

# SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

## High-throughput computation and evaluation of raman spectra

Qiaohao Liang<sup>1</sup>, Shyam Dwaraknath<sup>2</sup> & Kristin A. Persson <sup>1,2</sup>

Raman spectroscopy is used ubiquitously in the characterization of condensed materials, spanning from biomaterials, minerals to polymers, as it provides a unique fingerprint of local bonding and environment. In this work, we design and demonstrate a robust, automatic computational workflow for Raman spectra that employs first-principle calculations based on density functional perturbation theory. A set of computational results are compared to Raman spectra obtained from established experimental databases to estimate the accuracy of the calculated properties across chemical systems and structures. Details regarding the computational methodology and technical validation are presented along with the format of our publicly available data record.

### Background & Summary

Raman spectroscopy is an effective method for obtaining vibrational frequencies and local chemical bonding information in condensed materials. Due to its relative quick and non-destructive analytical nature, it is widely used in the analysis of crystalline materials and polymers - extending even into bio-medical materials and pharmaceuticals<sup>1-3</sup>, with advanced technologies and portable devices developed over the years. Indispensable for a complete vibrational spectrum observation, Raman spectroscopy has established a unique role in obtaining “fingerprints” for modern material science characterization. The increase in open access of experimentally measured Raman spectra databases, such as the RRUFF<sup>4</sup> Project and Bio-Rad’s KnowItAll Raman Spectral Library, has incurred increased popularity. These resources now contain thousands of Raman spectral data of materials ranging from minerals, semiconductors to prescription drugs, which greatly facilitates the referencing process and identification of novel materials.

Compared to an experimental Raman database, a computational Raman Spectrum database from ab initio electronic structure calculations has many advantages<sup>5-7</sup>. Cheaper and more abundant computing resources enable thousands of calculations for a larger number of materials in an automated fashion, significantly reducing human effort. Systematic calculations result in a uniform data type, enabling accelerated classification of modes, discovery of correlations, and screening of useful materials. However, despite these advantages, there has been limited effort in constructing a computational Raman Spectrum database or a systematic workflow for its development. The WURM Project<sup>8</sup> was an early prototype, but has shown little to no activity since 2012.

In this paper, we develop a computational workflow to calculate Raman spectra using ab initio density functional perturbation theory (DFPT): an accessible and fairly accurate tool for describing lattice dynamics<sup>9</sup>. Following the development of the dielectric tensor components in *Petousis et al.*<sup>10</sup>, computed directly using DFPT as implemented in VASP, the Raman tensors of 55 inorganic compounds are derived. They are then used to generate the Raman spectra data reported, which is benchmarked against experimental Raman spectra. These Raman spectra are integrated into the Materials Project (MP)<sup>11</sup>, a free, online database with computed properties to enable accelerated materials design. While Raman spectra are still computationally expensive, this work establishes a systematic workflow and associated benchmarking of numerical parameters for automated Raman spectra calculation that can be utilized by the community at large, and help increase the availability of reference Raman spectra for a wide variety of structures and chemical systems.

### Methods

**Theory and definitions.** In Raman spectroscopy measurements of crystalline systems, incident laser photons with frequency  $\omega_L$  interact with lattice vibrations, which can be described in the form of phonons. The inelastically scattered photons either exhibit a decrease or increase in frequency, resulting in symmetric Stokes or Anti-Stokes shifts, respectively. The recorded spectrum shows the intensity of the scattered light as a function of its frequency

<sup>1</sup>Department of Materials Science and Engineering, University of California, Berkeley, California, 94720, USA.

<sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA. Correspondence and requests for materials should be addressed to S.D. (email: [shyamd@lbl.gov](mailto:shyamd@lbl.gov))

change, often converted to wavenumbers  $\nu$  in units of  $\text{cm}^{-1}$ , known as Raman shifts. The Raman spectrum therefore provides a “fingerprint” description of the normal modes in the crystal through the comparison of incident and scattered photon frequencies. In our work, a normal mode’s wavenumber  $\nu_{mode}$  is directly obtained through a DFPT phonon calculations. The intensity of the mode  $i$  with polarization along  $\beta$  and electric field along  $\gamma$  is represented<sup>12–15</sup> as

$$I_{i\beta\gamma} = \frac{2\pi\hbar(\omega_L - \omega_i)^4}{c^4\omega_i} [n(\omega_i) + 1] \alpha_{i\beta\gamma}^2 \quad (1)$$

where  $\omega_i$  is the vibrational mode frequency and  $\omega_L$ , the laser frequency, is usually arbitrarily set in computational methods since the shift in photon frequency is independent of incident photon frequency due to the nature of the interaction.  $n(\omega_i) + 1 = \left(1 - e^{-\frac{\hbar\omega_i}{k_B T}}\right)^{-1}$  is the Bose occupation number that is dependent on both temperature and frequency. In order to accommodate the most common experimental conditions, we set  $T = 300$  K and  $\omega_L$  to the wavenumber corresponding to a 532 nm wavelength laser. For each mode, the Raman tensor<sup>12,16</sup>

$$\alpha_{i\beta\gamma} = \frac{\sqrt{\Omega}}{4\pi} \sum_{n\gamma} \frac{\partial \varepsilon_{\beta\gamma}}{\partial u_{in\nu}} e_{in\gamma} m_n^{-\frac{1}{2}} \quad (2)$$

is obtained from the finite difference derivative of the dielectric tensor,  $\varepsilon_{\beta\gamma}$ , with respect to the displacement  $u_{in\nu}$ .  $m_n$  is the mass of atom  $n$ ,  $\nu$  the direction of displacement,  $\Omega$  the unit cell volume, and  $e_{in\gamma}$  the eigenvector of the dynamical matrix. More specifically, the central difference scheme is employed as atoms are moved independently with displacement of 0.005 Å in both the positive (+) and negative (−) directions to calculate the differential of dielectric tensor with respect to displacement. However, the intensity expression in Eq. (1) is ideal for single crystals. Quite often experimental data use polycrystalline mineral specimens or powdered samples, in which case the intensity must be averaged over all possible orientations of the crystals which is accomplished by separating the total intensity  $I = I_{\perp} + I_{\parallel}$ <sup>17–19</sup> into the depolarized light intensity

$$I_{\perp} \sim (\omega_L - \omega_i)^4 \frac{n(\omega_i) + 1}{30\omega_i} [5G_{i,1} + 3G_{i,2}] \quad (3)$$

and the polarized light intensity

$$I_{\parallel} \sim (\omega_L - \omega_i)^4 \frac{n(\omega_i) + 1}{30\omega_i} [10G_{i,0} + 4G_{i,1}] \quad (4)$$

where the rotation invariants<sup>17,19</sup> are defined as

$$G_{i,0} = \frac{1}{3} (\alpha_{ixx} + \alpha_{iyy} + \alpha_{izz})^2 \quad (5)$$

$$G_{i,1} = \frac{1}{2} [(\alpha_{ixy} - \alpha_{iyx})^2 + (\alpha_{ixz} - \alpha_{izx})^2 + (\alpha_{izy} - \alpha_{iyz})^2] \quad (6)$$

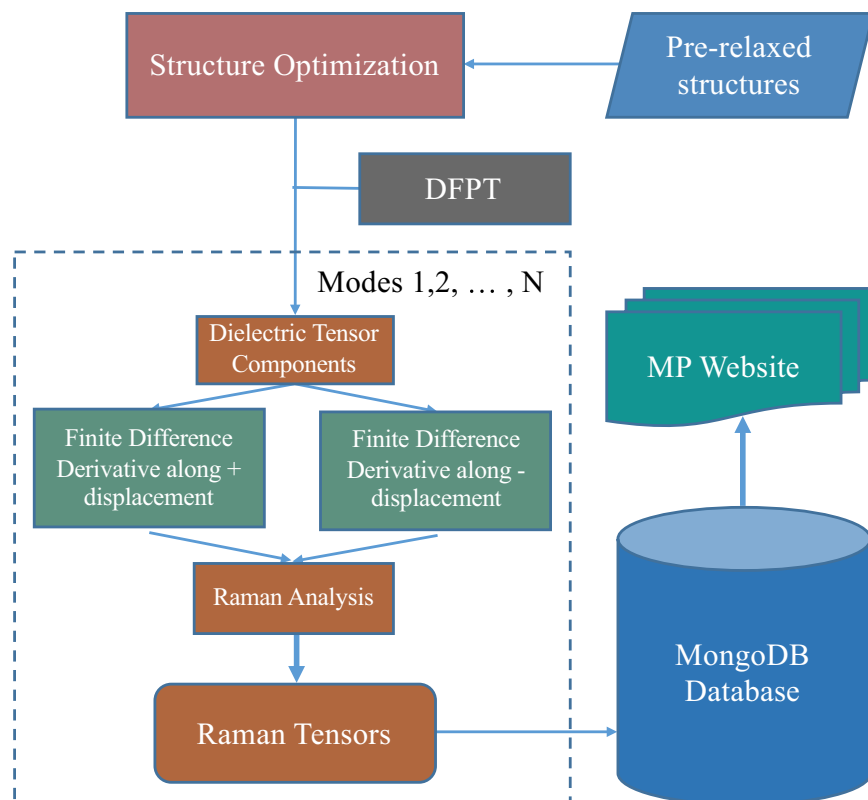
$$G_{i,2} = \frac{1}{2} [(\alpha_{ixy} + \alpha_{iyx})^2 + (\alpha_{ixz} + \alpha_{izx})^2 + (\alpha_{izy} + \alpha_{iyz})^2] + \frac{1}{3} [(\alpha_{ixx} - \alpha_{iyy})^2 + (\alpha_{ixx} - \alpha_{izz})^2 + (\alpha_{izz} - \alpha_{iyy})^2] \quad (7)$$

For the Raman spectra of each compound, the intensities of each mode are normalized by setting the max intensity in the spectra to 100 and then scaling other mode intensities accordingly.

**Computational workflow.** The workflow designed for automatic Raman Spectrum calculations is outlined in Fig. 1. The relaxed material structures are obtained from the Materials Project Database<sup>11,20,21</sup> and serve as input structures to the following DFPT simulation. For all calculations of vibrational normal modes, DFPT simulations were run using the Vienna Ab-Initio Simulation Package, also known as VASP<sup>22–25</sup>, using the GGA/PBE<sup>26</sup> + U<sup>27,28</sup> exchange-correlation functional and projector augmented wave (PAW) pseudopotentials<sup>29,30</sup>. The k-point density was set at 3,000 per reciprocal atom and the plane wave energy cut-off at 600 eV, same as those in the calculation of dielectric tensors<sup>10</sup>. The structure is then displaced along the calculated normal mode eigenvectors for dielectric tensor calculations. Raman susceptibility tensors are calculated from the finite difference derivative of the dielectric tensor as shown in Eq. (2) and the collection of computational Raman analysis data is stored with all associated metadata from the Materials Project.

## Data Records

**Computational data record.** The computational data is stored in a lightweight data-interchange format, and the JSON document can be downloaded directly from the Figshare repository<sup>31</sup>. The nested key/value pairs can be navigated using Tables 1 and 2. The document includes the original structure data from the Materials Project, normal modes information, tensors, and calculated intensities.



**Fig. 1** Computational workflow scheme used to calculate the Raman Tensors.

Key	Datatype	Length/Size	Description
perturbed_structures	dictionary	nmodes	Perturbed crystal structures for Raman analysis
nmodes	number	—	Number of vibrational modes
eigenvals	array	nmodes	List of eigenvalues
eigenvecs	array	nmodes	List of eigenvectors
dielectric_tensor	array	nmodes	List of Dielectric tensors
raman_tensor	array	nmodes	List of Raman tensors
wavenumbers	array	nmodes	List of Raman shifts $\text{cm}^{-1}$
intensity	array	nmodes	List of calculated intensities post-normalization
metadata	dictionary	7	Crystal structure before perturbation. Refer to Table 2.

**Table 1.** Description of the nested JSON structure for the ‘Computational Data’.

**Experimental data record.** The experimental data extracted from the RRUFF database is provided in a separate JSON document with nested structure as shown in Table 3. The file includes mineral type, peak locations, intensity values, RRUFF\_id, and matched Materials Project IDs. To ensure identical structures, each mineral is matched to a compound from the Materials Project by fitting Crystallographic Information Files (CIF) using pymatgen tools or manually referencing to mineralogy data<sup>32</sup>. The RRUFF spectra data and all CIF format files used are included in the Figshare repository<sup>31</sup> for increased transparency and ease in reproduction of the benchmarking results.

### Technical Validation

Unlike evaluating the accuracy of scalar material properties, comparison of the Raman spectra requires a matching process between the calculated vibrational modes and experimental Raman spectra peaks from the RRUFF database<sup>4</sup>. Hence, a linearly modelled cost metric,  $u = w_1(v_{mode} - v_{peak}) + w_2(I_{mode} - I_{peak}) = w_1\Delta v + w_2\Delta I$ , considering the proximity in both wavenumber and normalized intensity is used in the matching process: within the same compound, each of the computed vibrational modes is paired with a corresponding experimental peak that minimizes the overall cost. To ensure the matching process produces more accurate results, we process the calculated Raman spectra data in the following two steps.

First, for each compound’s computed Raman spectra, a signal to noise threshold of 0.356% is used to remove vibrational modes with experimentally undetectable intensities. This threshold is the average noise intensity to

Key	Datatype	Description
mp_id	string	Materials Project ID
formula	string	Chemical Formula
structure	dictionary	Unperturbed crystal structure from MP
point_group	string	Hermann-Mauguin Notation
space_group	number	Space_group number defined by International Union of Crystallography
crystal_system	string	Crystal system
nsites	number	Number of atoms in primitive cells

**Table 2.** Description of the nested JSON structure for ‘metadata’.

Key	Datatype	Description
type	array	Mineral Type(s)
formula	string	Chemical Formula and Mineral Name
RRUFF_id	string	ID from RRUFF database
noise	number	Background noise intensity (before normalization)
start	number	Smallest wavenumber value recorded in the Raman spectra
wavenumbers	array	List of peak locations
intensities	array	List of peak intensities (before normalization)
mp_id	string	Matched ID from the Materials Project

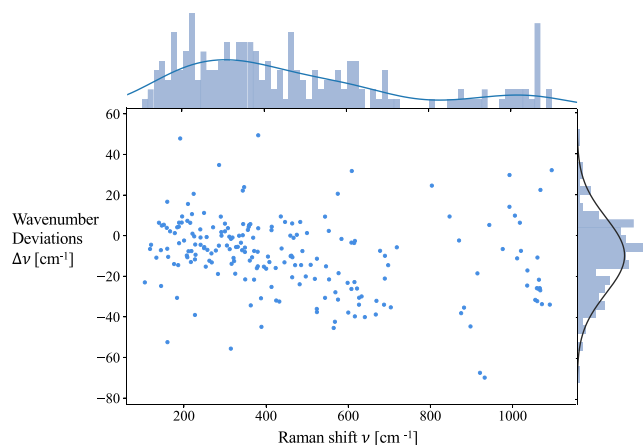
**Table 3.** Description of the nested JSON structure for extracted ‘Experimental Data’.

maximum intensity obtained from reference experimental data. Second, for each compound, the low wavenumber region is removed from the computational Raman spectrum for both experimental and computational reasons. From an experimental point of view, the RRUFF database records different initial probing locations for different compounds’ Raman spectra, which presents a noticeable challenge for the peak matching process. The average distance between the starting location of RRUFF Raman spectra and its first peak location is  $113\text{ cm}^{-1}$ , which provides a location reference for the computational Raman spectra of each compound. From a computational perspective, DFPT tend to exhibit larger errors at lower wavenumbers; these modes correspond to long range oscillations that are difficult to compute within limited size of periodic unit cells. For example, these low wavenumber modes, corresponding to phonon wavelengths on the orders of microns, can be influenced by long range features such as defects, dislocations and grain boundaries that are not captured in the simulations of perfectly periodic, single crystal structures.

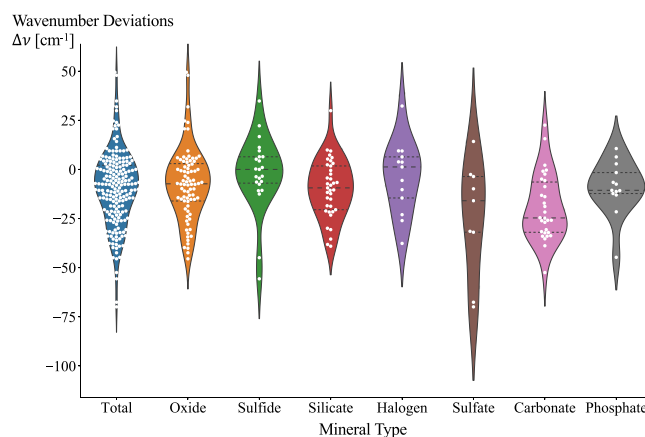
The data consists of 55 compounds and 205 pairs of matched peaks/modes, where the total calculation cost amounted to 9.5 Million CPU hours. The accuracy of the calculated phonon modes’ wavenumber values  $\nu_{mode}$  is represented in Fig. 2, where the compounds range from simple oxides, sulfides to carbonates. The average wavenumber deviation  $\Delta\nu$  for all the compounds is  $-9.66\text{ cm}^{-1}$  with a standard deviation of  $18.58\text{ cm}^{-1}$ , which indicates that our calculations tend to slightly underestimate the Raman peak locations. With a mean absolute relative error (MARE) of  $-2\%$ , we conclude that our test set is in reasonably good agreement with the experimental data available. In Fig. 3, violin plots are used to depict the wavenumber deviations grouped by mineral type. The oxide family consists 40% of the data, and shows weak underestimation of peak locations with average absolute wavenumber deviation of  $-7.5\text{ cm}^{-1}$ . The sulfate and carbonate families exhibit the largest wavenumber deviations. While the sulfate family’s larger deviation is more likely due to the limited number of sulfate compounds used in this work, the carbonate family exhibits more abundant data for distribution analysis, showcasing in an average wavenumber deviation of  $-19.3\text{ cm}^{-1}$ . Most notably, we observe large wavenumber deviations of carbonates’ modes at Raman shifts around  $1000\text{ cm}^{-1}$ . These modes are all generated from the vibrations of oxygen anions, which could be affected by small (unknown) amounts of water in mineralogical samples, which are not accounted for in the calculations.

The intensity deviations do not show a clear bias. Most often, as seen in Fig. 4, carefully tuned weights  $w_1$  and  $w_2$  can yield well paired peaks and modes, yet discrepancies in intensity still remain. It should be noted that our structures from MP assume a perfectly ordered bulk crystal at 0 K, and therefore extrinsic factors associated with temperature, pressure, defects, and phonon anharmonicity are not considered. These effects on the dielectric tensors are described in detail in Petousis *et al.*<sup>33</sup>, which carries over to the Raman tensors and eventually affects the calculated intensities. Interfaces and defects in particular are expected to be a strong scattering center for phonons which could account for much of the intensity variation as the phonon population deviates from a simple Bose-Einstein distribution.

Besides the extrinsic effects mentioned above, there are other factors that could contribute to discrepancies between computational and experimental values. One such effect is that of the use of the GGA/PBE exchange-correlation function, which has been shown to exhibit lattice under binding<sup>34</sup>, yielding larger bond lengths, lattice parameters and unit cell volumes of the relaxed structures. We investigated the correlation between the unit cell volume expansions % and Raman peaks’ wavenumber discrepancies ( $\text{cm}^{-1}$ ) to determine if the GGA/



**Fig. 2** Plot of wavenumber deviations  $\Delta\nu$  of computed modes versus their wavenumbers. Each data point belongs to a matched pair of calculated phonon mode and experimental peak. The x axis shows the wavenumbers of phonon modes,  $\nu = \nu_{mode}$ , while the y axis shows how much these wavenumbers of phonon modes differ from those of their matched experimental peaks, namely  $\Delta\nu = \nu_{mode} - \nu_{peak}$ . The Gaussian kernel density estimates of  $\nu$  and the normal distribution of  $\Delta\nu$  are shown respectively on the x and y axis.

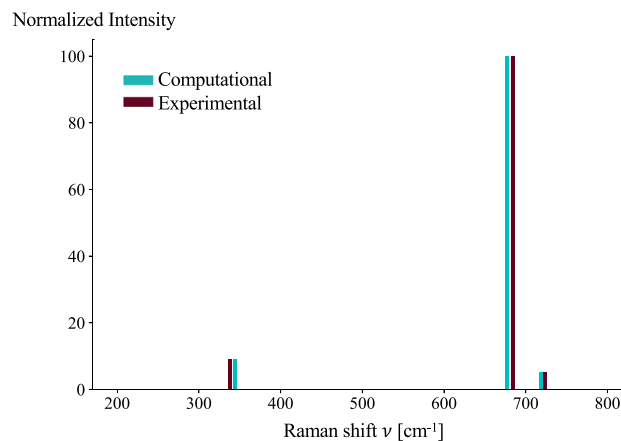


**Fig. 3** Distribution of wavenumber deviations  $\Delta\nu$  by mineral type. The violin outlines are Gaussian kernel density estimates of the data points in the middle. The quartiles are labelled with dashed lines.

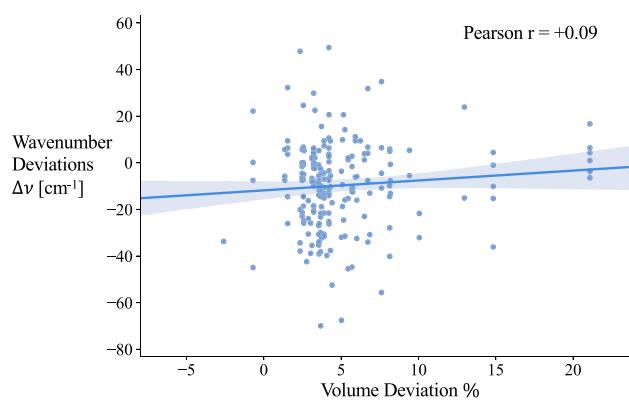
PBE known under-binding causes a systematic and significant deviation of the computed Raman data. Despite an average volume expansion of 4.94%, the weak correlation between the two (+0.09), as seen in Fig. 5, suggested that the effects of GGA/PBE is not significant in this work. Therefore, we do not attribute the observed wavenumber discrepancies to the use of GGA/PBE exchange correlation function. Other salient differences between computational assumptions and experimental conditions may also contribute to the discrepancies. For example, the applied rotation invariant intensity equation is an average result considering every possible scattering geometry, and therefore assumes a powdered or isotropic poly-crystalline sample. However, the quality of mineral samples do not always satisfy these ideal conditions, which may allow specific sample orientations or laser polarization directions in experimental measurements to create discrepancies in Raman spectra.

### Usage Notes

An automatic workflow for ab initio electronic structure Raman spectra is expected to be of interest to the broad community of materials science and chemistry. We expect this growing dataset to be used in the understanding of Raman characterization and provide a reference for experimentally measured Raman data. Furthermore, our work opens up opportunities in the development of a data intensive approaches. For example, the application of machine learning techniques on existing spectra may be utilized to enhance the data set as well as to improve the peak matching process. Such work provide the basis for quick identification by screening through compounds on the Materials Project website, comparing their computational Raman spectra data to the reference experimental counterpart. Still, we emphasize that the computational cost for developing this database is rather significant. While this work illustrates the promise of a computational Raman spectra database, further work is necessary to reduce the computational cost to access truly large data sets needed for machine learning approaches.



**Fig. 4** The computational and experimental Raman spectra of BeO (Bromellite). The direct comparison of computational and experimental Raman spectra shows that even when the peaks and modes are well paired, intensity deviation can still be observed.



**Fig. 5** Plot of wavenumber deviation  $\Delta\nu$  of computed modes versus volume deviation % of the compounds they belong to. The volume deviation % of each compound is calculated as  $\frac{V_{MP} - V_{RRUFF}}{V_{MP}} \cdot 100\%$ .

### Code Availability

The proprietary VASP-code is primarily used in the DFPT calculations. The processing and modifications of the simulations were implemented using Pymatgen<sup>20</sup> and FireWorks<sup>35</sup>. Pymatgen (Python Materials Genomics) is an open-source Python library under Massachusetts Institute of Technology (MIT) license for materials analysis. The workflow shown in Fig. 1 is implemented using FireWorks in Atomate<sup>36</sup>, which stores, executes, and manages calculation workflows and is free to public through Atomate's Github site under a modified GNU General Public License.

### References

1. Das, R. S. & Agrawal, Y. Raman spectroscopy: recent advancements, techniques and applications. *Vib. Spectrosc* **57**, 163–176 (2011).
2. Schrader, B. *Infrared and Raman spectroscopy: methods and applications* (John Wiley & Sons, 2008).
3. Parker, F. S. *Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry* (Springer Science & Business Media, 1983).
4. Lafuente, B., Downs, R. T., Yang, H. & Stone, N. The power of databases: The RRUFF project. *Highlights in Mineralogical Crystallography* **1**, 1–29 (2016).
5. Sluiter, M. H., Simonovic, D. & Tasci, E. S. Materials databases for the computational materials scientist. *Int J Min Met Mater* **18**, 303–308 (2011).
6. Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **17**, 1571–1586 (1996).
7. Curtarolo, S. *et al.* The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191 (2013).
8. Caracas, R. & Bobocioiu, E. Theoretical modelling of Raman spectra. In *Raman spectroscopy applied to Earth sciences and cultural heritage* (Mineralogical Society of Great Britain and Ireland, 2012).
9. Baroni, S., De Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys* **73**, 515 (2001).
10. Petousis, I. *et al.* High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. *Sci. Data* **4**, 160134 (2017).
11. Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater* **1**, 011002 (2013).
12. Prosandeev, S., Waghmare, U., Levin, I. & Maslar, J. First-order raman spectra of  $AB'_{1/2}B''_{1/2}O_3$  double perovskites. *Phys. Rev. B* **71**, 214307 (2005).

13. 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 (2003).
14. Petzelt, J. & Dvorak, V. Changes of infrared and raman spectra induced by structural phase transitions. ii. examples. *J. Phys. Condens. Matter* **9**, 1587 (1976).
15. Born, M. & Huang, K. *Dynamical theory of crystal lattices* (Clarendon press, 1954).
16. Ceriotti, M., Pietrucci, F. & Bernasconi, M. Ab initio study of the vibrational properties of crystalline  $\text{TeO}_2$ : the  $\alpha$ ,  $\beta$ , and  $\gamma$  phases. *Phys. Rev. B* **73**, 104304 (2006).
17. Placzek, G. *Handbuch der Radiologie* (Akademische Verlagsgesellschaft, 1934).
18. Hayes, W. & Loudon, R. *Scattering of light by crystals* (Wiley, 1978).
19. Hamaguchi, H.-O. The resonance effect and depolarization in vibrational raman scattering. *Advances in infrared and Raman spectroscopy* **12**, 273–310 (1985).
20. Ong, S. P. *et al.* Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
21. Ong, S. P. *et al.* The materials application programming interface (api): A simple, flexible and efficient api for materials data based on representational state transfer (rest) principles. *Comput. Mater. Sci.* **97**, 209–215 (2015).
22. Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558 (1993).
23. Kresse, G. & Hafner, J. *Ab initio* molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251 (1994).
24. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
25. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
26. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
27. Dudarev, S., Botton, G., Savrasov, S., Humphreys, C. & Sutton, A. Electron-energy-loss spectra and the structural stability of nickel oxide: An lsd + u study. *Phys. Rev. B* **57**, 1505 (1998).
28. Jain, A. *et al.* A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* **50**, 2295–2310 (2011).
29. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
30. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
31. Liang, Q., Dwaraknath, S. & Persson, K. A. High-throughput Computation and Evaluation of Raman Spectra. *Figshare*, <https://doi.org/10.6084/m9.figshare.7427393> (2018).
32. Anthony, J. W., Bideaux, R. A., Bladh, K. W. & Nichols, M. C. Handbook of mineralogy, volume iv, arsenates, phosphates, vanadates. 1–680, *Mineralogical Society of America, Chantilly, Virginia* (2000).
33. Petousis, I. *et al.* Benchmarking density functional perturbation theory to enable high-throughput screening of materials for dielectric constant and refractive index. *Phys. Rev. B* **93**, 115151 (2016).
34. Van de Walle, A. & Ceder, G. Correcting overbinding in local-density-approximation calculations. *Phys. Rev. B* **59**, 14992 (1999).
35. Jain, A. *et al.* Fireworks: A dynamic workflow system designed for high-throughput applications. *Concurr Comput: Pract E.* **27**, 5037–5059 (2015).
36. Mathew, K. *et al.* Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **139**, 140–152 (2017).

## Acknowledgements

This work was funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05CH11231 (Materials Project program KC23MP). This work made use of resources of the National Energy Research Scientific Computing Center (NERSC), supported by the Office of Basic Energy Sciences of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## Author Contributions

Q.L. performed the phonon and Raman spectra data calculations, obtained experimental data from online database, developed peak matching algorithm and the code for visualization, performed the data benchmarking and analysis, and wrote the paper. S.D. contributed to supervision of the calculations, the data analysis and planning of the paper. K.P. was involved in supervising and planning the work and its integration with the Materials Project effort.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**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 Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2019