



A library of ab initio Raman spectra for automated identification of 2D materials

Alireza Taghizadeh ^{1,2,3}✉, Ulrik Leffers³, Thomas G. Pedersen ^{1,2} & Kristian S. Thygesen ^{3,4}

Raman spectroscopy is frequently used to identify composition, structure and layer thickness of 2D materials. Here, we describe an efficient first-principles workflow for calculating resonant first-order Raman spectra of solids within third-order perturbation theory employing a localized atomic orbital basis set. The method is used to obtain the Raman spectra of 733 different monolayers selected from the Computational 2D Materials Database (C2DB). We benchmark the computational scheme against available experimental data for 15 known monolayers. Furthermore, we propose an automatic procedure for identifying a material based on an input experimental Raman spectrum and apply it to the cases of MoS₂ (H-phase) and WTe₂ (T'-phase). The Raman spectra of all materials at different excitation frequencies and polarization configurations are freely available from the C2DB. Our comprehensive and easily accessible library of ab initio Raman spectra should be valuable for both theoreticians and experimentalists in the field of 2D materials.

¹Department of Materials and Production, Aalborg University, Aalborg, Øst 9220, Denmark. ²Center for Nanostructured Graphene (CNG), Aalborg, Øst 9220, Denmark. ³Computational Atomic-scale Materials Design (CAMD), Department of Physics, Technical University of Denmark (DTU), Lyngby 2800 Kgs, Denmark. ⁴Center for Nanostructured Graphene (CNG), Technical University of Denmark (DTU), Lyngby 2800 Kgs, Denmark. ✉email: ata@nano.aau.dk

Following the discovery of graphene in 2004¹, the field of two-dimensional (2D) materials has grown tremendously during the last decade. Today, more than 50 different monolayer compounds including metals^{2,3}, semiconductors^{4–6}, insulators⁷, ferromagnets⁸, and superconductors^{9,10}, have been chemically grown or mechanically exfoliated from layered bulk crystals¹¹. The enormous interest in 2D materials has mainly been driven by their unique and easily tunable properties (as compared to 3D bulk crystals), which make them attractive for both fundamental research and technological applications in areas such as energy conversion/storage, (opto)-electronics, and photonics^{6,12,13}. Among the various experimental techniques used for characterizing 2D materials, Raman spectroscopy plays a pivotal role¹⁴ thanks to its simplicity, non-destructive nature, and high sensitivity towards key materials properties such as chemical composition, layer thickness (number of layers), inter-layer coupling, strain, crystal symmetries and sample quality^{15–17}.

Raman spectroscopy is a versatile technique for probing the vibrational modes of molecules and crystals from inelastically scattered light, and is widely used for identifying materials through their unique vibrational fingerprints¹⁸. There are various types of Raman spectroscopies that differ in the number of photons or phonons involved in the scattering process¹⁸. Here we focus on the first-order Raman processes in which only a single phonon is involved. Typically, this is the dominant scattering process in defect-free samples (which are considered here). Note that Raman processes involving defect states or several phonons may also play important roles in some 2D crystals such as graphene¹⁹. As shown schematically in Fig. 1(a), the light scattered from a crystal appears in three distinct frequency bands: A strong resonance at the incident frequency ω_{in} due to Rayleigh (elastic) scattering, and weaker resonances due to Raman (inelastic) scattering at $\omega_{in} - \omega_v$ and $\omega_{in} + \omega_v$ forming Stokes and anti-Stokes bands, respectively. Here, ω_v is the frequency of a (Raman active) vibrational mode of the crystal, i.e. a phonon. Depending on the symmetry of the phonon modes and polarization of the electromagnetic fields, a phonon mode may be active or inactive in the Raman spectrum.

While semi-classical theories of Raman spectroscopy can provide some qualitative insight¹⁸, a full quantum mechanical treatment is necessary for a quantitatively accurate description. In particular, ab initio techniques have been employed successfully to calculate Raman spectra of both molecules^{18,20} and solids^{21,22} typically showing good agreement with experimental spectra. The parameter-free nature of such computational schemes endow them with a high degree of predictive power, although their computational cost can be significant, thus, in practice limiting

them to relatively simple, i.e. crystalline, materials. In the realm of 2D materials, ab initio Raman studies have been limited to a handful of the most popular 2D crystals including graphene¹⁹, hBN²³, WTe₂²⁴, SnS, and SnSe²⁵, as well as MoS₂ and WS₂²⁶. In view of the significant experimental efforts currently being devoted to the synthesis and application of future 2D materials and the important role of Raman spectroscopy as a main characterization tool, it is clear that the compilation of a comprehensive library of Raman spectra of 2D materials across different crystal structures and chemical compositions is a critical and timely endeavor.

Recently, we have introduced the open Computational 2D Materials Database (C2DB)¹¹, which contains various calculated properties for several thousands 2D crystals using state of the art ab initio methods. The properties currently provided in the C2DB include the relaxed crystal structures, thermodynamic phase diagrams (convex hull), electronic band structures and related quantities (effective masses, deformation potentials, etc.), elastic properties (stiffness tensors, phonon frequencies), and optical conductivity/absorbance spectra. We stress that the materials in the C2DB comprise both experimentally known as well as hypothetical materials, i.e. materials that may or may not be possible to synthesize in reality.

In this paper, we present an ab initio high-throughput computation of the resonant first-order Raman spectra of more than 700 monolayers selected as the most stable 2D crystals from the C2DB. The calculations are based on an efficient density functional theory (DFT) implementation of the first-order Raman process employing a localized atomic orbital (LCAO) basis set²⁷. We describe the implementation and the automated workflow for computing the Raman spectra at three different excitation frequencies and nine polarization setups. All calculated Raman spectra are provided in Supplementary Figs. 2–734, and can be found at the C2DB website (<http://c2db.fysik.dtu.dk>). In addition, the applied computational routines are freely available online through the website. Our numerical results are benchmarked against available experimental data for selected 2D crystals (15 different monolayers) such as MoS₂, MoSSe, and MoSe₂. The calculated spectra show excellent agreement with experiments for the Raman peak positions and acceptable agreement for the relative peak intensities. Finally, we analyze the inverse problem of identifying a material based on an input (experimental) Raman spectrum as shown schematically in Fig. 1(b). Using MoS₂ (H-phase) and WTe₂ (T'-phase) as two examples, we find that a simple descriptor consisting of the first and second moments of the Raman spectrum combined with the Euclidean distance measure suffices to identify the correct material among the 700+ candidate materials in the database. In particular, this

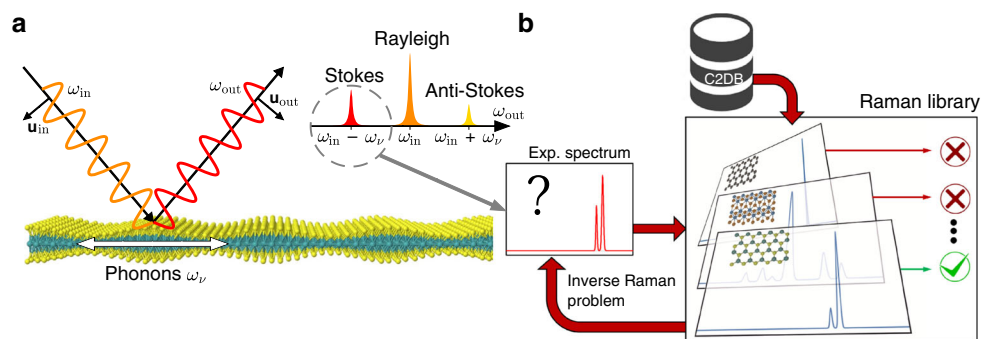


Fig. 1 Schematic view of Raman scattering process and inverse Raman problem. **a** Raman scattering processes, in which incident photons of polarization \mathbf{u}_{in} and frequency ω_{in} are scattered into \mathbf{u}_{out} and ω_{out} under emission (or absorption) of a phonon with frequency ω_v . Only zero momentum phonons contribute to first-order Raman processes but, for illustrative purposes, a finite momentum phonon is shown here. In a typical output spectrum, the Rayleigh (elastic), Stokes and anti-Stokes lines are observed. **b** Given an experimental spectrum, the Raman library based on the open Computational 2D Materials Database (C2DB) can be used to tackle the inverse Raman problem, i.e. identifying the underlying material based on its Raman spectrum.

procedure can be used to differentiate clearly the distinct structural phases of MoS₂ and WTe₂. Incidentally, the library of calculated Raman spectra provides a useful dataset for training machine learning algorithms^{28,29}. As such, our work is not only a valuable reference for experimentalists and theoreticians working in the field of 2D materials, but also represents a step in the direction of autonomous (in situ) characterization of materials.

Results

Theory of Raman Scattering. We first briefly review the theory of Raman scattering in the context of third-order perturbation theory. As discussed above, accurate modeling of Raman processes requires a quantum mechanical treatment to obtain the electronic properties. Regarding the electromagnetic field, it can be shown that a classical description of the field^{30,31} yields the same results as the full quantum mechanical theory that quantizes the photon field^{18,32}. The most common approach to Raman calculations is the Kramers–Heisenberg–Dirac approach³¹, in which the Raman tensor is obtained as a derivative of the electric polarizability with respect to the vibrational normal modes^{21,26,30,31}. Nonetheless, here we employ a more direct and much less-explored approach based on time-dependent third-order perturbation theory to obtain the rate for coherent electronic processes involving creation/annihilation of two photons and one phonon. While the two approaches can be shown to be equivalent¹⁸, at least when local field effects can be ignored as is the case for 2D materials, the third-order perturbative approach can be readily extended to higher order Raman processes (e.g. scattering on multiple phonons), and provides a more transparent physical picture of the Raman processes in terms of individual scattering events³³. Hence, our computational framework is prepared for future extensions to multi-phonon processes. Note that in terms of computational effort, the perturbative approach is comparable to the polarizability derivative method for typical crystals, for which the matrix element calculation dominates the computation time. In this case, both approaches scale as $N_\nu N_b^2$, where N_ν and N_b denote the number of phonon modes and electronic bands, respectively.

To derive an expression for the Raman intensity, both electron–light and electron–phonon Hamiltonians are treated as perturbations (the exact forms of these Hamiltonians are given in the method section). A general time-dependent perturbation can be written as $\hat{H}(t) \equiv \sum_{\omega_1} \hat{H}(\omega_1) \exp(-i\omega_1 t)$ (ω_1 runs over positive and negative frequencies and can also be zero). Note that, in our study, there are three distinct frequency components in $\hat{H}(t)$: input and output frequencies (ω_{in} and ω_{out}) due to the electron–light interaction and zero frequency (i.e. time-independent) for electron–phonon coupling. Within third-order perturbation theory, the transition rate $P_{i \rightarrow f}^{(3)}$ from an initial state $|\Psi_i\rangle$ to a final state $|\Psi_f\rangle$ due to the perturbative Hamiltonian $\hat{H}(t)$, is given by³⁴

$$P_{i \rightarrow f}^{(3)} = \frac{2\pi}{\hbar} \left| \sum_{ab} \sum_{(\omega_1, \omega_2, \omega_3)} \frac{\langle \Psi_f | \hat{H}(\omega_1) | \Psi_a \rangle \langle \Psi_a | \hat{H}(\omega_2) | \Psi_b \rangle \langle \Psi_b | \hat{H}(\omega_3) | \Psi_i \rangle}{(E_i - E_a + \hbar\omega_2 + \hbar\omega_3)(E_i - E_b + \hbar\omega_3)} \right|^2 \delta(E_f - E_i - \hbar\omega). \quad (1)$$

Here, a, b summations are performed over all eigenstates of the unperturbed system (here a set of electrons and phonons) and the sums over ω_n with $n = 1, 2, 3$ are over all three involved frequencies in the perturbative Hamiltonian $\hat{H}(t)$. The notation $(\omega_1 \omega_2 \omega_3)$ indicates that, in performing the summation over ω_n , the sum $\omega_1 + \omega_2 + \omega_3 = \omega$ is to be held fixed. In addition, E_α with $\alpha \in \{i, f, a, b\}$ denote the energies associated with $|\Psi_\alpha\rangle$ and

the Dirac delta ensures energy conservation. The light field is written as $\mathcal{F}(t) = \mathcal{F}_{\text{in}} \mathbf{u}_{\text{in}} \exp(-i\omega_{\text{in}} t) + \mathcal{F}_{\text{out}} \mathbf{u}_{\text{out}} \exp(-i\omega_{\text{out}} t) + \text{complex conjugate}$, where $\mathcal{F}_{\text{in/out}}$ and $\omega_{\text{in/out}}$ are the amplitudes and frequencies of the input/output electromagnetic fields, respectively, see Fig. 1(a). In addition, $\mathbf{u}_{\text{in/out}} = \sum_{\alpha} u_{\text{in/out}}^{\alpha} \mathbf{e}_{\alpha}$ denote the corresponding polarization vectors, where \mathbf{e}_{α} is the unit vector along the α -direction with $\alpha \in \{x, y, z\}$.

We now specialize to the case where the initial and final states of the system are given by $|\Psi_i\rangle = |0\rangle \otimes |n_\nu\rangle$ and $|\Psi_f\rangle = |0\rangle \otimes |n_\nu + 1\rangle$, respectively³² so that $E_f - E_i = \hbar\omega_\nu$. Here, $|0\rangle$ denotes the ground state of the electronic system and $|n_\nu\rangle$ is a state with n_ν phonons at frequency of ω_ν . In this case, the intensity of the Stokes Raman process for a phonon mode is proportional to $P_{i \rightarrow f}^{(3)}$, in which the transition rate involves a photon absorption, followed by an emission of a single phonon and photon. For this type of processes, $(\omega_1, \omega_2, \omega_3)$ are any permutation of $(\omega_{\text{in}}, -\omega_{\text{out}}, 0)$, e.g. $\omega_1 = \omega_{\text{in}}, \omega_2 = -\omega_{\text{out}}, \omega_3 = 0$ and five similar terms (all six terms contribute to the response at frequency of $\omega = \omega_{\text{in}} - \omega_{\text{out}}$). The total Raman intensity $I(\omega)$ is then obtained by summing over all possible final states, i.e. phonon modes ν . Inserting the perturbative Hamiltonians [c.f. Eqs. (6)–(8) in method section] in Eq. (1), the expression for the Stokes Raman intensity involving scattering events by only one phonon can be written

$$I(\omega) = I_0 \sum_{\nu} \frac{n_\nu + 1}{\omega_\nu} \left| \sum_{\alpha\beta} u_{\text{in}}^{\alpha} R_{\alpha\beta}^{\nu} u_{\text{out}}^{\beta} \right|^2 \delta(\omega - \omega_\nu). \quad (2)$$

Here, I_0 is an unimportant constant (since Raman spectra are always reported normalized) that is proportional to the input intensity and depends on the input frequency, and n_ν is given by the Bose–Einstein distribution, i.e. $n_\nu \equiv (\exp[\hbar\omega_\nu/k_B T] - 1)^{-1}$ at temperature T . Due to momentum conservation, only phonons at the center of the Brillouin zone contribute to the one-phonon Raman processes¹⁹. Furthermore, $R_{\alpha\beta}^{\nu}$ denotes the Raman tensor for phonon mode ν , see method section. Eq. (2) is used for computing the Raman spectra in this work for a given excitation frequency and polarization setup. It may be noted that one can derive a similar expression for the anti-Stokes Raman intensity by replacing $n_\nu + 1$ by n_ν in Eq. (2) and ω_ν by $-\omega_\nu$ in Eq. (10) in method section. Note, also, that the Raman shift $\hbar\omega$ is expressed in cm^{-1} with 1 meV equivalent to 8.0655 cm^{-1} .

Computational workflow. An overview of the automated workflow for computing the Raman tensor of the materials in the C2DB is shown in Fig. 2. First, the relaxed structures are extracted from the database. In this work, we consider only compounds that are dynamically stable. Next, the electronic band energies and wavefunctions are obtained from a DFT calculation. In parallel, a zone-center phonon calculation is performed to obtain the optical vibrational modes. From the obtained electronic states and phonon modes, the momentum and electron–phonon matrix elements are evaluated and stored. In the final step, for a given excitation frequency and input/output polarization vectors, the Raman spectrum is calculated using Eq. (2). The key feature of the approach outlined here is that the calculation process can be automatized, allowing one to perform thousands of calculations in parallel without human intervention.

For simplicity, we have restricted the study to non-magnetic materials, but our routines can be readily extended to include magnetic materials. The Raman spectra presented in this paper are computed for in-plane polarization, where the incoming and outgoing photons are polarized along the x - or y -directions,

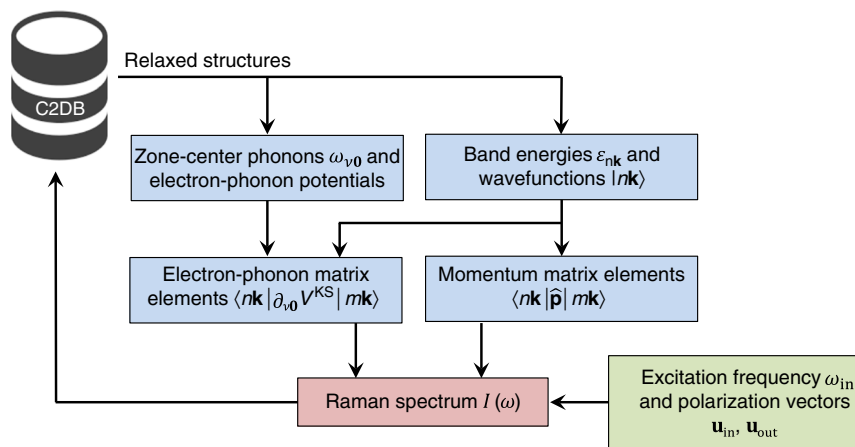


Fig. 2 Computational workflow. The diagram illustrates the steps necessary to calculate the Raman tensor of a material.

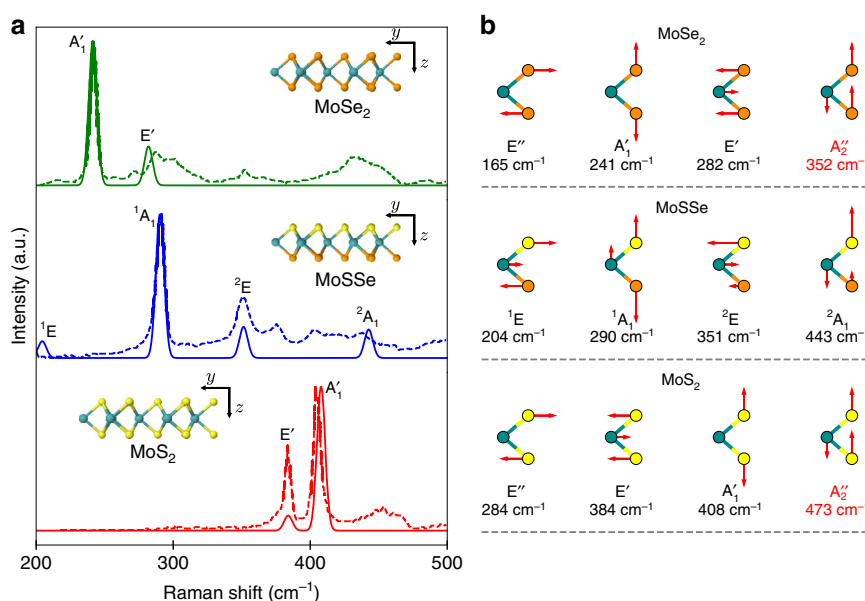


Fig. 3 Evolution of Raman spectra from MoSe₂ over MoSSe to MoS₂. **a** Comparison of the computed Raman spectra (solid) with the experimental results in ref. ³⁶ (dashed) for MoSe₂ (top), MoSSe (middle) and MoS₂ (bottom). The excitation wavelength is 532 nm, and both input and output electromagnetic fields are polarized along the *y*-direction. **b** Optical phonon modes for MoSe₂ (top), MoSSe (middle) and MoS₂ (bottom) labeled by the irreducible representations of the respective point groups. Note that A₂' modes (shown in red) are Raman inactive.

i.e. $\mathbf{u}_{in/out}$ are either $[1, 0, 0]$ or $[0, 1, 0]$. The four possible combinations are referred to as *xx*, *xy*, *yx*, and *yy* polarization setups.

Raman spectra and comparison with experiments. Figure 3 compares the calculated Raman spectrum of three different monolayer transition metal dichalcogenides (TMDs), namely MoSe₂, MoSSe and MoS₂, with the experimental data extracted from ref. ³⁵. For all three monolayers, a good agreement is observed both for the peak positions and relative amplitudes of the main peaks. Additional peaks in the experimental spectra presumably originate from the substrate or defects in the samples. The differences between the Raman spectra of the three materials provide valuable information about the crystal structure. Symmetry and Raman activity of phonon modes are determined by the irreducible point group representations. MoS₂ and MoSe₂ are members of point group D_{3h} , whereas MoSSe lacking a horizontal mirror plane σ_h belongs to the point group C_{3v} . In Mulliken

notation, the irreducible representation of MoS₂ and MoSe₂ is $2A_2' + A_1' + 2E' + E''$, whereas for MoSSe the lowered symmetry leads to $3A_1 + 3E$. For MoS₂ and MoSe₂, one member of both A₂' and E' is an acoustic mode, and the other A₂' mode is Raman inactive. For MoSSe, A₁ and E each contain an acoustic mode, and all other modes are Raman active. The relevant modes are shown schematically in Fig. 3(b). In general, a Raman active mode will only appear in certain polarization configurations. The tensorial Raman selection rules follow from the irreducible point group representations^{36,37} as shown for point groups D_{3h} and C_{3v} in Supplementary Note 1.

Next, we focus on the case of MoS₂, and investigate the dependency of the Raman spectrum on the excitation frequency and polarization, see Figs. 4(a) and 4(b), respectively. In Fig. 4(a), the Raman spectra are computed for three commonly used wavelengths of blue, green and red laser sources. In this case both in- and outgoing polarization vectors are along the *y*-direction (or *x*-direction). While the relative strength of the first Raman active

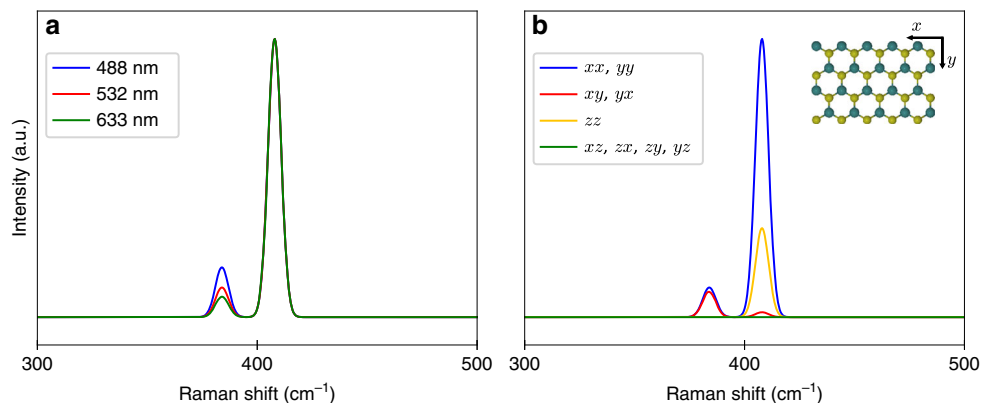


Fig. 4 Polarization and frequency dependent Raman spectra. **a** Raman spectra of MoS₂ evaluated at three different excitation wavelengths, blue (488 nm), green (532 nm), and red (633 nm) for the *xx*-polarization setup. **b** Polarized Raman spectra of MoS₂ for various input and output polarization directions at 532 nm excitation wavelength. The inset shows a top view of the crystal structure.

peak in the spectrum is enhanced slightly for shorter wavelengths, the shape of the spectrum does not change significantly. Note that, in reality, the relative amplitudes of the E' and A_1' modes may change considerably if the excitation frequency coincides with an exciton resonance^{38,39}. This is because excitons can selectively enhance specific Raman modes due to their symmetry^{40,41}. Although this effect is not captured properly in our independent-electron model, it is in principle straightforward to include by using the many-body eigenstates obtained by diagonalizing the Bethe–Salpeter equation (BSE) when evaluating the matrix elements in Eq. (1)^{41–45}. Moreover, the absolute magnitudes of the Raman peaks can vary substantially by changing the excitation wavelength due to the possible resonance with electronic states (resonance Raman spectroscopy)⁴⁰. Nonetheless, the overall magnitude of the Raman spectra is usually of little practical importance compared to the spectral positions and spectra are typically normalized as done here. Changing the polarization of electromagnetic fields not only influences the relative amplitudes of Raman peaks, but may switch certain modes on and off as shown in Fig. 4(b). For instance, the MoS₂ E' mode becomes completely inactive for the perpendicular polarization setup (zz) due to symmetry²⁶. This is easily confirmed using Supplementary Eq. (1) of Supplementary Note 1 predicting an inactive E' mode for *zz*-polarization. Note that, although the E'' mode is Raman active for *xz*-, *yz*-, *zx*-, and *zy*-polarizations, the intensity is too small to be observed in Fig. 4(b).

We have assessed the quality of the Raman library for a wide range of material compositions and crystal structures. Fig. 5 compares experimental and calculated Raman spectra for 12 monolayers including graphene, hBN, several conventional TMDs in the H- or T'-phase as well as anisotropic crystals such as phosphorene and Pd₂Se₄. In general, the number of Raman active modes increases with the number of atoms in the unit cell, as expected. For instance, there are more than eight peaks in the Raman spectrum of Pd₂Se₄. Furthermore, as a rule of thumb, Raman modes of materials containing heavier atoms are at lower frequencies and vice versa, e.g. the Raman peaks for graphene and hBN appear at frequencies above 1000 cm⁻¹. The experimental data are obtained under various experimental conditions such as different excitation wavelengths and polarizations or diverse sample substrates. Note that if polarized Raman spectra were not available (or in the case of unspecified polarization), an average of all four in-plane polarization settings, i.e. *xx*, *xy*, *yx*, and *yy*, has been used for generating the theoretical spectra. In general, there is quite good agreement between our calculations and experimental results in all cases, particularly, for the peak positions. The deviations can be attributed to various factors such as substrate

and excitonic effects, which are not captured in our calculations, as well as the quality of the experimental samples and other experimental uncertainties, all of which can influence the spectra considerably.

Identifying materials from their Raman spectra. At this point, we turn to a critical test of the ab initio Raman library: given an experimental Raman spectrum, is it possible to identify the underlying material by comparing the experimental spectrum to a library of calculated spectra? The answer to this question will depend on several factors including: (1) the quality of the experimental spectrum. (2) The quality of the calculated spectra, i.e. the ability of theory to reproduce a (high quality) experimental spectrum for a given material. (3) The size/density of the calculated Raman spectrum database. Obviously, a more densely populated database increases the chances that the experimental sample is, in fact, contained in the database. But, at the same time, this increases the risk of obtaining a false positive, i.e. matching the experimental spectrum by a calculated spectrum of a different material.

Putting the above idea into practice requires a quantitative measure for comparing Raman spectra. In the present work, we use the two lowest moments to fingerprint the Raman spectrum. In general, the N th Raman moment of the spectrum is given by

$$\langle \omega^N \rangle \equiv \int_0^\infty I(\omega) \omega^N d\omega = \sum_\nu I_\nu \omega_\nu^N, \quad (3)$$

where I_ν denotes the amplitude of mode ν , i.e. $I_\nu = I_0(n_\nu + 1) |\sum_{\alpha\beta} u_{in}^\alpha R_{\alpha\beta}^\nu u_{out}^\beta|^2 / \omega_\nu$. Note that, for these calculations, we normalize the Raman spectrum such that its zeroth moment becomes one, i.e. $\int_0^\infty I(\omega) d\omega = \sum_\nu I_\nu = 1$. Therefore, the first Raman moment corresponds to the mean value of the spectrum. Rather than using the second moment, we use the standard deviation of the spectrum as the selected measure, given by

$$\delta\omega = \sqrt{\langle \omega^2 \rangle - \langle \omega \rangle^2}. \quad (4)$$

Figure 6 shows a scatter plot of $\langle \omega \rangle$ and $\delta\omega/\langle \omega \rangle$ for the 733 monolayers at an excitation wavelength of 532 nm and *xx*-polarization setup obtained at the room temperature. In this plot, crystals composed of lighter elements appear further to the right because their optical phonons generally have higher energies. Furthermore, crystals with fewer atoms in the unit cell and/or higher degree of symmetry, appear in the bottom of the plot because they have fewer (non-degenerate) phonons and thus fewer peaks in their Raman spectrum resulting in a reduced

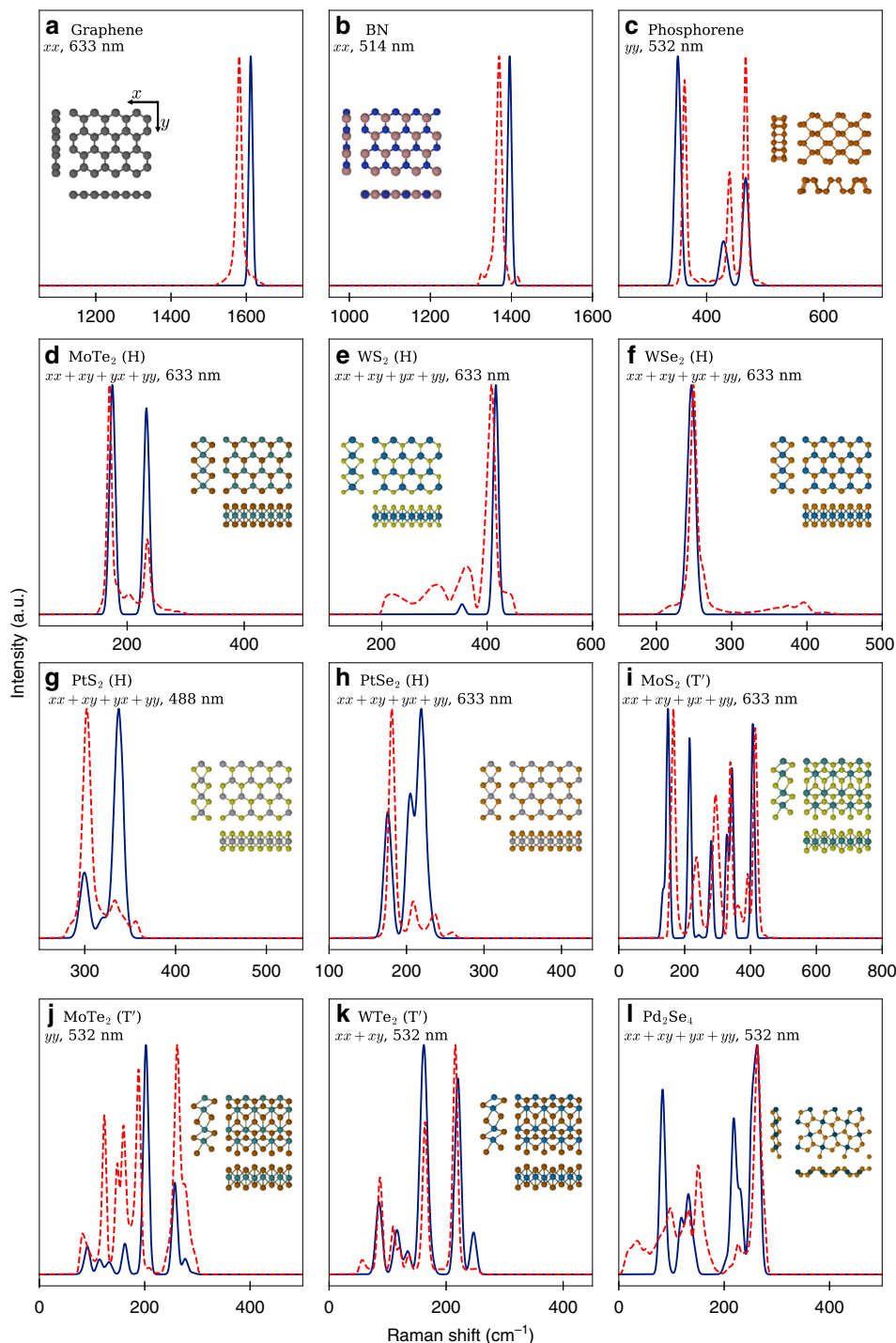


Fig. 5 Raman spectra of 12 monolayers. Comparison of computed Raman spectra (solid lines) with available experimental results (dashed lines). The experimental data are extracted from Refs. [23,69–79](#), for (a) to (l), respectively. The temperature is set to 300 K (room temperature) and excitation wavelength is specified in each case, see the main text. The crystal structures are shown in the insets including top view and cross sectional views. For all crystal structures the x - and y -directions are along the horizontal and vertical directions, respectively, as shown for graphene.

frequency spread. In particular, $\delta\omega$ vanishes for materials with only a single Raman peak such as graphene and hBN.

To test the feasibility of inverse Raman mapping, we evaluate the lowest Raman moment fingerprint for five experimental Raman spectra of MoS₂ (H-phase) and three spectra of WTe₂ (T'-phase) obtained from independent studies, see stars in Fig. 6. Similar analyses have been performed for the eleven additional crystals found in Fig. 5, and is provided in Supplementary Note 2.

Firstly, note that the fingerprint of MoS₂ in the T'-phase (WTe₂ in H-phase) is located relatively far from the H-phase (T'-phase) fingerprint in the plot, which suggests that the lowest Raman moments are indeed able to distinguish different structural phases of the same material. The insets highlight the regions surrounding the experimental data. The variation in the experimental fingerprints is due to small differences in the Raman spectra, originating from the variations in sample quality, substrate

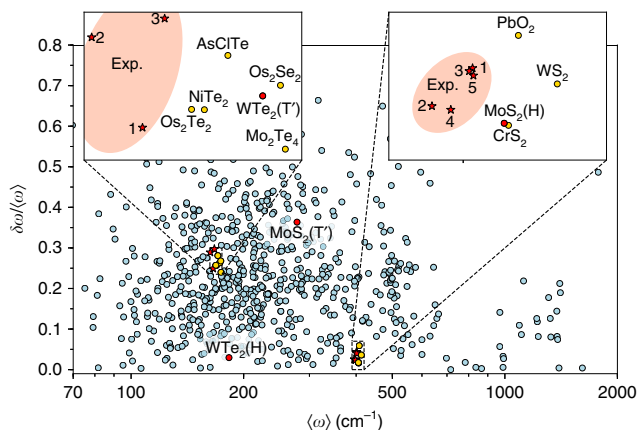


Fig. 6 Calculated Raman moments. Scatter plot of the first Raman moment and normalized standard deviation for 733 calculated spectra at excitation wavelength of 532 nm and xx -polarization setup (circles). For comparison, several independent experimental spectra for monolayer MoS₂ (in H-phase) and WTe₂ (in T'-phase) are also shown (stars). We highlight the points corresponding to MoS₂ and WTe₂ in red (H-phase, T'-phase, and experiments). The insets are zooms of the vicinity of the experimental data for MoS₂ and WTe₂. For MoS₂, 1–5 correspond to the experimental spectra obtained from refs. 36,46,79,80, and 41, respectively, whereas 1–3 for WTe₂ are adopted from refs. 47,81, and 75, respectively.

effects, measurement techniques/conditions, etc. Consequently, the precise peak positions and, in particular, their amplitudes can vary from one experiment to another. Clearly, the fingerprints of the calculated spectra for both MoS₂ and WTe₂ lie close to the experimental data. In a few cases, such as Pd₂Se₄, the experimental fingerprints lie further from the theoretical predictions, as illustrated in Supplementary Fig. 1. This may partly be due to insufficient sample quality for these less-explored 2D crystals. In fact, the deviation between theory and experiments is comparable to the variation between the different experiments. Importantly, only a few other materials show a similar agreement with the experimental data. This suggests that fingerprints including higher order moments could single out the correct material with even higher precision. For instance, the skewness (based on the third Raman moment) can be used to distinguish MoS₂ from CrS₂. By manual inspection of the Raman spectra, one readily confirms that the calculated spectra of MoS₂ and WTe₂ are in fact the best match to the experimental spectra, e.g. other candidates have Raman peaks that are not observed in the experimental spectra or the relative amplitudes of the peaks are completely different from the experimental data. Nonetheless, the procedure of manual inspection can be replaced by a more rigorous and unbiased approach as discussed below.

To compare the experimental and calculated Raman spectra quantitatively, we focus on the experimental data of Tongay et al.⁴⁶ and Cao et al.⁴⁷ for MoS₂ and WTe₂, respectively. The experimental spectra for MoS₂ are obtained without any polarizer at 77 K at an excitation wavelength of 488 nm. For WTe₂ in Cao et al.⁴⁷, the experiment is performed at room temperature using a 532 nm laser linearly polarized in-plane. To account for the unspecified polarization, we take the average of Raman spectra for the xx and xy polarization setups in the case of WTe₂, while for MoS₂ the average of all Raman spectra for transverse components (xx , xy , yx , and yy) is used as the theoretical spectrum. For quantitative comparison with the experimental data, one can use Euclidean distances between the experimental and theoretical spectra as a measure. For two Raman spectra $I_1(\omega)$ and $I_2(\omega)$, the Euclidean distance (or L^2 -

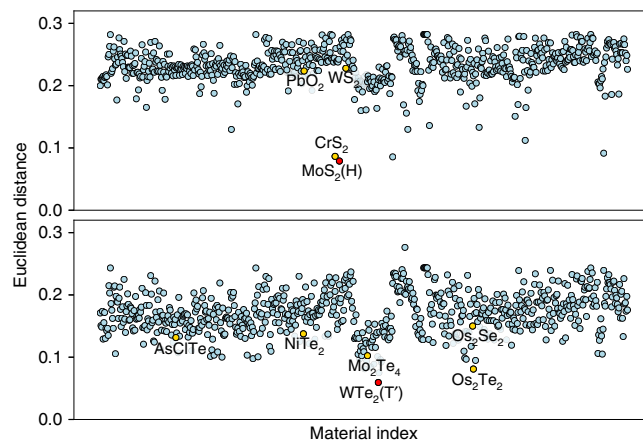


Fig. 7 Euclidean distances between Raman spectra. Distances (see main text for details) are calculated between theoretical Raman spectra and the experimental data of Tongay et al.⁴⁶ and Cao et al.⁴⁷ for MoS₂ (top) and WTe₂ (bottom). For comparison purposes, we highlight the points corresponding to materials in the insets of Fig. 6 by yellow.

norm) $\|I_1 - I_2\|$ is defined as

$$\|I_1 - I_2\| \equiv \left(\int_0^\infty |I_1(\omega) - I_2(\omega)|^2 d\omega \right)^{1/2}. \quad (5)$$

Note that the spectra are normalized such that the total area is unity. Figure 7 shows the computed Euclidean distances from the calculated Raman spectra to the experimental data for both MoS₂ and WTe₂. We highlight the points corresponding to the materials in the insets of Fig. 6. In both cases, identifying the smallest Euclidean distance confirms that the Raman spectra closest to the experimental data are indeed the calculated spectra of MoS₂ and WTe₂. This shows that the quality and accuracy of, respectively, the experimental and computed 2D materials Raman spectra, is sufficient for automatic structure identification.

Discussion

We have introduced a comprehensive library of ab initio computed Raman spectra for more than 700 2D materials spanning a variety of chemical compositions and crystal structures. The 2D materials comprise both experimentally known and hypothetical compounds, all dynamically stable and with low formation energies. Using an efficient first-principles implementation of third-order perturbation theory, the full resonant first-order Raman tensor was calculated including all nine possible combinations for polarization vectors of the input/output photons and three commonly used excitation wavelengths. All spectra are freely available as part of the C2DB and should comprise a valuable reference for both theoreticians and experimentalists in the field. The reliability of the computational approach was demonstrated by comparison with experimental spectra for 15 monolayers such as graphene, hBN, phosphorene and several TMDs in the H-, T-, and T'-phases.

We carefully tested the feasibility of inverse Raman mapping, i.e. to what extent the library of computed Raman spectra can be used to identify the composition and crystal structure of an unknown material from its Raman spectrum. For the specific cases of MoS₂ in H-phase and WTe₂ in T'-phase, we showed that a simple fingerprint based on the lowest moments of the Raman spectrum is sufficient to identify the materials from their experimental Raman spectrum. This represents a significant step in the direction of autonomous identification/characterization of materials. In addition, apart from being a useful reference for 2D materials research, the Raman library can be used to train

machine learning algorithms to predict Raman spectra directly from the atomic structure similarly to recent work on prediction of linear optical spectra for molecules⁴⁸. This is of particular importance in the currently attractive trend of employing machine learning algorithms in materials science^{28,29}.

In the present work, we have focused on Raman processes involving only a single phonon, i.e. first-order Raman processes, since these are typically the dominant contributions to the Raman spectrum. Nonetheless, the presented methodology can be readily extended to include two-phonon scattering processes, although the computational cost will be significantly increased. Excitonic effects in the Raman spectrum have been neglected since most experimental Raman spectra are recorded off-resonance where excitons play a minor role. The inclusion of excitonic effects can be achieved within the presented methodology by employing the many-body eigenstates obtained from the BSE^{42,49,50} instead of Slater determinantal electron-hole excitations. However, this will mainly affect the amplitude of the Raman peaks which is of secondary importance in practice. We only compute the Raman spectra of monolayers in the present work, but the library can be extended to multi-layer structures. For some 2D materials such as graphene or MoS₂ this can be done by employing existing exchange-correlation functionals capable of accurate modeling of van der Waals forces. But for other 2D systems such as phosphorene, further development of exchange-correlation functionals is required to describe the complex inter-layer couplings, particularly for low-frequency Raman modes⁵¹. The symmetry of phonons modes have previously been investigated for graphene⁵², the TMD family³⁷ and phosphorene⁵³ using group theory analysis. Based on the Raman library, such analysis could be performed for a much wider range of materials in future work. Finally, the current work has been restricted to non-magnetic materials, and the ab initio Raman response of magnetic materials is an interesting future research field.

Methods

Theory. In the independent-particle approximation, the Hamiltonian of a system of electrons interacting with phonons and electromagnetic fields takes the form $\hat{H} = \hat{H}_0 + \hat{H}_{ey} + \hat{H}_{ev}$, where \hat{H}_0 is the unperturbed Hamiltonian of the electrons (e) and phonons (v), \hat{H}_{ey} describes the electron-light interaction (here written in the velocity or minimal coupling gauge^{54,55}), and \hat{H}_{ev} describes the electron-phonon coupling. In second quantization, they are given by⁵⁶

$$\hat{H}_0 \equiv \sum_{nk} \epsilon_{nk} \hat{c}_{nk}^\dagger \hat{c}_{nk} + \sum_{vq} \hbar\omega_{vq} \left(\hat{a}_{vq}^\dagger \hat{a}_{vq} + \frac{1}{2} \right), \tag{6}$$

$$\hat{H}_{ey}(t) = \frac{e}{m} \mathcal{A}(t) \cdot \sum_{nmk} \mathbf{p}_{nmk} \hat{c}_{nk}^\dagger \hat{c}_{mk}, \tag{7}$$

$$\hat{H}_{ev} = \sum_{nmv} \sqrt{\frac{\hbar}{\omega_{vq}}} g_{nmk}^{vq} \hat{c}_{nk}^\dagger \hat{c}_{mk} \left(\hat{a}_{vq} + \hat{a}_{v(-q)}^\dagger \right). \tag{8}$$

Here, \hat{c}^\dagger/\hat{c} and \hat{a}^\dagger/\hat{a} are the creation/annihilation operators of electrons and phonons, respectively, \mathcal{A} denotes the vector potential ($\mathcal{F} = -\partial\mathcal{A}/\partial t$), ϵ_{nk} is the energy of the single-particle electronic state $|nk\rangle$, and $\hbar\omega_{vq}$ denotes the phonon energy of normal mode v and wavevector q . Furthermore, $\mathbf{p}_{nmk} = \langle nk|\hat{\mathbf{p}}|mk\rangle$ and $g_{nmk}^{vq} = \langle nk + \mathbf{q}|\partial_{vq} V^{KS}|mk\rangle$ are the momentum and electron-phonon matrix elements (to the first order in the atomic displacements⁵⁶), respectively, with the Kohn-Sham potential V^{KS} . The summation over \mathbf{k} implies an integral over the first Brillouin zone, i.e. $(2\pi)^D \sum_{\mathbf{k}} \rightarrow V \int_{\text{BZ}} d^D \mathbf{k}$ where V is the D -dimensional volume ($D = 2$ for 2D systems). Note that the \mathcal{A}^2 term does not contribute to the linear Raman response and, hence, is absent here. Moreover, we neglect the Coulomb interaction between electrons and holes, i.e. excitonic effects. If the Raman spectroscopy is performed with an excitation frequency that matches the exciton energy³⁸, the electron-hole interactions should be included, ideally within the GW and BSE framework⁴¹⁻⁴³.

We now insert the Hamiltonians given in Eqs. (6)-(8) in the third-order perturbation rate, Eq. (1). As mentioned in the main text, for the Stokes processes involving one phonon, six permutations of $(\omega_{in}, -\omega_{out}, 0)$ are used for $(\omega_1, \omega_2, \omega_3)$. Furthermore, the eigenstates of the unperturbed Hamiltonian $|\Psi_a\rangle$ (or $|\Psi_b\rangle$) can be

written as $|\Psi_e\rangle \otimes |n_{\nu}\rangle$, where $|\Psi_e\rangle$ and $|n_{\nu}\rangle$ are the many-body electronic and phononic states, respectively (the index e runs only over many-body electronic states). Since \hat{H}_{ev} and \hat{H}_{ey} are, respectively, linear in and independent of the phononic operator, the phonon state $|n_{\nu}\rangle$ contributes to the Stokes response only if $|n_{\nu}\rangle$ is either $|n_{\nu}\rangle$ or $|n_{\nu} \pm 1\rangle$. Consequently, $\langle \Psi_a | \hat{H}_{ey} | \Psi_i \rangle = \langle \Psi_e | \hat{H}_{ey} | 0 \rangle \delta_{\nu\nu'} \delta_{n_{\nu} n_{\nu'}}$ and $\langle \Psi_a | \hat{H}_{ev} | \Psi_i \rangle = \langle \Psi_e | \hat{H}_{ev} | 0 \rangle \delta_{\nu\nu'} \delta_{n_{\nu}(n_{\nu} \pm 1)}$ (δ_{ij} denotes the Kronecker delta). The total Raman intensity I is obtained by summing over all final states, i.e. phonon modes, and given by $I(\omega) = I_0 \sum_{\nu} (n_{\nu} + 1) |P_{\nu}|^2 \delta(\omega - \omega_{\nu}) / \omega_{\nu}$, where P_{ν} is defined as

$$P_{\nu} \equiv \sum_{ed} \left[\frac{\langle 0 | \mathbf{u}_{in} \cdot \hat{\mathbf{P}} | \Psi_e \rangle \langle \Psi_e | \hat{G}_{\nu} | \Psi_d \rangle \langle \Psi_d | \mathbf{u}_{out} \cdot \hat{\mathbf{P}} | 0 \rangle}{(\hbar\omega_{in} - \mathcal{E}_e)(\hbar\omega_{out} - \mathcal{E}_d)} + \frac{\langle 0 | \mathbf{u}_{in} \cdot \hat{\mathbf{P}} | \Psi_e \rangle \langle \Psi_e | \mathbf{u}_{out} \cdot \hat{\mathbf{P}} | \Psi_d \rangle \langle \Psi_d | \hat{G}_{\nu} | 0 \rangle}{(\hbar\omega_{in} - \mathcal{E}_e)(\hbar\omega_{\nu} - \mathcal{E}_d)} \right. \\ \left. + \frac{\langle 0 | \mathbf{u}_{out} \cdot \hat{\mathbf{P}} | \Psi_e \rangle \langle \Psi_e | \hat{G}_{\nu} | \Psi_d \rangle \langle \Psi_d | \mathbf{u}_{in} \cdot \hat{\mathbf{P}} | 0 \rangle}{(-\hbar\omega_{out} - \mathcal{E}_e)(-\hbar\omega_{in} - \mathcal{E}_d)} + \frac{\langle 0 | \mathbf{u}_{out} \cdot \hat{\mathbf{P}} | \Psi_e \rangle \langle \Psi_e | \mathbf{u}_{in} \cdot \hat{\mathbf{P}} | \Psi_d \rangle \langle \Psi_d | \hat{G}_{\nu} | 0 \rangle}{(-\hbar\omega_{out} - \mathcal{E}_e)(\hbar\omega_{\nu} - \mathcal{E}_d)} \right. \\ \left. + \frac{\langle 0 | \hat{G}_{\nu} | \Psi_e \rangle \langle \Psi_e | \mathbf{u}_{in} \cdot \hat{\mathbf{P}} | \Psi_d \rangle \langle \Psi_d | \mathbf{u}_{out} \cdot \hat{\mathbf{P}} | 0 \rangle}{(-\hbar\omega_{\nu} - \mathcal{E}_e)(\hbar\omega_{out} - \mathcal{E}_d)} + \frac{\langle 0 | \hat{G}_{\nu} | \Psi_e \rangle \langle \Psi_e | \mathbf{u}_{out} \cdot \hat{\mathbf{P}} | \Psi_d \rangle \langle \Psi_d | \mathbf{u}_{in} \cdot \hat{\mathbf{P}} | 0 \rangle}{(-\hbar\omega_{\nu} - \mathcal{E}_e)(-\hbar\omega_{in} - \mathcal{E}_d)} \right]. \tag{9}$$

Here, $\mathcal{E}_{e/d}$ denote the electronic energies (with respect to the electronic ground state), the summations over e and d include only electronic states, and $\hat{\mathbf{P}}$ and \hat{G}_{ν} are many-body electronic operators given by $\hat{\mathbf{P}} \equiv \sum_{nmk} \mathbf{p}_{nmk} \hat{c}_{nk}^\dagger \hat{c}_{mk}$ and $\hat{G}_{\nu} \equiv \sum_{nmk} g_{nmk}^{v0} \hat{c}_{nk}^\dagger \hat{c}_{mk}$. Note that the momentum conservation implies that only phonons at $\mathbf{q} = \mathbf{0}$ contribute to the response here¹⁹, i.e. $\omega_{\nu} \equiv \omega_{\mathbf{0}}$. Since both $\hat{\mathbf{P}}$ and \hat{G}_{ν} are bi-linear in the electronic operator, for a non-vanishing matrix elements, $|\Psi_{e/d}\rangle$ must include singly-excited states, i.e. terms in the form $\hat{c}_{\mathbf{k}}^\dagger \hat{c}_{\mathbf{k}'} |0\rangle$ (indices c and v imply conduction and valence bands, respectively)⁵⁰. Excitonic effects can readily be introduced at this stage by incorporating the BSE solution^{45,50}. However, we neglect the excitonic effects in the present work, and hence, each singly-excited state contributes individually to the response, i.e. $|\Psi_{e/d}\rangle = \hat{c}_{\mathbf{k}}^\dagger \hat{c}_{\mathbf{k}'} |0\rangle$ with an energy of $\mathcal{E}_{e/d} = \epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}'}$. At finite temperature, the expression for $|\Psi_{e/d}\rangle$ should be taken as $|\Psi_{e/d}\rangle = f_i (1 - f_j) \hat{c}_{\mathbf{k}}^\dagger \hat{c}_{\mathbf{k}'} |0\rangle$, where $f_i \equiv (1 + \exp[(\epsilon_{i\mathbf{k}} - \mu)/k_B T])^{-1}$ is the Fermi-Dirac distribution with chemical potential μ .

Rewriting P_{ν} in terms of the single-particle variables and polarization vectors leads to Eq. (2) for the Raman intensity, where the Raman tensor component, $R_{\alpha\beta}^{\nu}$, reads

$$R_{\alpha\beta}^{\nu} \equiv \sum_{ijmkn} \left[\frac{p_{ij}^{\alpha} (g_{jm}^{\nu} \delta_{in} - g_{im}^{\nu} \delta_{jn}) p_{mn}^{\beta}}{(\hbar\omega_{in} - \epsilon_{ji})(\hbar\omega_{out} - \epsilon_{mn})} + \frac{p_{ij}^{\alpha} (p_{jm}^{\beta} \delta_{in} - p_{im}^{\beta} \delta_{jn}) g_{mn}^{\nu}}{(\hbar\omega_{in} - \epsilon_{ji})(\hbar\omega_{\nu} - \epsilon_{mn})} \right. \\ \left. + \frac{p_{ij}^{\beta} (g_{jm}^{\nu} \delta_{in} - g_{im}^{\nu} \delta_{jn}) p_{mn}^{\alpha}}{(-\hbar\omega_{out} - \epsilon_{ji})(-\hbar\omega_{in} - \epsilon_{mn})} + \frac{p_{ij}^{\beta} (p_{jm}^{\alpha} \delta_{in} - p_{im}^{\alpha} \delta_{jn}) g_{mn}^{\nu}}{(-\hbar\omega_{out} - \epsilon_{ji})(\hbar\omega_{\nu} - \epsilon_{mn})} \right. \\ \left. + \frac{g_{ij}^{\alpha} (p_{jm}^{\alpha} \delta_{in} - p_{im}^{\alpha} \delta_{jn}) p_{mn}^{\beta}}{(-\hbar\omega_{\nu} - \epsilon_{ji})(\hbar\omega_{out} - \epsilon_{mn})} + \frac{g_{ij}^{\alpha} (p_{jm}^{\beta} \delta_{in} - p_{im}^{\beta} \delta_{jn}) p_{mn}^{\alpha}}{(-\hbar\omega_{\nu} - \epsilon_{ji})(-\hbar\omega_{in} - \epsilon_{mn})} \right] f_i (1 - f_j) f_n (1 - f_m). \tag{10}$$

Here, $\epsilon_{ij} \equiv \epsilon_{i\mathbf{k}} - \epsilon_{j\mathbf{k}}$, $p_{ij}^{\alpha} \equiv \langle i\mathbf{k} | \hat{p}^{\alpha} | j\mathbf{k} \rangle$, $g_{ij}^{\nu} \equiv \langle i\mathbf{k} | \partial_{v0} V^{KS} | j\mathbf{k} \rangle$, and $(i, j, m, n)/\nu$ are the electron/phonon band index. The line-shape broadening is accounted for by adding a small phenomenological imaginary part, $i\eta$, to the photon frequencies $\omega_{in/out} \rightarrow \omega_{in/out} + i\eta$. We set the frequency broadening to $\eta = 200$ meV in our calculations.

First-principles calculations. All DFT calculations are performed with the projector-augmented wave code, GPAW^{57,58}, in combination with the atomic simulation environment (ASE)⁵⁹. The Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional is used⁶⁰ and the Kohn-Sham orbitals are expanded using the double zeta polarized (dzp) basis set²⁷. Despite its fairly small size, the dzp basis set provides sufficiently accurate phonon modes. This has been tested by benchmarking the phonon frequencies obtained from this basis set against the results using the commonly-employed plane waves for 700+ monolayers (more than 7000 phonon modes). We confirm that for approximately 80% of all phonons, the discrepancy between the two approaches is less than 5%. Also, the choice of exchange-correlation functional may slightly influence the Raman spectra. For instance, it is known that the PBE functional tends to overestimate the lattice parameters and underestimate the phonon frequencies in crystals⁶¹, whereas the opposite occurs for the local-density approximation (LDA) functionals. Nonetheless, this choice only slightly influences our calculated Raman spectra, and PBE usually provides sufficiently accurate phonon frequencies in the range of theoretical and experimental uncertainties⁶². The monolayers are placed between two vacuum regions with thicknesses of 15 Å. A convergence test of Raman spectra with respect to the wavevector density is performed for several materials, and a mesh with the density of 25 Å⁻¹ for ground state calculations was chosen. The phonon modes are obtained using the standard approach based on calculating the dynamical matrices in the harmonic approximation⁶³. The dynamical matrix is evaluated using the small-displacement method⁶⁴, where the change of forces on a specific atom caused by varying the position of neighboring atoms is computed. Since only the zone-centered (Γ -point) phonons are required, the phonon modes can be computed based on the crystal unit cell. A k -mesh with a density of 12 Å⁻¹ is used for phonon calculations, and the forces are converged within 10⁻⁶ eVÅ⁻¹. Since the wavefunctions and Kohn-Sham potentials in GPAW are

evaluated on a real-space grid⁵⁷, a convergence test with respect to this grid spacing is performed and a real-space grid of 0.2 Å is chosen for calculations. The electron–phonon matrix elements are then obtained within the adiabatic approximation using a finite difference technique for evaluating the derivative of the Kohn–Sham potential⁶⁵. Similarly, the momentum matrix elements are calculated using the finite difference technique and the correction terms due to projector-augmented waves⁶⁶ are added. The width of the Fermi–Dirac occupations is set to $k_B T = 50$ meV for faster convergence of the DFT results. For generating the Raman spectra, a Gaussian $[G(\omega) = (\sigma\sqrt{2\pi})^{-1} \exp(-\omega^2/2\sigma^2)]$ with a variance $\sigma = 3$ cm⁻¹ is used to replace the Dirac delta function, which accounts for the inhomogeneous broadening of phonon modes. The temperature of the Bose–Einstein distributions is set to 300 K for all calculations except for the results in top panel of Fig. 7, where a temperature of 77 K is used. The calculations are submitted, managed, and received using the simple MyQueue workflow tool⁶⁷, which is a Python front-end to job scheduler.

Experimental Raman spectra. The experimental Raman spectra are extracted from the figures in the corresponding references using a common plot digitizer. To remove the noise in the experimental data, they are filtered using a Savitzky–Golay filter⁶⁸ of order three with a filter window length of eleven. For a fair comparison with our theoretical spectra in Fig. 5, we have convolved the experimental spectra with a Gaussian function with variance of 10 cm⁻¹ to reduce the effect of possible but unimportant small frequency shifts between the experimental and theoretical spectra. Furthermore, the Raman moments have been calculated over a frequency range where the main Raman peaks appear, from 350 to 450 cm⁻¹ for MoS₂ and from 75 to 260 cm⁻¹ for WTe₂. For calculating the Euclidean distance, both the experimental and theoretical spectra are convolved with a Gaussian function with variance of 6 cm⁻¹.

Data availability

All calculated Raman spectra are freely available online through <https://cmdrb.fysik.dtu.dk/c2db/>. Other data is available from the corresponding author upon reasonable request.

Code availability

GPWAW is an open-source DFT Python code based on the projector-augmented wave method and the ASE, which is available at <https://wiki.fysik.dtu.dk/gpaw/>. The Raman code used for generating Raman spectra in this work will be available in future releases of the code.

Received: 24 January 2020; Accepted: 6 May 2020;

Published online: 15 June 2020

References

- Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
- Anasori, B., Lukatskaya, M. R. & Gogotsi, Y. 2D metal carbides and nitrides (MXenes) for energy storage. *Nat. Rev. Mater.* **2**, 16098 (2017).
- Xi, X. et al. Strongly enhanced charge-density-wave order in monolayer NbSe₂. *Nat. Nanotechnol.* **10**, 765–769 (2015).
- Liu, H. et al. Phosphorene: An unexplored 2D semiconductor with a high hole mobility. *ACS Nano* **8**, 4033–4041 (2014).
- Mak, K. F., Lee, C., Hone, J., Shan, J. & Heinz, T. F. Atomically thin MoS₂: a new direct-gap semiconductor. *Phys. Rev. Lett.* **105**, 136805 (2010).
- Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nat. Nanotechnol.* **7**, 699–712 (2012).
- Ci, L. et al. Atomic layers of hybridized boron nitride and graphene domains. *Nat. Mater.* **9**, 430–435 (2010).
- Huang, B. et al. Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit. *Nature* **546**, 270–273 (2017).
- Wang, H. et al. High-quality monolayer superconductor NbSe₂ grown by chemical vapour deposition. *Nat. Commun.* **8**, 394 (2017).
- Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).
- Hastrup, S. et al. The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. *2D Mater.* **5**, 042002 (2018).
- Bonaccorso, F., Sun, Z., Hasan, T. & Ferrari, A. C. Graphene photonics and optoelectronics. *Nat. Photon.* **4**, 611–622 (2010).
- Gjerding, M. N., Petersen, R., Pedersen, T. G., Mortensen, N. A. & Thygesen, K. S. Layered van der Waals crystals with hyperbolic light dispersion. *Nat. Commun.* **8**, 320 (2017).
- Ferrari, A. C. Raman spectroscopy of graphene and graphite: Disorder, electron–phonon coupling, doping and nonadiabatic effects. *Solid State Commun.* **143**, 47–57 (2007).
- Ferrari, A. C. & Basko, D. M. Raman spectroscopy as a versatile tool for studying the properties of graphene. *Nat. Nanotechnol.* **8**, 235–246 (2013).
- Neumann, C. et al. Raman spectroscopy as probe of nanometre-scale strain variations in graphene. *Nat. Commun.* **6**, 8429 (2015).
- Li, X.-L. et al. Layer-number dependent optical properties of 2D materials and their application for thickness determination. *Adv. Funct. Mater.* **27**, 1604468 (2017).
- Long, D. A. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*. (John Wiley & Sons Ltd, Chichester, 2002).
- Saito, R., Hofmann, M., Dresselhaus, G., Jorio, A. & Dresselhaus, M. S. Raman spectroscopy of graphene and carbon nanotubes. *Adv. Phys.* **60**, 413–550 (2011).
- Wilson E. B., C. P., Decius J. C. (ed.) *Molecular Vibrations. The Theory of Infrared and Raman vibrational Spectra* (Dover Publications, New York, 1980).
- Umari, P. & Pasquarello, A. First-principles analysis of the Raman spectrum of vitreous silica: comparison with the vibrational density of states. *J. Phys. Condens. Matter* **15**, S1547–S1552 (2003).
- Liang, Q., Dwaraknath, S. & Persson, K. A. High-throughput computation and evaluation of Raman spectra. *Sci. Data* **6**, 135 (2019).
- Cai, Q. et al. Raman signature and phonon dispersion of atomically thin boron nitride. *Nanoscale* **9**, 3059–3067 (2017).
- Jiang, Y. C., Gao, J. & Wang, L. Raman fingerprint for semi-metal WTe₂ evolving from bulk to monolayer. *Sci. Rep.* **6**, 19624 (2016).
- Dewandre, A., Verstraete, M. J., Grobert, N. & Zanolli, Z. Spectroscopic properties of few-layer tin chalcogenides. *J. Phys.: Mater.* **2**, 044005 (2019).
- Liang, L. & Meunier, V. First-principles Raman spectra of MoS₂, WS₂ and their heterostructures. *Nanoscale* **6**, 5394 (2014).
- Larsen, A. H., Vanin, M., Mortensen, J. J., Thygesen, K. S. & Jacobsen, K. W. Localized atomic basis set in the projector augmented wave method. *Phys. Rev. B* **80**, 195112 (2009).
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
- Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 83 (2019).
- Albrecht, A. C. On the theory of Raman intensities. *J. Chem. Phys.* **34**, 1476–1484 (1961).
- Lee, S.-Y. & Heller, E. J. Time-dependent theory of Raman scattering. *J. Chem. Phys.* **71**, 4777 (1979).
- Ganguly, A. K. & Birman, J. L. Theory of lattice Raman scattering in insulators. *Phys. Rev.* **162**, 806–816 (1967).
- Yu, P. Y. & Cardona, M. *Fundamentals of Semiconductors: Physics and Materials Properties* (Springer, Berlin, 2010).
- Grosso, G. & Parravicini, G. P. *Solid State Physics*. (Academic Press, San Diego, CA, 2000).
- Zhang, J. et al. Janus monolayer transition-metal dichalcogenides. *ACS Nano* **11**, 8192–8198 (2017).
- Loudon, R. The Raman effect in crystals. *Adv. Phys.* **13**, 423–482 (1964).
- Ribeiro-Soares, J. et al. Group theory analysis of phonons in two-dimensional transition metal dichalcogenides. *Phys. Rev. B* **90**, 115438 (2014).
- Carvalho, B. R., Malard, L. M., Alves, J. M., Fantini, C. & Pimenta, M. A. Symmetry-dependent exciton-phonon coupling in 2D and bulk MoS₂ observed by resonance Raman scattering. *Phys. Rev. Lett.* **114**, 136403 (2015).
- Scheuschner, N., Gillen, R., Staiger, M. & Maultzsch, J. Interlayer resonant Raman modes in few-layer MoS₂. *Phys. Rev. B* **91**, 235409 (2015).
- Lee, J.-U., Kim, K. & Cheong, H. Resonant Raman and photoluminescence spectra of suspended molybdenum disulfide. *2D Mater.* **2**, 044003 (2015).
- del Corro, E. et al. Atypical exciton–phonon interactions in WS₂ and WSe₂ monolayers revealed by resonance Raman spectroscopy. *Nano Lett.* **16**, 2363–2368 (2016).
- Gillet, Y., Giantomassi, M. & Gonze, X. First-principles study of excitonic effects in Raman intensities. *Phys. Rev. B* **88**, 094305 (2013).
- Wang, Y., Carvalho, B. R. & Crespi, V. H. Strong exciton regulation of Raman scattering in monolayer MoS₂. *Phys. Rev. B* **98**, 161405 (2018).
- Taghizadeh, A. & Pedersen, T. G. Nonlinear optical selection rules of excitons in monolayer transition metal dichalcogenides. *Phys. Rev. B* **99**, 235433 (2019).
- Taghizadeh, A. & Pedersen, T. G. Nonlinear excitonic spin Hall effect in monolayer transition metal dichalcogenides. *2D Mater.* **7**, 015003 (2019).
- Tongay, S. et al. Two-dimensional semiconductor alloys: Monolayer Mo_{1-x}W_xSe₂. *Applied Physics Letters*, **104**, 012101 (2014).
- Cao, Y. et al. Anomalous vibrational modes in few layer WTe₂ revealed by polarized Raman scattering and first-principles calculations. *2D Materials*, **4**, 035024 (2017).
- Ghosh, K. et al. Deep learning spectroscopy: Neural networks for molecular excitation spectra. *Adv. Sci.* **6**, 1801367 (2019).

49. Cudazzo, P. et al. Exciton band structure in two-dimensional materials. *Phys. Rev. Lett.* **116**, 066803 (2016).
50. Taghizadeh, A. & Pedersen, T. G. Gauge invariance of excitonic linear and nonlinear optical response. *Phys. Rev. B* **97**, 205432 (2018).
51. Liang, L. et al. Low-frequency shear and layer-breathing modes in Raman scattering of two-dimensional materials. *ACS Nano* **11**, 11777–11802 (2017).
52. Malard, L. M., Guimarães, M. H. D., Mafra, D. L., Mazzoni, M. S. C. & Jorio, A. Group-theory analysis of electrons and phonons in *n*-layer graphene systems. *Phys. Rev. B* **79**, 125426 (2009).
53. Ribeiro-Soares, J., Almeida, R. M., Cañado, L. G., Dresselhaus, M. S. & Jorio, A. Group theory for structural analysis and lattice vibrations in phosphorene systems. *Phys. Rev. B* **91**, 205421 (2015).
54. Taghizadeh, A., Hipolito, F. & Pedersen, T. G. Linear and nonlinear optical response of crystals using length and velocity gauges: Effect of basis truncation. *Phys. Rev. B* **96**, 195413 (2017).
55. Reichardt, S. & Wirtz, L. Theory of resonant Raman scattering: Towards a comprehensive *ab initio* description. *Phys. Rev. B* **99**, 174312 (2019).
56. Giustino, F. Electron-phonon interactions from first principles. *Rev. Mod. Phys.* **89**, 015003 (2017).
57. Mortensen, J. J., Hansen, L. B. & Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **71**, 035109 (2005).
58. Enkovaara, J. et al. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Condens. Matter Phys.* **22**, 253202 (2010).
59. Larsen, A. H. et al. The atomic simulation environment: a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
60. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
61. Zhang, G.-X., Reilly, A. M., Tkatchenko, A. & Scheffler, M. Performance of various density-functional approximations for cohesive properties of 64 bulk solids. *N. J. Phys.* **20**, 063020 (2018).
62. Skelton, J. M. et al. Influence of the exchange-correlation functional on the quasi-harmonic lattice dynamics of II–VI semiconductors. *J. Chem. Phys.* **143**, 064710 (2015).
63. Fritsch, J. Density functional calculation of semiconductor surface phonons. *Phys. Rep.* **309**, 209–331 (1999).
64. Alfè, D. PHON: A program to calculate phonons using the small displacement method. *Comput. Phys. Commun.* **180**, 2622–2633 (2009).
65. Kaasbjerg, K., Thygesen, K. S. & Jacobsen, K. W. Phonon-limited mobility in *n*-type single-layer MoS₂ from first principles. *Phys. Rev. B* **85**, 115317 (2012).
66. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
67. Mortensen, J., Gjerding, M. & Thygesen, K. MyQueue: task and workflow scheduling system. *J. Open Source Softw.* **5**, 1844 (2020).
68. Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
69. Wu, J. B., Lin, M. L., Cong, X., Liu, H. N., & Tan, P. H. Raman spectroscopy of graphene-based materials and its applications in related devices. *Chemical Society Reviews*, **47**, 1822–1873 (2018).
70. Xia, F., Wang, H., & Jia, Y. Rediscovering black phosphorus as an anisotropic layered material for optoelectronics and electronics. *Nature communications*, **5**, 1–6 (2014).
71. Ruppert, C., Aslan, O. B., & Heinz, T. F. Optical properties and band gap of single- and few-layer MoTe₂ crystals. *Nano Letters*, **14**, 6231–6236 (2014).
72. Liu, H. L. et al. Deep-ultraviolet Raman scattering spectroscopy of monolayer WS₂. *Scientific reports*, **8**, 1–10 (2018).
73. Yang, M. et al. Anharmonicity of monolayer MoS₂, MoSe₂, and WSe₂: A Raman study under high pressure and elevated temperature. *Applied Physics Letters*, **110**, 093108 (2017).
74. Zhao, Y. et al. Extraordinarily strong interlayer interaction in 2D layered PtS₂. *Advanced materials*, **28**, 2399–2407 (2016).
75. Yan, M. et al. High quality atomically thin PtSe₂ films grown by molecular beam epitaxy. *2D Materials*, **4**, 045015 (2017).
76. Gupta, U. et al. Characterization of few-layer 1T-MoSe₂ and its superior performance in the visible-light induced hydrogen evolution reaction. *APL Materials*, **2**, 092802 (2014).
77. Chen, K. et al. A Simple Method for Synthesis of High-Quality Millimeter-Scale 1T' Transition-Metal Telluride and Near-Field Nanooptical Properties. *Advanced Materials*, **29**, 1700704 (2017).
78. Yu, J. et al. Direct observation of the linear dichroism transition in two-dimensional palladium diselenide. *Nano Letters*, **20**, 1172–1182 (2020).
79. Li, H. et al. From bulk to monolayer MoS₂: evolution of Raman scattering. *Advanced Functional Materials*, **22**, 1385–1390 (2012).
80. Lanzillo, N. A. et al. Temperature-dependent phonon shifts in monolayer MoS₂. *Applied Physics Letters*, **103**, 093102 (2013).
81. Zhou, J. et al. Large-area and high-quality 2D transition metal telluride. *Advanced Materials*, **29**, 1603471 (2017).

Acknowledgements

The authors thank M. N. Gjerding for helpful discussions throughout the project. This work was supported by the Center for Nanostructured Graphene (CNG) under the Danish National Research Foundation (project DNR103). K.S.T. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 773122, LIMA).

Author contributions

U.L. wrote the initial computer code. A.T. developed the formalism, modified the code and carried out the calculations. All authors discussed the results. A.T., T.G.P., and K.S.T. wrote the manuscript. T.G.P. and K.S.T. supervised the project.

Competing Interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16529-6>.

Correspondence and requests for materials should be addressed to A.T.

Peer review information *Nature Communications* thanks Ado Jorio, Vincent Meunier, and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020