime, first-principle GW-BSE method is widely adopted for *ab initio* calculations of optical absorption spectrum and is known to work well for semiconductors. The screened exchange interaction in the BSE kernel is assumed to be static. As the plasmon frequency is much higher than the exciton binding energy, the dynamical screening effects can be safely ignored. The functional dependence of the screened exchange with respect to  $G$  is also ignored as they lead to higher-order correction to electron-hole interaction vertex [45]. Moreover, the resultant BSE are solved with the Tamm-Dancoff approximations. Numerical

tests have been conducted and shown that the approximation is generally valid for cold semiconductors [45].

#### IV. OPTICAL RESPONSES WITH EXCITONIC EFFECTS

In this section, we derived expressions for optical conductivity up to the second-order following the prescription introduced in Sec. III A.

##### A. First-order conductivity

For semiconductors, we ignore the first diagram in Eq. (18), which corresponds to the Drude weight. The second term, which describes the interband transition can be expressed in terms of bare vertices and the interacting two-particle correlation function. Replacing the free e-h bubble with Eq. (26), the expression for the resonant part reads

$$\begin{aligned} \sigma_{eh}^{\mu\alpha}(\omega; \omega_1) &= \frac{iC_1}{\hbar\omega} \sum_s h_{ab}^{\mu,*} \frac{Y_{abk}^s Y_{ijk'}^{s*}}{\hbar\omega_1 - E_s + i\eta} h_{ij}^\alpha \\ &= \frac{iC_1}{\hbar\omega} \sum_s \frac{d_s^{\mu*} d_s^\alpha}{\hbar\omega_1 - E_s + i\eta}. \end{aligned} \quad (30)$$

where we defined the excitonic velocity  $d_s^\alpha$ ,

$$d_s^\alpha = Y_{abk}^{s*} h_{abk}^\alpha, \quad (31)$$

where  $\alpha$  is the polarization direction and  $s$  is the exciton state index. The real part of  $\sigma^{\mu\alpha}(\omega; \omega)$  is related to the optical absorption spectrum, which is often represented by the imaginary part of the dielectric function  $\epsilon_2(\omega)$ . We have  $\epsilon_2(\omega) = \text{Re}[\sigma(\omega)]/(\epsilon_0\omega)$  so

$$\epsilon_2^{\mu\alpha}(\omega) = \frac{\pi e^2 g}{\epsilon_0 \omega^2 V_{tot}} \sum_s d_s^{\mu*} d_s^\alpha \delta(\hbar\omega - E_s), \quad (32)$$

where  $\epsilon_0$  is the vacuum permittivity. Equation (32) is the familiar expression for optical absorption spectrum in the literature [50].

Equivalently, we can consider the excitonic effect by dressing one of the electron-photon vertex and evaluate the following diagram:

$$\sigma_{eh}^{\mu\alpha}(\omega; \omega_1) = \text{Diagram} \quad (33)$$

where we replace the interacting bubble diagram with the bare one but introduce the dressed vertex  $\tilde{h}$ . The mathematical expression for the above diagram is

$$\begin{aligned} \sigma_{eh}^{\mu\alpha}(\omega; \omega_1) &= \frac{iC_1}{\hbar\omega} \tilde{h}_{ab}^\alpha L_{0,abk}(\omega_1) h_{ab}^{\mu*} \\ &= \frac{iC_1}{\hbar\omega} L_{0,abk}^{-1}(\omega_1) L_{abk,ijk'}(\omega_1) h_{ij}^\alpha h_{ab}^{\mu*} L_{0,abk}(\omega_1), \end{aligned} \quad (34)$$

where in the second line we insert the dressed vertex from Eq. (26). The equivalence of these two derivations can be

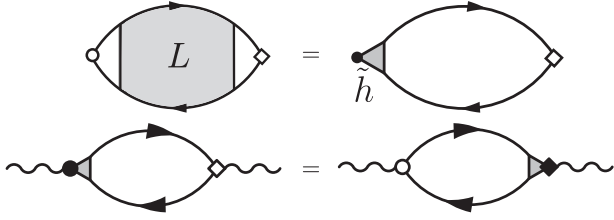


FIG. 4. (Top) The transformation of an interacting electron-hole bubble diagram into a bare bubble with a dressed vertex. (Bottom) Equivalence of dressing vertex on either side of the bubble.

expressed diagrammatically as shown in the top panel in Fig. 4, which demonstrates that we can “push” the shaded area into one of the vertex. We can also show that it is flexible to dress either the incoming photon vertex or the outgoing one.

## B. Second-order conductivity

The topologically inequivalent diagrams for the second-order optical conductivity including excitonic effects are shown in Eq. (35).

$$\begin{aligned}
 \sigma_{eh}^{\mu\alpha\beta}(\omega; \omega_1, \omega_2) = & \text{Diagram 1} + \text{Diagram 2} \\
 & + \text{Diagram 3} + \text{Diagram 4} \\
 & + \text{Diagram 5} + [(\alpha, \omega_1) \leftrightarrow (\beta, \omega_2)].
 \end{aligned} \tag{35}$$

Compared with diagrams for IP conductivity, the diagram with a three-photon line on the outgoing vertex is ignored since there are no corresponding corrections within the Tamm-Dancoff approximation and we are dealing with cold semiconductors. The first two terms are bubble diagrams with an either incoming or outgoing two-photon line. Excitonic effects are included by dressing one vertex, which is similar to the treatment for the linear-order diagram and is equivalent to replacing the free electron-hole correlation function with the interacting one as shown in Sec. IV A. At each vertex, the energy conservation of the photon and the electron-hole pair can be checked from the frequency argument of the single-particle Green’s functions. The indices  $a$  and  $b$  run over pairs between valence and conduction bands.

The interacting triangle diagram acquires different corrections from excitonic effects. Following the discussion in Sec. III A, it is approximated by the sum of three diagrams, each of which has two dressed vertices. Again, energy conservation can be readily checked at each vertex. We observe that the bare vertex in each diagram couples the electron or the hole states of the two correlated electron-hole pairs associated with the two dressed vertices. Moreover, the bare vertex only couples conduction or valence electrons in the same manifold.

After some tedious but straightforward algebra, we obtain an expression for the general second-order optical response including excitonic effects. The details will be given in the Appendix D. We write down the final expression, which roughly corresponds to the five diagrams above up to the symmetrization over photon frequency and polarizations,

$$\begin{aligned}
 \sigma_{eh}^{\mu\alpha\beta}(\omega; \omega_1, \omega_2) = & \frac{-C_2}{\hbar^2 \omega_1 \omega_2} \sum_{\lambda} \left[ \frac{d_{\lambda}^{\alpha} d_{\lambda}^{\mu\beta*}}{\hbar\omega_1 - E_{\lambda} + i\eta} - \frac{d_{\lambda}^{\alpha*} d_{\lambda}^{\mu\beta}}{\hbar\omega_1 + E_{\lambda} + i\eta} \right] + \frac{-C_2}{2\hbar^2 \omega_1 \omega_2} \sum_{\lambda} \left[ \frac{d_{\lambda}^{\alpha\beta} d_{\lambda}^{\mu*}}{\hbar\omega - E_{\lambda} + i\eta} - \frac{d_{\lambda}^{\alpha\beta*} d_{\lambda}^{\mu}}{\hbar\omega + E_{\lambda} + i\eta} \right] \\
 & + \frac{C_2}{\hbar^2 \omega_1 \omega_2} \sum_{s\lambda} \left[ \frac{d_s^{\alpha} \Pi_{s\lambda}^{\mu*} d_{\lambda}^{\beta*}}{(\hbar\omega_2 + E_{\lambda} + i\eta)(\hbar\omega_1 - E_s + i\eta)} - \frac{d_{\lambda}^{\mu*} \Pi_{\lambda s}^{\alpha} d_s^{\beta}}{(\hbar\omega - E_{\lambda} + i\eta)(\hbar\omega_2 - E_s + i\eta)} \right] \\
 & - \frac{C_2}{\hbar^2 \omega_1 \omega_2} \sum_{s\lambda} \frac{d_{\lambda}^{\mu} \Pi_{s\lambda}^{\alpha} d_s^{\beta*}}{(\hbar\omega + E_{\lambda} + i\eta)(\hbar\omega_2 + E_s + i\eta)} + [(\alpha, \omega_1) \leftrightarrow (\beta, \omega_2)],
 \end{aligned} \tag{36}$$



where we define the second-order excitonic velocity matrix element,  $d_{\lambda}^{\alpha\beta} = h_{ab}^{\alpha\beta} Y_{abk}^{\lambda*}$ , the interexciton coupling matrix elements,

$$\Pi_{\lambda s}^{\beta} = h_{cb}^{\beta} Y_{cak}^{\lambda*} Y_{bak}^s - h_{ba}^{\beta} Y_{cak}^{\lambda*} Y_{cbk}^s, \quad (37)$$

and  $C_2 = ge^3 \hbar^2 / V_{tot}$ . The first term in  $\Pi_{\lambda s}$  describes the coupling between electrons in two excitons while the second term is for the coupling between holes. The interexciton coupling term originates from the bare vertex that appears in the triangle diagram.

Equation (36) is the main result of this paper. The physical meaning of each term can be read out directly with the help of diagrams. Terms in the first parenthesis in Eq. (36) correspond to the first diagram where an incoming photon created an electron-hole pair with energy  $E_{\lambda}$ , which later absorbed and at the same time emitted a photon through the second-order velocity operator. The second term in the first parenthesis describes an anti-resonance process. The first term in the second parenthesis describes the excitation of an electron-hole pair by absorbing two photons, which emit a photon later. The  $1/2$  factor comes from the fact that this diagram originates from the second-order expansion of the coupling Hamiltonian in the functional integral and an exchange of  $\alpha \leftrightarrow \beta$  and  $\omega_1 \leftrightarrow \omega_2$  gives the same result.

The last three terms come from symmetrization and rearrangement of the three triangle diagrams. We can interpret the third diagram in Eq. (35) by starting from the bottom-left corner. An electron-hole pair was generated by absorbing a photon. The hole state was scattered to another hole state and emitted part of its energy at the bottom-right corner. The electron and the hole absorbed light but deexcited at the vertex on the top-left corner. In the fourth diagram, an electron-hole pair is generated by absorbing a photon on the top-left corner. The hole is scattered by another photon in the bottom-left corner. The pair recombines and emits a photon with frequency  $\omega$  at the right vertex. In the fifth diagram, which has a correspondence to the symmetrized partner term in the last line in Eq. (36), an antiresonance process generates an electron-hole pair then the hole state is scattered at the bare vertex. Finally, the e-h pair recombines and emits a photon.

It is interesting to compare Eq. (36) with other reported derivations. A comparison between Eq. (36) and Eq. (B1c) in Ref. [32] will be given in the next section. We will show that there is a one-to-one correspondence for each term between that expression and Eq. (36). However, the two derivations differ in how the velocity operators are defined. In the next section, we will present the numerical results of the second harmonic generation (SHG) for a simple two-band model.

## V. APPLICATION ON A MONOLAYER HEXAGONAL BORON NITRIDE

In this section, we apply our method to compute the linear absorption and the SHG spectrum for a model of monolayer hexagonal boron nitride (h-BN). Since monolayer h-BN is a large band gap semiconductor it is known to have strong excitonic effects as we shall also demonstrate later.

### A. Two-band tight-binding model

We employ a two-band tight-binding model for monolayer h-BN [51]. In the local basis of B and N atoms, the tight-binding Hamiltonian in the momentum space reads

$$H_{\mathbf{k}}^{\text{hBN}} = \begin{pmatrix} \Delta & t_0 f_{\mathbf{k}} \\ t_0 f_{\mathbf{k}}^* & -\Delta \end{pmatrix}, \quad (38)$$

where we define the structure factor  $f_{\mathbf{k}} = 1 + e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2}$  with the primitive lattice vectors  $\mathbf{a}_1 = a_0(\frac{\sqrt{3}}{2}\hat{x} - \frac{1}{2}\hat{y})$ ,  $\mathbf{a}_2 = a_0(\frac{\sqrt{3}}{2}\hat{x} + \frac{1}{2}\hat{y})$ , and the lattice constant  $a_0 = 2.46 \text{ \AA}$ . The asymmetric on-site energies are denoted as  $\Delta$  and  $-\Delta$  for B and N atoms, respectively. We choose  $\Delta = 3.9 \text{ eV}$  and the nearest-neighbor-hopping strength  $t_0 = 2.7 \text{ eV}$  in our calculations below. The eigenenergies and eigenstates read

$$\begin{aligned} \varepsilon_{c/v\mathbf{k}} &= \pm \sqrt{\Delta^2 + (t_0 |f_{\mathbf{k}}|)^2}, \quad |c\mathbf{k}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{1 + \frac{\Delta}{\varepsilon_{c\mathbf{k}}}} \\ \sqrt{1 - \frac{\Delta}{\varepsilon_{c\mathbf{k}}}} \frac{f_{\mathbf{k}}}{|f_{\mathbf{k}}|} \end{pmatrix}, \\ |v\mathbf{k}\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} -\sqrt{1 - \frac{\Delta}{\varepsilon_{c\mathbf{k}}}} \frac{f_{\mathbf{k}}}{|f_{\mathbf{k}}|} \\ \sqrt{1 + \frac{\Delta}{\varepsilon_{c\mathbf{k}}}} \end{pmatrix}. \end{aligned} \quad (39)$$

With our choices of parameters, the band gap is  $2\Delta = 7.8 \text{ eV}$ .

Berry connection plays a crucial role in calculating optical matrix elements in both length gauge and velocity gauge. It is defined as  $\xi_{nm\mathbf{k}} = iA^{-1} \int_A d\mathbf{r} u_{nk}^* \nabla_{\mathbf{k}} u_{m\mathbf{k}}$ , where  $A$  is the unit cell area and  $u_{n\mathbf{k}}$  is the cell-periodic part of the Bloch state, which is distinct from the eigenstates in Eq. (39). Following Ref. [51], we compute

$$\xi_{nm\mathbf{k}} = i(\nabla_{\mathbf{q}} U_{n\mathbf{k};m\mathbf{k}+\mathbf{q}})|_{\mathbf{q}=0}, \quad (40)$$

where the overlap matrix element  $U$  is defined as

$$U_{n\mathbf{k};m\mathbf{k}+\mathbf{q}} = \sum_{j=A,B} \langle n\mathbf{k}|j\rangle \langle j|m\mathbf{k}+\mathbf{q}\rangle e^{-i\mathbf{q}\cdot\boldsymbol{\tau}_j}, \quad (41)$$

where  $\langle n\mathbf{k}|j\rangle$  is the wavefunction coefficients of the  $j$ th sublattice orbital and the atom positions are given by  $\boldsymbol{\tau}_A = \mathbf{0}$  and  $\boldsymbol{\tau}_B = \frac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2)$  for  $A$  and  $B$  sublattice, respectively. The off-diagonal part of  $\xi_{nm\mathbf{k}}$  is the matrix elements of the position operator, which describes interband optical transition amplitude. The Hermiticity of the overlap matrix can be seen from the definition.

We numerically solve BSE for excitons in this model. In our implementation, the repulsive exchange term  $V$  is ignored and the attractive screened Coulomb interaction  $W$  is modeled by a Yukawa form [52]

$$W_{vc,\mathbf{k}\mathbf{k}'} = -\frac{1}{8\pi^2 \varepsilon \varepsilon_0} U_{v\mathbf{k}',v\mathbf{k}} U_{c\mathbf{k},c\mathbf{k}'} \frac{e^{-l|\mathbf{k}-\mathbf{k}'|}}{|\mathbf{k}-\mathbf{k}'|}, \quad (42)$$

where  $U_{n\mathbf{k},m\mathbf{k}'}$  is the overlap matrix elements defined above, we set the effective thickness  $l$  to be  $1 \text{ \AA}$  and the dielectric constant  $\varepsilon = 1.5$ . The unit-cell volume  $V = Al$  is used in the calculation. Excitonic velocity matrix elements  $d_s^{\mu}$  from Eq. (31), and the interexciton coupling matrix element  $\Pi_{\lambda s}^{\alpha}$  defined in Eq. (37) are then computed from exciton envelope functions for optical responses. The contraction over band indices is trivial for a two-band model.

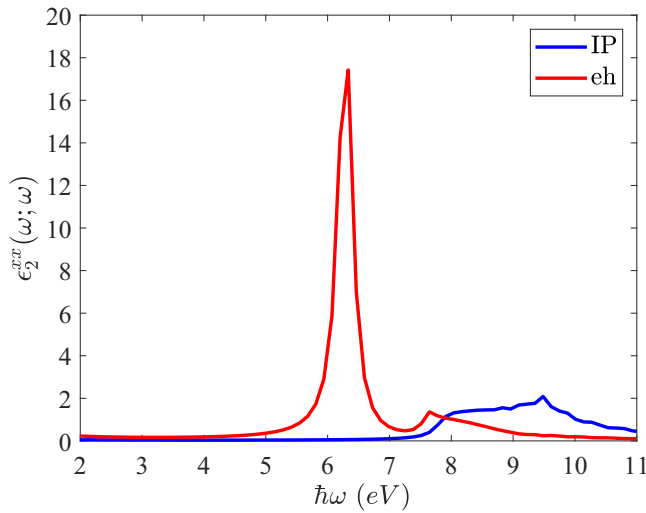


FIG. 5. The  $xx$  tensor component of the imaginary part of the dielectric function computed without (blue-solid line) and with (red-solid line) excitonic effects.

### B. Optical response from the h-BN model

We show the numerical results of absorption and SHG tensors for the two-band Hamiltonian in Figs. 5 and 6, respectively. Excitonic effects on these optical responses are demonstrated by comparing the IP results with those including e-h interactions. In the IP case, we use the expression given in Ref. [41], which is derived in the velocity gauge within the density matrix formalism.

For the SHG conductivity tensor, we evaluate the expressions by setting  $\omega_1 = \omega_2 = \omega$ . Lattice symmetry constrains that the  $xxx$  tensor component is the only independent tensor component for second-order conductivity tensors. Other components are either equal or differ by a minus sign. Hence, we only show the  $xxx$  component below. In our numerical calculations, a uniform  $48 \times 48$   $\mathbf{k}$  grid and a broadening of 0.01 Ry is used. With the chosen parameters, we obtain a binding energy of 1.4 eV for the lowest exciton state. The imaginary part of the dielectric function, which represents the absorption spectrum is shown in Fig. 5. For IP results, we see a step function like a spectrum close to the band edge and a peak at 9.5 eV, which reflects the joint density of state of the monolayer h-BN in the same energy range. Excitonic effects qualitatively change the spectrum. We observe that the spectrum manifests two peaks compared to the IP case, which can be attributed to A and B exciton excitations at 6.4 and 7.8 eV, respectively [29]. This result suggests substantial enhancement of the response due to excitonic effects. Our results generally agree with the reported results from model calculations [52]. We also confirm numerically that the results of Eq. (36) reduced to the IP results if we set the kernel to zero when solving the BSE Hamiltonian.

In Fig. 6, we show the  $xxx$  tensor component of SHG conductivity, which are comparable to the results in Ref. [52]. For the IP response shown in Fig. 6(a), higher responses are observed between 4–5 eV and 8–10 eV. The real part of the SHG responses between 8–10 eV resembles those in the absorption spectrum, which are due to the single-photon

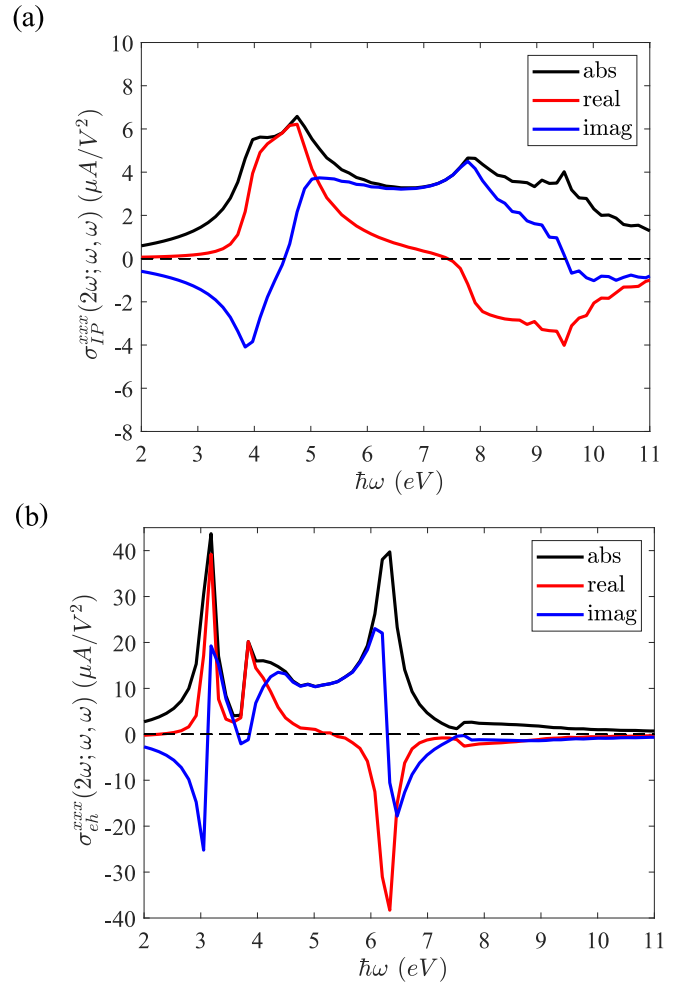


FIG. 6. SHG conductivity spectrum,  $\sigma_{eh}^{xxx}(2\omega; \omega, \omega)$ , (a) computed with IPA and (b) computed from Eq. (36) including excitonic effects. Black-, red-, and blue-solid lines are for the absolute value, the real part, and the imaginary part of the conductivity tensor, respectively.

resonance structure in the denominator of Eq. (36) while those between 4–5 eV can be understood as the corresponding two-photon resonance contribution.

Comparing Figs. 6(a) and 6(b), we can see that excitonic effects greatly enhance the SHG intensity and shift the spectrum to the low energy side. Moreover, two pronounced peaks appear; one is at 6.4 eV, which is identical to that in the absorption spectrum, and the second peak is located at 3.2 eV, which is half the energy of the first peak. The presence of these peaks is the consequence of either a one-photon or a two-photon resonance to the A exciton from the pole structure shown in Eq. (36). Secondary peaks due to B excitons can also be identified at 7.8 eV and 3.9 eV. A detailed analysis reveals that the excitonic enhancement is a result from strong excitonic dipole matrix elements and large interexciton couplings between A and B excitons [31].

### C. Comparison with the density matrix formulation

We compare our expression Eq. (36) with those derived from the density matrix formulation. In Ref. [32], expressions

of second-order conductivity tensors including excitonic effects in different gauges are derived from the density matrix formulation. In particular, Eq. (B1c) in Ref. [32], which is also derived in the velocity gauge reads

$$\begin{aligned}
 & \sigma_{eh}^{\mu\alpha\beta}[\boldsymbol{\pi}_{\lambda_s}](\omega; \omega_1, \omega_2) \\
 &= \frac{C_2}{i\hbar^2\omega_1\omega_2} \sum_{\lambda} \left[ \frac{A_{\lambda}^{\mu\beta} \pi_{\lambda}^{\alpha*}}{\hbar\omega_1 - E_{\lambda}} + \frac{A_{\lambda}^{\mu\beta*} \pi_{\lambda}^{\alpha}}{\hbar\omega_1 + E_{\lambda}} \right] \\
 & - \frac{C_2}{2i\hbar^2\omega_1\omega_2} \sum_{\lambda} \left[ \frac{\pi_{\lambda}^{\mu} A_{\lambda}^{\alpha\beta*}}{\hbar\omega - E_{\lambda}} + \frac{\pi_{\lambda}^{\mu*} A_{\lambda}^{\alpha\beta}}{\hbar\omega + E_{\lambda}} \right] \\
 & + \frac{C_2}{\hbar^2\omega_1\omega_2} \sum_{\lambda_s} \left[ \frac{\pi_{\lambda}^{\mu} \pi_{\lambda_s}^{\alpha} \pi_s^{\beta*}}{(\hbar\omega - E_{\lambda})(\hbar\omega_2 - E_s)} \right. \\
 & + \frac{\pi_{\lambda}^{\mu*} \pi_{\lambda_s}^{\alpha*} \pi_s^{\beta}}{(\hbar\omega + E_{\lambda})(\hbar\omega_2 + E_s)} \\
 & \left. - \frac{\pi_{\lambda}^{\beta} \pi_{\lambda_s}^{\mu} \pi_s^{\alpha*}}{(\hbar\omega_2 + E_{\lambda})(\hbar\omega_1 - E_s)} \right], \quad (43)
 \end{aligned}$$

where  $C_2$  is defined earlier and  $\boldsymbol{\pi}_{\lambda}$ ,  $\boldsymbol{\pi}_{\lambda_s}$ , and  $\boldsymbol{A}_s$  are the excitonic velocity matrix elements, the interexciton velocity matrix elements, and the second-order velocity matrix elements, respectively. They are defined as follows. We first define the excitonic dipole matrix element,  $\boldsymbol{\xi}_{\lambda} = Y_{cv\mathbf{k}}^{\lambda} \boldsymbol{\xi}_{v\mathbf{k}}$ . The definition of the matrix element of the excitonic velocity operator follows from the commutator between the BSE Hamiltonian and the excitonic dipole operator,  $\boldsymbol{\pi}_{\lambda} = -iE_{\lambda} \boldsymbol{\xi}_{\lambda}$ . Similarly, the interexciton velocity operator  $\boldsymbol{\pi}_{\lambda_s}$  is defined as

$$\boldsymbol{\pi}_{\lambda_s} = i(E_{\lambda} - E_s) \boldsymbol{\xi}_{\lambda_s}. \quad (44)$$

The interexciton dipole matrix element  $\boldsymbol{\xi}_{\lambda_s}$  consists of inter- and intraband part,

$$\boldsymbol{\xi}_{\lambda_s} = \boldsymbol{Q}_{\lambda_s} + \boldsymbol{R}_{\lambda_s}, \quad (45)$$

where  $\boldsymbol{R}_{\lambda_s}$  represents the interband contribution within electrons or holes states in excitons,

$$\boldsymbol{R}_{\lambda_s} = \sum_{cv\mathbf{k}} Y_{cv\mathbf{k}}^{\lambda,*} \left( \sum_{c_1 \neq c} Y_{c_1 v \mathbf{k}}^s \boldsymbol{\xi}_{c_1 \mathbf{k}} - \sum_{v_1 \neq v} Y_{cv_1 \mathbf{k}}^s \boldsymbol{\xi}_{v_1 \mathbf{k}} \right) \quad (46)$$

while  $\boldsymbol{Q}_{\lambda_s}$  corresponds to the intraband contribution described by the covariant derivative defined earlier,

$$\boldsymbol{Q}_{\lambda_s} = i \sum_{\mathbf{k}} Y_{cv\mathbf{k}}^{\lambda,*} \boldsymbol{D}_g(Y_{cv\mathbf{k}}^s), \quad (47)$$

where  $D_g^{\alpha}$  is the generalized derivative, defined as  $D_g^{\alpha} O_{ab} = \frac{\partial O_{ab}}{\partial k^{\alpha}} - i(\xi_{aa}^{\alpha} - \xi_{bb}^{\alpha}) O_{ab}$ . In a two-band model,  $\boldsymbol{R}_{\lambda_s}$  vanishes so we have  $\boldsymbol{\xi}_{\lambda_s} = \boldsymbol{Q}_{\lambda_s}$ . The matrix element of the second-order velocity operator  $A_s^{\alpha\beta}$  is defined by taking the commutator of the excitonic dipole operator and  $\boldsymbol{\pi}_s$ , which is expressed as  $\boldsymbol{A}_s = \sum_{\lambda} (\boldsymbol{\xi}_{\lambda} \boldsymbol{\pi}_{\lambda_s} - \boldsymbol{\pi}_{\lambda} \boldsymbol{\xi}_{\lambda_s})$ .

In contrast to our excitonic operators defined earlier, the excitonic velocity operators here are defined through the commutator between the excitonic dipole operators and the exciton Hamiltonian. By comparing Eqs. (36) and (43), we can see that the two expressions can be related to each other by the following substitutions,  $d_{\lambda}^{\alpha} \rightarrow \pi_{\lambda}^{\alpha*}$ ,  $d_{\lambda}^{\alpha\beta} \rightarrow iA_{\lambda}^{\alpha\beta*}$ , and  $\Pi_{\lambda_s}^{\alpha} \rightarrow \pi_{\lambda_s}^{\alpha}$ . For the correspondence between  $\Pi_{\lambda_s}^{\alpha}$  and  $\pi_{\lambda_s}^{\alpha}$ ,

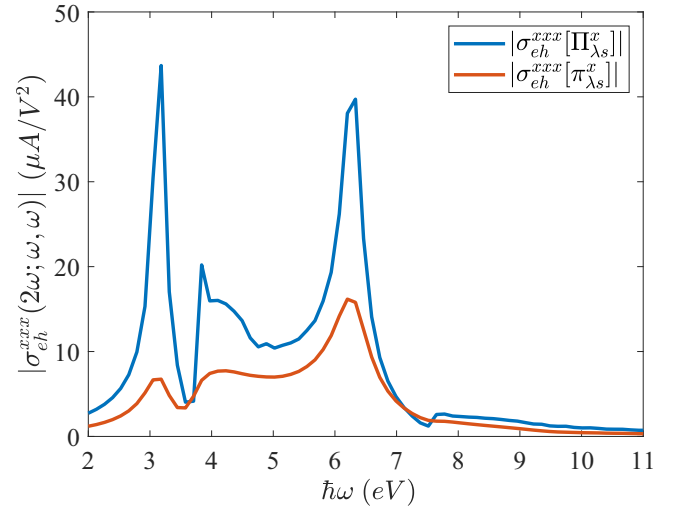


FIG. 7. The absolute value of SHG conductivity tensors computed with Eq. (36) (blue-solid line),  $\sigma_{eh}^{xxx}[\Pi_{\lambda_s}^x]$  and Eq. (43) (orange-solid line),  $\sigma_{eh}^{xxx}[\pi_{\lambda_s}^x]$ .

we note the similarity between Eqs. (37) and (46). The IP velocity matrix elements is related to the dipole matrix elements through Eq. (7). In our definition, the excitonic velocity matrix elements are defined by contracting the IP velocity matrix elements with exciton envelope functions while it is defined as the two-particle generalization of the velocity operator in Ref. [32]. A similar comparison also applies to the interexciton velocity and the second-order excitonic velocity operators. Such difference originates from the fact that we start with a single-particle formulation and treat electron-hole interactions as a perturbation, while in Ref. [32] the derivation starts with the single-particle density matrix formulation but is generalized to excitonic operators defined from the two-particle Hamiltonian in the end. In Fig. 7, we compare the numerical results from the two expressions. We observe that although the two results are qualitatively similar to each other, their absolute intensity differs. Our expression tends to give a higher conductivity than the result from Eq. (43) and the first peak around 3 eV is more pronounced. To further understand this difference, we analyze the contribution of each term separated by parenthesis in Eq. (36). We find that the first three terms have the dominant contributions to the SHG conductivity tensor. Their contributions are shown in Figs. 8(a) and 8(b) for Eqs. (36) and (43), respectively. From the peak position in each panel, we can identify their single- or two-photon resonance origins. A term-by-term comparison shows that the frequency dependence and the sign of corresponding terms from the two derivations qualitatively agree. However, the difference in their relative magnitude leads to the difference in total responses. Specifically, the large deviation at around 3.0 eV is due to the cancellation between the contribution of the second and the third term as shown in Fig. 8(b). In contrast, the cancellation is less effective from our expression as shown in Fig. 8(a). We further compare the different matrix elements defined above. For the interexciton coupling matrices, by definition the diagonal elements of  $\pi_{\lambda_s}^{\alpha}$  vanish while  $\Pi_{\lambda_s}^{\alpha}$  can have finite diagonal elements. We find that the off-diagonal elements of  $\Pi_{\lambda_s}^{\alpha}$  is larger than  $\pi_{\lambda_s}^{\alpha}$  in magnitude in

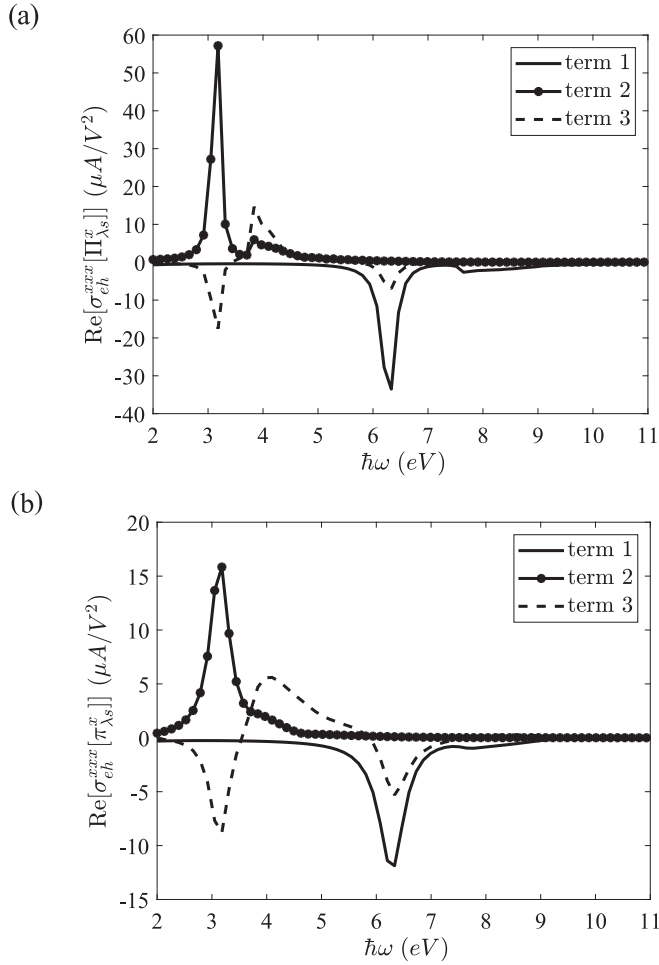


FIG. 8. The real part of the three dominant terms to SHG. They correspond to the first term through the third term in (a) Eq. (36) and (b) Eq. (43).

our calculation. The comparison of the second-order velocity matrix elements also shows a similar trend.

We caution that the comparison of the two expressions should be made carefully. Although the enhancement from excitonic effects have been demonstrated in previous calculations, we observe that the enhancement is stronger when using

the expression derived in the present study as shown in Fig. 7. One possible issue is the small number of bands used in our calculations as we are working with a two-band model. It is shown that a large number of bands are needed to converge the calculations in the velocity gauge [5,6,48]. Another possible reason for the discrepancy of the two derivations is the lack of self-consistency in our paper. Since we use the bare Green's function in all calculations, accordingly the velocity operator is defined from the corresponding IP Hamiltonian. For a self-consistent theory, one would include electron-hole interaction effects through the self-energy, which could lead to a correction term to the velocity operator from the effective single-particle Hamiltonian. A fully self-consistent treatment is beyond this work and will be left in the future work.

## VI. CONCLUSIONS

In summary, we derive the expression for second-order optical responses including excitonic effects with a diagrammatic approach. Our approach extends the previous derivation for the IP cases by dressing the electron-photon coupling vertex with the electron-hole ladder diagrams. It is known that excitonic effects are strong for low dimensional materials. Hence, we expect that excitonic effects are essential to describe nonlinear optical responses. First-principle calculations of second-order optical responses including excitonic effects can be straightforwardly performed by implementing the derived expression. All the ingredients can be obtained from standard density functional theory packages and software with BSE solvers implemented. Although we only focus on the second-order response, the diagrammatic rules provided here can be readily applied to higher-order responses.

## ACKNOWLEDGMENTS

Y.-H.C. and Y.-T.C. thank Hsin Lin and Prof. T. G. Pedersen for the discussion. Y.-H.C. thanks J. Ruan and Prof. S. G. Louie for collaborations on related projects. This work was supported by the National Science and Technology Council of Taiwan under Grant No. 112-2112-M-001-048-MY3. We acknowledge the use of computational resources at the National Center for High-performance Computing (NCHC) in Taiwan.

## APPENDIX A: DERIVATION OF THE SECOND-ORDER CONDUCTIVITY TENSOR

In the first part of this section, we give the detailed derivations of the second-order optical conductivity within IP approximation. The derivation including electron-hole interactions is given in Appendix D. These derivations demonstrate the Feynmann rules listed in Ref. [33] and in our paper.

For the second-order response, the conductivity tensor  $\sigma^{\mu\alpha\beta}(\omega; \omega_1, \omega_2)$  can be computed from the second derivative of the current density with respect to the external fields,  $\frac{\delta^2 J^\mu(t)}{\delta E^\alpha(t_1) \delta E^\beta(t_2)}|_{E=0}$ . From Eq. (9) we can identify that there are two operators, which consist of expansion of the external field. One is the velocity operator associated with the current operator and the other is the coupling Hamiltonian  $H_E$ . In total there are four terms contributing to the second-order conductivity,

$$\frac{\delta^2 v_E^\mu(t)}{\delta E^\beta(t_2) \delta E^\alpha(t_1)} - v_E^\mu(t) \frac{\delta^2 \int dt' H_E(t')}{\delta E^\beta(t_2) \delta E^\alpha(t_1)} - \frac{\delta v_E^\mu(t)}{\delta E^\beta(t_2)} \frac{\delta \int dt' H_E(t')}{\delta E^\alpha(t_1)} + \frac{1}{2!} v_E^\mu(t) \frac{\delta^2 (\int dt'' H_E(t''))^2}{\delta E^\alpha(t_1) \delta E^\beta(t_2)}, \quad (\text{A1})$$

where  $\mathbf{E} = 0$  is set after taking the derivatives. As we discussed in the main text, we identify the vertex associated with the current operator as the outgoing vertex and those associated with the  $H_E$  as the incoming vertex. The number of photon lines on a vertex is determined by the order of derivatives with respect to the external field. Therefore the first term correspond

to the diagram with three photon lines on the outgoing vertex, the third term is the diagram with two photon lines on the outgoing vertex and one photon line on the incoming vertex. The second and fourth term both have two incoming vertex and one outgoing vertex. While the second term has two photon lines on the incoming vertex, the fourth term has one photon line on each incoming vertex. We note that the functional derivative can be taken in a different order, which corresponds to the symmetry of exchanging the two external fields.

We are interested in the second and the third term, which generate diagrams with two incoming or two outgoing photons on the same vertex. The third term can be obtained by combining terms we calculated for the first-order conductivity. Using  $\frac{\delta}{\delta E(t)} = \int d\omega' e^{i\omega't} \frac{\delta}{\delta E(\omega')}$ , we have

$$\begin{aligned} -\frac{\delta v_E^\mu(t)}{\delta E^\beta(t_2)} \frac{\delta}{\delta E^\alpha(t_1)} \int dt' H_E(t') &= -\int d\omega_1 e^{i\omega_1 t_2} \frac{\delta}{\delta E^\beta(\omega_1)} \left[ \sum_{n=0}^{\infty} \frac{1}{n!} \prod_{k=1}^n \int \frac{d\omega_k}{2\pi} e^{-i\omega_k t} \left( \frac{ie}{\hbar\omega_k} \right) E^{\alpha_k}(\omega_k) \hat{h}^{\mu\alpha_1 \dots \alpha_n} \right] \\ &\quad \times \int dt' \int d\omega_2 e^{i\omega_2 t_1} \frac{\delta}{\delta E^\alpha(\omega_2)} \left[ \sum_{n=1}^{\infty} \frac{1}{n!} \prod_{k=1}^n \int \frac{d\omega_k}{2\pi} e^{-i\omega_k t'} \left( \frac{ie}{\hbar\omega_k} \right) E^{\alpha_k}(\omega_k) \hat{h}^{\alpha_1 \dots \alpha_n} \right] \\ &= -\int dt' \int \frac{d\omega_1}{2\pi} \int \frac{d\omega_2}{2\pi} e^{-i\omega_2(t'-t_1)} e^{-i\omega_1(t-t_2)} \left( \frac{ie}{\hbar\omega_1} \right) \hat{h}^{\mu\beta}(t) \left( \frac{ie}{\hbar\omega_2} \right) \hat{h}^\alpha(t'). \end{aligned} \quad (\text{A2})$$

The term with the second derivative in the second term of Eq. (A1) reads

$$\begin{aligned} \frac{\delta^2}{\delta E^\beta(t_2) \delta E^\alpha(t_1)} \int dt' H_E(t') &= \int d\omega'' e^{i\omega'' t_2} \frac{\delta}{\delta E^\beta(\omega'')} \int d\omega' e^{i\omega' t_1} \frac{\delta}{\delta E^\alpha(\omega')} \int dt' \sum_{n=1}^{\infty} \frac{1}{n!} \prod_{k=1}^n \int \frac{d\omega_k}{2\pi} e^{-i\omega_k t'} \left( \frac{ie}{\hbar\omega_k} \right) E^{\alpha_k}(\omega_k) \hat{h}^{\alpha_1 \dots \alpha_n} \\ &= \frac{1}{2!} \int d\omega'' e^{i\omega'' t_2} \frac{\delta}{\delta E^\beta(\omega'')} \int d\omega' e^{i\omega' t_1} \frac{\delta}{\delta E^\alpha(\omega')} \int dt' \int \frac{d\omega_1}{2\pi} \int \frac{d\omega_2}{2\pi} \\ &\quad \times e^{-i\omega_1 t'} e^{-i\omega_2 t'} \left( \frac{ie}{\hbar\omega_1} \right) \left( \frac{ie}{\hbar\omega_2} \right) E^{\alpha_1}(\omega_1) E^{\alpha_2}(\omega_2) \hat{h}^{\alpha_1 \alpha_2} \\ &= \frac{1}{2!} \int dt' \int \frac{d\omega''}{2\pi} e^{i\omega''(t_2-t')} \int \frac{d\omega'}{2\pi} e^{i\omega'(t_1-t')} \left( \frac{ie}{\hbar\omega''} \right) \left( \frac{ie}{\hbar\omega'} \right) (\hat{h}^{\alpha\beta}(t') + \hat{h}^{\beta\alpha}(t')). \end{aligned} \quad (\text{A3})$$

Using  $v_E^\mu = h^\mu$  at the zeroth order, the second term is

$$-h^\mu(t) \frac{1}{2!} \int dt' \int \frac{d\omega''}{2\pi} e^{i\omega''(t_2-t')} \int \frac{d\omega'}{2\pi} e^{i\omega'(t_1-t')} \left( \frac{ie}{\hbar\omega''} \right) \left( \frac{ie}{\hbar\omega'} \right) (\hat{h}^{\alpha\beta}(t') + \hat{h}^{\beta\alpha}(t')).$$

The Fourier transformation of the expectation value of the third term is

$$\begin{aligned} \sigma_{IP}^{\mu\alpha\beta,3}(\omega; \omega_1, \omega_2) &\frac{\delta(\omega - \omega_1 - \omega_2)}{2\pi} \\ &= -e \int dt e^{i\omega t} \int \frac{dt_1}{2\pi} e^{-i\omega_1 t_1} \int \frac{dt_2}{2\pi} e^{-i\omega_2 t_2} \int dt' \int \frac{d\omega_3}{2\pi} \int \frac{d\omega_4}{2\pi} e^{-i\omega_4(t'-t_1)} e^{-i\omega_3(t-t_2)} \left( \frac{ie}{\hbar\omega_3} \right) \left( \frac{ie}{\hbar\omega_4} \right) \langle \hat{h}^{\mu\beta}(t) \hat{h}^\alpha(t') \rangle \\ &= -e \int dt \int \frac{dt_1}{2\pi} e^{i(-\omega_1 + \omega_4)t_1} \int \frac{dt_2}{2\pi} e^{i(-\omega_2 + \omega_3)t_2} \int dt' \int \frac{d\omega_3}{2\pi} \int \frac{d\omega_4}{2\pi} e^{-i\omega_4 t'} e^{-i(\omega_3 - \omega)t} \left( \frac{ie}{\hbar\omega_3} \right) \left( \frac{ie}{\hbar\omega_4} \right) \langle \hat{h}^{\mu\beta}(t) \hat{h}^\alpha(t') \rangle \\ &= \frac{1}{(2\pi)^2} \frac{e^3}{\hbar^2 \omega_2 \omega_1} \int dt \int dt' e^{i\omega_1 t'} e^{-i(\omega_2 - \omega)t} h_{ab}^{\mu\beta} h_{cd}^\alpha \langle c_a^\dagger(t) c_b(t) c_c^\dagger(t') c_d(t') \rangle \\ &= -\frac{1}{(2\pi)^2} \frac{e^3}{\hbar^2 \omega_2 \omega_1} \int dt \int dt' e^{i\omega_1 t'} e^{-i(\omega_2 - \omega)t} h_{ab}^{\mu\beta} h_{ba}^\alpha G_b(t, t') G_a(t', t) \\ &= -\frac{1}{(2\pi)^2} \frac{e^3}{\hbar^2 \omega_2 \omega_1} \int dt \int dt' e^{i\omega_1 t'} e^{-i(\omega_2 - \omega)t} h_{ab}^{\mu\beta} h_{ba}^\alpha \int \frac{d\omega'}{2\pi} G_b(\omega') e^{-i\omega'(t-t')} \int \frac{d\omega''}{2\pi} e^{-i\omega''(t'-t)} G_a(\omega'') \\ &= -\frac{1}{(2\pi)^4} \frac{e^3}{\hbar^2 \omega_2 \omega_1} \int d\omega' \int d\omega'' \int dt \int dt' e^{i(\omega_1 + \omega' - \omega'')t'} e^{-i(\omega_2 - \omega + \omega' - \omega'')t} h_{ab}^{\mu\beta} h_{ba}^\alpha G_b(\omega') G_a(\omega'') \\ &= -\frac{e^3}{\hbar^2 \omega_2 \omega_1} \frac{\delta(\omega - \omega_1 - \omega_2)}{2\pi} \int \frac{d\omega'}{2\pi} h_{ab}^{\mu\beta} h_{ba}^\alpha G_b(\omega') G_a(\omega_1 + \omega'), \end{aligned} \quad (\text{A4})$$

where the superscript 3 indicates that it is the third term in Eq. (A1) and in the sixth line only the connected diagram is considered. For the second term, we have

$$\begin{aligned}
& \sigma_{IP}^{\mu\alpha\beta,2}(\omega; \omega_1, \omega_2) \frac{\delta(\omega - \omega_1 - \omega_2)}{2\pi} \\
&= -\frac{e}{2!} \int dt e^{i\omega t} \int \frac{dt_1}{2\pi} \int \frac{dt_2}{2\pi} \int dt' \int \frac{d\omega''}{2\pi} \int \frac{d\omega'}{2\pi} e^{i(-\omega_2+\omega'')t_2} e^{i(-\omega_1+\omega')t_1} e^{-i(\omega'+\omega'')t'} \left(\frac{ie}{\hbar\omega''}\right) \left(\frac{ie}{\hbar\omega'}\right) \langle \hat{h}^\mu(t) \hat{h}^{\alpha\beta}(t') \rangle \\
&= \frac{e^3}{\hbar^2 \omega_1 \omega_2} \frac{1}{2!(2\pi)^2} \int dt e^{i\omega t} \int dt' e^{-i(\omega_1+\omega_2)t'} \langle \hat{h}^\mu(t) \hat{h}^{\alpha\beta}(t') \rangle \\
&= \frac{1}{2!(2\pi)^2} \frac{e^3}{\hbar^2 \omega_1 \omega_2} \int dt e^{i\omega t} \int dt' e^{-i(\omega_1+\omega_2)t'} h_{ab}^\mu h_{cd}^{\alpha\beta} \langle c_a^\dagger(t) c_b(t) c_c^\dagger(t') c_d(t') \rangle \\
&= -\frac{1}{2!(2\pi)^2} \frac{e^3}{\hbar^2 \omega_1 \omega_2} \int dt e^{i\omega t} \int dt' e^{-i(\omega_1+\omega_2)t'} h_{ab}^\mu h_{cd}^{\alpha\beta} G_{bc}(t, t') G_{da}(t', t) \\
&= -\frac{1}{2!(2\pi)^2} \frac{e^3}{\hbar^2 \omega_1 \omega_2} \int dt e^{i\omega t} \int dt' e^{-i(\omega_1+\omega_2)t'} h_{ab}^\mu h_{ba}^{\alpha\beta} \int \frac{d\omega'}{2\pi} G_b(\omega') e^{-i\omega'(t-t')} \int \frac{d\omega''}{2\pi} e^{-i\omega''(t'-t)} G_a(\omega'') \\
&= -\frac{1}{2!(2\pi)^2} \frac{e^3}{\hbar^2 \omega_1 \omega_2} \int dt \int dt' e^{i(-\omega_1-\omega_2+\omega-\omega'')t} h_{ab}^\mu h_{ba}^{\alpha\beta} \int \frac{d\omega'}{2\pi} \int \frac{d\omega''}{2\pi} e^{i(\omega''+\omega-\omega')t} G_b(\omega') G_a(\omega'') \\
&= -\frac{e^3}{2\hbar^2 \omega_1 \omega_2} \frac{\delta(\omega - \omega_1 - \omega_2)}{2\pi} h_{ab}^\mu h_{ba}^{\alpha\beta} \int \frac{d\omega''}{2\pi} G_b(\omega'' + \omega_1 + \omega_2) G_a(\omega''). \tag{A5}
\end{aligned}$$

These results agree with Eq. (42) in Ref. [33]

## APPENDIX B: EQUATION OF MOTION OF THE TWO-PARTICLE CORRELATION FUNCTION

### 1. Perturbative expansion

In this Appendix we show how the ladder diagrams can be derived from the perturbative expansion of Coulomb interactions when evaluating Eq. (21). The infinite sum of ladder diagrams was performed in Ref. [47] to describe the exciton absorption spectrum. In the first-order expansion over  $H_{\text{int}}$  for both the numerator and the denominator of Eq. (21), we obtain [40]

$$\begin{aligned}
& \langle c_{ka}^\dagger(t) c_{kb}(t) c_{kc}^\dagger(t') c_{kd}(t') \rangle_{\text{int}} \\
&= \langle c_{ka}^\dagger(t) c_{kb}(t) c_{kc}^\dagger(t') c_{kd}(t') \rangle \\
&\quad - i \left\langle c_{ka}^\dagger(t) c_{kb}(t) c_{kc}^\dagger(t') c_{kd}(t') \right. \\
&\quad \times \left. \int dt'' V_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}}^{abcd} c_{\mathbf{k}_1+\mathbf{q}a}^\dagger c_{\mathbf{k}_2-\mathbf{q}b}^\dagger c_{\mathbf{k}_2c} c_{\mathbf{k}_1d} \right\rangle^{\text{conn.}}, + \dots, \tag{B1}
\end{aligned}$$

where the superscript conn. indicates that we only consider connected diagrams since the disconnected diagrams are canceled by the expansion of the denominator. The bubble diagram in Fig. 9 corresponds to the first term in Eq. (B1) while the other diagram represents one of the non-self-energy diagrams from the second term.

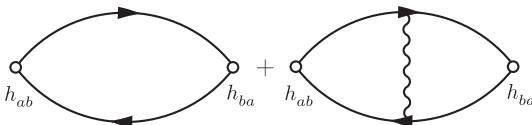


FIG. 9. Zero-order and first-order expansion.

In the next-order expansion over  $V_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}}^{abcd}$ , we have the term  $\langle c^\dagger c c^\dagger c H_{\text{int}}^2 \rangle$ , which involves twelve creation and annihilation operators. Excluding the self-energy and disconnected diagrams, two of the new diagrams are depicted in Fig. 10 [40]. We could continue with higher-order expansions. However, in most cases, only approximate solutions can be attained. It has been shown by Mahan [42] that for semiconductors, additional vertex corrections, such as “entangled” interactions (cross Coulomb line diagrams in Fig. 10) beyond ladder diagrams is proportional to the density of excited electrons or holes and can be safely ignored in the weak field limit. Keeping only the ladder diagrams leads to the Dyson-like equation. Symbolically, we have

$$\begin{aligned}
L &= G_0 G_0 + G_0 G_0 V_q G_0 G_0 + G_0 G_0 V_q G_0 G_0 V_q G_0 G_0 + \dots \\
&= L_0 + L_0 V_q L_0 + L_0 V_q L_0 V_q L_0 + \dots \\
&= L_0 + L_0 V_q L, \tag{B2}
\end{aligned}$$

where  $G_0$  is defined in Eq. (16), and  $L_0 \equiv G_0 G_0$ , and the interacting two-particle correlation function (B1) is denoted as  $L$ . The infinite sum of ladder diagrams is illustrated in Fig. 11. Equation (B2) is the BSE with kernel  $V_q$  for the two-particle correlation function. For a static interaction  $V_q$ , the Bethe-Salpeter equation can be reformulated as an eigenvalue problem, from the solutions of which we can evaluate the two-particle correlation function Eq. (B1) and the velocity-velocity correlation function, Eq. (21). In the above derivation,

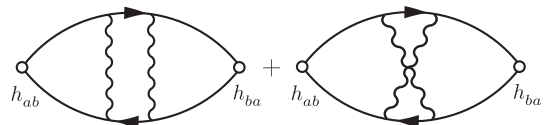


FIG. 10. Second-order expansion.

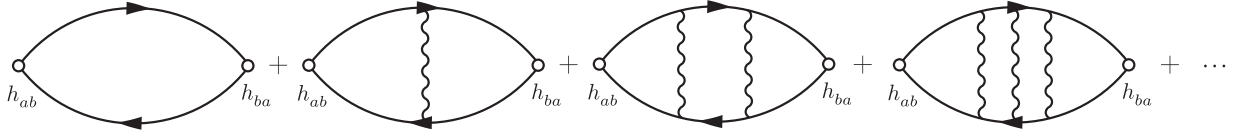


FIG. 11. Ladder diagrams of the two-particle correlation function.

the bubble diagrams are not included for simplicity. In Appendix B 2, we will show a derivation, which systematically includes both bubble and ladder diagrams [43,44,53].

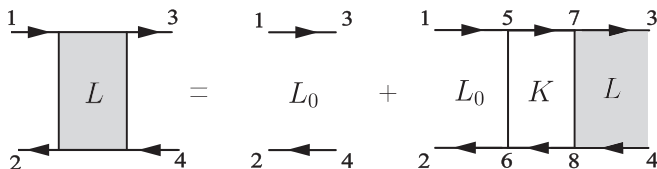
## 2. Functional derivative approach

We show that ladder diagrams can be derived using a functional derivative approach with the choice of GW self-energy. Moreover, the derivation ensures that there are no new light-matter coupling vertices introduced in the two-particle level. A detailed derivation can be found in [43,44,53]. Here, we reproduce the key steps. The Dyson's equation for a single-particle Green's function  $G$  is  $G = G_0 + G_0 h G_0 + G_0 \Sigma G$ , where  $G_0$  is the bare Green's function,  $h$  is the electron-photon coupling vertex and  $\Sigma$  is the electron self-energy due to Coulomb interactions. The equation of motion of the two-particle correlation can be derived from the definition

$$L(1, 4; 2, 3) \equiv \frac{\delta G(1, 2)}{\delta h(3, 4)}. \quad (\text{B3})$$

$$\begin{aligned} L(1, 4; 2, 3) &= - \int d5d6 G(1, 5) \frac{\delta}{\delta h(3, 4)} [G_0^{-1}(5, 6) - h(5, 6) - \Sigma(5, 6)] G(6, 2) \\ &= - \int d5d6 G(1, 5) \left[ -\delta(5, 3)\delta(6, 4) - \frac{\delta \Sigma(5, 6)}{\delta h(3, 4)} \right] G(6, 2) \\ &= G(1, 3)G(4, 2) + \int d5d6 G(1, 5) \frac{\delta \Sigma(5, 6)}{\delta h(3, 4)} G(6, 2) \\ &= G(1, 3)G(4, 2) + \int d5d6 G(1, 5) G(6, 2) \int d7d8 \frac{\delta \Sigma(5, 6)}{\delta G(7, 8)} \frac{\delta G(7, 8)}{\delta h(3, 4)} \\ &= L_0(1, 4; 2, 3) + \int d5d6d7d8 L_0(1, 6; 2, 5) K(5, 8; 6, 7) L(7, 4; 8, 3), \end{aligned} \quad (\text{B8})$$

where in the last line the kernel  $K$  is defined as  $\frac{\delta \Sigma}{\delta G}$ . We obtain the BSE for the two-particle correlation  $L = L_0 + L_0 \frac{\delta \Sigma}{\delta G} L$ . In our work we choose the Hartree and the GW self-energy, where  $W$  is the screened Coulomb potential. Ignoring the higher-order corrections from  $\frac{\delta W}{\delta G}$ , the equation of motion of the two-particle correlation function is described by the ladder diagrams shown in Fig. 12 with  $V_{\mathbf{q}}$  replaced by the kernel  $K$ . As we can see from the final equation of motion that no


 FIG. 12. Two-particle correlation function  $L(1, 4; 2, 3)$ .

Use functional integral identity [43],

$$\begin{aligned} \int dz F[\psi(\xi); x, z] F^{-1}[\psi(\xi); z, y] &= \delta(x, y), \quad (\text{B4}) \\ \frac{\delta}{\delta \psi(z)} F[\psi(\xi); x, y] &= - \int d\xi d\zeta F[\psi(\xi); x, \xi] \frac{\delta}{\delta \psi(z)} F^{-1} \\ &\quad \times [\psi(\xi); \xi, \zeta] F[\psi(\xi); \zeta, y] \end{aligned} \quad (\text{B5})$$

and

$$\frac{\delta F}{\delta \psi(y)} = \int dy \frac{\delta F}{\delta G(y)} \frac{\delta G(y)}{\delta \psi(y)}, \quad (\text{B6})$$

we rewrite

$$L(1, 4; 2, 3) = - \int d5d6 G(1, 5) \frac{\delta}{\delta h(3, 4)} G^{-1}(5, 6) G(6, 2). \quad (\text{B7})$$

Inserting Dyson equation for the single-particle Green's function,  $G(1, 2) = G_0(1, 2) + \int d3d4 G_0(1, 3)[h(3)\delta(3, 4) + \Sigma(3, 4)]G(4, 2)$ , to the equation above, we have

new electron-photon vertex are generated, which justifies the Feynman rule we introduced at the two-particle level.

## APPENDIX C: GREEN'S FUNCTION AND CORRELATION FUNCTION

Throughout Appendix C we set  $\hbar = 1$  to simplify the book-keeping and the Einstein summation convention is implied for band and momentum indices. Following Ref. [33], the noninteracting single-particle Green's function is

$$G_b(\omega) = \frac{1}{\omega - \epsilon_b}. \quad (\text{C1})$$

Its frequency integral is

$$I_1(\omega_1) = \int \frac{d\omega}{2\pi} G_a(\omega) = f(\epsilon_a), \quad (\text{C2})$$

where  $f(\epsilon_a)$  is the Fermi-Dirac function. The convolution of two and three Green's function are

$$I_2(\omega_1) = \int \frac{d\omega}{2\pi} G_a(\omega) G_b(\omega + \omega_1), \quad (C3)$$

$$I_3(\omega_1, \omega_2) = \int \frac{d\omega}{2\pi} G_a(\omega) G_b(\omega + \omega_1) G_c(\omega + \omega_1 + \omega_2), \quad (C4)$$

respectively. The evaluation of these integrals can be done by working with Matsubara frequencies then perform analytical continuation back to the real frequencies [47]. For  $I_2$ , we consider

$$S_2(i\omega_1) = \frac{1}{\beta} \sum_n \frac{1}{z_n - \epsilon_a} \frac{1}{z_n + i\omega_1 - \epsilon_b}, \quad (C5)$$

where  $z_n = i(2n+1)\pi/\beta$ . The summation can be done by considering a contour integral of a function

$$0 = \int \frac{dz}{2\pi i} f(z) F(z) \quad (C6)$$

with  $f(z) = \frac{1}{e^{\beta z} + 1}$  and

$$F(z) = \frac{1}{z - \epsilon_a} \frac{1}{z + i\omega_1 - \epsilon_b}. \quad (C7)$$

The function  $f(z)F(z)$  has poles and contribute to the contour integral at

$$z = z_n, \quad R_1 = -\frac{1}{\beta} F(z_n), \quad (C8)$$

$$z = \epsilon_a, \quad R_2 = \frac{f(\epsilon_a)}{\epsilon_a - \epsilon_b + i\omega_1}, \quad (C9)$$

$$z = \epsilon_b - i\omega_1, \quad R_3 = \frac{f(\epsilon_b)}{\epsilon_b - \epsilon_a - i\omega_1}. \quad (C10)$$

We note that the residue of Fermi function is  $-1/\beta$  so we have

$$\begin{aligned} S_2(i\omega_1) &= \frac{1}{\beta} F(z_n) = \frac{f(\epsilon_a)}{\epsilon_a - \epsilon_b + i\omega_1} - \frac{f(\epsilon_b)}{\epsilon_a - \epsilon_b + i\omega_1} \\ &= -\frac{f_{ba}}{i\omega_1 - \epsilon_{ba}}, \end{aligned} \quad (C11)$$

where we define  $f_{ab} = f(\epsilon_a) - f(\epsilon_b)$  and  $\epsilon_{ab} = \epsilon_a - \epsilon_b$ . For  $I_3$ , we consider the following function:

$$F_3(z) = \frac{1}{z - \epsilon_a} \frac{1}{z + i\omega_1 - \epsilon_b} \frac{1}{z + i\omega_1 + i\omega_2 - \epsilon_c}. \quad (C12)$$

The contour integral has three pieces

$$z = z_n, \quad R_1 = -\frac{1}{\beta} F_3(z_n), \quad (C13)$$

$$z = \epsilon_a, \quad R_2 = \frac{f(\epsilon_a)}{(\epsilon_{ab} + i\omega_1)(\epsilon_{ac} + i\omega_{12})}, \quad (C14)$$

$$z = \epsilon_b - i\omega_1, \quad R_3 = \frac{f(\epsilon_b)}{(\epsilon_{ba} - i\omega_1)(\epsilon_{bc} + i\omega_2)}, \quad (C15)$$

$$z = \epsilon_c - i\omega_1 - i\omega_2, \quad R_4 = \frac{f(\epsilon_c)}{(\epsilon_{ca} - i\omega_{12})(\epsilon_{cb} - i\omega_2)}. \quad (C16)$$

where  $\omega_{12} = \omega_1 + \omega_2$ . The sum of them gives

$$\begin{aligned} I_3(i\omega_1, i\omega_2) &= \frac{f(\epsilon_a)}{(\epsilon_{ba} - i\omega_1)(\epsilon_{ca} - i\omega_{12})} + \frac{f(\epsilon_b)}{(\epsilon_{ba} - i\omega_1)(\epsilon_{bc} + i\omega_2)} + \frac{-f(\epsilon_c)}{(\epsilon_{ca} - i\omega_{12})(\epsilon_{bc} + i\omega_2)} \\ &= \frac{f(\epsilon_a)(\epsilon_{bc} + i\omega_2) + f(\epsilon_b)(\epsilon_{ca} - i\omega_{12}) - f(\epsilon_c)(\epsilon_{ba} - i\omega_1)}{(\epsilon_{ba} - i\omega_1)(\epsilon_{bc} + i\omega_2)(\epsilon_{ca} - i\omega_{12})} \\ &= \frac{f_{ab}(i\omega_2 - \epsilon_{cb}) + f_{cb}(i\omega_1 - \epsilon_{ba})}{(i\omega_1 - \epsilon_{ba})(i\omega_2 - \epsilon_{cb})(i\omega_{12} - \epsilon_{ca})}. \end{aligned} \quad (C17)$$

### 1. Two-particle correlation function

To compute the noninteracting two particle correlation function, we start from the Matsubara component,

$$L_{ab}^M(i\omega_p) = \frac{i}{\beta} \sum_m G_b^M(i\omega_m) G_a^M(i\omega_m + i\omega_p). \quad (C18)$$

Using the result in the previous section, we have

$$L_{ab}^M(i\omega_p) = i \frac{f_{ba}}{i\omega_p - \epsilon_{ab}}. \quad (C19)$$

The retarded component can be obtained by an analytical continuation, which reads

$$L_{0,ab}^R(\omega) = i \frac{f_{ba}}{\omega + i\eta - \epsilon_{ab}}, \quad (C20)$$

where the subscript 0 indicates that it is the noninteracting correlation function.

Interacting electron-hole correlation function can be constructed from the BSE solutions [35,45,46,54]. With Tamm-Dancoff approximation, we have

$$H_{cv'c'} Y_{c'v'}^\lambda = (\omega_{cv} - K_{cv'c'}) Y_{c'v'}^\lambda = \Omega_\lambda Y_{c'v'}^\lambda, \quad (C21)$$

$$H_{vcv'c'} X_{v'c'}^\lambda = (\omega_{vc} + K_{vcv'c'}) X_{v'c'}^\lambda = -\Omega_\lambda X_{v'c'}^\lambda, \quad (C22)$$

where  $\Omega_\lambda$  is the eigenvalue,  $Y^\lambda$  and  $X^\lambda$  are eigenvectors for resonant and antiresonant sectors, respectively. We see that  $H_{vcv'c'} = -H_{cv'c'}^*$  and  $X_{v'c'}^\lambda = Y_{c'v'}^{\lambda*}$ . The interacting correlation function can then be constructed as

$$\begin{aligned} L_{ijknmk'}^R(\omega) &= i \sum_\lambda \left[ \bar{f}_i \bar{f}_j \bar{f}_n \bar{f}_m \frac{Y_{ijk}^\lambda Y_{nmk'}^{\lambda*}}{\omega - \Omega_\lambda + i\eta} \right. \\ &\quad \left. - \bar{f}_i \bar{f}_j \bar{f}_m \bar{f}_n \frac{Y_{jik}^{\lambda*} Y_{mkn}^\lambda}{\omega + \Omega_\lambda + i\eta} \right], \end{aligned} \quad (C23)$$

where we define  $\bar{f} = 1 - f$ .



**APPENDIX D: DERIVATION OF THE SECOND-ORDER CONDUCTIVITY TENSOR  
WITH ELECTRON-HOLE INTERACTIONS**

For the second-order response, there are four diagrams and their permutations. Vertex corrections of the first three diagrams in Ref. [33] are similar to those in the linear response. The second term in Eq. (42) in Ref. [33] gets vertex correction on the out-going vertex with two photon lines,

$$\begin{aligned}\sigma_{eh}^{\mu\alpha\beta,2} &= \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{ab}^\alpha \int d\omega' G_b(\omega') G_a(\omega' + \omega_1) \tilde{h}_{ab}^{\mu\beta}(\omega_1) = \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{ab}^\alpha L_{0,ab}(\omega_1) h_{cd}^{\mu\beta*} L_{cd,ab}(\omega_1) L_{0,ab}^{-1}(\omega_1) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{ab}^\alpha L_{cd,ab}(\omega_1) h_{cd}^{\mu\beta*} = \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_\lambda h_{ab}^\alpha \left[ \bar{f}_c \bar{f}_d \bar{f}_a \bar{f}_b \frac{Y_{cd\mathbf{k}}^\lambda Y_{ab\mathbf{k}'}^{\lambda*}}{\hbar\omega_1 - \Omega_\lambda + i\eta} - f_c \bar{f}_d \bar{f}_a \bar{f}_b \frac{Y_{dc\mathbf{k}}^{\lambda*} Y_{bak'}^\lambda}{\hbar\omega_1 + \Omega_\lambda + i\eta} \right] h_{cd}^{\mu\beta*} \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_\lambda \left[ \frac{d_\lambda^\alpha d_\lambda^{\mu\beta*}}{\hbar\omega_1 - \Omega_\lambda + i\eta} - \frac{d_\lambda^{\alpha*} d_\lambda^{\mu\beta}}{\hbar\omega_1 + \Omega_\lambda + i\eta} \right],\end{aligned}\quad (D1)$$

where we denote dressed operators with a tilde and the superscript 2 indicates that it is the second term in Eq. (42) in Ref. [33]; In the third line we use the inverse of the bare two particle correlation function,

$$L_{0,ab}^{-1} = \frac{\hbar\omega - \epsilon_{ab}}{f_{ba}}, \quad (D2)$$

and in the last line we define  $d_\lambda^{\alpha\beta} = h_{ab}^\alpha Y_{ab\mathbf{k}}^{\lambda*}$ . We can also dress the incoming vertex,

$$\begin{aligned}\sigma_{eh}^{\mu\alpha\beta,2} &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \tilde{h}_{ab}^\alpha(\omega_1) \int d\omega' G_b(\omega') G_a(\omega' + \omega_1) h_{ba}^{\mu\beta} = \frac{-e^3}{\hbar^2\omega_1\omega_2} L_{0,ab}^{-1}(\omega_1) L_{ab,cd}(\omega_1) h_{cd}^\alpha L_{0,ab}(\omega_1) h_{ba}^{\mu\beta} \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{cd}^\alpha L_{ab,cd}(\omega_1) h_{ba}^{\mu\beta},\end{aligned}\quad (D3)$$

which is equivalent to Eq. (D1).

For the third diagram in Eq. (42) in Ref. [33], we dress the incoming vertex and get

$$\begin{aligned}\sigma_{eh}^{\mu\alpha\beta,3} &= \frac{-e^3}{2\hbar^2\omega_1\omega_2} \tilde{h}_{ab}^{\alpha\beta}(\omega) \int d\omega' G_b(\omega') G_a(\omega' + \omega_{12}) h_{ba}^\mu = \frac{-e^3}{2\hbar^2\omega_1\omega_2} L_{0,ab}^{-1}(\omega_{12}) L_{ab,cd}(\omega_{12}) h_{cd}^{\alpha\beta} L_{0,ab}(\omega_{12}) h_{ba}^\mu \\ &= \frac{-e^3}{2\hbar^2\omega_1\omega_2} h_{cd}^{\alpha\beta} L_{ab,cd}(\omega_{12}) h_{ba}^\mu = \frac{-e^3}{2\hbar^2\omega_1\omega_2} \sum_s \left[ h_{cd}^{\alpha\beta} \frac{Y_{ab\mathbf{k}}^\lambda Y_{cd\mathbf{k}'}^{\lambda*}}{\hbar\omega - \Omega_\lambda + i\eta} h_{ba}^\mu - h_{cd}^{\alpha\beta} \frac{Y_{bak'}^{\lambda*} Y_{dck'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} h_{ba}^\mu \right] \\ &= \frac{-e^3}{2\hbar^2\omega_1\omega_2} \sum_s \left[ \frac{d_\lambda^{\alpha\beta} d_\lambda^{\mu*}}{\hbar\omega - \Omega_\lambda + i\eta} - \frac{d_\lambda^{\alpha\beta*} d_\lambda^\mu}{\hbar\omega + \Omega_\lambda + i\eta} \right].\end{aligned}\quad (D4)$$

The new diagram at the second-order response is the triangle diagram with three vertices. As discussed in the main text, we approximate the interacting three particle correlation function with the e-h ladder diagram and obtain three diagrams, each of which has two dressed vertices and one bare vertex. For the first derived diagram we dress one of the incoming vertices and the outgoing vertex,

$$\begin{aligned}\sigma_{eh}^{\mu\alpha\beta,4.1} &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \tilde{h}_{ba}^\alpha(\omega_1) h_{cb}^\beta \tilde{h}_{ca}^\mu(\omega) I_{abc}(\omega_1, \omega_2) = \frac{-e^3}{\hbar^2\omega_1\omega_2} L_{0,ba}^{-1}(\omega_1) L_{ba,ij}(\omega_1) h_{ij}^\alpha h_{cb}^\beta h_{ns}^{\mu*} L_{ns,ca}(\omega) L_{0,ca}^{-1}(\omega) I_{abc}(\omega_1, \omega_2) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \frac{\omega_1 - \epsilon_{ba}}{f_{ab}} L_{ba,ij}(\omega_1) h_{ij}^\alpha h_{cb}^\beta h_{ns}^{\mu*} L_{ns,ca}(\omega) \frac{\omega - \epsilon_{ca}}{f_{ac}} \frac{f_{ab}(\omega_2 - \epsilon_{cb}) + f_{cb}(\omega_1 - \epsilon_{ba})}{(\omega_1 - \epsilon_{ba})(\omega - \epsilon_{ca})(\omega_2 - \epsilon_{cb})} \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} h_{ij}^\alpha h_{cb}^\beta h_{ns}^{\mu*} \frac{1}{f_{ab}} \frac{1}{f_{ac}} \frac{f_{ab}(\omega_2 - \epsilon_{cb}) + f_{cb}(\omega_1 - \epsilon_{ba})}{(\omega_2 - \epsilon_{cb})} \\ &\quad \times \left( \frac{Y_{bak}^s Y_{ijk'}^{\lambda*}}{(\hbar\omega_1 - \Omega_s + i\eta) \hbar\omega - \Omega_\lambda + i\eta} + \frac{Y_{abk}^{s*} Y_{jik'}}{\hbar\omega_1 + \Omega_s + i\eta} \frac{Y_{snk}^{\lambda*} Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} \right) \\ &= \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} h_{ij}^\alpha h_{cb}^\beta h_{ns}^{\mu*} \left( -\frac{Y_{bak}^s Y_{ijk'}^{\lambda*}}{\hbar\omega_1 - \Omega_s + i\eta} \frac{Y_{nsk}^{\lambda*} Y_{cak'}^\lambda}{\hbar\omega - \Omega_\lambda + i\eta} + \frac{Y_{abk}^{s*} Y_{jik'}}{\hbar\omega_1 + \Omega_s + i\eta} \frac{Y_{snk}^{\lambda*} Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} \right) \\ &= \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} h_{cb}^\beta \left( -\frac{Y_{bak}^s d_s^\alpha}{\hbar\omega_1 - \Omega_s + i\eta} \frac{d_\lambda^{\mu*} Y_{cak'}^{\lambda*}}{\hbar\omega - \Omega_\lambda + i\eta} + \frac{Y_{abk}^{s*} d_s^{\alpha*}}{\hbar\omega_1 + \Omega_s + i\eta} \frac{d_\lambda^\mu Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} \right),\end{aligned}\quad (D5)$$

where the superscript 4.1 indicates that it is the first diagram derived from the triangle diagram, in the third line, we replace the bare three particle correlation function  $I_{abc}$  with Eq. (C17) introduced in Appendix. A, in the sixth line we set the occupations to their equilibrium values, and in the fifth line, we replace the product of the two correlation functions with the following:

$$L_{ba,ij}(\omega_1)L_{ns,ca}(\omega) = \sum_{s\lambda} \left( \frac{Y_{bak}^s Y_{ijk'}^{s*}}{\hbar\omega_1 - \Omega_s + i\eta} \frac{Y_{nsk}^\lambda Y_{cak'}^{\lambda*}}{\hbar\omega - \Omega_\lambda + i\eta} + \frac{Y_{abk}^{s*} Y_{jik'}^s}{\hbar\omega_1 + \Omega_s + i\eta} \frac{Y_{snk}^{\lambda*} Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} \right). \quad (D6)$$

For the second derived diagram, we dressed the other incoming vertex and the outgoing vertex,

$$\begin{aligned} \sigma_{eh}^{\mu\alpha\beta,4.2} &= \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{ba}^\alpha \tilde{h}_{cb}^\beta(\omega_2) \tilde{h}_{ca}^\mu(\omega) I_{abc}(\omega_1, \omega_2) = \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{ba}^\alpha L_{0,cb}^{-1}(\omega_2) L_{cb,lm}(\omega_2) h_{lm}^\beta h_{ns}^{\mu*} L_{ns,ca}(\omega) L_{0,ca}^{-1}(\omega) I_{abc}(\omega_1, \omega_2) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} h_{ba}^\alpha \frac{\omega_2 - \epsilon_{cb}}{f_{bc}} h_{lm}^\beta h_{ns}^{\mu*} \frac{\omega - \epsilon_{ca}}{f_{ac}} \frac{f_{ab}(\omega_2 - \epsilon_{cb}) + f_{cb}(\omega_1 - \epsilon_{ba})}{(\omega_1 - \epsilon_{ba})(\omega - \epsilon_{ca})(\omega_2 - \epsilon_{cb})} L_{cb,lm}(\omega_2) L_{ns,ca}(\omega) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} h_{ba}^\alpha h_{lm}^\beta h_{ns}^{\mu*} \frac{1}{f_{bc}} \frac{1}{f_{ac}} \frac{f_{ab}(\omega_2 - \epsilon_{cb}) + f_{cb}(\omega_1 - \epsilon_{ba})}{(\omega_1 - \epsilon_{ba})} \\ &\quad \times \left( \frac{Y_{cbk}^\lambda Y_{lmk'}^{\lambda*}}{\hbar\omega_2 - \Omega_\lambda + i\eta} \frac{Y_{nsk}^s Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} + \frac{Y_{bck}^{\lambda*} Y_{mlk'}^\lambda}{\hbar\omega_2 + \Omega_\lambda + i\eta} \frac{Y_{snk}^{s*} Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \right) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} h_{ba}^\alpha h_{lm}^\beta h_{ns}^{\mu*} \left( -\frac{Y_{cbk}^\lambda Y_{lmk'}^{\lambda*}}{\hbar\omega_2 - \Omega_\lambda + i\eta} \frac{Y_{nsk}^s Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} + \frac{Y_{bck}^{\lambda*} Y_{mlk'}^\lambda}{\hbar\omega_2 + \Omega_\lambda + i\eta} \frac{Y_{snk}^{s*} Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \right) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} h_{ba}^\alpha \left( -\frac{Y_{cbk}^\lambda d_\lambda^\beta}{\hbar\omega_2 - \Omega_\lambda + i\eta} \frac{d_s^{\mu*} Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} + \frac{Y_{bck}^{\lambda*} d_\lambda^{\beta*}}{\hbar\omega_2 + \Omega_\lambda + i\eta} \frac{d_s^\mu Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \right). \quad (D7) \end{aligned}$$

In the above, we replace the product of the correlation functions with

$$L_{cb,lm}(\omega_2)L_{ns,ca}(\omega) = \sum_{s\lambda} \left( \frac{Y_{cbk}^\lambda Y_{lmk'}^{\lambda*}}{\hbar\omega_2 - \Omega_\lambda + i\eta} \frac{Y_{nsk}^s Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} + \frac{Y_{bck}^{\lambda*} Y_{mlk'}^\lambda}{\hbar\omega_2 + \Omega_\lambda + i\eta} \frac{Y_{snk}^{s*} Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \right). \quad (D8)$$

For the third derived diagram we dress both incoming vertices,

$$\begin{aligned} \sigma_{eh}^{\mu\alpha\beta,4.3} &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \tilde{h}_{ba}^\alpha(\omega_1) \tilde{h}_{cb}^\beta(\omega_2) h_{ca}^\mu I_{abc}(\omega_1, \omega_2) = \frac{-e^3}{\hbar^2\omega_1\omega_2} L_{0,ba}^{-1}(\omega_1) L_{ba,ij}(\omega_1) h_{ij}^\alpha L_{0,cb}^{-1}(\omega_2) L_{cb,lm}(\omega_2) h_{lm}^\beta h_{ac}^\mu I_{abc}(\omega_1, \omega_2) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \frac{\omega_1 - \epsilon_{ba}}{f_{ab}} L_{ba,ij}(\omega_1) h_{ij}^\alpha h_{lm}^\beta L_{cb,lm}(\omega_2) \frac{\omega_2 - \epsilon_{cb}}{f_{bc}} h_{ac}^\mu \frac{f_{ab}(\omega_2 - \epsilon_{cb}) + f_{cb}(\omega_1 - \epsilon_{ba})}{(\omega_1 - \epsilon_{ba})(\omega - \epsilon_{ca})(\omega_2 - \epsilon_{cb})} \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{ca}^{\mu*} h_{ij}^\alpha h_{lm}^\beta \frac{f_{ab}(\omega_2 - \epsilon_{cb}) + f_{cb}(\omega_1 - \epsilon_{ba})}{(\omega - \epsilon_{ca}) f_{ab} f_{bc}} \\ &\quad \times \sum_{s\lambda} \left( -\frac{Y_{bak}^s Y_{ijk'}^{s*}}{\hbar\omega_1 - \Omega_s + i\eta} \frac{Y_{bck}^{\lambda*} Y_{mlk'}^\lambda}{\hbar\omega_2 + \Omega_\lambda + i\eta} - \frac{Y_{abk}^{s*} Y_{jik'}^s}{\hbar\omega_1 + \Omega_s + i\eta} \frac{Y_{cbk}^\lambda Y_{lmk'}^{\lambda*}}{\hbar\omega_2 - \Omega_\lambda + i\eta} \right) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} h_{ca}^{\mu*} h_{ij}^\alpha h_{lm}^\beta \sum_{s\lambda} \left( \frac{Y_{bak}^s Y_{ijk'}^{s*}}{\hbar\omega_1 - \Omega_s + i\eta} \frac{Y_{bck}^{\lambda*} Y_{mlk'}^\lambda}{\hbar\omega_2 + \Omega_\lambda + i\eta} - \frac{Y_{abk}^{s*} Y_{jik'}^s}{\hbar\omega_1 + \Omega_s + i\eta} \frac{Y_{cbk}^\lambda Y_{lmk'}^{\lambda*}}{\hbar\omega_2 - \Omega_\lambda + i\eta} \right) \\ &= \frac{-e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} \left[ \frac{h_{ca}^{\mu*} Y_{bak}^s Y_{bck}^{\lambda*} d_s^\alpha d_\lambda^{\beta*}}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega_2 + \Omega_\lambda + i\eta)} - \frac{h_{ca}^{\mu*} Y_{abk}^{s*} Y_{cbk}^\lambda d_s^{\alpha*} d_\lambda^\beta}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega_2 - \Omega_\lambda + i\eta)} \right]. \quad (D9) \end{aligned}$$

In the fourth line we replace the product of two correlation functions using

$$L_{ba,ij}(\omega_1)L_{cb,lm}(\omega_2) = \sum_{s\lambda} \left( -\frac{Y_{bak}^s Y_{ijk'}^{s*}}{\hbar\omega_1 - \Omega_s + i\eta} \frac{Y_{bck}^{\lambda*} Y_{mlk'}^\lambda}{\hbar\omega_2 + \Omega_\lambda + i\eta} - \frac{Y_{abk}^{s*} Y_{jik'}^s}{\hbar\omega_1 + \Omega_s + i\eta} \frac{Y_{cbk}^\lambda Y_{lmk'}^{\lambda*}}{\hbar\omega_2 - \Omega_\lambda + i\eta} \right). \quad (D10)$$

Combining all three derived diagrams and explicitly symmetrizing them by adding terms with exchanged indices and frequencies,  $\alpha \leftrightarrow \beta$ ,  $\omega_1 \leftrightarrow \omega_2$ , we get

$$\begin{aligned}
 \sigma_{eh}^{\mu\alpha\beta,4} = & \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} \left( -\frac{h_{cb}^\beta Y_{bak}^s d_s^\alpha}{\hbar\omega_1 - \Omega_s + i\eta} \frac{d_\lambda^{\mu*} Y_{cak'}^{\lambda*}}{\hbar\omega - \Omega_\lambda + i\eta} + \frac{h_{cb}^\beta Y_{abk}^{s*} d_s^{\alpha*}}{\hbar\omega_1 + \Omega_s + i\eta} \frac{d_\lambda^\mu Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} \right) \\
 & + \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} \left( -\frac{h_{cb}^\alpha Y_{bak}^s d_s^\beta}{\hbar\omega_2 - \Omega_s + i\eta} \frac{d_\lambda^{\mu*} Y_{cak'}^{\lambda*}}{\hbar\omega - \Omega_\lambda + i\eta} + \frac{h_{cb}^\alpha Y_{abk}^{s*} d_s^{\beta*}}{\hbar\omega_2 + \Omega_s + i\eta} \frac{d_\lambda^\mu Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} \right) \\
 & - \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} \left( -\frac{h_{ba}^\alpha Y_{cbk}^\lambda d_s^\beta}{\hbar\omega_2 - \Omega_\lambda + i\eta} \frac{d_s^{\mu*} Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} + \frac{h_{ba}^\alpha Y_{bck}^{\lambda*} d_s^{\beta*}}{\hbar\omega_2 + \Omega_\lambda + i\eta} \frac{d_s^\mu Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \right) \\
 & - \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} \left( -\frac{h_{ba}^\beta Y_{cbk}^\lambda d_s^\alpha}{\hbar\omega_1 - \Omega_\lambda + i\eta} \frac{d_s^{\mu*} Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} + \frac{h_{ba}^\beta Y_{bck}^{\lambda*} d_s^{\alpha*}}{\hbar\omega_1 + \Omega_\lambda + i\eta} \frac{d_s^\mu Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \right) \\
 & - \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} \left[ \frac{h_{ca}^{\mu*} Y_{bak}^s Y_{bck}^{\lambda*} d_s^\alpha d_\lambda^{\beta*}}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega_2 + \Omega_\lambda + i\eta)} - \frac{h_{ca}^{\mu*} Y_{abk}^{s*} Y_{cbk}^\lambda d_s^{\alpha*} d_\lambda^\beta}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega_2 - \Omega_\lambda + i\eta)} \right] \\
 & - \frac{e^3}{\hbar^2\omega_1\omega_2} \sum_{s\lambda} \left[ \frac{h_{ca}^{\mu*} Y_{bak}^s Y_{bck}^{\lambda*} d_s^\beta d_\lambda^{\alpha*}}{(\hbar\omega_2 - \Omega_s + i\eta)(\hbar\omega_1 + \Omega_\lambda + i\eta)} - \frac{h_{ca}^{\mu*} Y_{abk}^{s*} Y_{cbk}^\lambda d_s^{\beta*} d_\lambda^\alpha}{(\hbar\omega_2 + \Omega_s + i\eta)(\hbar\omega_1 - \Omega_\lambda + i\eta)} \right]. \tag{D11}
 \end{aligned}$$

Grouping terms with the same denominator, we have

$$\begin{aligned}
 \sigma_{eh}^{\mu\alpha\beta,4}/C = & - \sum_{s\lambda} \frac{h_{cb}^\beta Y_{bak}^s d_s^\alpha}{\hbar\omega_1 - \Omega_s + i\eta} \frac{d_\lambda^{\mu*} Y_{cak'}^{\lambda*}}{\hbar\omega - \Omega_\lambda + i\eta} + \sum_{s\lambda} \frac{h_{ba}^\beta Y_{cbk}^\lambda d_s^\alpha}{\hbar\omega_1 - \Omega_\lambda + i\eta} \frac{d_s^{\mu*} Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} \\
 & + \sum_{s\lambda} \frac{h_{cb}^\beta Y_{abk}^{s*} d_s^{\alpha*}}{\hbar\omega_1 + \Omega_s + i\eta} \frac{d_\lambda^\mu Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} - \sum_{s\lambda} \frac{h_{ba}^\beta Y_{bck}^{\lambda*} d_s^{\alpha*}}{\hbar\omega_1 + \Omega_\lambda + i\eta} \frac{d_s^\mu Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \\
 & - \sum_{s\lambda} \frac{h_{cb}^\alpha Y_{bak}^s d_s^\beta}{\hbar\omega_2 - \Omega_s + i\eta} \frac{d_\lambda^{\mu*} Y_{cak'}^{\lambda*}}{\hbar\omega - \Omega_\lambda + i\eta} + \sum_{s\lambda} \frac{h_{ba}^\alpha Y_{cbk}^\lambda d_s^\beta}{\hbar\omega_2 - \Omega_\lambda + i\eta} \frac{d_s^{\mu*} Y_{cak'}^{s*}}{\hbar\omega - \Omega_s + i\eta} \\
 & + \sum_{s\lambda} \frac{h_{cb}^\alpha Y_{abk}^{s*} d_s^{\beta*}}{\hbar\omega_2 + \Omega_s + i\eta} \frac{d_\lambda^\mu Y_{ack'}^\lambda}{\hbar\omega + \Omega_\lambda + i\eta} - \sum_{s\lambda} \frac{h_{ba}^\alpha Y_{bck}^{\lambda*} d_s^{\beta*}}{\hbar\omega_2 + \Omega_\lambda + i\eta} \frac{d_s^\mu Y_{ack'}^s}{\hbar\omega + \Omega_s + i\eta} \\
 & - \sum_{s\lambda} \frac{h_{ca}^{\mu*} Y_{bak}^s Y_{bck}^{\lambda*} d_s^\alpha d_\lambda^{\beta*}}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega_2 + \Omega_\lambda + i\eta)} + \sum_{s\lambda} \frac{h_{ca}^{\mu*} Y_{abk}^{s*} Y_{cbk}^\lambda d_s^{\alpha*} d_\lambda^\beta}{(\hbar\omega_2 + \Omega_s + i\eta)(\hbar\omega_1 - \Omega_\lambda + i\eta)} \\
 & + \sum_{s\lambda} \frac{h_{ca}^{\mu*} Y_{abk}^{s*} Y_{cbk}^\lambda d_s^{\beta*} d_\lambda^\alpha}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega_2 - \Omega_\lambda + i\eta)} - \sum_{s\lambda} \frac{h_{ca}^{\mu*} Y_{bak}^s Y_{bck}^{\lambda*} d_s^\beta d_\lambda^{\alpha*}}{(\hbar\omega_2 - \Omega_s + i\eta)(\hbar\omega_1 + \Omega_\lambda + i\eta)}, \tag{D12}
 \end{aligned}$$

where  $C = \frac{e^3}{\hbar^2\omega_1\omega_2}$ . We note that both  $\lambda$  and  $s$  are dummy indices so we can redefine them in the summation,

$$\begin{aligned}
 \sigma_{eh}^{\mu\alpha\beta,4}/C = & - \sum_{s\lambda} \frac{d_s^\alpha d_\lambda^{\mu*} (h_{cb}^\beta Y_{bak}^s Y_{cak'}^{\lambda*} - h_{ba}^\beta Y_{cbk}^\lambda Y_{cak'}^{\lambda*})}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega - \Omega_\lambda + i\eta)} + \sum_{s\lambda} \frac{d_s^{\alpha*} d_\lambda^\mu (h_{cb}^\beta Y_{abk}^{s*} Y_{ack'}^\lambda - h_{ba}^\beta Y_{bck}^{\lambda*} Y_{ack'}^\lambda)}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega + \Omega_\lambda + i\eta)} \\
 & - \sum_{s\lambda} \frac{d_s^\beta d_\lambda^{\mu*} (h_{cb}^\alpha Y_{bak}^s Y_{cak'}^{\lambda*} - h_{ba}^\alpha Y_{cbk}^\lambda Y_{cak'}^{\lambda*})}{(\hbar\omega_2 - \Omega_s + i\eta)(\hbar\omega - \Omega_\lambda + i\eta)} + \sum_{s\lambda} \frac{d_s^{\beta*} d_\lambda^\mu (h_{cb}^\alpha Y_{abk}^{s*} Y_{ack'}^\lambda - h_{ba}^\alpha Y_{bck}^{\lambda*} Y_{ack'}^\lambda)}{(\hbar\omega_2 + \Omega_s + i\eta)(\hbar\omega + \Omega_\lambda + i\eta)} \\
 & - \sum_{s\lambda} \frac{(h_{ca}^{\mu*} Y_{bak}^s Y_{bck}^{\lambda*} - h_{ca}^{\mu*} Y_{abk}^{s*} Y_{cbk}^\lambda) d_s^\alpha d_\lambda^{\beta*}}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega_2 + \Omega_\lambda + i\eta)} + \sum_{s\lambda} \frac{d_s^{\alpha*} d_\lambda^\beta (h_{ca}^{\mu*} Y_{abk}^{s*} Y_{cbk}^\lambda - h_{ca}^{\mu*} Y_{bak}^s Y_{bck}^{\lambda*})}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega_2 - \Omega_\lambda + i\eta)} \\
 = & - \sum_{s\lambda} \frac{d_s^\alpha d_\lambda^{\mu*} \Pi_{\lambda s}^\beta}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega - \Omega_\lambda + i\eta)} - \sum_{s\lambda} \frac{d_s^{\alpha*} d_\lambda^\mu \Pi_{s\lambda}^\beta}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega + \Omega_\lambda + i\eta)} \\
 & - \sum_{s\lambda} \frac{d_s^\beta d_\lambda^{\mu*} \Pi_{\lambda s}^\alpha}{(\hbar\omega_2 - \Omega_s + i\eta)(\hbar\omega - \Omega_\lambda + i\eta)} - \sum_{s\lambda} \frac{d_s^{\beta*} d_\lambda^\mu \Pi_{s\lambda}^\alpha}{(\hbar\omega_2 + \Omega_s + i\eta)(\hbar\omega + \Omega_\lambda + i\eta)} \\
 & + \sum_{s\lambda} \frac{\Pi_{s\lambda}^{\mu*} d_s^\alpha d_\lambda^{\beta*}}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega_2 + \Omega_\lambda + i\eta)} + \sum_{s\lambda} \frac{d_s^{\alpha*} d_\lambda^\beta \Pi_{\lambda s}^{\mu*}}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega_2 - \Omega_\lambda + i\eta)}, \tag{D13}
 \end{aligned}$$

where we use Eq. (37) for interexciton couplings in the last three lines. It is easy to check that  $\Pi_{\lambda s}^\beta = \Pi_{s\lambda}^{\beta*}$ .

Finally, combining all three terms and symmetrized  $\sigma_{eh}^{\mu\alpha\beta,2}$  and  $\sigma_{eh}^{\mu\alpha\beta,3}$ , we have

$$\begin{aligned}
& \sigma_{eh}^{\mu\alpha\beta}(\omega; \omega_1, \omega_2) + \sigma_{eh}^{\mu\beta\alpha}(\omega; \omega_2, \omega_1) \\
&= -\frac{e^3}{\hbar^2 \omega_1 \omega_2} \sum_{\lambda} \left[ \left( \frac{d_{\lambda}^{\alpha} d_{\lambda}^{\mu\beta*}}{\hbar\omega_1 - \Omega_{\lambda} + i\eta} - \frac{d_{\lambda}^{\alpha*} d_{\lambda}^{\mu\beta}}{\hbar\omega_1 + \Omega_{\lambda} + i\eta} \right) + \left( \frac{d_{\lambda}^{\beta} d_{\lambda}^{\mu\alpha*}}{\hbar\omega_2 - \Omega_{\lambda} + i\eta} - \frac{d_{\lambda}^{\beta*} d_{\lambda}^{\mu\alpha}}{\hbar\omega_2 + \Omega_{\lambda} + i\eta} \right) \right] \\
&\quad - \frac{e^3}{2\hbar^2 \omega_1 \omega_2} \sum_{\lambda} \left[ \left( \frac{d_{\lambda}^{\alpha\beta} d_{\lambda}^{\mu*}}{\hbar\omega - \Omega_{\lambda} + i\eta} - \frac{d_{\lambda}^{\alpha\beta*} d_{\lambda}^{\mu}}{\hbar\omega + \Omega_{\lambda} + i\eta} \right) + \left( \frac{d_{\lambda}^{\beta\alpha} d_{\lambda}^{\mu*}}{\hbar\omega - \Omega_{\lambda} + i\eta} - \frac{d_{\lambda}^{\beta\alpha*} d_{\lambda}^{\mu}}{\hbar\omega + \Omega_{\lambda} + i\eta} \right) \right] \\
&\quad + \frac{e^3}{\hbar^2 \omega_1 \omega_2} \sum_{s\lambda} \left[ -\frac{d_{\lambda}^{\mu*} \Pi_{\lambda s}^{\beta} d_s^{\alpha}}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega - \Omega_{\lambda} + i\eta)} - \frac{d_{\lambda}^{\mu} \Pi_{s\lambda}^{\beta} d_s^{\alpha*}}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega + \Omega_{\lambda} + i\eta)} \right] \\
&\quad + \frac{e^3}{\hbar^2 \omega_1 \omega_2} \sum_{s\lambda} \left[ -\frac{d_{\lambda}^{\mu*} \Pi_{\lambda s}^{\alpha} d_s^{\beta}}{(\hbar\omega - \Omega_{\lambda} + i\eta)(\hbar\omega_2 - \Omega_s + i\eta)} - \frac{d_{\lambda}^{\mu} \Pi_{s\lambda}^{\alpha} d_s^{\beta*}}{(\hbar\omega + \Omega_{\lambda} + i\eta)(\hbar\omega_2 + \Omega_s + i\eta)} \right] \\
&\quad + \frac{e^3}{\hbar^2 \omega_1 \omega_2} \sum_{s\lambda} \left[ \frac{d_s^{\alpha} \Pi_{s\lambda}^{\mu*} d_{\lambda}^{\beta*}}{(\hbar\omega_1 - \Omega_s + i\eta)(\hbar\omega_2 + \Omega_{\lambda} + i\eta)} + \frac{d_s^{\alpha*} \Pi_{s\lambda}^{\mu} d_{\lambda}^{\beta}}{(\hbar\omega_1 + \Omega_s + i\eta)(\hbar\omega_2 - \Omega_{\lambda} + i\eta)} \right]. \tag{D14}
\end{aligned}$$

Without explicit symmetrization, we can write

$$\begin{aligned}
& \sigma_{eh}^{\mu\alpha\beta}(\omega; \omega_1, \omega_2) \\
&= \frac{-e^3}{\hbar^2 \omega_1 \omega_2} \sum_{\lambda} \left[ \frac{d_{\lambda}^{\alpha} d_{\lambda}^{\mu\beta*}}{\hbar\omega_1 - \Omega_{\lambda} + i\eta} - \frac{d_{\lambda}^{\alpha*} d_{\lambda}^{\mu\beta}}{\hbar\omega_1 + \Omega_{\lambda} + i\eta} \right] + \frac{-e^3}{2\hbar^2 \omega_1 \omega_2} \sum_{\lambda} \left[ \frac{d_{\lambda}^{\alpha\beta} d_{\lambda}^{\mu*}}{\hbar\omega - \Omega_{\lambda} + i\eta} - \frac{d_{\lambda}^{\alpha\beta*} d_{\lambda}^{\mu}}{\hbar\omega + \Omega_{\lambda} + i\eta} \right] \\
&\quad + \frac{e^3}{\hbar^2 \omega_1 \omega_2} \sum_{s\lambda} \left[ -\frac{d_{\lambda}^{\mu*} \Pi_{\lambda s}^{\alpha} d_s^{\beta}}{(\hbar\omega - \Omega_{\lambda} + i\eta)(\hbar\omega_2 - \Omega_s + i\eta)} - \frac{d_{\lambda}^{\mu} \Pi_{s\lambda}^{\alpha} d_s^{\beta*}}{(\hbar\omega + \Omega_{\lambda} + i\eta)(\hbar\omega_2 + \Omega_s + i\eta)} \right] \\
&\quad + \frac{e^3}{\hbar^2 \omega_1 \omega_2} \sum_{s\lambda} \frac{d_s^{\alpha} \Pi_{s\lambda}^{\mu*} d_{\lambda}^{\beta*}}{(\hbar\omega_2 + \Omega_{\lambda} + i\eta)(\hbar\omega_1 - \Omega_s + i\eta)}, \tag{D15}
\end{aligned}$$

which is the part explicitly shown in Eq. (36) in the main text.

- 
- [1] E. Blount, Formalisms of band theory, *Solid State Phys.* **13**, 305 (1962).
- [2] R. D. King-Smith and D. Vanderbilt, Theory of polarization of crystalline solids, *Phys. Rev. B* **47**, 1651 (1993).
- [3] R. Resta, Macroscopic polarization in crystalline dielectrics: The geometric phase approach, *Rev. Mod. Phys.* **66**, 899 (1994).
- [4] J. E. Sipe and A. I. Shkrebtii, Second-order optical response in semiconductors, *Phys. Rev. B* **61**, 5337 (2000).
- [5] G. B. Ventura, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, Gauge covariances and nonlinear optical responses, *Phys. Rev. B* **96**, 035431 (2017).
- [6] A. Taghizadeh, F. Hipolito, and T. G. Pedersen, Linear and nonlinear optical response of crystals using length and velocity gauges: Effect of basis truncation, *Phys. Rev. B* **96**, 195413 (2017).
- [7] R. W. Boyd, *Nonlinear Optics* (Elsevier, Amsterdam, 2008).
- [8] A. Taghizadeh and T. G. Pedersen, Nonlinear excitonic spin Hall effect in monolayer transition metal dichalcogenides, *2D Mater.* **7**, 015003 (2019).
- [9] H. Xu, H. Wang, J. Zhou, and J. Li, Pure spin photocurrent in non-centrosymmetric crystals: Bulk spin photovoltaic effect, *Nat. Commun.* **12**, 4330 (2021).
- [10] Y. Onishi and L. Fu, Quantum geometry, optical absorption and topological gap bound, *Phys. Rev. X* **14**, 011052 (2024).
- [11] T. Holder, D. Kaplan, and B. Yan, Consequences of time-reversal-symmetry breaking in the light-matter interaction: Berry curvature, quantum metric, and diabatic motion, *Phys. Rev. Res.* **2**, 033100 (2020).
- [12] J. Ahn, G.-Y. Guo, N. Nagaosa, and A. Vishwanath, Riemannian geometry of resonant optical responses, *Nat. Phys.* **18**, 290 (2022).
- [13] T. Morimoto and N. Nagaosa, Topological nature of nonlinear optical effects in solids, *Sci. Adv.* **2**, e1501524 (2016).
- [14] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Optical spectrum of MoS<sub>2</sub>: Many-body effects and diversity of exciton states, *Phys. Rev. Lett.* **111**, 216805 (2013).
- [15] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. Li, B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, Exciton binding energy and nonhydrogenic Rydberg series in monolayer WS<sub>2</sub>, *Phys. Rev. Lett.* **113**, 076802 (2014).

- [16] G. Wang, A. Chernikov, M. M. Glazov, T. F. Heinz, X. Marie, T. Amand, and B. Urbaszek, Colloquium: Excitons in atomically thin transition metal dichalcogenides, *Rev. Mod. Phys.* **90**, 021001 (2018).
- [17] T. Smoleński, P. E. Dolgirev, C. Kuhlenkamp, A. Popert, Y. Shimazaki, P. Back, X. Lu, M. Kroner, K. Watanabe, T. Taniguchi *et al.*, Signatures of Wigner crystal of electrons in a monolayer semiconductor, *Nature (London)* **595**, 53 (2021).
- [18] Y. Xu, K. Kang, K. Watanabe, T. Taniguchi, K. F. Mak, and J. Shan, A tunable bilayer Hubbard model in twisted WSe<sub>2</sub>, *Nat. Nanotechnol.* **17**, 934 (2022).
- [19] J. Cai, E. Anderson, C. Wang, X. Zhang, X. Liu, W. Holtzmann, Y. Zhang, F. Fan, T. Taniguchi, K. Watanabe, Y. Ran *et al.*, Signatures of fractional quantum anomalous Hall states in twisted MoTe<sub>2</sub>, *Nature (London)* **622**, 63 (2023).
- [20] L. M. Malard, T. V. Alencar, A. P. M. Barboza, K. F. Mak, and A. M. de Paula, Observation of intense second harmonic generation from MoS<sub>2</sub> atomic crystals, *Phys. Rev. B* **87**, 201401(R) (2013).
- [21] G. Wang, X. Marie, I. Gerber, T. Amand, D. Lagarde, L. Bouet, M. Vidal, A. Balocchi, and B. Urbaszek, Giant enhancement of the optical second-harmonic emission of WSe<sub>2</sub> monolayers by laser excitation at exciton resonances, *Phys. Rev. Lett.* **114**, 097403 (2015).
- [22] L. Lafeta, A. Corradi, T. Zhang, E. Kahn, I. Bilgin, B. R. Carvalho, S. Kar, M. Terrones, and L. M. Malard, Second- and third-order optical susceptibilities across excitons states in 2D monolayer transition metal dichalcogenides, *2D Mater.* **8**, 035010 (2021).
- [23] E. K. Chang, E. L. Shirley, and Z. H. Levine, Excitonic effects on optical second-harmonic polarizabilities of semiconductors, *Phys. Rev. B* **65**, 035205 (2001).
- [24] R. Leitsmann, W. G. Schmidt, P. H. Hahn, and F. Bechstedt, Second-harmonic polarizability including electron-hole attraction from band-structure theory, *Phys. Rev. B* **71**, 195209 (2005).
- [25] E. Luppi and V. Vénard, A review of recent theoretical studies in nonlinear crystals: Towards the design of new materials, *Semicond. Sci. Technol.* **31**, 123002 (2016).
- [26] C. Attacalite and M. Grüning, Nonlinear optics from an *ab initio* approach by means of the dynamical Berry phase: Application to second- and third-harmonic generation in semiconductors, *Phys. Rev. B* **88**, 235113 (2013).
- [27] M. Grüning and C. Attacalite, Second harmonic generation in *h*-BN and MoS<sub>2</sub> monolayers: Role of electron-hole interaction, *Phys. Rev. B* **89**, 081102(R) (2014).
- [28] Y.-H. Chan, D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Giant exciton-enhanced shift currents and direct current conduction with subbandgap photo excitations produced by many-electron interactions, *Proc. Natl. Acad. Sci. USA* **118**, e1906938118 (2021).
- [29] Y.-S. Huang, Y.-H. Chan, and G.-Y. Guo, Large shift currents via in-gap and charge-neutral excitons in a monolayer and nanotubes of BN, *Phys. Rev. B* **108**, 075413 (2023).
- [30] F. Xuan, M. Lai, Y. Wu, and S. Y. Quek, Exciton-enhanced spontaneous parametric down-conversion in 2D crystals, [arXiv:2305.08345](https://arxiv.org/abs/2305.08345).
- [31] J. Ruan, Y. hao Chan, and S. G. Louie, Excitonic effects in nonlinear optical responses: Exciton-state formalism and first-principles calculations, [arXiv:2310.09674](https://arxiv.org/abs/2310.09674).
- [32] A. Taghizadeh and T. G. Pedersen, Gauge invariance of excitonic linear and nonlinear optical response, *Phys. Rev. B* **97**, 205432 (2018).
- [33] D. E. Parker, T. Morimoto, J. Orenstein, and J. E. Moore, Diagrammatic approach to nonlinear optical response with application to Weyl semimetals, *Phys. Rev. B* **99**, 045121 (2019).
- [34] S. Gassner and E. J. Mele, Regularized lattice theory for spatially dispersive nonlinear optical conductivities, *Phys. Rev. B* **108**, 085403 (2023).
- [35] S. Reichardt and L. Wirtz, Nonadiabatic exciton-phonon coupling in Raman spectroscopy of layered materials, *Sci. Adv.* **6**, eabb5915 (2020).
- [36] G. Antonius and S. G. Louie, Theory of exciton-phonon coupling, *Phys. Rev. B* **105**, 085111 (2022).
- [37] H.-Y. Chen, D. Sangalli, and M. Bernardi, Exciton-phonon interaction and relaxation times from first principles, *Phys. Rev. Lett.* **125**, 107401 (2020).
- [38] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, Reading, PA, 1995).
- [39] P. Coleman, *Introduction to Many-Body Physics* (Cambridge University Press, Cambridge, 2015).
- [40] A. Altland and B. D. Simons, *Condensed Matter Field Theory*, 2nd ed. (Cambridge University Press, Cambridge, 2010).
- [41] D. J. Passos, G. B. Ventura, J. M. Viana Parente Lopes, J. M. B. Lopes dos Santos, and N. M. R. Peres, Nonlinear optical responses of crystalline systems: Results from a velocity gauge analysis, *Phys. Rev. B* **97**, 235446 (2018).
- [42] G. D. Mahan, Excitons in degenerate semiconductors, *Phys. Rev.* **153**, 882 (1967).
- [43] G. Strinati, Application of the Green's functions method to the study of the optical properties of semiconductors, *Riv. Nuovo Cim.* **11**, 1 (1988).
- [44] G. Stefanucci and R. van Leeuwen, *Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction* (Cambridge University Press, Cambridge, 2013).
- [45] M. Rohlfing and S. G. Louie, Electron-hole excitations and optical spectra from first principles, *Phys. Rev. B* **62**, 4927 (2000).
- [46] G. Onida, L. Reining, and A. Rubio, Electronic excitations: Density-functional versus many-body Green's-function approaches, *Rev. Mod. Phys.* **74**, 601 (2002).
- [47] G. D. Mahan, *Many-Particle Physics* (Springer, New York, 2000).
- [48] C. Aversa and J. E. Sipe, Nonlinear optical susceptibilities of semiconductors: Results with a length-gauge analysis, *Phys. Rev. B* **52**, 14636 (1995).
- [49] S. Ismail-Beigi, E. K. Chang, and S. G. Louie, Coupling of nonlocal potentials to electromagnetic fields, *Phys. Rev. Lett.* **87**, 087402 (2001).

- [50] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, Berkeleygw: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures, *Comput. Phys. Commun.* **183**, 1269 (2012).
- [51] J. L. Cheng, N. Vermeulen, and J. E. Sipe, Numerical study of the optical nonlinearity of doped and gapped graphene: From weak to strong field excitation, *Phys. Rev. B* **92**, 235307 (2015).
- [52] T. G. Pedersen, Intraband effects in excitonic second-harmonic generation, *Phys. Rev. B* **92**, 235432 (2015).
- [53] L. Hedin, New method for calculating the one-particle Green's function with application to the electron-gas problem, *Phys. Rev.* **139**, A796 (1965).
- [54] E. Perfetto, D. Sangalli, A. Marini, and G. Stefanucci, First-principles approach to excitons in time-resolved and angle-resolved photoemission spectra, *Phys. Rev. B* **94**, 245303 (2016).