This is the accepted manuscript made available via CHORUS. The article has been published as:

Strong exciton regulation of Raman scattering in monolayer MoS$_2$
Yuanxi Wang, Bruno R. Carvalho, and Vincent H. Crespi
Phys. Rev. B 98, 161405 — Published 17 October 2018
DOI: 10.1103/PhysRevB.98.161405
Strong exciton regulation of Raman scattering in monolayer MoS$_2$

Yuanxi Wang
Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA
2-Dimensional Crystal Consortium, Pennsylvania State University, University Park, Pennsylvania 16802 and Material Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802

Bruno R. Carvalho
Departamento de Física, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte 59078-970, Brazil

Vincent H. Crespi
Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA
Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, USA and 2-Dimensional Crystal Consortium, Pennsylvania State University, University Park, Pennsylvania 16802

The weakly screened electron-hole interactions in an atomically thin semiconductor not only downshift its excitation spectrum from a quasiparticle one, but also redistribute excitation energies and wavefunction characters with profound effects on diverse modes of material response, including the exciton-phonon scattering processes accessible to resonant Raman measurements. Here we develop a first-principles framework to calculate frequency-dependent resonant Raman intensities that includes excitonic effects and goes beyond the Placzek approximation. We show how excitonic effects in MoS$_2$ strongly regulate Raman scattering amplitudes and thereby explain the puzzling near-absence of resonant Raman response around the A and B excitons (band-edge excitations which produce very strong signals in optical absorption), and also the pronounced strength of the resonant Raman response from the C exciton (a higher-energy excitation arising from parallel valence and conduction bands). Furthermore, this efficient perturbative approach reduces the number of GW-BSE calculations from two per Raman mode (in finite displacement) to one for all modes and affords natural extension to higher-order resonant Raman processes.

Low-energy excitations of two-dimensional semiconductors are dominated by very strong excitonic effects, as exemplified by the A and B excitons arising from band-edge excitations in monolayer MoS$_2$ and the C exciton arising from near-parallel valence and conduction bands [1, 2]. While excitonic resonances are evident from absorption spectroscopy [3], resonant Raman spectroscopy offers a more multifaceted perspective: the Raman intensity of a phonon mode plotted against laser energies (Raman excitation profile) not only reveals excitonic resonances with resolutions on par with absorption, but also reflects exciton-phonon coupling strengths [4, 5]. Raman features emerging upon reaching resonance furthermore capture finite-momentum processes such as higher-order Raman scattering and defect scattering [6–10], both key processes in valleytronics [5, 11, 12]. The appeal of these rich outputs, combined with the procedural simplicity of Raman measurements (perhaps best attested by Raman’s original discovery using sunlight, distilled liquids, and a human eye [13]) contribute to its wide usage. Spectral features in Raman excitation profiles are generally aligned in energy with absorption features for molecules [14–16] and 3D bulk solids [17], with intensities of similar orders of magnitude, as modulated by electron-phonon interactions. This expectation is confounded by the puzzling near-absence of Raman intensity measured at the A/B exciton energies and the disproportionately higher Raman intensity measured at the C exciton in MoS$_2$ [6, 18]. This anomaly in 2D semiconductors suggests an unidentified regulating effect by excitons.

Despite the rich experimental data on Raman measurements of 2D solids, the role of excitons on Raman spectra is rarely modeled at a first-principles level beyond calculating shifted resonance energies, because of the high computational cost of many-body perturbation theory calculations and the sparsity of implementations that consolidate electron-phonon and many-body phenomena. One recent important theoretical advance [19] (implemented in [20, 21]) used finite differences through solution of the Bethe-Salpeter equation (BSE) on the quasiparticle (GW) band structure, but employed a quasistatic Placzek approximation that is only valid in the non-resonant regime. Here we follow the generalized approach of Ref. [22], valid for solids in the resonant regime, to develop a perturbation framework that goes beyond the Placzek approximation and includes electron-hole interactions. Both ingredients are crucial to capturing exciton-regulated Raman scattering in MoS$_2$, including dramatic differences in the strength of the resonant response in the vicinity of the A/B and C excitons that agree with experiments. We show that band-extrema electron-hole pairs such as the A/B excitons suppress Raman response due to their energies being well separated from the rest of the exciton spectrum, and that parallel-band electron-hole pairs such as the C exciton amplify Raman response due to their bunching of energies caus-
ing strong rehybridization during atomic vibration.

We first explain the detailed theoretical and computational basis of the calculations; readers interested primarily in the results and physical interpretation of excitonic effects in resonant Raman spectra may advance to Fig. 1 and the associated discussion. First-principles Raman spectra calculations are most straightforward for Raman shifts, routinely achieving excellent agreement with experiments [23–25]. Raman intensities are usually computed within the non-resonant Placzek approximation: since the scattered light intensity is proportional to the electronic susceptibility $\chi(\omega)$ periodically modified by atomic vibrations ($\omega$ is the incident light frequency), a product-to-sum identity converts the scattered $\cos(\omega_{\text{phonon}}t)\cos(\omega t)$ wave into Stokes and anti-Stokes components [26]. The scattering amplitude depends on how strongly $\chi$ is changed by vibrations $R$, i.e. $|d\chi/dR|^2 \propto |d\omega/dR|^2$, where $\omega(\omega)$ is the dielectric function $\epsilon(\omega) = 1 + \sum_S |\langle 0 | r | S \rangle|^2 / (\omega_S - \omega - i\gamma)$, $\gamma$ is the lifetime broadening of the excitons, $\omega_S$ is the energy of state $|S\rangle$, and $S$ runs over all excitations (the “negative frequency” contribution is suppressed for clarity but is included in all calculations). This derivative has been calculated using the second derivative of the electronic density matrix [27, 28], the “$2n+1$” theorem [29], or finite differences of the static dielectric tensor [24, 30] (calculated from density functional perturbation theory [31]). The derivative can also be expanded by treating $\omega_S$ and the matrix elements separately,

$$\frac{d\epsilon(\omega)}{dR} = \sum_S \left( \frac{\partial\epsilon(\omega)}{\partial\omega_S} \frac{d\omega_S}{dR} + \frac{1}{\omega_S - \omega - i\gamma} \frac{d}{dR} \left| \langle 0 | r | S \rangle \right|^2 \right) \equiv d_2 + d_3. \tag{1}$$

Using first-order perturbation theory [32], the former group of “two-band terms”

$$d_2 = -\sum_S \frac{\langle 0 | r | S \rangle \langle S | \partial_R H | S \rangle \langle S | r | 0 \rangle}{(\omega_S - \omega - i\gamma)^2} \tag{2}$$

involves only transitions between pairs of bands. The latter group of “three-band terms” (see Supplemental Materials)

$$d_3 = \sum_{S' \neq S} \frac{-1}{\omega_S - \omega - i\gamma} \frac{\langle 0 | r | S' \rangle \langle S' | \partial_R H | S \rangle \langle S | r | 0 \rangle}{\omega_{S'} - \omega_S} + \sum_{S' \neq S} \frac{1}{\omega_{S'} - \omega - i\gamma} \frac{\langle 0 | r | S \rangle \langle S | \partial_R H | S' \rangle \langle S' | r | 0 \rangle}{\omega_{S'} - \omega_S}$$

$$= \sum_{S' \neq S} \frac{-\langle 0 | r | S' \rangle \langle S' | \partial_R H | S \rangle \langle S | r | 0 \rangle}{(\omega_{S'} - \omega - i\gamma)(\omega_S - \omega - i\gamma)} \tag{3}$$

contains transitions between three states. No phonon-momentum dependence appears since we only focus on first-order Raman. So far the expressions are general: if all quantities are calculated at the DFT level, the Hamiltonian $H = H_{\text{DFT}}$ [33] and $|S\rangle$ are free electron-hole transitions separated by $\omega_S$; if calculated at the BSE level, $H = H_{\text{BSE}}$ [34] and $|S\rangle$ are excitonic wavefunctions with eigenvalues $\omega_S$. Physically, two- and three-band terms respectively represent contributions from the oscillating excitation eigenvalues (Kohn-Sham eigenvalues or BSE eigenvalues) and the oscillatory rehybridization of wavefunctions (Kohn-Sham orbitals or BSE eigenvectors) [35]. By combining the $d_2$ and $d_3$ terms we recover the usual perturbation expression for Raman susceptibility $\alpha$, ignoring small phonon energies,

$$\alpha_{\text{perturb.}} \propto \sum_{S,S',}\frac{\langle 0 | r | S' \rangle \langle S' | \partial_R H | S \rangle \langle S | r | 0 \rangle}{(\omega_{S'} - \omega - i\gamma)(\omega_S - \omega - i\gamma)}. \tag{4}$$

Phonon frequencies appear explicitly in the more rigorous many-body formulation, where resonant terms becomes equivalent to Eqn. 4 in the limit that the phonon frequencies are zero [26, 36]. Three-band terms are often neglected due to the apparent squared denominator of the two-band terms (see Eqs. 1 and 2) [20]; the final expanded expression shows that three-band terms become important when the intervals between excitation energies are small.

So long as laser energies $\omega$ are away from excitation levels so that $\omega_{\text{phonon}} \ll |\omega - \omega_S + i\gamma|$, the Placzek approximation holds [22] and finite-displacement calculations using static dielectric tensors [24, 37] agree qualitatively with Raman intensities measured at finite (but sub-bandgap) $\omega$, due to the near-constant dielectric function in this regime. The use of Placzek approximation in the resonant regime [19, 38] was argued to be problematic in Ref. [22], where a more rigorous expression is derived that is equivalent to keeping only the three-band terms $d_3$. These $d_3$ terms correspond to the so-called “Albrecht B/C terms” (or Herzberg-Teller terms) in the vibronic theory for resonant Raman intensities in molecules accounting for nuclear wavefunctions [39–41]. The seemingly missing “Albrecht A terms” (or Condon terms) [41] only arise for excitations with finite Frank-Condon shifts and is negligible for delocalized vibrations in solids [42, 43] (and even for localized vibrations near certain common defects in MoS2 [44]).

Since $d_3$ readily separates from $d_2$ in the perturbation approach, we derive the single-particle expansion for both at the BSE level and numerically verify that their sum matches the spectra obtained from finite displacements within the Placzek approximation and that, for $\omega \to 0$, $d_3$ (general) and $d_2 + d_3$ (Placzek) converge to the same value, i.e. $d_2$ goes to zero. With the optical matrix elements in Eqn. 4 readily available in existing GW-BSE codes within the the dipole approximation, we focus on evaluating the exciton-phonon coupling matrix elements $\langle S' | \partial_R H | S \rangle$. For the derivative of the exciton
Hamiltonian $\partial_R H^{\text{BSE}}$ within the Tamm-Dancoff approximation, we neglect the contribution from the derivative of the BSE kernel $\partial K$ [45], neglect the derivative of the quasiparticle correction by using $\partial_R H^{\text{QP}} \approx \partial_R H^{\text{DFT}}$ (as validated in Refs. [45, 46]) so that

$$d_2 = \sum_{S, v, c, k} \frac{|\langle 0 | r | S \rangle|^2 |\langle S | v, c, k \rangle|^2}{(\omega_S - \omega - i\gamma)^2} \times \left[ \langle c | \partial_R H^{\text{DFT}} | c \rangle - \langle v | \partial_R H^{\text{DFT}} | v \rangle \right]$$

(5)

where $S_{v, c, k} = \langle S | v, c, k \rangle$. Here we neglect $c \neq c'$ and $v \neq v'$ terms in Ref. [45] (DFT-level “three-band” terms) since they only contribute significantly when the energy separation between bands is similar to phonon energies; for the low-energy electronic structure of MoS$_2$, most bandpairs of small separation are up-down spin copies forbidding interband scattering, with the exception of the valence band top being split by spin-orbit interaction. Although in general bands split by spin-orbit coupling allow interband scattering (yielding significant DFT-level “three-band” terms [33, 47]), the spin-orbit Hamiltonian near the valleys in MoS$_2$ only involves $\sigma_2$ so that spins components are decoupled [1, 48]. This approximation is numerically justified later. The $d_3$ terms involve

$$d_3 = \sum_{S, v, c, k} \frac{|\langle 0 | r | S \rangle|^2 |\langle S | v, c, k \rangle|^2}{(\omega_S - \omega - i\gamma)^2} \times \left[ \langle c | \partial_R H^{\text{DFT}} | c \rangle - \langle v | \partial_R H^{\text{DFT}} | v \rangle \right] \delta_{v, c, k}$$

(6)

Again neglecting $c \neq c'$ and $v \neq v'$ terms and substituting into Eqn. 3 gives

$$d_3 = \sum_{S \neq S', v, c, k} \frac{|\langle 0 | r | S' \rangle|^2 |\langle S' | r | 0 \rangle S_{v, c, k}}{\omega_{S} - \omega - i\gamma} \left( \omega_{S'} - \omega - i\gamma \right) \times \left[ \langle c | \partial_R H^{\text{DFT}} | c \rangle - \langle v | \partial_R H^{\text{DFT}} | v \rangle \right]$$

(7)

All calculations will follow Eqns. 5 and 7.

All GW-BSE calculations are performed using the BerkeleyGW package [49, 50] based on Kohn-Sham eigenvalues and orbitals obtained within the local density approximation, using Quantum ESPRESSO [51]. An energy cut-off of 24 Ry, 500 empty bands, and a $12 \times 12 \times 1$ k-point grid was used for the dielectric matrix and quasiparticle self-energy, where the Coulomb interaction is truncated in the out-of-plane direction [52]. The static remainder technique [53] accelerates convergence of the quasiparticle gap. BSE matrix elements are assembled using 3 valence bands and 4 conduction bands on the same grid and interpolated onto a $40 \times 40 \times 1$ grid for diagonalization (Haydock iteration is not used because BSE eigenvectors are needed). The Supplemental Material contains details on convergence tests for the above parameters and all calculations involving phonons. Finally, summation over $S \neq S'$ terms are limited to eigenvalue pairs no further apart than 0.3 eV; exciton pairs separated further contribute negligibly due to large denominators in Eqn. 7 and their constituent single-particle transitions being from different bands. Increasing this convergence parameter to 0.4 eV changes Raman intensities by at most 2% (for any laser frequency). We include 800 excitonic states to converge Raman intensities within the 0–3.5 eV spectral range.

The calculated Raman intensities $|\alpha(\omega)|^2$ for the out-of-plane $A_1'$ mode in Fig. 1 shows that combining $|d_2|^2$ (blue solid) and $|d_3|^2$ terms (red solid) from the perturbation approach into $|d_2 + d_3|^2$ (green solid) correctly matches the result from finite displacements (filled light green). Only the three-band plot is to be compared with experiments: Raman intensity is suppressed at the A/B excitons compared with the C exciton. The lower panel shows A/B exciton eigenvalues far below all others and eigenvalues near the C exciton bunched together.

![FIG. 1. Resonant Raman intensities of the out-of-plane $A_1'$ mode in MoS$_2$ calculated as a function of the laser energy. Combining two-band $|d_2|^2$ (blue dashed) and three-band $|d_3|^2$ terms (red solid) calculated from perturbation theory into $|d_2 + d_3|^2$ (green solid) correctly matches the result from finite displacements (filled light green). Only the three-band plot is to be compared with experiments: Raman intensity is suppressed at the A/B excitons compared with the C exciton. The lower panel shows A/B exciton eigenvalues far below all others and eigenvalues near the C exciton bunched together.](image-url)
fects, these transitions are constituents of the C excitons with BSE eigenvalues bunched near 2.4 eV [1] (red hollow in Fig. 2, truncated within its range of convergence). This bunching does not cause an order-of-magnitude change in the absorbance spectral features (whose integral is constrained by the f-sum rule [34]), apart from an overall redshift due to the exciton binding energy and a redistribution of spectral weight rendering exciton resonances sharper than single-particle features. However, as in standard perturbation theory where smaller eigenvalue intervals lead to wavefunctions being more strongly perturbed, bunched BSE eigenvalues cause strong rehybridization of excitonic states during atomic vibration (i.e. decreased denominators $\omega_S - \omega_S$ in the first line of Eqn. 3) and regroups what used to be independent transitions at different $k$-points (which cannot scatter into each other by a $\Gamma$ phonon) into excitonic states all with zero momenta (which allows inter-scattering i.e. increased numerator in Eqn. 7). Therefore, three-band terms contribute an order-of-magnitude amplification in Raman intensities around the C exciton resonance. This can be seen even in the results from finite displacements in Fig. 2, where Raman intensities without electron-hole interaction near 4 eV (blue filled) are amplified to form the highest Raman peak with electron-hole interaction near 2.4 eV (red filled); comparing the more rigorous three-band spectra would yield the same conclusion. In stark contrast, the A and B excitons – each doubly degenerate (two valleys) – are well separated from other excitations, so they only contribute to two-band terms (dashed blue in Fig. 1). Since only three-band terms are valid for on-resonance frequencies, the orphaned A and B states should not appear in an experimental measurement. Thus the final frequency-dependent Raman intensity $|d_1|^2$ (red in Fig. 1) is suppressed at the A/B excitons and amplified at the C exciton. In this way, our perturbation method reveals how spectral features in resonant Raman characterize not only the exciton spectrum and wavefunction character, but also how exciton-phonon coupling enables inter-state scattering.

We now compare with experiments in Fig. 3 and demonstrate that agreement is only achieved for the beyond-Placzek treatment of Raman intensity including excitonic effects. Two sets of experimental data on the frequency-dependent $A'_1$ mode intensity from Refs. [6, 18] are aligned at the 2.8 eV data point and normalized in intensity by the Raman peak of silicon at 520 cm$^{-1}$ (which has its own known frequency dependence) to yield the modulus-squared of the Raman susceptibility $|\alpha(\omega)|^2$ (to be distinguished from Raman cross-section, which has an additional $\omega^2$ frequency dependence [35, 47]), which can be directly compared with the calculated results. The calculated three-band intensity from Fig. 1 is broadened by 0.2 eV to reflect more realistic C exciton lifetimes estimated from those of free carriers in MoS$_2$ [55]. Good agreement is achieved for the Raman intensity suppression around the A and B excitons, as clearly resolved by the red points (not missing potential resonances) and for the Raman intensity amplification near the C exciton. The two very small resonances measured at the A/B exciton energies and a scissors shift applied are discussed in the Supplemental Materials. In all prior comparisons between finite displacement BSE calculations (Placzek) and experiments known to us, satisfactory agreements were achieved for few laser frequencies [21] or for limited spectral region (e.g. the lowest excitonic peak in [19], WS$_2$ A/B excitons in [20], and the WSe$_2$ C exciton in [20]). Going beyond Placzek allows us to achieve agreement over the energy range of all three excitons.

This analysis has broader implications. For the band structure of a generic solid, every exciton bound state from the solution of the BSE consists of electron-hole pairs with matching group velocities, either at band extrema (zero velocity, spanning a direct gap) or along parallel bands (finite velocity, more common in indirect band gap materials). We expect band-extrema excitons in general to suppress Raman response: by construction these excitons have energies well below parallel-band excitons, giving large denominators in Eqn. 7. Even when there are multiple degenerate valleys as in the case of MoS$_2$, the localized (in k-space) nature of band-extrema excitons allows us to approximate the electron-phonon coupling matrix elements to be constants in Eqn. 7, so (focusing on one $vc$ pair) the sum $\sum_k (S'|k) (k|S)$ can be contracted to zero due to the orthogonality of $S$ and $S'$, giving a vanishing numerator. By contrast, we expect parallel-band excitons in general to amplify Raman response: by construction, parallel pairs of conduction and
valence bands span larger Brillouin zone areas (often emanating from high symmetry points, which gives them a further multiplicative degeneracy factor) and therefore allow abundant ways of assembling into excitons with similar energies bunching in a narrow energy window (as many as there are sampled k-points in the parallel-band areas). The resonant Raman intensity of silicon amplified by excitonic effects (compared with the independent quasiparticle case) in Ref. [19] is presumably attributed to this mechanism, given the abundance of parallel bands in silicon [56]. As a consequence of the general validity of the three-band dominance demonstrated here, resonant Raman measurements can directly probe how excitons undergo inter-state scattering by phonons, which affects exciton population dynamics and lifetimes [57].

The perturbation framework developed here not only allows us to go beyond the classical Placzek approximation and include excitonic effects, but also to achieve better scaling behavior: the GW-BSE routine is only performed once statically (at the slight expense of calculating electron-phonon coupling matrix elements for all Raman active modes), compared with finite differences methods where at least two GW-BSE runs are needed for each Raman active mode. This advantage can be exploited to accelerate Raman intensity calculations for low-symmetry materials such as ReS$_2$ [58], with 18 Raman modes. For second-order Raman intensities, the computational demand for finite differences is even higher, requiring evaluating the BSE dielectric function $N^2$ times, $N$ being the number of Raman modes. In addition, finite-momentum phonon displacements need to be performed on supercells compatible with phonon wavevectors. Despite the computational challenge, second-order Raman intensities were successfully calculated from first-principles recently [59]. Our perturbation treatment can be naturally extended to calculate second-order Raman, where the electron-phonon coupling matrix elements would also be calculated for finite-momenta phonons, but without employing supercells thanks to density functional perturbation theory. The key challenge would be in efficiently calculating finite momentum excitons [60, 61] (exciton dispersions), which may be overcome using accurate tight-binding based models (fitted to GW band structures) [62].

This work is supported by computational time on the LSU-superMIC through the XSEDE allocation TG-DMR170050 and by the National Science Foundation Materials Innovation Platform under DMR-1539916. B.R.C. acknowledges the financial support from the Brazilian agencies CNPq and CAPES.

![Graph](image_url)

**FIG. 3.** Experimental Raman excitation profile for the out-of-plane $A_1$ mode from Ref. [18] (red) and Ref. [6] (black), compared with the calculated three-band terms in Fig. 1 with broadening increased to 0.2 eV to reflect more realistic exciton lifetimes as estimated from free electron lifetimes.


[32] expanding to order n will give scattering rates for n-phonon Raman processes [50].

[33] \[ H_{DFT}^{\text{GW}} = \left( \epsilon_{c,v} - \epsilon_{c,v}^{\text{GW}} \right) \delta_{c',v'} - \delta_{v',c} \], where \( \epsilon_{c,v} \) are conduction and valence band energies respectively.

[34] \[ H_{\text{BSE}}^{\text{GW}} = \left( \epsilon_{c,GW} - \epsilon_{c,v}^{\text{GW}} \right) \delta_{c,v'} - \delta_{v',c} + K_{c,v',c}^{\text{GW}} \], where \( \epsilon_{c,GW} \) are quasiparticle energies and \( K_{c,v',c}^{\text{GW}} \) is the BSE kernel [50].


