Nonlinear optical selection rules of excitons in monolayer transition metal dichalcogenides

Taghizadeh, Alireza; Pedersen, Thomas Garm

Published in:
Physical Review B

DOI (link to publication from Publisher):
10.1103/PhysRevB.99.235433

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Nonlinear optical selection rules of excitons in monolayer transition metal dichalcogenides

Alireza Taghizadeh* and T. G. Pedersen
Department of Physics and Nanotechnology, Aalborg University, DK-9220 Aalborg Øst, Denmark
and Center for Nanostructured Graphene (CNG), DK-9220 Aalborg Øst, Denmark

(Received 27 February 2019; revised manuscript received 19 June 2019; published 28 June 2019)

Monolayer transition metal dichalcogenides (TMDs) are characterized by strong light-matter interactions due to enhanced excitonic effects, which make them exciting materials for fundamental physics and optoelectronics applications. Moreover, the valley-dependent chirality of the band structure in TMDs significantly modifies the optical selection rules for single- and multiphoton processes. Here, we propose an analytical approach for calculating the linear and nonlinear optical (NLO) responses of monolayer TMDs, including excitonic effects at low photon energies. Based on this approach, we provide an informative diagram which encompasses all excitonic selection rules. The diagram enables us to identify main transitions for the first-, second-, and third-order optical responses. As a case study, we calculate the optical conductivity and second-/third-harmonic generation responses of monolayer MoS_2 and demonstrate that the analytical approach accurately reproduces the spectra obtained using the Bethe-Salpeter equation (BSE). Moreover, the analytical approach enables us to obtain valuable physical insight into the fundamental transitions responsible for individual resonances, which is not straightforward in the full BSE method. Our analytical approach can readily be extended to higher-order nonlinearities and, hence, provides a simple but accurate tool for analyzing experimental NLO spectra of monolayer TMDs.

DOI: 10.1103/PhysRevB.99.235433

I. INTRODUCTION

The nonlinear optical (NLO) response of materials includes various multiphoton processes such as second-/third-harmonic generation (SHG/THG), optical rectification, optical Kerr, sum-/difference-frequency generation, etc., which find numerous important applications in optical communication, biomolecular detection, and surface characterization [1,2]. Recently, the interest in the NLO response of two-dimensional (2D) materials has grown drastically due to their large nonlinear coefficients, which facilitate a vast range of applications in optoelectronic devices [3,4] as well as fundamental physics, e.g., nonlinear charge and spin Hall effects [5–7]. The NLO responses of semimetallic graphene have been extensively studied both theoretically and experimentally [8–14]. In centrosymmetric graphene, even-order NLO processes vanish in the dipole approximation [9]. In contrast, the intrinsic broken inversion symmetry in monolayer transition metal dichalcogenides (TMDs) leads to nonvanishing even-order NLO responses. Furthermore, strong excitonic effects due to reduced screening as well as huge spin-orbit coupling (SOC) in monolayer TMDs make their optical and electronic properties unique [15].

The linear optical response of monolayer TMDs has been extensively studied theoretically using the Bethe-Salpeter equation (BSE) in combination with quasiparticle band structures obtained by *ab initio* techniques [16–18] or tight-binding (TB) methods [19–21]. Besides these rigorous calculations, analytical approaches have also been developed by employing the so-called massive Dirac Hamiltonian to obtain physical insight [19,20,22,23]. Employing the analytical approach not only reduces the computational burden of excitonic optical response calculations but also leads to the identification of excitonic selection rules for the linear optical response in monolayer TMDs [24,25]. Optical selection rules are fundamental principles which determine allowed and forbidden transitions mediated by light. These rules are imposed by various symmetries of the system such as time-reversal, spatial, or rotational symmetries. For instance, the brightness of s and darkness of p excitons in the linear optical spectrum of monolayer TMDs have been explained by selection rules [24,25]. Note that although excitons in monolayer TMDs are of Wannier-Mott type, which are also found in conventional semiconductors, they possess different selection rules due to the chirality of the band structure at the band gap [25]. Regarding the NLO response, a few calculations for monolayer TMDs can be found in which the BSE solution is employed for computing the second-harmonic susceptibilities [26–28]. While the linear optical response originates only from the transition between the ground state and excited states, the NLO responses depend also on the transition between excited states [29], i.e., interexcitonic transitions. Therefore, more complicated selection rules are expected to govern NLO processes. The NLO selection rules have been studied qualitatively in Ref. [30] and quantitatively using a Wannier model (with $1/r$ Coulomb potential) in a recent work [31]. Nonetheless, the qualitative work cannot provide any estimation for the excitonic matrix elements, while the Wannier model fails to accurately capture the nonlinear excitonic spectra due to the missing Bloch band curvature [25]. Therefore, a systematic study of excitonic selection rules and their effects on the NLO response

*ata@nano.aau.dk

including the Bloch band signature monolayer TMDs is missing.

In this paper, we develop an analytical approach for systematic study of the NLO response in monolayer TMDs. By using an effective Hamiltonian, which includes the trigonal warping (TW) term, allowed and forbidden transitions between excitonic states are identified in the dipole approximation. A diagrammatic method is then suggested as a means for predicting the dominant transitions and their origin for NLO processes up to any required order. We apply the proposed analytical approach to monolayer MoS_2 as a case study and confirm the validity and accuracy of the analytical approach by comparing with the full BSE results. Our approach can be safely employed for computing the NLO response of other monolayer TMDs due to similar physics. This paper is organized as follows. We begin by reviewing the theoretical framework for the NLO response in monolayer TMDs using the effective TW Hamiltonian. Based on the developed framework, optical selection rules for the first-, second-, and third-order processes are discussed in Sec. III. In Sec. IV, the linear and NLO conductivities computed for monolayer MoS_2 are reported. Finally, a summary of the main findings is provided in Sec. V. A set of appendices explain the details of the derived expressions.

II. THEORETICAL FRAMEWORK

In this section, we present the theoretical framework for calculating the NLO response of monolayer TMDs using our analytical approach. First, general expressions for the optical conductivity (OC), SHG, and THG are reviewed. Then, the TW Hamiltonian for monolayer TMDs is introduced. Finally, analytical expressions for the required excitonic matrix elements are presented. It should be noted that throughout the text, all vectors and tensors are indicated by bold letters, and the single-particle/many-body operators and matrix elements are denoted by lowercase/uppercase letters, respectively.

A. Excitonic nonlinear optical response

The linear and NLO responses of a material can be characterized by linear and NLO conductivities (or equivalently susceptibilities) [2]. In the present paper, we assume that the material is excited by a normal-incident monochromatic electric field written as $\boldsymbol{\mathcal{E}}(t)=\boldsymbol{\mathcal{E}}_\omega\exp(-i\omega t)+\text{c.c.}$, where $\boldsymbol{\mathcal{E}}_\omega=\mathcal{E}^x_\omega\boldsymbol{e}_x+\mathcal{E}^y_\omega\boldsymbol{e}_y$ is the electric field phasor ($\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ are unit vectors along the x and y directions, respectively). The light-matter interaction is then studied using a density matrix approach, which has been explained in detail in our previous papers, cf. Refs. [32,33]. In this approach, a set of coupled equations for the reduced density matrix $\rho_{ji\mathbf{k}}=\langle 0|\hat{c}^\dagger_{i\mathbf{k}}\hat{c}_{j\mathbf{k}}|0\rangle$ is derived, where $|0\rangle$ denotes the many-body ground state. Furthermore, i and j are band indices, $\mathbf{k}$ is the wave vector, and $\hat{c}^\dagger$ ($\hat{c}$) is the fermionic creation (annihilation) operator. Neglecting exchange terms, the dynamical equations for $\rho_{ij\mathbf{k}}$ for a two-band semiconductor within the mean-field approximation (MFA) are presented in Eqs. (A3a) and (A3b) of Appendix A. These equations are then solved perturbatively up to any required order of the field at vanishing temperature, and the Nth-order density matrix elements $\rho^{(N)}_{ij\mathbf{k}}$ are determined. In the present paper, we only focus on the NLO processes of SHG/THG for simplicity, and the required density matrices are found in Eqs. (A5a)–(A5f).

Upon determining the density matrix, the Nth harmonic current density is evaluated as $\mathbf{J}^{(N)}(t)=\text{Tr}[\hat{\rho}^{(N)}\hat{\mathbf{J}}]$, where the charge current density operator reads $\hat{\mathbf{J}}=-e\hat{\boldsymbol{\Pi}}/(mA)$ (for a crystal area A) [33]. Here, $\hat{\boldsymbol{\Pi}}$ is the Heisenberg momentum defined as $\hbar\hat{\boldsymbol{\Pi}}=im[\hat{\mathcal{H}}_0,\hat{\mathbf{R}}]$, where $\hat{\mathcal{H}}_0$ and $\hat{\mathbf{R}}$ are the total unperturbed Hamiltonian and many-body position operators, respectively [33]. In turn, the Nth harmonic generated current density reads

$$\mathbf{J}^{(N)}(t)=\boldsymbol{\sigma}^{(N)}\underbrace{\boldsymbol{\mathcal{E}}_\omega\ldots\boldsymbol{\mathcal{E}}_\omega}_{N\text{ times}}\exp(-iN\omega t)+\text{c.c.}, \tag{1}$$

where $\boldsymbol{\sigma}^{(N)}$ is the Nth-order conductivity tensor ($N\geqslant 1$). Note that linear and NLO susceptibilities, $\boldsymbol{\chi}^{(N)}$ are related to their corresponding conductivities by $\boldsymbol{\sigma}^{(N)}=-iN\omega\epsilon_0\boldsymbol{\chi}^{(N)}$. The expressions of the first-, second-, and dominant part of the third-order conductivity tensors at vanishing temperature are given by [33]

$$\sigma^{(1)}_{\alpha\beta}=C_1\sum_m\left[\frac{\Pi^\alpha_{0m}X^\beta_{m0}}{\hbar\omega-E_m}-\frac{\Pi^\alpha_{m0}X^\beta_{0m}}{\hbar\omega+E_m}\right], \tag{2a}$$

$$\sigma^{(2)}_{\eta\alpha\beta}=C_2\sum_{n,m}\left[\frac{\Pi^\eta_{0n}Q^\alpha_{nm}X^\beta_{m0}}{(2\hbar\omega-E_n)(\hbar\omega-E_m)}+\frac{\Pi^\eta_{n0}Q^\alpha_{mn}X^\beta_{0m}}{(2\hbar\omega+E_n)(\hbar\omega+E_m)}-\frac{X^\alpha_{0n}\Pi^\eta_{nm}X^\beta_{m0}}{(\hbar\omega+E_n)(\hbar\omega-E_m)}\right], \tag{2b}$$

$$\sigma^{(3)}_{\mu\eta\alpha\beta}=C_3\sum_{l,n,m}\left[\frac{\Pi^\mu_{0l}Q^\eta_{ln}Q^\alpha_{nm}X^\beta_{m0}}{(3\hbar\omega-E_l)(2\hbar\omega-E_n)(\hbar\omega-E_m)}-\frac{\Pi^\mu_{l0}Q^\eta_{nl}Q^\alpha_{mn}X^\beta_{0m}}{(3\hbar\omega+E_l)(2\hbar\omega+E_n)(\hbar\omega+E_m)}\right.$$
$$\left.+\frac{X^\alpha_{0l}X^\beta_{m0}}{(\hbar\omega+E_l)(\hbar\omega-E_m)}\left(\frac{Q^\eta_{ln}\Pi^\mu_{nm}}{2\hbar\omega+E_n}-\frac{\Pi^\mu_{ln}Q^\eta_{nm}}{2\hbar\omega-E_n}\right)\right]. \tag{2c}$$

Here, $C_3\equiv eC_2\equiv e^2C_1\equiv -e^4\hbar/(m^2A)$, $\mu,\eta,\alpha,\beta=\{x,y\}$ are the Cartesian coordinates, E_n is the exciton energy, and the summations run over excited states. In addition, the excitonic matrix elements are defined as $\Pi^\alpha_{0n}=(\Pi^\alpha_{n0})^*\equiv -iE_nX^\alpha_{0n}$,

$\Pi^{\alpha}_{nm} \equiv i(E_n - E_m)Q^{\alpha}_{nm}$, and

$$X^{\alpha}_{0n} = \left(X^{\alpha}_{n0}\right)^* \equiv -\sum_{\mathbf{k}} \psi^{(n)}_{\mathbf{k}} \frac{p^{\alpha}_{vc\mathbf{k}}}{i\varepsilon_{cv\mathbf{k}}}, \tag{3a}$$

$$Q^{\alpha}_{nm} \equiv i\sum_{\mathbf{k}} \psi^{(n)*}_{\mathbf{k}} \left[\psi^{(m)}_{\mathbf{k}}\right]_{;k_\alpha}, \tag{3b}$$

where $\varepsilon_{cv\mathbf{k}} \equiv \varepsilon_{c\mathbf{k}} - \varepsilon_{v\mathbf{k}}$ (indices c and v imply conduction and valence bands, respectively), and $p^{\alpha}_{vc\mathbf{k}} \equiv \langle v\mathbf{k}|\hat{p}_\alpha|c\mathbf{k}\rangle$ are the transition energy and momentum matrix element between single-particle states, respectively. Moreover, $\psi^{(n)}_{\mathbf{k}}$ is the exciton wave function projected onto a band-to-band transition, i.e., the exciton state is expressed as $|\psi^{(n)}\rangle = \sum_{\mathbf{k}} \psi^{(n)}_{\mathbf{k}} \hat{c}^{\dagger}_{c\mathbf{k}} \hat{c}_{v\mathbf{k}}|0\rangle$ [34] (the exciton center-of-mass motion is neglected due to the negligible photon momentum). The summations over $\mathbf{k}$ should be understood as integrals over the Brillouin zone (BZ), i.e., $(2\pi)^2 \sum_{\mathbf{k}} \to A \int_{\mathrm{BZ}} d^2\mathbf{k}$. Finally, $[\psi^{(n)}_{\mathbf{k}}]_{;k_\alpha} \equiv \partial\psi^{(n)}_{\mathbf{k}}/\partial k_\alpha - i(\Omega^{\alpha}_{cc\mathbf{k}} - \Omega^{\alpha}_{vv\mathbf{k}})\psi^{(n)}_{\mathbf{k}}$ denotes the generalized derivative of the exciton wave function with the Berry connections $\Omega^{\alpha}_{nm\mathbf{k}} \equiv \langle n\mathbf{k}|i\partial/\partial k_\alpha|m\mathbf{k}\rangle$. The generalized derivative stems from the intraband part of the position operator [35,36]. Equations (2a) and (2b) have been reported in our previous paper [33] while Eq. (2c) can be obtained in a similar manner by extending the general approach to the third order in the electric field. Although these expressions are obtained for a two-band semiconductor, they can be used for multiband semiconductors by redefining the Q^{α}_{nm} matrix elements (see Ref. [33] for more details). Note that, with a minor modification, these expressions can also be employed for calculating the nonlinear spin conductivities [7].

To determine excitonic energies and wave functions, we solve the BSE [34,37], given by

$$\varepsilon_{cv\mathbf{k}}\psi^{(n)}_{\mathbf{k}} - \sum_{\mathbf{k}'} V^{\mathbf{k},\mathbf{k}'}_{cv}\psi^{(n)}_{\mathbf{k}'} = E_n\psi^{(n)}_{\mathbf{k}}, \tag{4}$$

where $V^{\mathbf{k},\mathbf{k}'}_{cv}$ is the direct Coulomb matrix elements defined in Eq. (A4) of Appendix A. The Coulomb matrix elements include band signatures through the Bloch overlaps, $\langle n\mathbf{k}|n\mathbf{k}'\rangle$. For 2D materials, a screened Coulomb interaction in the Keldysh form is employed, which accurately describes the excitonic spectrum [20,33,38,39]. The dielectric screening of the surrounding environment is included in this model via the ϵ_s parameter in the Coulomb matrix elements, see Eq. (A4). The ϵ_s parameter is given by the average of the substrate and superstrate dielectric constants. The BSE solutions contain both the bound and unbounded excitons.

B. Trigonal warping Hamiltonian for monolayer TMDs

It is well-known that the valence and conduction bands of monolayer TMDs in the vicinity of the Dirac points are mainly formed by $(|d_{x^2-y^2}\rangle + |d_{xy}\rangle)/\sqrt{2}$ and $|d_{z^2}\rangle$ orbitals of the metal atom, respectively [22,40,41]. Hence, we can construct an effective 2×2 TB Hamiltonian [7,19], which reproduces the low-energy band structure of monolayer TMDs. Including the SOC, the TB Hamiltonian in $\mathbf{k}$-space reads [7,42,43]

$$H_0(\mathbf{k}) = \begin{bmatrix} \Delta + \lambda g(\mathbf{k})s & -\gamma f^*(\mathbf{k}) \\ -\gamma f(\mathbf{k}) & -\Delta - \lambda g(\mathbf{k})s \end{bmatrix}, \tag{5}$$

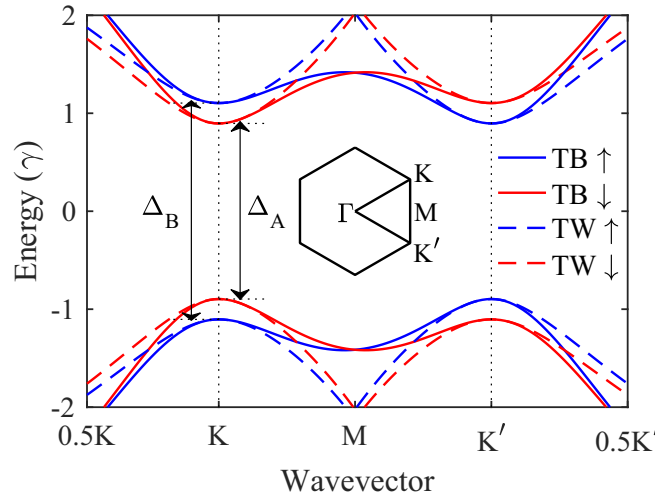

FIG. 1. Typical band structure of monolayer TMD near the band gap obtained using the full TB (solid lines) or TW Hamiltonian (dashed lines) for $\gamma = \Delta = 50\lambda$. Blue and red lines show spin-up ($s = 1$) and -down ($s = -1$) bands, respectively. The inset shows the BZ and its high symmetry points.

where $s = \pm 1$, Δ, γ and λ are the spin index, on-site energy, effective hopping, and SOC strength, respectively. The wavevector-dependent functions read

$$f(\mathbf{k}) \equiv e^{ik_x a/\sqrt{3}} + 2e^{-ik_x a/(2\sqrt{3})}\cos(k_y a/2), \tag{6a}$$

$$g(\mathbf{k}) \equiv 2\sin(k_y a) - 4\sin(k_y a/2)\cos(k_x a/\sqrt{3}), \tag{6b}$$

with the lattice constant a. A typical band structure obtained for this Hamiltonian is illustrated in Fig. 1. The SOC lifts the degeneracy of the bands at each valley, while the time-reversal symmetry (TRS) relates the bands in separate valleys, i.e., spin-up states at the K valley are degenerate with spin-down states at the K′ valley. Note that the band structure in Fig. 1 is slightly shifted compared to the typical band structure of monolayer TMDs, where the SOC mainly lifts the valence band degeneracy. For instance, in monolayer MoS_2, the valence band splitting is ∼150 meV, whereas the conduction band splitting is only ∼3 meV [40]. However, this vertical shift of the band structure is not important in our calculations, since only energy differences (transitions) matter.

The massive Dirac Hamiltonian can be derived from Eq. (5) by a Taylor expansion of $f(\mathbf{k})$ and $g(\mathbf{k})$ around the Dirac points to linear order in $\mathbf{k}$. Although this approximate Hamiltonian has proven useful for characterizing the linear optical properties of TMD monolayers, it fails to account for any even-order nonlinear response (in the dipole approximation) due to the presence of full rotation symmetry, C_∞ [7]. This issue can be resolved by including terms up to second order in $\mathbf{k}$ for $f(\mathbf{k})$, which leads to restored threefold rotation symmetry, C_3, and the TW of isoenergy contours [44]. Hence, we form the TW Hamiltonian by expanding $g(\mathbf{k})$ and $f(\mathbf{k})$ at the K/K′ valleys as $g(\mathbf{k}) \approx 3\sqrt{3}\tau$ and

$$f(\mathbf{k}) \approx \frac{\sqrt{3}}{2} e^{i\pi/3}[-i(\kappa_x - i\kappa_y\tau) + \zeta(\kappa_x + i\kappa_y\tau)^2]. \tag{7}$$

Here, $\kappa_\alpha \equiv a(k_\alpha - K_\alpha)$ is the dimensionless wave vector measured with respect to K or K′, $\zeta \equiv \sqrt{3}/12$ multiplies the TW term, and $\tau = \pm 1$ is the valley index. Note that the second-order term in κ in the expansion of $g(\mathbf{k})$ is neglected here, since its contribution is much smaller than the equivalent term in $f(\mathbf{k})$ due to $\lambda \ll \gamma$. The band structure obtained from the TW Hamiltonian is compared with the TB band structure in Fig. 1. Although the TW band structure deviates from the TB band structure away from K or K′, it accurately reproduces the optical response at low photon frequency [7]. Using the TW Hamiltonian, we derive analytical expressions for the momentum matrix elements and Berry connections to first order in ζ as presented in Eqs. (B1a)–(B1e) of Appendix B. These equations have previously been employed to obtain analytical expressions for the linear and NLO responses in the independent-particle approximation, which are in excellent agreement with the full TB results [7].

C. Analytical approach for excitonic optical response

The impact of excitons on the optical response can be accurately studied by incorporating the BSE solution. Although solving the BSE for a given single-particle band structure is possible, in principle, it is a numerically demanding practice. Furthermore, the physical interpretation of the optical response is not straightforward. Fortunately, the approximate solution of the BSE using the massive Dirac Hamiltonian provides not only accurate estimates of the excitonic energies and wave functions but also valuable physical insight [20,45]. In the massive Dirac approximation, the BSE is transformed to a modified Schrödinger equation for a 2D hydrogen-like atom [45], hereafter referred to as the *analytical BSE*, see Eq. (C1) in Appendix C. Due to the full rotational symmetry, the solution of the analytical BSE yields an exciton wave function with a distinct angular momentum, given by $\psi^{(n)}_{\mathbf{k}} = \exp(il\theta)\phi^{(n)}_l(\kappa)$. Here, θ and κ are the phase and magnitude of a dimensionless wave vector, respectively, l denotes the angular momentum of the exciton, and $\phi^{(n)}_l(\kappa)$ is the radial part of the exciton wave function. Note that the angular momentum l is defined in reciprocal space rather than real coordinate space.

In the BSE, the band overlaps $\langle n\mathbf{k}|n\mathbf{k}'\rangle$, appearing in the integrand, are sensitive to the chosen phase of Bloch states. Although the choice of phase does not influence any measurable quantity such as the optical response, it can change the interpretation of angular momenta in the solutions of the analytical BSE. In analogy to the hydrogen atom orbitals, Bloch phases can be chosen in such a way that the excitons with $l = -2, -1, 0, +1, +2$ correspond to orbitals d^-, p^-, s, p^+, d^+, respectively [24,25,46]. In the present paper, however, we have chosen a different phase convention for the Bloch states, which leads to a -1 shift of the exciton angular momenta at the K valley, i.e., the $l = -3, -2, -1, 0, +1$ states are labeled as d^-, p^-, s, p^+, d^+ orbitals, respectively. At the K′ valley, the angular momentum shift is $+1$, while the plus and minus orbitals interchange their roles due to the TRS, e.g., p^- ($l = -2$) at the K valley is degenerate with p^+ ($l = +2$) at the K′ valley. Transforming from our phase convention to the conventional hydrogen atom phase convention is straightforward by relabeling excitons, yet it simplifies our analytical expressions considerably. Note that, analogous to the hydrogen orbitals, excitons are labeled as ns, $np^\pm$, $nd^\pm$, etc., where n is the principal quantum number starting from 1, 2, and 3 for s, $p^\pm$ and $d^\pm$ orbitals, respectively.

Besides the issue of exciton labeling, the calculated excitonic spectrum of the analytical BSE deviates from the conventional Rydberg series known for hydrogen atoms due to the presence of Bloch band signatures and screened Coulomb interaction. In particular, the degeneracy between orbitals with the same principal quantum number is lifted. This has several implications for the energetic ordering of exciton states. First, for the same principal quantum number, orbitals with lower angular momenta have higher energies (or smaller binding energies), e.g., $E_{3d^\pm} < E_{3p^\pm} < E_{3s}$ [47]. This can be explained by considering the screening property of the Keldysh potential. In fact, the 2D Coulomb interaction is only effectively screened for charges with small in-plane separation [38,39]. The Keldysh potential has a weak logarithmic divergence at short distances, while it approaches the $1/r$ dependence of the conventional Coulomb potential at large distances [38]. For excitons with lower angular momenta, the wave function is confined closer to the origin, i.e., the hole and electron are closer to each other. Therefore, excitons with lower angular momenta experience less Coulomb interaction due to the pronounced short-range screening of the Keldysh potential and, hence, possess smaller binding energies. Second, orbitals with positive and negative angular momenta, e.g., $3p^+$ and $3p^-$, are not degenerate in a single valley, due to the nonzero Berry curvature [46]. Nonetheless, due to the TRS, orbitals with positive angular momenta at the K valley are degenerate with the negative ones at the K′ valley and vice versa. Note that, if the TW Hamiltonian is used for the analytical BSE, a group of orbitals with angular momenta $l, l \pm 3, l \pm 6, \cdots$ are coupled [25], e.g., s states are coupled to $f^\pm$ states. This coupling is typically weak ($\zeta < 1$) and, hence, ignored here, which enables us to derive analytical expressions for the excitonic matrix elements.

Combining the analytical BSE solution with the matrix elements obtained using the TW Hamiltonian, i.e., Eqs. (B1a)–(B1e), both linear and NLO responses can be determined. This is done by performing the angular integral (θ) for the exciton matrix elements between the ground state $|0\rangle$ and the excited states $|n\rangle$, i.e., X^α_{0n}, or between the different excited states, i.e., Q^α_{nm}. Allowed and forbidden transitions are identified based on the properties of these matrix elements. The analytical expressions for X^x_{0n} at the K valley for s, $p^\pm$, and $d^\pm$ excitons are given by

$$X^x_{0n}(s) = X_0 \int_0^\infty \phi^{(n)}_{-1} \frac{\eta_\kappa + 1}{i\varepsilon_\kappa} \kappa d\kappa, \tag{8a}$$

$$X^x_{0n}(p^+) = 0, \tag{8b}$$

$$X^x_{0n}(p^-) = X_0 \zeta \int_0^\infty \phi^{(n)}_{-2} \frac{\eta^3_\kappa + 4\eta_\kappa + 3}{2\varepsilon_\kappa} \kappa^2 d\kappa, \tag{8c}$$

$$X^x_{0n}(d^+) = X_0 \int_0^\infty \phi^{(n)}_{+1} \frac{\eta_\kappa - 1}{i\varepsilon_\kappa} \kappa d\kappa, \tag{8d}$$

$$X^x_{0n}(d^-) = 0. \tag{8e}$$

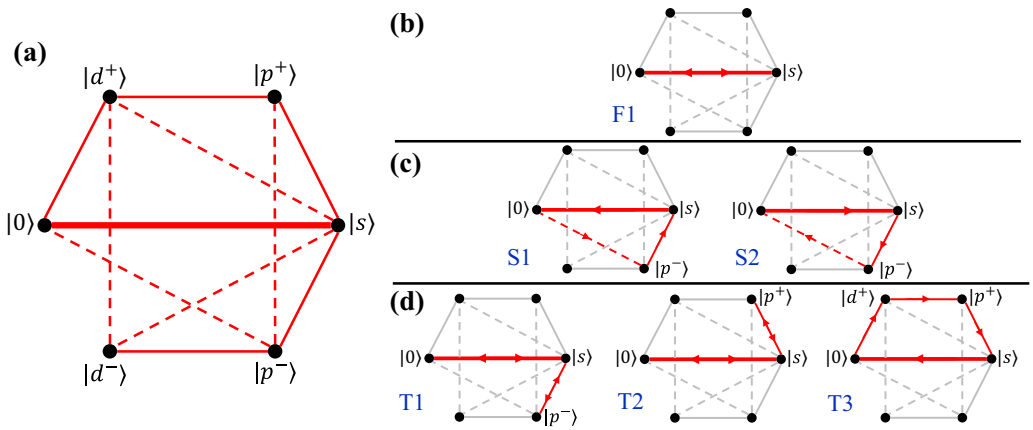

FIG. 2. (a) Schematic diagram of optical selection rules at the K valley. Black dots represent the ground state, $|0\rangle$, and excited states, $|s\rangle$, $|p^{\pm}\rangle$, and $|d^{\pm}\rangle$. Solid/dashed lines represent the possible transitions between the states, which are allowed without/with the TW. The thicker line for coupling between $|0\rangle$ and $|s\rangle$ states indicates larger matrix elements compared to the rest. Note that the transition may be in both directions, e.g., $|0\rangle \to |s\rangle$ or $|s\rangle \to |0\rangle$. The dominant transition paths for the (b) first-, (c) second-, (d) third-order processes are shown in red. Note that a similar diagram can readily be provided for the K′ valley by interchanging the plus and minus states.

Here, $X_0 = mv_F\pi/(2a^2)$ and the Fermi velocity is given by $v_F \equiv \sqrt{3}a\gamma/(2\hbar)$. The wave-vector-dependent functions η_κ and ε_κ are defined as $\eta_\kappa \equiv \Delta_{s\tau}/\varepsilon_\kappa$ and $2\varepsilon_\kappa \equiv \sqrt{4\Delta_{s\tau} + 3\gamma^2\kappa^2}$ with $\Delta_{s\tau} \equiv \Delta + 3\sqrt{3}\lambda s\tau$. Moreover, if the initial excitonic state possesses an angular momentum of l, Q^x_{nm} reads

$$\frac{Q^x_{nm}(l \to l \pm 1)}{\pi a} = i\int_0^\infty \phi_l^{(n)} \frac{\partial \phi_{l\pm1}^{(m)}}{\partial \kappa} \kappa d\kappa \pm i\int_0^\infty \phi_l^{(n)} \phi_{l\pm1}^{(m)} (\eta_\kappa + l \pm 1) d\kappa, \quad (9a)$$

$$\frac{Q^x_{nm}(l \to l \pm 2)}{\pi a} = \frac{\zeta}{2}\int_0^\infty \phi_l^{(n)} \phi_{l\pm2}^{(m)} (5\eta_\kappa - \eta_\kappa^3) \kappa d\kappa, \quad (9b)$$

$$Q^x_{nm}(l \to l \pm 3) = 0, \quad (9c)$$

$$\frac{Q^x_{nm}(l \to l \pm 4)}{\pi a} = \frac{\zeta}{2}\int_0^\infty \phi_l^{(n)} \phi_{l\pm4}^{(m)} (\eta_\kappa^3 - 3\eta_\kappa) \kappa d\kappa. \quad (9d)$$

Similar expressions can be obtained for the y direction, which differ from the above equations by factors of $\pm i$. In addition, excitonic matrix elements at the K′ valley are found simply by interchanging plus and minus orbitals. These expressions clearly exhibit the excitonic selection rules as discussed in the next section.

III. OPTICAL SELECTION RULES

Incorporation of Eqs. (8a)–(8e) and (9a)–(9d) into the conductivity tensors of Eqs. (2a)–(2c) can be performed by utilizing the schematic diagram presented in Fig. 2(a). In this diagram, the black dots represent the ground state $|0\rangle$ and first few excited states $|s\rangle$, $|p^{\pm}\rangle$, and $|d^{\pm}\rangle$. The lines represent transitions that are allowed by the optical selection rules, whereas the missing links indicate the forbidden transitions, for which *excitonic matrix elements vanish*. This diagram is obtained for the K valley, but a similar one can readily be provided for the K′ valley by interchanging the plus and minus states. From a physical point of view, allowed and forbidden transitions are identified by examining the conservation of the angular momentum [30], i.e., preserving the total angular momentum including the photon, lattice, and exciton momenta during a light-matter interaction. If the initial and final states carry angular momenta of l_i and l_f, respectively, the transition between them is allowed in the Dirac approximation if $l_f - l_i = \pm 1$ (solid lines), while it is allowed by the TW if $l_f - l_i \pm 3 = \pm 1$ (dashed lines). For the TW-enabled transitions, any combination of plus and minus signs is allowed, leading to a total of four possibilities. With our phase convention for Bloch eigenstates, the angular momentum of the ground state should be set to 0. Note that a ± 3 unit of angular momentum can be generated by the honeycomb lattice due to the C_3 symmetry. Moreover, a linearly polarized photon can transfer ± 1 unit of angular momentum, since it is a linear combination of left- and right-handed circularly polarized light. Using the momentum conservation rule, the transitions between $|0\rangle$ and $|p^+\rangle$ states are forbidden at the K valley (allowed by the TW at the K′ valley), whereas the $|s\rangle \leftrightarrow |p^+\rangle$ transitions are allowed in both valleys. Among all allowed transitions, the TW-enabled transitions are typically weaker than the rest, because they are scaled by ζ. In addition, since $\eta_\kappa \approx 1$ for small κ, the matrix elements of $|0\rangle \leftrightarrow |d^+\rangle$ transitions are considerably smaller than $|0\rangle \leftrightarrow |s\rangle$ transitions, i.e., $|X^x_{0n}(d^+)| < |X^x_{0n}(s)|$. Therefore, the coupling between $|0\rangle$ and $|s\rangle$ states is generally stronger than the remaining transitions, which is illustrated by a thicker line for $|0\rangle \leftrightarrow |s\rangle$ transitions in Fig. 2.

Using the diagram in Fig. 2(a), we can systematically identify the dominant transitions for linear and NLO processes. The many-body system is initially in its ground state with all electrons occupying the valence band. Depending on the number of photons N involved in the NLO process, precisely N intermediate states are visited before the system eventually returns to the ground state. Therefore, two, three and four intermediate states are involved for the first-, second-, and third-order processes, respectively, with the successive transitions

given by first order $|0\rangle \to |m\rangle \to |0\rangle$, second order $|0\rangle \to |m\rangle \to |n\rangle \to |0\rangle$, third order $|0\rangle \to |m\rangle \to |n\rangle \to |l\rangle \to |0\rangle$. In each case, $|l\rangle$, $|n\rangle$, and $|m\rangle$ can be any of the excited states. Hereafter, any set of possible transitions is referred to as a (transition) *path*.

For a first-order process (such as linear absorption), the dominant transition path is $|0\rangle \to |s\rangle \to |0\rangle$ and, hence, mainly s excitons are bight in the OC spectrum [25]. In addition to s excitons, d^+ and p^- excitons can also be excited, but their brightness is much lower than the s ones, see Eqs. (8a)–(8e). Figure 2(c) shows the dominant transition paths for the second-order processes (such as SHG and optical rectification), i.e., S1: $|0\rangle \to |p^-\rangle \to |s\rangle \to |0\rangle$ and S2: $|0\rangle \to |s\rangle \to |p^-\rangle \to |0\rangle$. Note that the S1 and S2 paths lead to resonances at different frequencies, since the system is relaxed through distinct states, i.e., $|s\rangle$ and $|p^-\rangle$ for S1 and S2, respectively. The quadratic process requires at least one transition, which is enabled by the TW and, hence, any quadratic response in monolayer TMDs vanishes if the TW is not included [7]. Besides these two dominant paths, two other paths are conceivable for the second-order process, $|0\rangle \to |d^+\rangle \to |s\rangle \to |0\rangle$ and $|0\rangle \to |s\rangle \to |d^+\rangle \to |0\rangle$. However, their contributions are considerably smaller then the S1 and S2 paths due to weak coupling between the ground state and d excitons, as well as between the s and d^+ excitons [see Eq. (9b)]. Moving to the third-order processes (such as THG or optical Kerr), four states are involved and, hence, the number of possible paths grows considerably. However, we can still identify the dominant paths as shown with labels in Fig. 2(c). In contrast to the quadratic response, the third-order response does not necessitate the TW to survive. The present algorithm for identifying the dominant transitions can readily be extended to higher order NLO processes in monolayer TMDs.

Combining the above-mentioned selection rules with the information about exciton energies, we can predict the spectral position and nature of resonance peaks appearing in OC, SHG, and THG spectra. Starting with the linear optical process, it is expected that the first strong resonances in the OC spectrum correspond to $1s$ and $2s$ states as shown schematically in Fig. 3. This becomes more complicated for the NLO processes since contributions of all intermediate transitions should be summed to obtain the full spectrum, e.g., all transitions between s and p^- excitons should be considered for determining the SHG spectrum. Nonetheless, spectrally neighboring states dominate, due to the energy-dependent denominators of Eq. (2b) or (2c). Therefore, the lowest frequency resonances in the SHG spectrum are anticipated to be at frequencies coinciding with half of $1s$ and $2p^-$ energies, and they are dominated by the transitions indicated in Fig. 3. Note that probing p excitons in the linear optical spectrum of monolayer TMD is challenging due to the weak coupling between $|0\rangle$ and $|p^-\rangle$, whereas it can be explored experimentally using the SHG response as already demonstrated in Ref. [47]. Proceeding to the THG, we expect that the lowest frequency resonance occurs at one-third of the $1s$ state energy. In addition, multiple transitions contribute simultaneously to the resonance as illustrated in Fig. 3. In the next section, we numerically confirm the validity of these predictions for monolayer MoS_2.

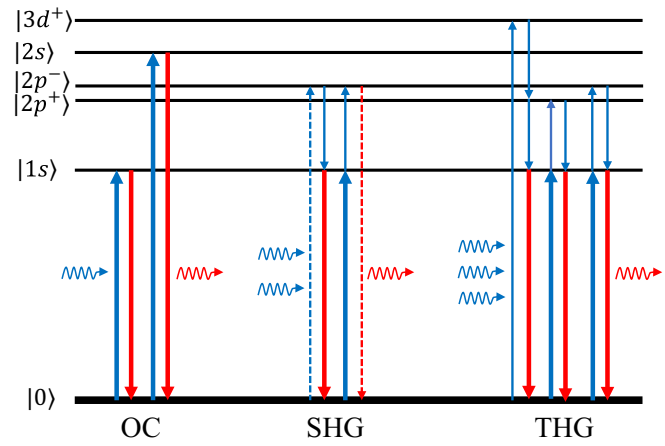

FIG. 3. Sketch of exciton energy levels at the K valley with the dominant transition paths (solid/dashed arrows) for the first resonances of OC, SHG, and THG spectra. Dashed arrows represent the TW-enabled transitions, see Fig. 2. Red arrows indicate transitions that relaxes the system to its ground state.

IV. NUMERICAL RESULTS FOR MONOLAYER MoS_2

In this section, we apply the proposed approach to compute OC, SHG, and THG spectra of monolayer MoS_2 as a representative member of the TMD family. Nonetheless, the main findings hold true for other monolayer TMDs due to similar physics. This is because monolayer TMDs share identical lattice structure (honeycomb lattice), which will lead to identical excitonic selection rules. Furthermore, the Dirac Hamiltonian has been successfully employed for modeling optical and electronic properties of various monolayer TMDs, see e.g., Ref. [22]. Analogously, one can add the TW term to the Dirac Hamiltonian for all monolayer TMDs with different values for ζ. The required parameters, Δ, γ, and λ, are determined by fitting to the experimental data or calculated quasiparticle band structures and, here, are set to 1.25 eV, 1.51 eV, 7.2 meV, respectively [7]. A lattice constant of $a = 3.18$ Å and screening length of $r_0 = 44.3$ Å are assumed [48]. Due to the symmetry of the honeycomb lattice, at normal incidence, it is sufficient to consider the diagonal components of the conductivity tensors, i.e., $\sigma^{(1)}_{xx}$, $\sigma^{(1)}_{xxx}$, and $\sigma^{(1)}_{xxxx}$ [36,49]. The line-shape broadening is accounted for by adding a small phenomenological imaginary part, $i\Gamma$, to the frequency, i.e., $\omega \to \omega + i\Gamma$. We set $\Gamma = 10$ meV for our calculations here [50]. For the full BSE calculations, a dense **k**-mesh with more than 17 000 points is employed for getting converged excitonic energies. For simplicity, we consider a suspended sample, i.e., $\epsilon_s = 1$, and neglect the effect of the substrate. Although the substrate dielectric screening will change the excitonic binding energy, the optical selection rules will be unaffected, and hence our conclusions hold true. Note that since the TB Hamiltonian, i.e., Eq. (5), cannot accurately capture the features of the band structure in the vicinity of Γ point, all spectral features that correspond to C excitons will be absent in the generated spectra. Nonetheless, the resonances at photon energies below the C exciton can be captured accurately.

TABLE I. Energies (in eV) of the first few excitons computed using the full BSE or analytical (or Dirac) approach for monolayer MoS_2.

Exciton	$\|1s\rangle$	$\|2p^+\rangle$	$\|2p^-\rangle$	$\|2s\rangle$	$\|3d^+\rangle$
Dirac, A	1.909	2.114	2.128	2.179	2.220
BSE, A	1.872	2.088	2.103	2.151	2.197
Dirac, B	2.050	2.258	2.271	2.323	2.364
BSE, B	2.017	2.228	2.242	2.291	2.345

We begin by comparing the exciton energies obtained using the full BSE with the ones computed by the analytical approach (in the massive Dirac approximation) in Table I. In the analytical approach, Eq. (C1) should be solved multiple times for different angular momenta and using two different band gaps, $\Delta \pm 3\sqrt{3}\lambda = \{\Delta_A, \Delta_B\}$, for the A and B excitons, whereas, in the full BSE method, solving Eq. (4) once with the TB Hamiltonian provides all required exciton energies. Nonetheless, the one-dimensional integral equation of the analytical BSE approach is solved with much less numerical effort compared to the 2D integral equation of the full BSE method. Results show that the binding energy of 1s exciton is approximately ~ 0.5 eV, which is in agreement with previously reported experimental [17,51] and theoretical works [18,48,52,53]. Excitonic energies obtained in the analytical approach are overestimated slightly by approximately 30 meV, which is due to the underestimation of the Coulomb potential. Although this minor difference can be compensated merely by a slight reduction of the screening parameter, we use the same screening for both approaches here. Precise identification of the angular momentum characteristics of excitons obtained using the full BSE requires investigating the spatial distribution of the exciton wave function [47], whereas it is automatically accomplished in the analytical approach.

Upon determining the exciton energies, the OC spectrum is computed using Eq. (2a) and shown in Fig. 4. The excitonic OC spectrum displays several distinct resonances inside the band gap due to the fundamental and higher order excitons and agrees with the previous results for monolayer MoS_2 [19,20]. Apart from the minor energy shift, the analytical approach provides a very accurate replication of the BSE spectrum. Furthermore, we can readily identify the contributions of different excitons in the spectrum. The results show that the resonances in the OC spectrum mainly stem from the s excitons and, hence, the first two resonances in the spectrum appear at the frequencies of the $1s$ and $2s$ states at each valley. Although transitions mediated by p^- and d^+ excitons are allowed by the selection rules, they are barely observable in the spectrum due to their negligible contribution. Indeed, we numerically confirm that the magnitude of the matrix element between the ground state and $1s$ (or $2s$) exciton is at least ten times larger than that of the ground state and the $2p^-$ or $3d^+$ exciton.

For the second-order processes, we limit our analysis to SHG, but similar results can be obtained for other quadratic processes such as optical rectification [7]. Figure 5 illustrates the SHG spectrum computed for monolayer MoS_2 using the

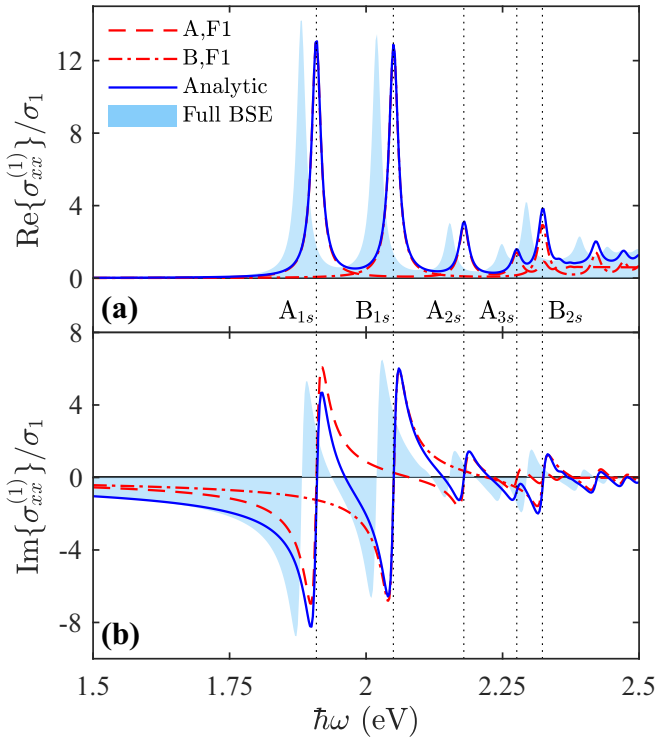

FIG. 4. Real (a) and imaginary (b) parts of excitonic OC spectrum for monolayer MoS_2 [$\sigma_1 \equiv e^2/(4\hbar)$]. Lines show the results obtained by the analytical approach, whereas the filled area is the full BSE spectrum. Solid blue lines indicate the total response including all paths, and red dashed/dashed-dotted lines represent the contributions of A/B excitons for the F1 path [see Fig. 2(b)].

full BSE and analytical approaches, which shows a good agreement between the two methods. As expected for a typical excitonic SHG spectrum, excitons appear as strong resonances at frequencies below half the band gap. The spectral positions of the fundamental peaks agree well with the experimental reports [26,54]. Using the analytical approach, we identify and label the resonance peaks according to the responsible exciton. The contributions of the two dominant paths in Fig. 2(c) are distinguished. These two paths generate resonances with similar strength but at different frequencies, corresponding to the s and p^- exciton energies. Therefore, the peaks corresponding to $2s$ and $2p^-$ states are not degenerate and we expect that they can be resolved experimentally for a clean sample at low temperatures. In the experimental results of Ref. [26] for monolayer MoS_2, only the peaks corresponding to A_{1s} and B_{1s} states were discerned. Nonetheless, a resonance due to the $2s/2p^-$ state has been observed experimentally in the SHG spectrum of monolayer WSe_2 [28]. Based on the present theory, this resonance is expected to split into two spectrally close peaks if a finer spectrum becomes available. Similar conclusions apply to other monolayer TMDs. We also numerically examine the magnitude of the matrix element between the $1s$ and $2p^-$ or $3d^+$ states, and confirm that the former is approximately ten times larger the latter. Combining this fact with the weak coupling of the ground state and d^+ excitons leads to a negligible contribution of the transition paths via the d^+ excitons in the total SHG response. Note that,

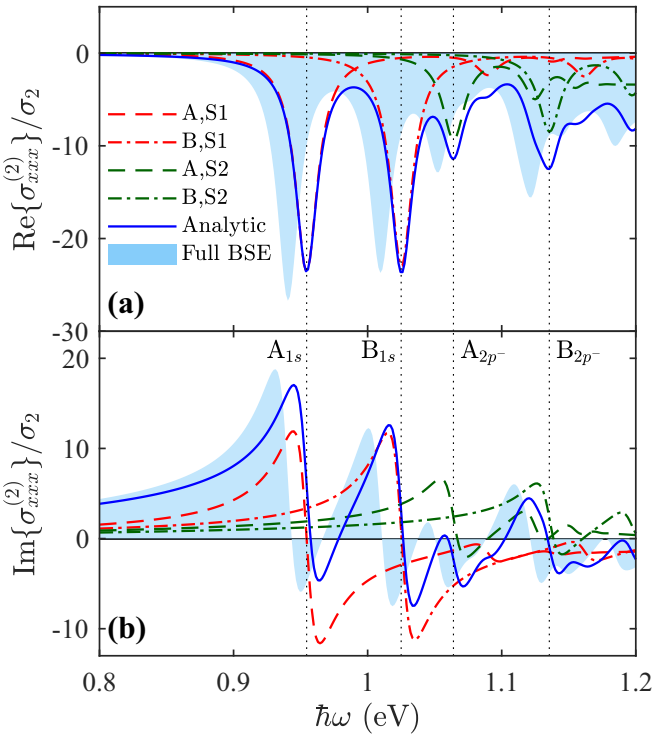

FIG. 5. Real (a) and imaginary (b) parts of excitonic SHG spectrum for monolayer MoS_2 ($\sigma_2 \equiv 1 \times 10^{-15}$ SmV^{-1}). Lines show the results obtained by the analytical approach, whereas the filled area is the full BSE spectrum. Solid blue lines indicate the total response including all paths, and dashed/dashed-dotted lines represent the contributions of A/B excitons. The contributions of the S1 and S2 paths [see Fig. 2(b)] are shown in red and green, respectively.

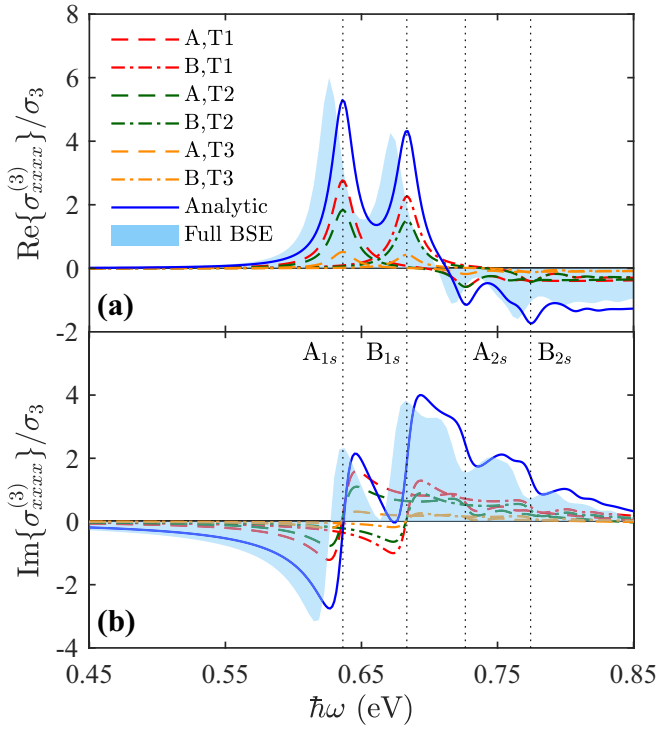

FIG. 6. Real (a) and imaginary (b) parts of excitonic THG spectrum for monolayer MoS_2 ($\sigma_3 \equiv 1 \times 10^{-23}$ Sm^{2}V^{-2}). Lines show the results obtained by the analytical approach, whereas the filled area is the full BSE spectrum. Solid blue lines indicate the total response including all paths, and dashed/dashed-dotted lines represent the contributions of A/B excitons. The contributions of the T1, T2, and T3 paths [see Fig. 2(b)] are shown in red, green, and brown, respectively.

for circularly polarized light, the resonances corresponding to s and p excitons possess different valley dichroism due to their distinct angular momenta [7].

In Fig. 6, we compare the THG conductivity computed using the full BSE with the analytical approach. The THG results follow the main trends observed in the OC and SHG responses. However, multiple transition paths should be included to obtain an accurate spectrum. In contrast to the OC and SHG responses, where each resonance is mainly formed by a single transition path, the resonances in THG spectrum have contributions from several distinct paths. For instance, the fundamental peaks, i.e., A_{1s} and B_{1s}, originate from three paths; T1: $|0\rangle \to |s\rangle \to |p^-\rangle \to |s\rangle \to |0\rangle$, T2: $|0\rangle \to |s\rangle \to |p^+\rangle \to |s\rangle \to |0\rangle$, and T3: $|0\rangle \to |d^+\rangle \to |p^+\rangle \to |s\rangle \to |0\rangle$. Note that, although the T3 path includes a weak transition between the ground and d^+ states, the coupling between d^+ and p^+ states can be sufficiently large so as to compensate this weak transition and, hence, contribute significantly to the THG response. Indeed, our calculations show that the magnitude of the coupling between the $2p^+$ and $3d^+$ states is almost three times larger than the coupling between the $1s$ and $2p^+$ states. The THG spectrum of monolayer MoS_2 has not been experimentally measured yet, to the best of our knowledge, and our results may provide the required guidelines for understanding the nature of resonances in future experimental studies.

V. CONCLUSION

In summary, we have elucidated an analytical approach for calculating the linear and NLO responses of intrinsic monolayer TMDs, including excitonic effects. Analytical expressions for excitonic matrix elements were derived and used to determine the NLO response of monolayer MoS_2. Using a diagrammatic approach, we clearly demonstrate the importance of s excitons in NLO responses and the crucial role of TW in even-order nonlinearities. With a much less numerical effort, the analytical approach generates almost identical spectra to that of the numerically demanding BSE, while it also provides a transparent picture of the excitonic transitions in the NLO response. For instance, the fundamental resonance in the THG spectrum appears at the frequency of the $1s$ exciton and, originates from three distinct transition paths including p and d excitons. Our approach can readily be extended to identify the main transitions for any other NLO process in monolayer TMDs.

ACKNOWLEDGMENTS

The authors thank F. Hipolito for helpful discussions throughout the project. This work was supported by the QUSCOPE center sponsored by the Villum Foundation and the

CNG center under the Danish National Research Foundation (Project No. DNRF103).

APPENDIX A: EQUATION OF MOTION AND ITS PERTURBATIVE SOLUTION

Here, we review the derivation of the dynamical equation for the density matrix and present its perturbative solution up to third order in the external field. We follow the procedure outlined in Refs. [32,33]. The many-body Hamiltonian of a system of electrons under the influence of an external perturbation reads $\hat{H}(t) \equiv \hat{\mathcal{H}}_0 + \hat{U}(t)$, where $\hat{\mathcal{H}}_0$ and $\hat{U}(t)$ are the many-body unperturbed (including Coulomb contribution) and light-matter interaction parts, respectively. In second quantization, they are given by

$$\hat{\mathcal{H}}_0 \equiv \sum_n \varepsilon_n^0 \hat{c}_n^\dagger \hat{c}_n + \frac{1}{2}\sum_{k,l,n,m} V_{klmn}\hat{c}_k^\dagger \hat{c}_l^\dagger \hat{c}_n \hat{c}_m, \tag{A1a}$$

$$\hat{U}(t) \equiv \sum_{n,m} u_{nm}(t)\hat{c}_n^\dagger \hat{c}_m, \tag{A1b}$$

where $u_{nm} = \langle n|\hat{u}(t)|m\rangle$ and V_{klmn} are the matrix elements of the time-dependent and Coulomb interactions, respectively. In the dipole approximation (long wavelength regime), the single-electron interaction Hamiltonian reads $\hat{u}(t) \equiv e\hat{\mathbf{r}} \cdot \boldsymbol{\mathcal{E}}(t)$. The Hamiltonian is written in the single-particle basis $|n\rangle$, i.e., $\hat{h}_0|n\rangle = \varepsilon_n^0|n\rangle$ with $\hat{h}_0$ the unperturbed Hamiltonian of a single electron. In the Heisenberg picture and within the MFA, this Hamiltonian leads to the usual equation of motion (quantum Liouville equation) for the density matrix $\rho_{ji} = \langle 0|\hat{c}_i^\dagger \hat{c}_j|0\rangle$ [32],

$$\begin{aligned} i\hbar\frac{\partial \rho_{ji}}{\partial t} - \varepsilon_{ji}\rho_{ji} - \sum_l (u_{jl}\rho_{li} - u_{li}\rho_{jl}) \\ = \sum_{l,m,n}(V_{mlni} - V_{lmni})(\rho_{nl} - \delta_{mi}\delta_{nl}\delta_{lv})\rho_{jm} \\ + \sum_{l,m,n}(V_{jlmn} - V_{jlnm})(\rho_{nl} - \delta_{mj}\delta_{nl}\delta_{lv})\rho_{mi}, \end{aligned} \tag{A2}$$

where $\varepsilon_{ji} \equiv \varepsilon_j - \varepsilon_i$, and $\varepsilon_i \equiv \varepsilon_i^0 + \sum_l (V_{ilil} - V_{illi})\delta_{lv}$ are the quasiparticle energies. Note that the Kronecker delta, δ_{lv}, serves to count occupied states only. To proceed, we make several reasonable assumptions to simplify the dynamical equation. In periodic systems, each index should run over both band index and wave vector. We assume that the density matrix is diagonal with respect to the wave vector, i.e., $\rho_{j\mathbf{k}_j i\mathbf{k}_i} \equiv \rho_{ji\mathbf{k}_i}\delta_{\mathbf{k}_i,\mathbf{k}_j}$ [32]. Furthermore, all exchange terms are ignored, since they have a negligible effect on the output spectrum [18].

Using the above-mentioned approximations for a two-band semiconductor, the dynamical equations for the population densities and coherences read [32,33]

$$\begin{aligned} i\hbar\frac{\partial \rho_{cv\mathbf{k}}}{\partial t} - \varepsilon_{cv\mathbf{k}}\rho_{cv\mathbf{k}} + \sum_{\mathbf{k}'} V_{cv}^{\mathbf{k},\mathbf{k}'}(\rho_{vv\mathbf{k}} - \rho_{cc\mathbf{k}})\rho_{cv\mathbf{k}'} - \sum_{\mathbf{k}'}\left[V_{vv}^{\mathbf{k},\mathbf{k}'}(\rho_{vv\mathbf{k}'} - 1) - V_{cc}^{\mathbf{k},\mathbf{k}'}\rho_{cc\mathbf{k}'}\right]\rho_{cv\mathbf{k}} \\ = (u_{cc\mathbf{k}} - u_{vv\mathbf{k}})\rho_{cv\mathbf{k}} + u_{cv\mathbf{k}}(\rho_{vv\mathbf{k}} - \rho_{cc\mathbf{k}}), \end{aligned} \tag{A3a}$$

$$i\hbar\frac{\partial \rho_{cc\mathbf{k}}}{\partial t} - \sum_{\mathbf{k}'}\left[V_{vc}^{\mathbf{k},\mathbf{k}'}\rho_{vc\mathbf{k}'}\rho_{cv\mathbf{k}} - V_{cv}^{\mathbf{k},\mathbf{k}'}\rho_{cv\mathbf{k}'}\rho_{vc\mathbf{k}}\right] = u_{cv\mathbf{k}}\rho_{vc\mathbf{k}} - u_{vc\mathbf{k}}\rho_{cv\mathbf{k}}. \tag{A3b}$$

The equation of motion for $\rho_{cv\mathbf{k}}$ is simply obtained by complex conjugating the first equation, since the density matrix operator is Hermitian. In addition, the population conservation, i.e., normalization of the density matrix, leads to $\rho_{vv\mathbf{k}} = 1 - \rho_{cc\mathbf{k}}$. In Eqs. (A3a) and (A3a), we ignore the exchange terms due to their negligible influence [18]. Moreover, $V_{nm}^{\mathbf{k},\mathbf{k}'}$ are the direct Coulomb matrix elements, given by

$$V_{nm}^{\mathbf{k},\mathbf{k}'} \approx \frac{e^2}{2\epsilon_0}\frac{\langle n\mathbf{k}|n\mathbf{k}'\rangle\langle m\mathbf{k}'|m\mathbf{k}\rangle}{|\mathbf{k}-\mathbf{k}'|(\epsilon_s + r_0|\mathbf{k}-\mathbf{k}'|)}, \tag{A4}$$

where a Keldysh potential with surrounding screening ϵ_s and screening length r_0 is assumed [7]. The approximation in Eq. (A4) is mainly because of the neglected terms for nonzero reciprocal vectors [33].

For a time-harmonic electric field, i.e., $\hat{u}(t) = \hat{u}_\omega \exp(-i\omega t) + \text{H.c.}$, the dynamical equation for the density matrix can be solved perturbatively up to any required order of the field using the Green's functions [32]. Here, we only focus on the expressions for SHG/THG for brevity, but similar results can be readily obtained for the optical rectification and Kerr processes. Hence, the Nth-order density matrix reads $\rho_{ji\mathbf{k}}^{(N)}(t) = \rho_{ji\mathbf{k}}^{(N\omega)}\exp(-iN\omega t) + \rho_{ji\mathbf{k}}^{(-N\omega)}\exp(iN\omega t)$, where $\rho_{ji\mathbf{k}}^{(N\omega)}$ for $N = 1, 2, 3$ are given by

$$\rho_{cv\mathbf{k}}^{(\omega)} = \sum_n \frac{\psi_{\mathbf{k}}^{(n)}U_{n0}}{\hbar\omega - E_n}, \tag{A5a}$$

$$\rho_{cv\mathbf{k}}^{(2\omega)} = \sum_{n,m}\frac{\psi_{\mathbf{k}}^{(n)}U_{nm}U_{m0}}{(2\hbar\omega - E_n)(\hbar\omega - E_m)}, \tag{A5b}$$

$$\rho_{cv\mathbf{k}}^{(3\omega)} = \sum_{l,n,m}\frac{\psi_{\mathbf{k}}^{(l)}U_{ln}U_{nm}U_{m0}}{(3\hbar\omega - E_l)(2\hbar\omega - E_n)(\hbar\omega - E_m)}, \tag{A5c}$$

$$\rho_{cc\mathbf{k}}^{(\omega)} = 0, \tag{A5d}$$

$$\rho_{cc\mathbf{k}}^{(2\omega)} = \rho_{cv\mathbf{k}}^{(\omega)}\rho_{vc\mathbf{k}}^{(\omega)}, \tag{A5e}$$

$$\rho_{cc\mathbf{k}}^{(3\omega)} = \rho_{cv\mathbf{k}}^{(2\omega)}\rho_{vc\mathbf{k}}^{(\omega)} + \rho_{cv\mathbf{k}}^{(\omega)}\rho_{vc\mathbf{k}}^{(2\omega)}. \tag{A5f}$$

Here, $U_{0n} = U_{n0}^* \equiv \sum_{\mathbf{k}}\psi_{\mathbf{k}}^{(n)}u_{vc\mathbf{k}}^{(\omega)}$ and $U_{nm} \equiv \sum_{\mathbf{k}}\psi_{\mathbf{k}}^{(n)*}\psi_{\mathbf{k}}^{(m)}[u_{cc\mathbf{k}}^{(\omega)} - u_{vv\mathbf{k}}^{(\omega)}]$ are the excitonic matrix elements of the interaction Hamiltonian, where $u_{nm\mathbf{k}}^{(\omega)} \equiv \langle n\mathbf{k}|\hat{u}_\omega|m\mathbf{k}\rangle$. Indeed, U_{0n}

corresponds to the coupling between the ground state $|0\rangle$ and excited state $|n\rangle$, whereas U_{nm} is the interexcitonic matrix element. Note that $\rho^{(N\omega)}_{vc\mathbf{k}} = [\rho^{(-N\omega)}_{cv\mathbf{k}}]^*$ and $\rho^{(N\omega)}_{vv\mathbf{k}} = -\rho^{(N\omega)}_{cc\mathbf{k}}$ due to the Hermiticity and normalization of the density matrix, respectively. Upon obtaining the density matrix, the expectation value of the Nth-order current density is determined straightforwardly as $\mathbf{J}^{(N)}(t) = \mathrm{Tr}[\hat{\rho}^{(N)}\hat{\mathbf{J}}]$ leading to Eqs. (2a)–(2c) for the conductivities. Note that U_{nm} includes the ill-defined intraband matrix elements of the position operator, which leads to appearance of the generalized derivative [35]. Therefore, by using the rule $(r^{\alpha}_{cc\mathbf{k}} - r^{\alpha}_{vv\mathbf{k}})\psi^{(m)}_{cv\mathbf{k}} = i[\psi^{(m)}_{cv\mathbf{k}}]_{;k_\alpha}$, Eq. (3b) can be obtained [33].

APPENDIX B: TRIGONAL-WARPING HAMILTONIAN

For the 2×2 TB Hamiltonian, Eq. (5), the eigenenergies for a given spin read $\varepsilon_{c\mathbf{k}} = -\varepsilon_{v\mathbf{k}} = \sqrt{\Delta_s^2 + \gamma^2 F^2}$ with $\Delta_s \equiv \Delta + s\lambda g(\mathbf{k})$. The corresponding eigenvectors are then given by $|c\mathbf{k}\rangle = [\cos(b/2), -\sin(b/2)e^{i\phi}]$ and $|v\mathbf{k}\rangle = [\sin(b/2), \cos(b/2)e^{i\phi}]$, where F and ϕ are defined using $f(\mathbf{k}) = F\exp(i\phi)$, and $\cos(b) \equiv \Delta_s/\varepsilon_{c\mathbf{k}}$. Furthermore, the momentum operator in the TB method is given as the $\mathbf{k}$ derivative of the Hamiltonian, $\hbar p^{\alpha} = m\partial H_0/\partial k_\alpha$. Hence, the momentum matrix elements can readily be determined using $p^{\alpha}_{nm\mathbf{k}} = \langle n\mathbf{k}|p^{\alpha}|m\mathbf{k}\rangle$, e.g., $\hbar p^{\alpha}_{vc\mathbf{k}} = -m\gamma[\cos(b)F_\alpha + iF\phi_\alpha]$ with $F_\alpha \equiv \partial F/\partial k_\alpha$ and $\phi_\alpha \equiv \partial\phi/\partial k_\alpha$.

Taylor-expanding $f(\mathbf{k})$ and $g(\mathbf{k})$ functions in the vicinity of the Dirac points, K: $2\pi[3^{-1/2}, 3^{-1}]/a$ and K′: $2\pi[3^{-1/2}, -3^{-1}]/a$, the massive Dirac and TW Hamiltonian are determined. Using Eq. (7) for $f(\mathbf{k})$, we calculate any required parameter and keep only zeroth and first-order terms in ζ. Hence, the transition energies, $\varepsilon_{cv\mathbf{k}}$, momentum matrix elements, $p^{\alpha}_{vc\mathbf{k}}$, and Berry connections, $\Omega^{\alpha}_{cc\mathbf{k}} - \Omega^{\alpha}_{vv\mathbf{k}}$, are given by

$$\varepsilon_{cv\mathbf{k}} \approx 2\varepsilon_\kappa\left[1 - \zeta\tau\kappa\left(1-\eta_\kappa^2\right)\sin(3\theta)\right], \tag{B1a}$$

$$\frac{p^{x}_{vc\mathbf{k}}}{p_0} \approx \eta_\kappa\cos(\theta) + i\tau\sin(\theta) + \frac{\zeta\kappa}{2}\left[3i\cos(2\theta) - i\cos(4\theta) - \tau\eta_\kappa\left(4+\eta_\kappa^2\right)\sin(2\theta) + \tau\eta_\kappa\left(2-\eta_\kappa^2\right)\sin(4\theta)\right], \tag{B1b}$$

$$\frac{p^{y}_{vc\mathbf{k}}}{p_0} \approx \eta_\kappa\sin(\theta) - i\tau\cos(\theta) - \frac{\zeta\kappa}{2}\left[3i\sin(2\theta) + i\sin(4\theta) + \tau\eta_\kappa\left(4+\eta_\kappa^2\right)\cos(2\theta) + \tau\eta_\kappa\left(2-\eta_\kappa^2\right)\cos(4\theta)\right], \tag{B1c}$$

$$\frac{\Omega^{x}_{cc\mathbf{k}} - \Omega^{x}_{vv\mathbf{k}}}{a} \approx \eta_\kappa\left[\tau\frac{\sin(\theta)}{\kappa} + \frac{\zeta}{2}\left(5-\eta_\kappa^2\right)\cos(2\theta) - \frac{\zeta}{2}\left(3-\eta_\kappa^2\right)\cos(4\theta)\right], \tag{B1d}$$

$$\frac{\Omega^{y}_{cc\mathbf{k}} - \Omega^{y}_{vv\mathbf{k}}}{a} \approx -\eta_\kappa\left[\tau\frac{\cos(\theta)}{\kappa} + \frac{\zeta}{2}\left(5-\eta_\kappa^2\right)\sin(2\theta) + \frac{\zeta}{2}\left(3-\eta_\kappa^2\right)\sin(4\theta)\right], \tag{B1e}$$

with $p_0 \equiv -mv_F$. In these expressions, the terms proportional to ζ stem from the TW effect, and well-known expressions in the massive Dirac approximation are obtained by setting $\zeta = 0$.

APPENDIX C: ANALYTICAL BSE

In the massive Dirac approximation, a simple expression can be obtained for the Coulomb matrix element. Then, using an ansatz $\psi^{(n)}_{\mathbf{k}} = \exp(il\theta)\phi^{(n)}_l(\kappa)$ and performing the angular integral, the BSE is reduced to the following 1D eigenvalue problem for the radial part of exciton wave function [20]:

$$\begin{aligned} E_n\phi^{(n)}_l(\kappa) = 2\varepsilon_\kappa\phi^{(n)}_l(\kappa) - \frac{e^2}{8(2\pi)^2\epsilon_0 a}\int_0^\infty d\kappa'\phi^{(n)}_l(\kappa')\kappa'\{&2\sin(b_\kappa)\sin(b_{\kappa'})D_l(\kappa,\kappa') \\ &+ [1+\cos(b_\kappa)][1+\cos(b_{\kappa'})]D_{l+\tau}(\kappa,\kappa') + [1-\cos(b_\kappa)][1-\cos(b_{\kappa'})]D_{l-\tau}(\kappa,\kappa')\}. \end{aligned} \tag{C1}$$

Here, $\cos(b_\kappa) \equiv \Delta_{s\tau}/\varepsilon_\kappa$ and $D_l(\kappa,\kappa')$ are defined as

$$D_l(\kappa,\kappa') \equiv \int_0^{2\pi}\frac{\cos(l\theta)d\theta}{q(\epsilon_s + r_0 q/a)}, \tag{C2}$$

with $q \equiv \sqrt{\kappa^2 + \kappa'^2 - 2\kappa\kappa'\cos(\theta)}$. From Eq. (C1), one can confirm that the l state at one valley is degenerate with the $-l$ state at the other valley [note that $D_l(\kappa,\kappa') = D_{-l}(\kappa,\kappa')$]. For each angular momentum, Eq. (C1) can be solved numerically by discretizing κ, using the technique explained in Ref. [55]. Upon determining the exciton eigenenergies and eigenvectors, the excitonic matrix elements are obtained using Eqs. (8a)–(8e) and (9a)–(9d).

[1] Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley-Interscience, Hoboken, NJ, 2002).

[2] R. W. Boyd, *Nonlinear Optics*, 3rd ed. (Elsevier Science Publishing Co Inc., Amsterdam, 2008).

[3] Z. Sun, A. Martinez, and F. Wang, Nat. Photon. **10**, 227 (2016).

[4] A. Autere, H. Jussila, Y. Dai, Y. Wang, H. Lipsanen, and Z. Sun, Adv. Mater. **30**, 1705963 (2018).

[5] Y. Zhang, J. van den Brink, C. Felser, and B. Yan, 2D Mater. **5**, 044001 (2018).

[6] M. Eginligil, B. Cao, Z. Wang, X. Shen, C. Cong, J. Shang, C. Soci, and T. Yu, Nat. Commun. **6**, 7636 (2015).

[7] A. Taghizadeh and T. G. Pedersen, arXiv:1812.02596.

[8] E. Hendry, P. J. Hale, J. Moger, A. K. Savchenko, and S. A. Mikhailov, Phys. Rev. Lett. **105**, 097401 (2010).

[9] S.-Y. Hong, J. I. Dadap, N. Petrone, P.-C. Yeh, J. Hone, and R. M. Osgood, Phys. Rev. X **3**, 021014 (2013).

[10] H. Rostami, M. I. Katsnelson, and M. Polini, Phys. Rev. B **95**, 035416 (2017).

[11] F. Hipolito and V. M. Pereira, 2D Mater. **4**, 021027 (2017).

[12] M. Baudisch, A. Marini, J. D. Cox, T. Zhu, F. Silva, S. Teichmann, M. Massicotte, F. Koppens, L. S. Levitov, F. J. G. de Abajo, and J. Biegert, Nat. Commun. **9**, 1018 (2018).

[13] T. Jiang, D. Huang, J. Cheng, X. Fan, Z. Zhang, Y. Shan, Y. Yi, Y. Dai, L. Shi, K. Liu, C. Zeng, J. Zi, J. E. Sipe, Y.-R. Shen, W.-T. Liu, and S. Wu, Nat. Photon. **12**, 430 (2018).

[14] B. Semnani, R. Jago, S. Safavi-Naeini, H. Majedi, E. Malic, and P. Tassin, 2D Mater. **6**, 031003 (2019).

[15] J. R. Schaibley, H. Yu, G. Clark, P. Rivera, J. S. Ross, K. L. Seyler, W. Yao, and X. Xu, Nat. Rev. Mater. **1**, 16055 (2016).

[16] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (2013).

[17] M. M. Ugeda, A. J. Bradley, S.-F. Shi, F. H. da Jornada, Y. Zhang, D. Y. Qiu, W. Ruan, S.-K. Mo, Z. Hussain, Z.-X. Shen, F. Wang, S. G. Louie, and M. F. Crommie, Nat. Mater. **13**, 1091 (2014).

[18] K. S. Thygesen, 2D Mater. **4**, 022004 (2017).

[19] G. Berghäuser and E. Malic, Phys. Rev. B **89**, 125309 (2014).

[20] F. Wu, F. Qu, and A. H. MacDonald, Phys. Rev. B **91**, 075310 (2015).

[21] E. Ridolfi, C. H. Lewenkopf, and V. M. Pereira, Phys. Rev. B **97**, 205409 (2018).

[22] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao, Phys. Rev. Lett. **108**, 196802 (2012).

[23] A. J. Chaves, R. M. Ribeiro, T. Frederico, and N. M. R. Peres, 2D Mater. **4**, 025086 (2017).

[24] T. Cao, M. Wu, and S. G. Louie, Phys. Rev. Lett. **120**, 087402 (2018).

[25] X. Zhang, W.-Y. Shan, and D. Xiao, Phys. Rev. Lett. **120**, 077401 (2018).

[26] M. L. Trolle, Y.-C. Tsao, K. Pedersen, and T. G. Pedersen, Phys. Rev. B **92**, 161409(R) (2015).

[27] M. Grüning and C. Attaccalite, Phys. Rev. B **89**, 081102(R) (2014).

[28] G. Wang, X. Marie, I. Gerber, T. Amand, D. Lagarde, L. Bouet, M. Vidal, A. Balocchi, and B. Urbaszek, Phys. Rev. Lett. **114**, 097403 (2015).

[29] B. Han, C. Robert, E. Courtade, M. Manca, S. Shree, T. Amand, P. Renucci, T. Taniguchi, K. Watanabe, X. Marie, L. E. Golub, M. M. Glazov, and B. Urbaszek, Phys. Rev. X **8**, 031073 (2018).

[30] J. Xiao, Z. Ye, Y. Wang, H. Zhu, Y. Wang, and X. Zhang, Light Sci. Appl. **4**, e366 (2015).

[31] D. B. S. Soh, C. Rogers, D. J. Gray, E. Chatterjee, and H. Mabuchi, Phys. Rev. B **97**, 165111 (2018).

[32] T. G. Pedersen, Phys. Rev. B **92**, 235432 (2015).

[33] A. Taghizadeh and T. G. Pedersen, Phys. Rev. B **97**, 205432 (2018).

[34] L. X. Benedict and E. L. Shirley, Phys. Rev. B **59**, 5441 (1999).

[35] C. Aversa and J. E. Sipe, Phys. Rev. B **52**, 14636 (1995).

[36] A. Taghizadeh, F. Hipolito, and T. G. Pedersen, Phys. Rev. B **96**, 195413 (2017).

[37] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, Phys. Rev. Lett. **80**, 4510 (1998).

[38] P. Cudazzo, I. V. Tokatly, and A. Rubio, Phys. Rev. B **84**, 085406 (2011).

[39] M. L. Trolle, T. G. Pedersen, and V. Véniard, Sci. Rep. **7**, 39844 (2017).

[40] A. Kormányos, V. Zólyomi, N. D. Drummond, P. Rakyta, G. Burkard, and V. I. Fal'ko, Phys. Rev. B **88**, 045416 (2013).

[41] S. Konabe and S. Okada, Phys. Rev. B **90**, 155304 (2014).

[42] Z. Qiao, H. Jiang, X. Li, Y. Yao, and Q. Niu, Phys. Rev. B **85**, 115439 (2012).

[43] T. G. Pedersen, Phys. Rev. B **98**, 165425 (2018).

[44] A. Kormányos, G. Burkard, M. Gmitra, J. Fabian, V. Zólyomi, N. D. Drummond, and V. Fal'ko, 2D Mater. **2**, 022001 (2015).

[45] J. Zhou, W.-Y. Shan, W. Yao, and D. Xiao, Phys. Rev. Lett. **115**, 166803 (2015).

[46] A. Srivastava and A. Imamoğlu, Phys. Rev. Lett. **115**, 166802 (2015).

[47] Z. Ye, T. Cao, K. O'Brien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, Nature **513**, 214 (2014).

[48] T. Olsen, S. Latini, F. Rasmussen, and K. S. Thygesen, Phys. Rev. Lett. **116**, 056401 (2016).

[49] F. Hipolito, A. Taghizadeh, and T. G. Pedersen, Phys. Rev. B **98**, 205420 (2018).

[50] F. Cadiz, E. Courtade, C. Robert, G. Wang, Y. Shen, H. Cai, T. Taniguchi, K. Watanabe, H. Carrere, D. Lagarde, M. Manca, T. Amand, P. Renucci, S. Tongay, X. Marie, and B. Urbaszek, Phys. Rev. X **7**, 021026 (2017).

[51] A. R. Klots, A. K. M. Newaz, B. Wang, D. Prasai, H. Krzyzanowska, J. Lin, D. Caudel, N. J. Ghimire, J. Yan, B. L. Ivanov, K. A. Velizhanin, A. Burger, D. G. Mandrus, N. H. Tolk, S. T. Pantelides, and K. I. Bolotin, Sci. Rep. **4**, 6608 (2014).

[52] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **88**, 245309 (2013).

[53] H. Shi, H. Pan, Y.-W. Zhang, and B. I. Yakobson, Phys. Rev. B **87**, 155304 (2013).

[54] L. M. Malard, T. V. Alencar, A. P. M. Barboza, K. F. Mak, and A. M. de Paula, Phys. Rev. B **87**, 201401(R) (2013).

[55] Calvin Yi-Ping Chao and S. L. Chuang, Phys. Rev. B **43**, 6530 (1991).