

Excitonic two-photon absorption in monolayer transition metal dichalcogenides: Impact of screening and trigonal warping

Thomas Garm Pedersen * and Alireza Taghizadeh 

Department of Materials and Production, Aalborg University, DK-9220 Aalborg Øst, Denmark



(Received 28 April 2021; revised 2 August 2021; accepted 12 August 2021; published 26 August 2021)

Two-photon spectroscopy is an important tool for detection of states that are dark in single-photon spectra. We calculate two-photon absorption spectra for monolayer transition-metal dichalcogenides including excitonic effects based on a Bethe-Salpeter equation approach. Both spin-orbit coupling and trigonal warping are included. Our results reveal that otherwise dark p states dominate the two-photon response. However, weakly allowed two-photon transitions coinciding with single-photon s -state resonances appear due to trigonal warping. The angular momentum character of participating excitons is analyzed and we study the sensitivity to screening by dielectric surroundings and discuss the free-carrier limit. Biexciton contributions to the high-energy two-photon response are estimated using a simplified Wannier approach. Finally, a brief comparison with recent experiments is given.

DOI: [10.1103/PhysRevB.104.085431](https://doi.org/10.1103/PhysRevB.104.085431)

I. INTRODUCTION

The two-photon absorption (TPA) process was first predicted by Göppert-Mayer [1] and since applied in a range of laser-based applications such as microfabrication [2], imaging [3], power limiting [4], and data storage [5]. In addition, TPA spectra provide important information on electronic states since the optical selection rules are distinct from single-photon absorption (SPA) transitions [6]. This follows from the fact that TPA may be seen as two successive SPA events proceeding via an intermediate state. In particular, in inversion-symmetric systems, one-photon transitions connect opposite-parity states, whereas TPA couples states with identical parity [6]. In low-dimensional semiconductors, SPA very precisely identifies the lowest optically bright excitons [7–10]. Yet, to determine their binding energy, excitonic resonances must be compared to the quasiparticle band gap, at which a continuum of transitions sets in. This onset is often difficult to locate in SPA spectra, particularly in low-dimensional semiconductors. In contrast, TPA probes higher excitons that are dark or only weakly bright in SPA. Hence, using Wannier excitons as an illustrative example, SPA mainly probes bright $1s$ excitons while TPA reveals the location of $2p$ and higher dark states. The combined information in these spectra has recently enabled the extraction of quantitative exciton binding energies and band gaps using TPA spectroscopy in carbon nanotubes [11] as well as three of the most important transition-metal dichalcogenides (TMDs) WSe_2 [12,13], WS_2 [14], and MoS_2 [15]. In addition, the very large TPA response demonstrated in MoS_2 [16,17] is promising for technological applications.

In the present paper, we explore excitonic TPA in monolayer TMDs using a four-band model of the band structure including spin-orbit (SO) coupling. Excitons are then modeled

on top of the band structure using the Keldysh potential to incorporate electron-hole interaction. We describe the non-linear optical response within a length-gauge approach that has previously been shown to accurately capture the response, even if the number of bands considered is severely truncated [18]. Moreover, semianalytical results are obtained using a simplified trigonal-warping Hamiltonian [19] and we discuss the free-carrier limit of the response. Our focus is on the role of screening in TPA and we find that TPA resonances remain visible even in highly screened cases such as hBN encapsulation. Theoretical analyses of the TPA response in TMDs including selection rules have previously been made in Refs. [20,21]. While no quantitative predictions for TPA were given in Ref. [20], a detailed velocity-gauge analysis of the spectral response was presented in Ref. [21]. In that work, however, the sum over intermediate states was approximated by replacing actual intermediate energies by their average. This replacement has the great advantage of allowing for an analytic evaluation of the sum. In turn, a Wannier-exciton model can be applied to express the TPA amplitude in terms of the final-state gradient [21]. Using instead length gauge, as in the present paper, the intensity of individual intermediate-state contributions can be accurately described. In length gauge, accurate inter-exciton matrix elements are obtained in terms of generalized derivatives of exciton wave functions, even if a very restricted set of bands is included [18]. We can therefore decompose the total TPA response into dominant transition paths. We also uncover the role of trigonal warping breaking Wannier-exciton selection rules based on definite angular momentum. In fact, we predict weak TPA features coinciding with SPA resonances due to trigonal warping. Furthermore, the role of screening by a dielectric environment is studied in detail, including the free-carrier limit appropriate for cases of very large screening. Finally, we include exciton-exciton interactions through biexciton contributions to the TPA response showing that these add significantly to the high-energy part of the spectrum.

*tgp@mp.aau.dk

II. TWO-PHOTON ABSORPTION

We will consider a two-dimensional (2D) material under excitation by a monochromatic electric field $\vec{\mathcal{E}}(t) = \frac{1}{2}(\vec{\mathcal{E}}_0 e^{-i\omega t} + \vec{\mathcal{E}}_0^* e^{i\omega t})$. In this case, the induced sheet current density oscillating as $e^{-i\omega t}$ is

$$\vec{\mathcal{J}}_\omega(t) = \frac{1}{2}\sigma^{(1)}\vec{\mathcal{E}}_0 e^{-i\omega t} + \frac{1}{2}\sigma^{(3)}\vec{\mathcal{E}}_0\vec{\mathcal{E}}_0^*\vec{\mathcal{E}}_0 e^{-i\omega t} + \text{higher order.} \quad (1)$$

Here, $\sigma^{(1)}$ is the linear conductivity tensor while $\sigma^{(3)} = \sigma^{(3)}(\omega; \omega, -\omega, \omega)$ is the third-order contribution, also denoted the optical Kerr effect. Both conductivities are sheet quantities obtained by integrating currents over the perpendicular (z)

dimension. Up to third order, the absorbed power density is $\mathcal{P}_{\text{abs}} = \frac{1}{2}\text{Re}\{\vec{\mathcal{E}}_0^*\sigma^{(1)}\vec{\mathcal{E}}_0 + \vec{\mathcal{E}}_0^*\sigma^{(3)}\vec{\mathcal{E}}_0\vec{\mathcal{E}}_0^*\vec{\mathcal{E}}_0\}$. For transmission through a fully isotropic system (scalar conductivities), the reduction of the optical intensity $I = \frac{1}{2}\varepsilon_0 c |\mathcal{E}_0|^2$ is determined by the transmittance $T = 1 - \mathcal{P}_{\text{abs}}/I = 1 - \alpha - \beta I$ with $\alpha = \text{Re}\sigma^{(1)}/(\varepsilon_0 c)$ and $\beta = 2\text{Re}\sigma^{(3)}/(\varepsilon_0 c)^2$ the single- and two-photon absorptance, respectively.

We apply the length-gauge excitonic density matrix approach developed in Refs. [18,19]. In those works, the only third-order process considered was third-harmonic generation $\mathcal{J}_{3\omega}$. Hence, for the present purpose, we extend our previous work by collecting the third-order response at the fundamental frequency, which is given by

$$\begin{aligned} \sigma_{\mu\eta\alpha\beta}^{(3)}(\omega) = & -\frac{e^4\hbar}{m^2 A} \sum_{m,n,l} \frac{X_{0l}^\alpha \Pi_{ln}^\eta Q_{nm}^\mu X_{m0}^\beta [\hbar^2\omega^2(4E_m + E_n) + \hbar\omega(E_l E_n - E_m E_n - 4E_l E_m) + 3E_l E_m E_n]}{(\hbar^2\omega^2 - E_l^2)(\hbar^2\omega^2 - E_m^2)(2\hbar\omega - E_n)E_n} \\ & + \frac{e^4\hbar}{m^2 A} \sum_{m,n,l} \frac{\Pi_{0l}^\mu Q_{ln}^\eta Q_{nm}^\alpha X_{m0}^\beta [\hbar\omega(4E_m - E_n) - 3E_m E_n]}{(\hbar\omega - E_l)(\hbar^2\omega^2 - E_m^2)(2\hbar\omega - E_n)E_n} + (\omega \rightarrow -\omega)^*. \end{aligned} \quad (2)$$

In this expression, greek indices indicate Cartesian directions, A is the sample area, E_i is the energy of the i th state measured relative to the ground state E_0 , and m, n , and l each run over all states including the ground state. In general, care should be taken due to the divergence encountered when $n = 0$ (see Ref. [22]). This term, however, does not contribute to TPA at finite frequency and, hence, can be ignored. The matrix elements are dipole type:

$$X_{0m}^\alpha = \sum_{\vec{k}} \frac{i p_{vc\vec{k}}^\alpha}{E_{cv\vec{k}}} \psi_{\vec{k}}^{(m)}, \quad Q_{mn}^\beta = i \sum_{\vec{k}} \psi_{\vec{k}}^{(m)*} [\psi_{\vec{k}}^{(n)}]_{;k_\beta}, \quad (3)$$

as well as momenta Π_{ij}^α . Here, $[\psi_{\vec{k}}^{(n)}]_{;k_\beta} \equiv d\psi_{\vec{k}}^{(n)}/dk_\beta - i(\Omega_{cc\vec{k}}^\beta - \Omega_{vv\vec{k}}^\beta)\psi_{\vec{k}}^{(n)}$ denotes the generalized derivative of the n th exciton wave function $\psi_{\vec{k}}^{(n)}$ expressed in terms of the Berry connection $\vec{\Omega}_{nm\vec{k}}$ [23]. Also, $p_{vc\vec{k}}^\alpha$ is a Cartesian component of the interband momentum matrix element and $E_{cv\vec{k}}$ is the corresponding transition energy for a pair of valence (v) and conduction (c) bands. The momenta are converted to dipole matrix elements using [19] $\Pi_{0l}^\mu = -iE_l X_{0l}^\mu$ and $\Pi_{ln}^\mu = i(E_l - E_n)Q_{ln}^\mu$. In TMDs, the nonvanishing third-order tensor elements are related via $\sigma_{xxxx}^{(3)} = \sigma_{xyxy}^{(3)} + \sigma_{xyyx}^{(3)} + \sigma_{yyxx}^{(3)}$ in addition to an identical set with $x \leftrightarrow y$ interchange. With circular polarization $\vec{\mathcal{E}}_0 = (\hat{x} \pm i\hat{y})\mathcal{E}_0/\sqrt{2}$, the response is $\vec{\mathcal{J}}_\omega = \sigma_{xxxx}^{(3)}(\hat{x} \pm i\hat{y})\mathcal{E}_0^3$. In the degenerate (i.e., monochromatic) case, such as the present, only the sum $\sigma_{xyxy}^{(3)} + \sigma_{xyyx}^{(3)} + \sigma_{yyxx}^{(3)}$ has physical meaning and we, therefore, need only consider the diagonal response

$$\sigma_{xxxx}^{(3)}(\omega) = \frac{ie^4\hbar}{m^2 A} \sum_{m,n,l} \frac{X_{0l}^x Q_{ln}^x Q_{nm}^x X_{m0}^x F_{mnl}(\omega)}{(\hbar^2\omega^2 - E_l^2)(\hbar^2\omega^2 - E_m^2)(2\hbar\omega - E_n)E_n} + (\omega \rightarrow -\omega)^*, \quad (4)$$

where

$$F_{mnl}(\omega) = \hbar^2\omega^2(8E_m E_l - 4E_m E_n - E_n^2) + \hbar\omega(E_m - E_l)E_n^2 - 3E_l E_m E_n^2. \quad (5)$$

From this expression, it is clear that both single-photon ($\hbar\omega = E_m$ or E_l) and two-photon resonances ($2\hbar\omega = E_n$) contribute to the response. For the present purposes, we focus on the two-photon resonance. Replacing E_n by $2\hbar\omega$ in the real part of the TPA response means that $F_{mnl}(\omega) = -4\hbar^2\omega^2(E_m + \hbar\omega)(E_l + \hbar\omega)$, such that

$$\text{Re}\sigma_{xxxx}^{(3)}(\omega) = \frac{2\pi e^4\hbar^2\omega}{m^2 A} \sum_n \left| \sum_m \frac{X_{0m}^x Q_{mn}^x}{E_m - \hbar\omega} \right|^2 \delta(E_n - 2\hbar\omega). \quad (6)$$

This relation is equivalent to derivations based on Fermi's "golden rule" [24]. In practical applications, we include line-shape broadening by adding a small imaginary part to the frequency $\omega \rightarrow \omega + i\Gamma$ and replace the Dirac delta function by a Lorentzian:

$$\sigma_{xxxx}^{(3)}(\omega) = \frac{2ie^4\hbar^2\omega}{m^2 A} \sum_n \left| \sum_m \frac{X_{0m}^x Q_{mn}^x}{E_m - \hbar\omega} \right|^2 \frac{1}{E_n - 2\hbar\omega}, \quad (7)$$

where ω is now understood to be complex. In this expression, m is the intermediate state while n is the final state. In the cylindrically symmetric Wannier approximation, the ground

state zero is only dipole coupled to s type excited states, which determine the SPA response. In TPA, the intermediate state m would be s type and, in turn, the final state n must be p type. As explained in Ref. [20] and demonstrated below, such simple selection rules apply only approximately due to trigonal warping.

III. EXCITONIC RESPONSE

The effective four-band model introduced in Ref. [25] captures the essential physics of monolayer TMDs in the low-energy regime. Briefly, a two-site hexagonal lattice with spin-orbit coupling is considered, leading to four bands with band gaps located at the K and K' Dirac points. Moreover, states are diagonal in spin such that, in fact, the model reduces to two decoupled two-band systems, one for each spin. However, only the combined four bands preserve time-reversal symmetry. The simplicity of the model makes it feasible to incorporate exciton effects with a dense k -point grid. Based on this band structure, we model exciton states by solving the Bethe-Salpeter (BSE) equation using a 100×100 k -point grid. We describe electron-hole interactions using the Keldysh potential [26] with a nonlocal screening $\varepsilon(q) = \kappa + r_0q$, where κ is the average of sub- and superstrate dielectric constants and r_0 is the TMD screening length. A slightly modified version of this model has been found to apply to TMD heterostructures as well [27]. Different values of κ represent (1) suspended TMDs $\kappa = 1$, (2) TMDs on SiO₂ substrates $\kappa = 1.55$, (3) hBN encapsulated TMDs $\kappa = 4.5$, and (4) free-carrier limit $\kappa = 100$. The large κ value applied to simulate free-carrier response can be compared to the analytic free-carrier limit derived in the Appendix.

Before discussing the optical response, we investigate the exciton levels in the representative TMDs MoS₂ and WSe₂. The former has a moderate SO splitting, while SO coupling in the latter is sufficient to shift B excitons into the continuum formed by A excitons. The band structure follows from diagonalizing the Hamiltonian [25]:

$$\vec{H} = \begin{pmatrix} \Delta + \sigma\lambda g & -\gamma f \\ -\gamma f^* & -\Delta - \sigma\lambda g \end{pmatrix} \quad (8)$$

with

$$\begin{aligned} f(\vec{k}) &= e^{ik_x a/\sqrt{3}} + 2e^{-ik_x a/2\sqrt{3}} \cos(k_y a/2), \\ g(\vec{k}) &= 2\{\sin(k_x a\sqrt{3}/2 + k_y a/2) - \sin(k_y a) \\ &\quad - \sin(k_x a\sqrt{3}/2 - k_y a/2)\}. \end{aligned} \quad (9)$$

Here, $\pm\Delta$ is the staggered on-site potential of the two sublattices, λ is the effective spin-orbit coupling, γ is the nearest-neighbor hopping parameter, and a is the lattice constant. The band gaps of the two valleys are $E_K = 2\Delta - \sigma 6\sqrt{3}\lambda$ and $E_{K'} = 2\Delta + \sigma 6\sqrt{3}\lambda$, respectively. Hence, for spin-up electrons ($\sigma = 1$), the true band gap $E_g = 2\Delta - 6\sqrt{3}\lambda$ is at the K point. All parameters are taken from Ref. [28] with the exception that we use $\Delta = 1.08\text{eV}$ for WSe₂ to match the experimental band gap. Also, in Ref. [28], Δ was determined without considering spin-orbit coupling, i.e., using $\Delta \approx E_g/2$. Hence, to ensure a more accurate value, a shift $\Delta \rightarrow \Delta + 3\sqrt{3}\lambda$ is applied in the present paper. Finally, the

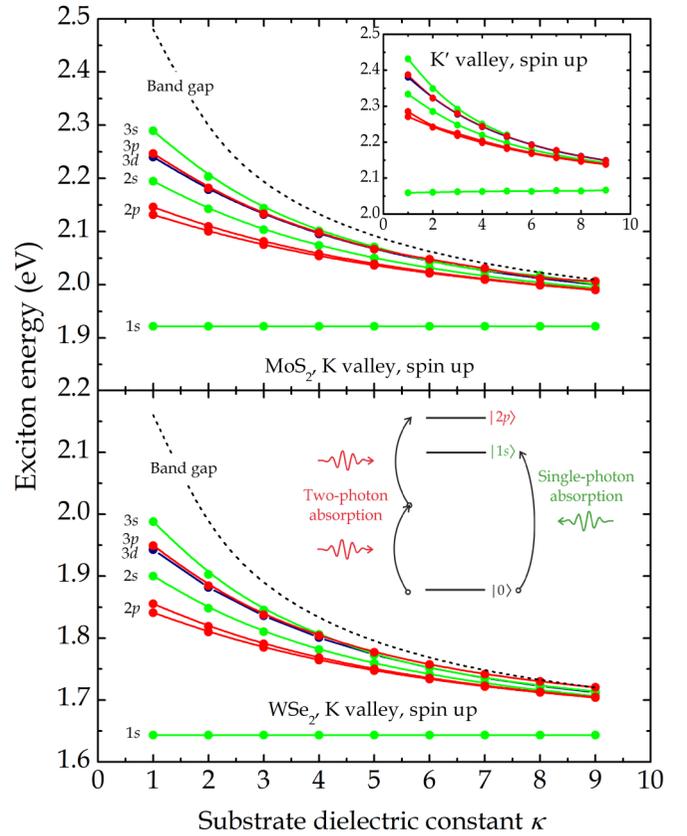


FIG. 1. Exciton levels in MoS₂ (upper panel) and WSe₂ (lower panel) vs substrate screening. Main panels show K valley spin-up spectra, while the upper inset depicts the corresponding K' states for MoS₂. The lower inset is a schematic illustration of TPA and SPA processes and their distinct selection rules.

role of a screening environment κ on the band gap must be addressed. Physically, the band gap contains a quasiparticle contribution, which is sensitive to κ . The bare band-gap parameter Δ corresponds to the suspended case $\kappa = 1$. To correct for screening, we will therefore rely on the experimental observation that the fundamental 1s exciton resonance is largely independent of κ , i.e., occurs at approximately the same energy irrespective of substrate [29]. This can be seen as a cancellation between shifts of quasiparticle and exciton binding energies. Hence, for $\kappa \neq 1$ we adjust Δ to keep the fundamental 1s exciton at the energy found in the unscreened case $\kappa = 1$.

The exciton energy levels of MoS₂ and WSe₂ are illustrated in Fig. 1 as functions of surrounding screening κ . The fundamental 1s exciton is located at approximately 1.92 and 1.64 eV in the two materials, respectively. Since the present model does not have exact rotational symmetry, angular momentum assignment, i.e., s , p , d , etc., is only approximate. In practice, we assign an angular momentum l by projecting exciton wave functions on circular eigenstates $\exp(il\theta)$ and locate the largest projection. Two distinct 2p states ($2p^\pm$) are found in this manner. These would have been degenerate in case of perfect circular symmetry but are split by trigonal warping. Above these, 2s and 3d states appear, with the latter nearly degenerate with 3p states. The corresponding states

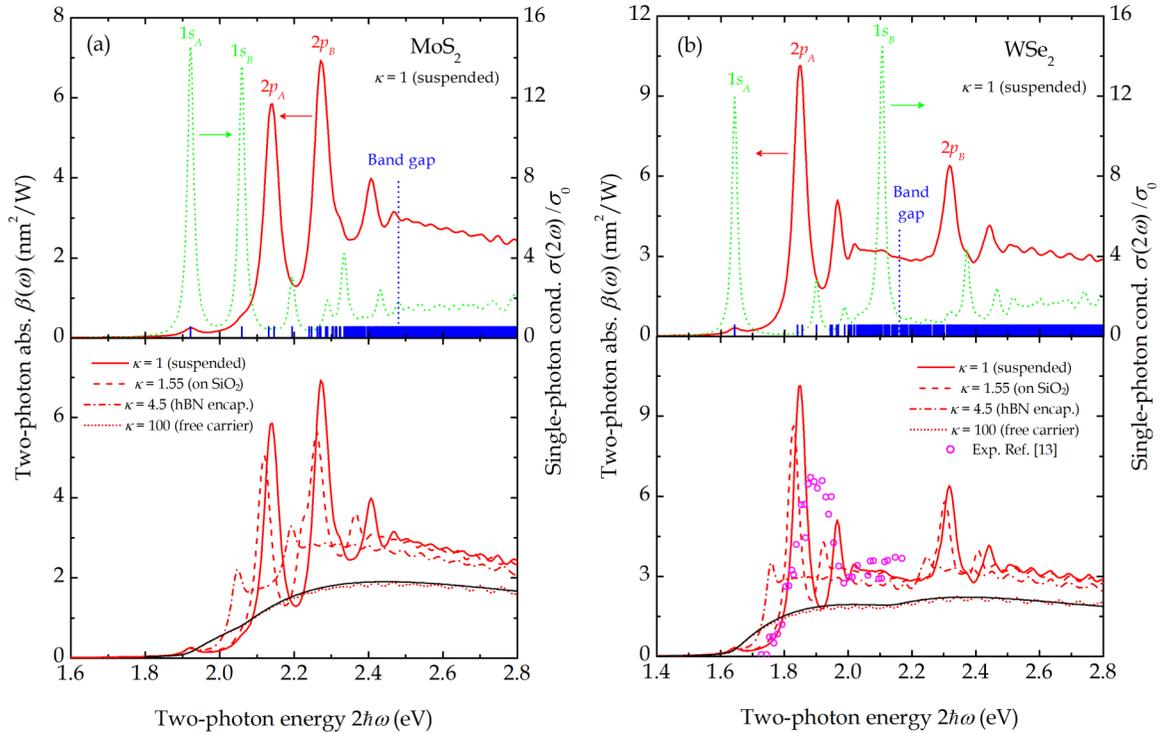


FIG. 2. SPA (right axes) and TPA (left axes) spectra of MoS_2 (a) and WSe_2 (b). Designations are included for the dominant resonances including angular momentum and valley. In the top panels, vertical blue lines indicate exciton levels and the dashed lines show the band gap. In the lower panels, the impact of substrate screening is illustrated in four cases. The solid black curves are analytical free-carrier spectra. For WSe_2 , an experimental spectrum from Ref. [13] is added (circles).

in the K' valley, as shown in the upper inset in Fig. 1, are similar but shifted upwards by the spin-orbit splitting. Also, if spin-down states are considered, K and K' valley states are interchanged.

The TPA process is shown schematically in the lower inset of Fig. 1. Because of the intermediate state, the selection rules are opposite to those of the SPA process. Hence, whereas SPA is dominated by s states, the final state in TPA is predominantly p type. With perfectly circular symmetry (and within the dipole approximation), these selection rules would be exact. Trigonal warping, however, relaxes symmetry requirements and in TPA small contributions from s and d states are expected. In Fig. 2, we compare SPA and TPA spectra for MoS_2 and WSe_2 . In both cases, a line broadening of $\hbar\Gamma = 10$ meV is applied. The single-photon response is reported in terms of the real part of the conductivity σ calculated in the standard manner [18,19] and normalized by the conductivity quantum $\sigma_0 = e^2/4\hbar$. Pronounced resonances in the SPA spectra at the $1s_A$ and $1s_B$ states of the two valleys are observed but higher s states are visible as well. Despite trigonal warping, it is seen that p states have little intensity in the linear response. In contrast, TPA spectra are dominated by p states. However, a small but clearly visible peak at the $1s_A$ energy is observed, as are shoulders on the high-energy side of $2p_B$. These features are consequences of trigonal warping, as demonstrated below.

The lower panels in Fig. 2 illustrate the evolution of the TPA spectra as screening by the surroundings is increased from the vacuum limit $\kappa = 1$ to the free-carrier limit $\kappa \gg 1$,

where exciton effects are completely suppressed. We also include two technologically relevant cases, namely, $\kappa = 1.55$ and 4.5, that describe screening by SiO_2 substrates and hBN encapsulation, respectively. We stress that, in all cases, the $1s_A$ resonance is kept fixed at the unscreened value. As a consequence, all higher excitons redshift as screening is increased (see Fig. 1). With the assumed broadening, discrete p resonances are still clearly visible even in cases of substantial screening such as hBN encapsulation. However, the background free-carrier contribution becomes increasingly dominant with increased screening.

Figure 2 also includes a comparison of theory with experimental TPA data for WSe_2 from Ref. [13] since no data for monolayer MoS_2 exist. The measurements are for WSe_2 on SiO_2 substrates and reported in terms of $\text{Im}\{\chi^{(3)}\}$, which is equivalent to $\beta(\omega)/\omega$. In the figure, experimental data have been scaled to match theory near the low-energy side of the main resonance around $2\hbar\omega = 1.8$ eV. In this spectral range, experiment and calculation are in good agreement. However, we note that the measured main resonance does not coincide with the calculated $2p_A$ position, in contradiction to theory. Rather, a broad resonance overlapping with both $2p_A$ and $3p_A$ is observed without a discernable dip in between. A similar conclusion was reached in Ref. [13] by applying the SPA spectrum to locate the band gap and a Keldysh model similar to the present one to compute p -state energies. We speculate that the discrepancy is at least partly due to anomalous homogeneous and inhomogeneous broadening that is not accounted for in the model, in addition to relatively poor spectral resolution

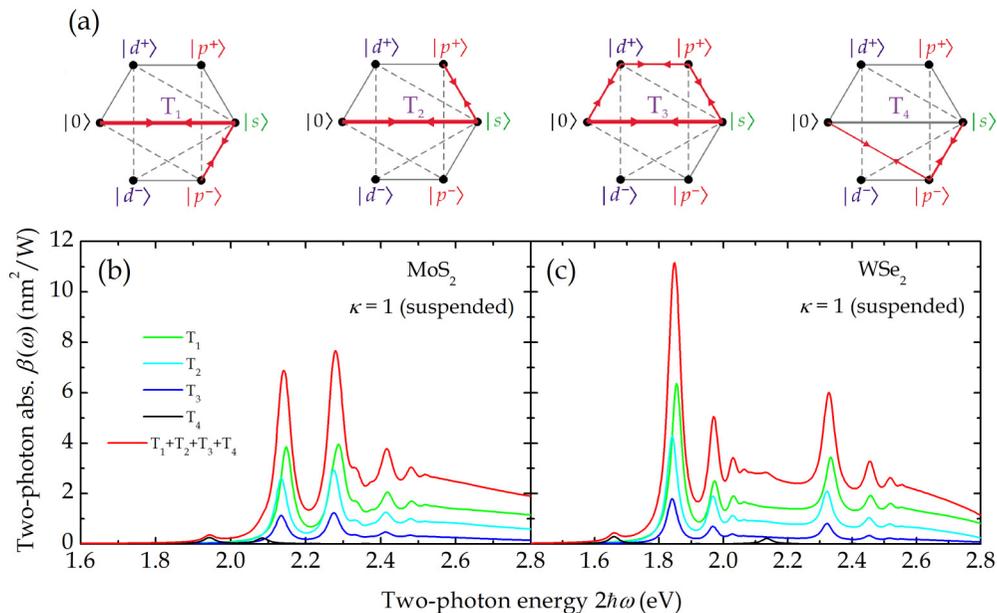


FIG. 3. Four dominant TPA paths in the K valley (a) with participating transitions shown in red with line thickness indicating intensity. (b), (c) TPA spectra of suspended MoS₂ (b) and WSe₂ (c) decomposed into dominant paths.

in the experiment. It is well known that a simple model with fixed broadening cannot quantitatively account for the SPA spectrum [30]. Thus, distinct $2s$ resonances are predicted from theory if the empirical $1s$ broadening is adopted for the full SPA spectrum. In contrast, measured $2s$ resonances only show up as very weak spectral features. This discrepancy can be largely resolved by including energy-dependent broadening mechanisms involving quasiparticle effects [30]. It is possible that similar state-dependent broadening could improve the agreement in the TPA case as well. Finally, we note that the magnitude of the TPA response, even at resonance, is somewhat smaller than the single-frequency absolute values reported in Refs. [16,17] for monolayer MoS₂.

In the limit of infinite screening, the excitonic theory should agree with usual free-carrier expressions. In practice, a formal demonstration of this agreement can be complicated, however. A useful strategy to this end is to express the response in terms of exciton Green's functions that become diagonal in k space if Coulomb effects are ignored. In the Appendix, we show that this procedure leads to a simple integral over k space [see Eq. (A3)]. Moreover, an excellent analytical approximation to the integral is given in Eqs. (A5) and (A6). These agree with previous results for gapped graphene [31] and semiconductor quantum wells [32]. In Fig. 2, we compare the results of Eq. (A3) with the numerical solution of the exciton model using very large screening $\kappa = 100$. These are shown as solid black and dotted red lines, respectively, and show excellent agreement. The exciton model therefore clearly converges to the correct free-carrier limit as screening increases.

The TPA response in TMDs directly probes a set of transitions that are not readily accessible in SPA. Hence, it is a potential source of important information about exciton states, in addition to the previously documented ability to estimate exciton binding energies [11–15]. To discuss the contributions from distinct excitons it is useful to consider

the semianalytical k -space model developed in Ref. [19]. In this model, excitons are expanded in states of definite angular momentum $\psi_l(k, \theta) = \phi_l(k) \exp(il\theta)$ that couple to $\psi_{l\pm 1}(k, \theta)$. States with $l = \{-3, -2, -1, 0, 1\}$ in the K valley are identified as $\{d^-, p^-, s, p^+, d^+\}$, respectively. A similar designation applies to the K' valley under the exchange $l^\pm \rightarrow -l^\mp$. Due to the coupling between angular momenta, however, the l designation is not exact but only represents the dominant character, and the degeneracy between l^\pm states is raised. In addition, optical excitation produces a complicated landscape of couplings, with trigonal warping responsible for coupling between angular momenta l and $l \pm 2, l \pm 4$. As a consequence, K valley dipole matrix elements between the ground state zero and $\{p^-, d^+\}$ become weakly allowed. When summed over spin, the two valley contributions are identical.

In Fig. 3, we decompose the full TPA response into dominant contributions using the semianalytical model. Exciton states are computed by discretizing on a radial k -grid writing [33] $k = \tan(\frac{\pi}{2}x)$ and applying a 300 point Gauss-Laguerre quadrature grid for $x \in [0, 1]$. For angular momentum l , Coulomb matrices of the form $D_l(k, q) = \int_0^{2\pi} \cos(l\theta) v[(k^2 + q^2 - 2kq \cos \theta)^{1/2}] d\theta$ with interaction $v(q) = 1/[q(\kappa + r_0q)]$ are required. We compute these by expanding in partial fractions $v(q) = 1/(\kappa q) - r_0/[\kappa(\kappa + r_0q)]$ such that the first term can be integrated analytically in terms of elliptic integrals, while the second is regular and easily handled using numerical integration. The first term has singular diagonal entries $D_l(k, k)$, which we avoid by using $D_l(k, k) \approx 2[1 + \ln(16k/\Delta k)]/(\kappa k)$ obtained by averaging over $q \in [k, k + \Delta k]$. This procedure ensures very fast convergence of exciton energy with k -grid size.

Comparing Figs. 2 and 3, it is immediately clear that the semianalytical model is in excellent agreement with the BSE results. Importantly, this allows us to assess the contributions from specific angular momenta in the TPA response.

In Fig. 3(a), the dominant transition paths are illustrated. The intense $|0\rangle \rightarrow |s\rangle$ transition is shown as a thick line. Similarly, weaker transitions with $\Delta l = \pm 1$ that do not require trigonal warping are indicated by thin lines. Finally, transitions with $\Delta l = \pm 2, \pm 4$ that only arise when angular warping is included are shown as thin dashed lines. Overall, the dominant TPA paths T_1 and T_2 proceed with $|s\rangle$ as the intermediate state and $|p^\pm\rangle$ as final states. From their decomposition it is seen that their resonances occur at slightly different two-photon energies, in agreement with the lifted degeneracy between the two p states.

From Eq. (7) it is tempting to conclude that all TPA resonances can be attributed to a single pair of intermediate m and final n states. However, it must be remembered that the sum over m runs over *all* intermediate states. Because the sum is squared it is perfectly possible for transition paths such as $|0\rangle \rightarrow |s\rangle \rightarrow |p^+\rangle \rightarrow |d^+\rangle \rightarrow |0\rangle$ and its reversed version to appear when the intermediate sum includes both s and d states. These two paths make up transition path T_3 , which adds a significant contribution to the full response. Finally, trigonal warping allows for transition path T_4 , in which the intermediate and final states are $|p^-\rangle$ and $|s\rangle$, respectively. These states are identical to those of T_1 , but taken in the opposite order. In particular, X_{0p^-} is only nonzero in the presence of trigonal warping, which makes the diagram inherently weak. Importantly, however, two-photon resonances appear at $1s_A$ and $1s_B$ for this diagram. It follows that weak TPA resonances are expected at the location of intense SPA peaks, in contrast to predictions based on cylindrical symmetry. These peaks agree very well with the full BSE spectra in Fig. 2. This is different from the case of third-harmonic generation, for which the T_4 diagram is negligible [19].

IV. WANNIER MODEL

We now apply the TPA model to Wannier excitons rather than full Bethe-Salpeter states. Wannier excitons are expected to be accurate approximations in cases of large screening. Moreover, the Wannier model readily allows for inclusion of other electron-hole complexes, in particular, biexcitons. Such four-particle states are prohibitively complicated at the Bethe-Salpeter level. A Wannier exciton in position space can be defined as

$$\psi_m(\vec{r}) = \frac{1}{\sqrt{A}} \sum_{\vec{k}} \psi_{\vec{k}}^{(m)} e^{i\vec{k}\cdot\vec{r}}, \quad (10)$$

with the inverse $\psi_{\vec{k}}^{(m)} = A^{-1/2} \int \psi_m(\vec{r}) e^{-i\vec{k}\cdot\vec{r}} d^2r$. Physically, \vec{r} describes the relative electron-hole coordinate and the prefactor ensures that $\psi_m(\vec{r})$ is normalized. To estimate exciton matrix elements, we now approximate transition energies and momenta by the values at the band gap, i.e., $E_{cv\vec{k}} \approx E_g$ and $p_{vc\vec{k}} \approx mv_F$ with Fermi velocity $v_F = \sqrt{3}a\gamma/(2\hbar)$. Hence, using a few elementary manipulations, it follows that

$$X_{0m}^x \approx \frac{imv_F}{E_g} \sqrt{A} \psi_m(0), \quad Q_{mn}^x = \int \psi_m^*(\vec{r}) x \psi_n(\vec{r}) d^2r. \quad (11)$$

The interpretation of the first of these is that electron and hole positions must coincide in optical interband transitions. Under the assumptions of isotropic effective-mass dispersion,

vanishing exchange interaction, and local screening $\varepsilon(q) \approx \kappa_{\text{eff}}$ with effective dielectric constant κ_{eff} , Wannier excitons in position space become 2D hydrogenlike eigenstates. For a 2D hydrogenlike $1s$ state $\psi_{1s}(\vec{r}) = \sqrt{8/\pi} \exp\{-2r/a_X\}/a_X$ with effective Bohr radius a_X , we get $\psi_m(0) = \sqrt{8/\pi}/a_X$. A more accurate result for X_{0m}^x is obtained if the full effective mass dispersion $E_{cv\vec{k}} \approx E_g(1 + k^2\rho^2/4)$ with $\rho = \sqrt{8}\hbar v_F/E_g$ is retained and leads to multiplication by an additional factor $1/(1 - \rho^2/a_X^2) - \rho/a_X \cos^{-1}(\rho/a_X)/(1 - \rho^2/a_X^2)^{3/2}$. Moreover, using explicit 2D hydrogenlike bound np states with $n = 2, 3, 4, \dots$ demonstrates that

$$Q_{1s,np}^x = a_X \frac{(n-1)^{n-5/2} (n - \frac{1}{2})^{5/2}}{n^{n+3/2}}. \quad (12)$$

While these expressions are less accurate than full Bethe-Salpeter results, a great advantage of the Wannier approximation is the straightforward inclusion of multiparticle electron-hole complexes such as trions and biexcitons. In particular, generalizing the Wannier exciton model, biexcitons are modeled as two electrons ($e1$ and $e2$) and two holes ($h1$ and $h2$) with effective-mass energy dispersion and interaction via a locally screened Coulomb potential. After elimination of the center of mass, these four-particle complexes are described by relative vectors $\vec{r} = \vec{r}_{e1} - \vec{r}_{h1}$, $\vec{r}' = \vec{r}_{e2} - \vec{r}_{h2}$ and $\vec{R} = \vec{r}_{h1} - \vec{r}_{h2}$. We consider a biexciton state formed by two m excitons $\psi_{mm}(\vec{r}, \vec{r}', \vec{R})$ and apply the Heitler approximation [34] $\psi_{mm}(\vec{r}, \vec{r}', \vec{R}) \approx \psi_m(\vec{r})\psi_m(\vec{r}')\Psi(\vec{R})$. In the Wannier model, the exciton-biexciton matrix element is found by setting one electron-hole separation to zero and integrating over the remaining degrees of freedom, i.e., [34]

$$\begin{aligned} X_{m,mm}^x &\approx \frac{imv_F}{E_g} \int \psi_m^*(\vec{r}) \psi_{mm}(\vec{r}, 0, \vec{R}) d^2r d^2R \\ &\approx \frac{imv_F}{E_g} \psi_m(0) \int \Psi(\vec{R}) d^2R. \end{aligned} \quad (13)$$

The remaining integral can be estimated assuming a Gaussian hole-hole distribution

$$\Psi(\vec{R}) = \frac{1}{\sqrt{\pi}R_{hh}} \exp\left\{-\frac{R^2}{2R_{hh}^2}\right\}, \quad (14)$$

where R_{hh} is the hole-hole root-mean-square distance, i.e., $\langle \Psi | R^2 | \Psi \rangle = R_{hh}^2$. In this way, $X_{m,mm}^x \approx 2\sqrt{\pi} imv_F \psi_m(0) R_{hh}/E_g$. The fact that matrix elements are proportional to R_{hh} is a signature of the ‘‘giant oscillator strength’’ effect associated with biexcitons. In fact, the result is ‘‘pathological’’ in the sense that $X_{m,mm}$ diverges in the limit of vanishing biexciton binding, i.e., as $R_{hh} \rightarrow \infty$ [35]. A more satisfactory model is based on the bipolariton concept, which explicitly accounts for Pauli-principle effects among the two electrons and two holes in the four-particle complex [35]. Effectively, the bipolariton model adds a multiplicative factor $\Delta E_{mm}/(2E_m - 2\hbar\omega)$, with $\Delta E_{mm} \equiv 2E_m - E_{mm}$ the biexciton binding energy, to the matrix elements [35]. The magnitude of this factor is unity exactly at resonance (for real-valued frequencies). However, the pathological limit at vanishing biexciton binding is removed due to the factor ΔE_{mm} . Hence, proper account of antisymmetry among identical fermions is clearly important in biexcitons. We

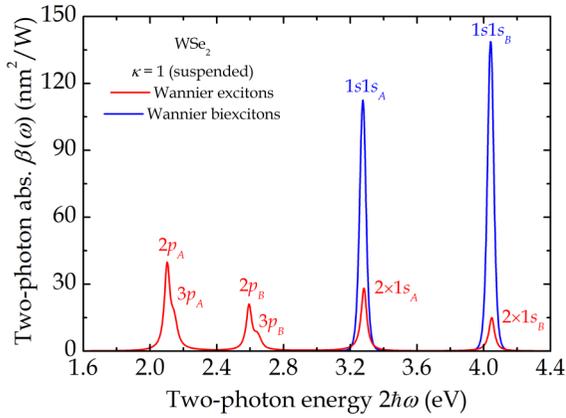


FIG. 4. TPA spectrum of suspended WSe₂ in the Wannier model split into exciton (red) and biexciton (blue) contributions.

emphasize that Pauli-principle effects do not affect excitons, since electrons and holes are distinguishable.

The path having m excitons and mm biexcitons as intermediate and final states, respectively, leads to the TPA contribution

$$\sigma_{xxxx}^{(3,\text{biex})}(\omega) = \frac{2ie^4\hbar^2\omega}{m^2A} \sum_m \left| \frac{X_{0m}^x X_{m,mm}^x}{E_m - \hbar\omega} \right|^2 \frac{1}{E_{mm} - 2\hbar\omega}. \quad (15)$$

Eventually, including only the lowest exciton $1s$ and its associated biexciton $1s1s$,

$$\sigma_{xxxx}^{(3,\text{biex})}(\omega) = \frac{8ie^4m^2\hbar^2\omega v_F^4}{\pi E_g^4 a_X^4} \frac{R_{hh}^2 \Delta E_{1s1s}^2}{|E_{1s} - \hbar\omega|^4 (E_{1s1s} - 2\hbar\omega)}. \quad (16)$$

Since $E_{1s1s} \sim 2E_{1s}$, this term can give a large (multiply resonant) contribution near $\hbar\omega = E_{1s}$. In the numerical results below, we use a slightly increased broadening of 20 meV in the biexciton response to account for the increased dephasing of these states. We apply the above results to suspended WSe₂, for which accurate computed values of R_{hh} and ΔE_{1s1s} (46.9 Å and 20.2 meV, respectively) exist [36]. We determine κ_{eff} by matching to the Bethe-Salpeter $1s$ energy, which, in turn, leads to $a_X = a_0\kappa_{\text{eff}}/\mu_r$ with a_0 the hydrogen Bohr radius and $\mu_r = E_g/(4mv_F^2)$ the relative reduced mass.

In Fig. 4, we illustrate the contributions from both Wannier excitons and biexcitons. As a consistency check, we have verified that the Wannier results agree with the semianalytical k -space model used above, provided trigonal warping, band nonparabolicity, and nonlocal screening are ignored. Comparing first the exciton spectrum to the Bethe-Salpeter one in Fig. 2(b), several differences are noted. Even though $1s$ energies agree by construction, p -state energies are overestimated by the Wannier model. This is a consequence of the well-known nonhydrogenic nature of the Keldysh potential model, which leads to increased p -state binding energies compared to hydrogenlike results. In addition, the intensity of $2p$ and $3p$ resonances in Wannier TPA spectra is overestimated, partly because the energy denominator $E_{1s} - \frac{1}{2}E_{np}$ in the sum over intermediate states is too small. Finally, since unbound p states are omitted, the high-energy continuum background is absent. The exciton spectrum in Fig. 4 also

feature peaks at twice the $1s$ energies. These are single-photon resonances in the transition between ground and intermediate state. Obviously, the simplified Eq. (6) cannot be applied in this frequency range and the full frequency dependence of Eqs. (4) and (5) must be retained.

Turning now to the biexciton contribution in Fig. 4, these are seen to add substantial peaks to the TPA response but only in the frequency range of single-photon resonances, i.e., at energy $\hbar\omega \approx E_{1s}$ and higher. This is expected, as the final-state energy is close to twice the energy of the intermediate $1s$ state. In fact, the slight redshift relative to $2E_{1s}$ due to the biexciton binding energy is discernable in Fig. 4. Notably, the intensity of biexciton peaks is quite large as a consequence of the multiply resonant nature [see Eq. (16)]. Hence, biexciton effects are of importance in the two-photon response of TMDs similarly to observations in traditional semiconductors. We stress, however, that their contribution is in the high-energy range of single-photon resonances and, hence, does not interfere with the low-energy range of two-photon resonances studied with the Bethe-Salpeter model in the previous section.

V. SUMMARY

In summary, we modeled the TPA response due to excitons in monolayer transition-metal dichalcogenides using three approaches: the full numerical Bethe-Salpeter equation, the k -space based semianalytical trigonal warping model, and the analytical Wannier approximation. The distinct TPA selection rules mean that otherwise dark equal-parity transitions dominate the response. Combined with single-photon spectra, this provides access to exciton series with different angular

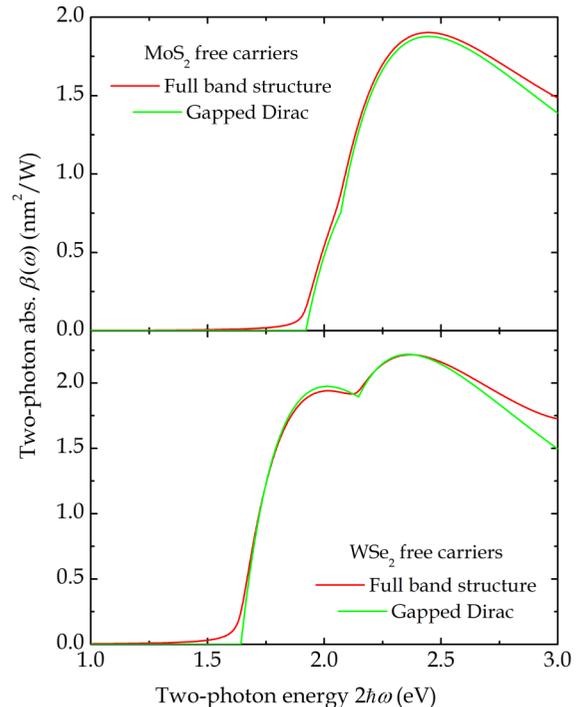


FIG. 5. Two-photon absorption without excitonic effects for MoS₂ (top) and WSe₂ (bottom). Red and green curves show results of numerical integration and analytic approximations, respectively.

momenta, including dark exciton binding energies. In fact, trigonal warping means that odd-parity transitions are weakly allowed in TPA spectra as well. We identify the dominant transitions and the role of trigonal warping in their intensity. The effect of dielectric screening including relevant experimental cases and the free-carrier limit is carefully analyzed. Moreover, a brief comparison with recent experiments for WSe₂ is made. Finally, biexciton transitions are described within the Wannier approach and shown to contribute significantly to the high-energy TPA response.

ACKNOWLEDGMENT

The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project No. DNR103.

APPENDIX: FREE-CARRIER TPA RESPONSE

In this appendix, we establish the free-carrier limit of the general excitonic expression. This limit is appropriate whenever screening is sufficiently large that Coulomb interaction can be ignored. It follows that excitonic theory should describe free-carrier response in this limit. A very convenient way of extracting the limit is by introducing the exciton Green's function [23]

$$G_{\vec{k},\vec{k}'}^{(\omega)} = \sum_n \frac{\psi_{\vec{k}}^{(n)} \psi_{\vec{k}'}^{(n)*}}{E_n - \hbar\omega} \quad (\text{A1})$$

that becomes $G_{\vec{k},\vec{k}'}^{(\omega)} \rightarrow \delta_{\vec{k},\vec{k}'} (E_{cv\vec{k}} - \hbar\omega)^{-1}$ in the free-carrier limit. Applying Eq. (3), it then follows that

$$\sum_m \frac{X_{0m}^\alpha Q_{mn}^\beta}{E_m - \hbar\omega} = - \sum_{\vec{k},\vec{k}'} \frac{P_{vc\vec{k}}^\alpha}{E_{cv\vec{k}}} G_{\vec{k},\vec{k}'}^{(\omega)} [\psi_{\vec{k}'}^{(n)}]_{;k_\beta} \rightarrow - \sum_{\vec{k}} \frac{P_{vc\vec{k}}^\alpha [\psi_{\vec{k}}^{(n)}]_{;k_\beta}}{E_{cv\vec{k}} (E_{cv\vec{k}} - \hbar\omega)}, \quad (\text{A2})$$

where the last expression is the free-carrier limit. Hence,

$$\begin{aligned} \sigma_{xxxx}^{(3)}(\omega) &\rightarrow \frac{2ie^4 \hbar^2 \omega}{m^2 A} \sum_{\vec{k},\vec{k}'} \frac{P_{vc\vec{k}}^x}{E_{cv\vec{k}} (E_{cv\vec{k}} - \hbar\omega)} \frac{P_{cv\vec{k}'}^x}{E_{cv\vec{k}'} (E_{cv\vec{k}'} - \hbar\omega)} \sum_n \frac{[\psi_{\vec{k}}^{(n)}]_{;k_x} [\psi_{\vec{k}'}^{(n)*}]_{;k'_x}}{E_n - 2\hbar\omega} \\ &= \frac{2ie^4 \hbar^2 \omega}{m^2 A} \sum_{\vec{k},\vec{k}'} \frac{P_{vc\vec{k}}^x}{E_{cv\vec{k}} (E_{cv\vec{k}} - \hbar\omega)} \frac{P_{cv\vec{k}'}^x}{E_{cv\vec{k}'} (E_{cv\vec{k}'} - \hbar\omega)} [G_{\vec{k},\vec{k}'}^{(2\omega)}]_{;k_x;k'_x} \rightarrow \frac{2ie^4 \hbar^2 \omega}{m^2 A} \sum_{\vec{k}} \left| \left(\frac{P_{vc\vec{k}}^x}{E_{cv\vec{k}} (E_{cv\vec{k}} - \hbar\omega)} \right)_{;k_x} \right|^2 \frac{1}{E_{cv\vec{k}} - 2\hbar\omega}. \end{aligned} \quad (\text{A3})$$

Here, integration by parts was applied twice in the last reformulation before inserting the free-carrier limit of the Green's function. Expansion near a Dirac point demonstrates that

$$\left| \left(\frac{P_{vc\vec{k}}^x}{E_{cv\vec{k}} (E_{cv\vec{k}} - \hbar\omega)} \right)_{;k_x} \right|^2 \simeq \frac{81m^2 a^4 \gamma^4 (E_{cv\vec{k}}^2 - E_{g\sigma}^2)}{2\hbar^2 E_{cv\vec{k}}^8} \quad (\text{A4})$$

with band gaps $E_{g\sigma} = 2\Delta + \sigma 6\sqrt{3}\lambda$. Then, integration yields

$$\text{Re}\sigma_{xxxx}^{(3)}(\omega) \simeq \frac{27e^4 a^2 \gamma^2}{128\hbar} \sum_{\sigma=\pm 1} \frac{E_{g\sigma}^2 - 4\hbar^2 \omega^2}{\hbar^6 \omega^6} \theta(2\hbar\omega - E_{g\sigma}). \quad (\text{A5})$$

It can be shown that a slightly more accurate result is

$$\text{Re}\sigma_{xxxx}^{(3)}(\omega) \simeq \frac{9e^4 a^2 \gamma^2}{512\hbar} \frac{(E_g^2 - 4\hbar^2 \omega^2)(3E_g^2 + 4\hbar^2 \omega^2)}{\hbar^8 \omega^8} \theta(2\hbar\omega - E_g). \quad (\text{A6})$$

This expression agrees with the real part of the gapped Dirac model result [31]

$$\sigma_{xxxx}^{(3)}(\omega) \propto \frac{i}{\hbar^8 \omega^8} (3E_g^2 + \hbar^2 \omega^2) \ln \left(\frac{E_g - \hbar\omega}{E_g + \hbar\omega} \right). \quad (\text{A7})$$

In Fig. 5, full band-structure and gapped Dirac model results, i.e., Eqs. (A3) and (A6), are compared for the cases of MoS₂ and WSe₂. It is seen that, apart from smoothing due to broadening applied in the numerical integral, the expressions are in very good agreement.

[1] M. Göppert-Mayer, *Ann. Phys. (NY)* **401**, 273 (1931).

[2] S. Kawata, H.-B. Sun, T. Tanaka, and K. Takada, *Nature (London)* **412**, 697 (2001).

- [3] Q. Liu, B. Guo, Z. Rao, B. Zhang, and J. R. Gong, *Nano Lett.* **13**, 2436 (2013).
- [4] J. E. Ehrlich, X. L. Wu, I.-Y. S. Lee, Z.-Y. Hu, H. Röckel, S. R. Marder, and J. W. Perry, *Opt. Lett.* **22**, 1843 (1997).
- [5] A. S. Dvornikov, E. P. Walker, and P. M. Rentzepis, *J. Phys. Chem. A* **113**, 13633 (2009).
- [6] M. Rumi and J. Perry, *Adv. Opt. Photon.* **2**, 451 (2010).
- [7] M. J. O'Connell, S. M. Bachilo, C. B. Huffman, V. C. Moore, M. S. Strano, E. H. Haroz, K. L. Rialon, P. J. Boul, W. H. Noon, C. Kittrell, J. Ma, R. H. Hauge, R. B. Weisman, and R. E. Smalley, *Science* **297**, 593 (2002).
- [8] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, *Phys. Rev. Lett.* **105**, 136805 (2010).
- [9] W. Zhao, Z. Ghorannevis, L. Chu, M. Toh, C. Kloc, P.-H. Tan, and G. Eda, *ACS Nano* **7**, 791 (2013).
- [10] C. Ruppert, O. B. Aslan, and T. F. Heinz, *Nano Lett.* **14**, 6231 (2014).
- [11] F. Wang, G. Dukovic, L. E. Brus, and T. F. Heinz, *Science* **308**, 838 (2005).
- [12] Z. Ye, T. Cao, K. O'Brien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, *Nature (London)* **513**, 214 (2014).
- [13] 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).
- [14] K. He, N. Kumar, L. Zhao, Z. Wang, K. F. Mak, H. Zhao, and J. Shan, *Phys. Rev. Lett.* **113**, 026803 (2014).
- [15] Y. Xie, S. Zhang, Y. Li, N. Dong, X. Zhang, L. Wang, W. Liu, I. M. Kislyakov, J.-M. Nunzi, H. Qi, L. Zhang, and J. Wang, *Photon. Res.* **7**, 762 (2019).
- [16] Y. Li, N. Dong, S. Zhang, X. Zhang, Y. Feng, K. Wang, L. Zhang, and J. Wang, *Laser Photon. Rev.* **9**, 427 (2015).
- [17] X. Dai, X. Zhang, I. M. Kislyakov, L. Wang, J. Huang, S. Zhang, N. Dong, and J. Wang, *Opt. Express* **27**, 13744 (2019).
- [18] A. Taghizadeh and T. G. Pedersen, *Phys. Rev. B* **97**, 205432 (2018).
- [19] A. Taghizadeh and T. G. Pedersen, *Phys. Rev. B* **99**, 235433 (2019).
- [20] M. M. Glazov, L. E. Golub, G. Wang, X. Marie, T. Amand, and B. Urbaszek, *Phys. Rev. B* **95**, 035311 (2017).
- [21] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, *Phys. Rev. B* **92**, 085413 (2015).
- [22] D. M. Bishop, *J. Chem. Phys.* **100**, 6535 (1994).
- [23] T. G. Pedersen, *Phys. Rev. B* **92**, 235432 (2015).
- [24] G. S. Agarwal and W. Harshawardhan, *Phys. Rev. Lett.* **77**, 1039 (1996).
- [25] T. G. Pedersen, *Phys. Rev. B* **98**, 165425 (2018).
- [26] L. V. Keldysh, *JETP Lett.* **29**, 658 (1979).
- [27] H. C. Kamban and T. G. Pedersen, *Sci. Rep.* **10**, 5537 (2020).
- [28] J. Have, N. M. R. Peres, and T. G. Pedersen, *Phys. Rev. B* **100**, 045411 (2019).
- [29] Y. F. Yu, Y. L. Yu, C. Xu, Y. Q. Cai, L. Q. Su, Y. Zhang, Y. W. Zhang, K. Gundogdu, and L. Y. Cao, *Adv. Func. Mater.* **26**, 4733 (2016).
- [30] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, *Phys. Rev. Lett.* **111**, 216805 (2013).
- [31] F. Hipolito, A. Taghizadeh, and T. G. Pedersen, *Phys. Rev. B* **98**, 205420 (2018).
- [32] A. Shimizu, *Phys. Rev. B* **40**, 1403 (1989).
- [33] C. Y.-P. Chao and S. L. Chuang, *Phys. Rev. B* **43**, 6530 (1991).
- [34] L. Banyai, I. Galbraith, and H. Haug, *Phys. Rev. B* **38**, 3931 (1988).
- [35] A. L. Ivanov, H. Haug, and L. V. Keldysh, *Phys. Rep.* **296**, 237 (1998).
- [36] D. W. Kidd, D. K. Zhang, and K. Varga, *Phys. Rev. B* **93**, 125423 (2016).