THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC
Article

Delta Machine Learning for Predicting Dielectric Properties and Raman Spectra

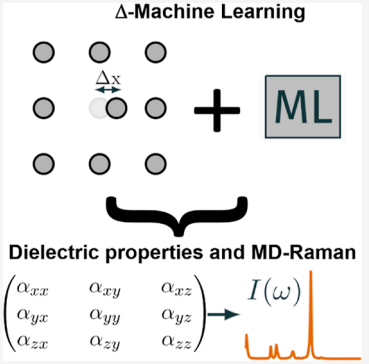
Manuel Grumet, Clara von Scarpatetti, Tomáš Bučko,* and David A. Egger*

Cite This: J. Phys. Chem. C 2024, 128, 6464−6470

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Raman spectroscopy is an important characterization tool with diverse applications in many areas of research. We propose a machine learning (ML) method for predicting polarizabilities with the goal of providing Raman spectra from molecular dynamics trajectories at a reduced computational cost. A linear-response model is used as a first step, and symmetry-adapted ML is employed for the higher-order contributions as a second step. We investigate the performance of the approach for several systems, including molecules and extended solids. The method can reduce the training-set sizes required for accurate dielectric properties and Raman spectra in comparison to a single-step ML approach.



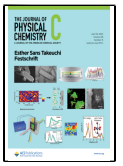
INTRODUCTION

Atomic motions are often key to physical and chemical phenomena occurring at finite temperature, in both solid-state and molecular systems. Experimentally, the dynamic behavior of such systems can be probed by Raman spectroscopy. It is a table-top technique that is available in many laboratories because it is less complicated and expensive than, for example, neutron scattering.[1] Raman spectroscopy plays an important role in many areas of research, including catalysis,[2−5] perovskite photovoltaics[6−9] and semiconductor physics more broadly,[10] and ionic conductors.[11−13] Computational predictions of Raman spectra using first-principles calculations are an important counterpart to experimental measurements. They provide further insights into the behavior of materials, facilitate the interpretation of measured spectra via theory−experiment comparisons, and enable predictions of dynamic properties in new compounds.

The central quantity for theoretical calculations of the Raman spectra is the polarizability tensor, $\boldsymbol{\alpha}$. It describes the first-order dielectric response of a system to external electric fields. In practice, the dielectric tensor, $\boldsymbol{\epsilon}$, can be used for periodic systems since it contains the same information;[14] in the following text, we shall use both quantities interchangeably. Raman spectra are commonly calculated within the harmonic approximation whereby the derivatives of $\boldsymbol{\alpha}$ with respect to atomic displacements, determining the intensity of peaks, are calculated along eigenvectors of harmonic modes.[15−17] But this method is limited since it cannot capture anharmonic effects, higher-order Raman scattering, or the explicit temperature dependencies of Raman modes. These effects are relevant in a variety of physical systems and scenarios. For example, a description of phase transitions in solid materials requires temperature-dependent phonon modes.[18] While a proper description of phase transitions and thermal behavior of vibrational modes is certainly not an easy task also in methods such as molecular dynamics (MD), a strictly harmonic phonon picture prevents the capture of these effects even in principle.

By contrast, MD simulations offer a way to include these effects and overcome the limitations of the harmonic approach. A Raman spectrum can be computed from an MD trajectory by calculating Fourier-transformed velocity autocorrelation functions (VACFs) of the components of $\boldsymbol{\alpha}$.[19−21] However, this requires computing a time series of $\boldsymbol{\alpha}$ values from multiple MD snapshots. The number of data points needed in such an MD-Raman approach depends on the desired frequency resolution and the total range of frequencies that needs to be covered, but typically, at least a few hundred points are needed. Such polarizability calculations can be done from first principles using density functional perturbation theory (DFPT),[22] but this renders MD-based Raman calculations computationally demanding. The cost of the DFPT computations usually far outweighs that of the MD simulation itself and forms the main bottleneck in the MD-Raman approach. These large computational efforts limit the range of

Received: February 9, 2024
Revised: March 16, 2024
Accepted: March 20, 2024
Published: April 3, 2024



https://doi.org/10.1021/acs.jpcc.4c00886
J. Phys. Chem. C 2024, 128, 6464−6470

physical scenarios and systems that one can investigate with the method.

The computational cost of first-principles calculations can be significantly reduced by machine learning (ML) techniques. Specifically, ML can exploit redundancies and symmetries of structures generated by MD and is now widely used for learning and predicting energies and forces.[23−25] Relevant for MD-Raman calculations, ML force fields can greatly accelerate the MD part of the simulations.[26] Yet, the cost of computing polarizability tensors from DFPT remains. Hence, more recently, ML methods have also been applied to predict tensorial properties including polarizabilities, which currently is a very active research area.[27] Both symmetry-adapted kernel-based methods[28−33] and neural-network approaches[34−39] have been used for this, as well as a physical based small parametric model.[40] In addition, ML methods have also been applied to compute aspects of Raman spectra directly, without explicit consideration of polarizabilities.[41−43] Delta ML (Δ-ML) is a combined approach to predicting physical quantities: a computationally inexpensive approximation is used as a first step and ML methods are then applied to learn only the differences between first-step predictions and true values.[44] The challenge in any Δ-ML method is the search for a physical model to achieve sound first-step predictions that can be seamlessly combined with the proceeding ML method. It has been previously shown that the prediction of polarizabilities in a molecular crystal can be improved by using the polarizability of the molecular monomer as a first step.[28] However, to the best of our knowledge, no systematic Δ-ML approach that can be applied beyond the case of molecular crystals has been proposed and investigated yet for the prediction of dielectric properties such as polarizabilities and Raman spectra.

In this paper, we propose a Δ-ML method for predicting dielectric properties of molecules and materials. Focusing on polarizabilities, we suggest a linear response model (LRM) that encodes key information about the dielectric response of the system as a first step. Combination of the LRM with ML for tensorial properties is assessed via inspection of polarizability components and Raman spectra. We find that the Δ-ML method increases accuracy and reduces the required training-set size compared to a direct ML approach where the same ML model is applied to the polarizability data directly, i.e., without using any first-step approximation. Applying Δ-ML to small molecules and extended solids, including more complicated materials, as well as discussing its inherent limitations and potential for further improvements, we demonstrate its predictive power for practical MD-Raman calculations across a broad range of physical systems.

RESULTS AND DISCUSSION

In order to develop our method, we start from a Taylor expansion of a component of $\boldsymbol{\alpha}$ with respect to atomic displacements ($\Delta\boldsymbol{x}$) from their respective equilibrium positions or any other reference structure ($\boldsymbol{x}_0$)

$$\alpha_{\alpha\beta}(\boldsymbol{x}) = \alpha_{\alpha\beta}(\boldsymbol{x}_0) + \sum_{i=1}^{3N_{at}} \frac{\partial \alpha_{\alpha\beta}}{\partial x_i}\Delta x_i + \dots \quad (1)$$

N_{at} is the number of atoms of the system and the index i enumerates the components of a $3N_{at}$ dimensional atomic position vector $\boldsymbol{x}$. Equation 1 is an exact formula which can be approximated to arbitrary order. Here, we consider the simplest, first-order variant: an LRM is constructed by determining the constant term, $\boldsymbol{\alpha}(\boldsymbol{x}_0)$, and the respective derivatives, $\partial\alpha_{\alpha\beta}/\partial x_i$. The constant term can be determined via a single DFPT calculation for the equilibrium structure. The first-order derivatives can be obtained through additional DFPT calculations of displaced structures. Applying a central difference method, two calculations are necessary for each atomic degree of freedom. We note that the problem of choosing the coordinate frame in systems with rotational degrees of freedom can easily be solved by employing rotationally and translationally invariant internal coordinates, as discussed in Section I.E of the Supporting Information. A numerical demonstration that rigid rotations do not significantly deteriorate predictions of our LRM model is given in Section I.H of the Supporting Information. Symmetry considerations can be used in this procedure in order to reduce the number of required calculations since any two symmetry-equivalent atoms imply derivatives that are related via similar symmetry operations as position vectors of atoms. Thus, a total of $6N + 1$ calculations are required to parametrize the LRM, where N is the number of symmetry-inequivalent atoms in the system. Further details about the LRM are presented in the Supporting Information, including an overview of the number of DFPT calculations performed in Table S5.

We parametrize the LRM via DFPT calculations[45] using VASP[46] and the PBE exchange−correlation functional.[47] This allows for predicting the polarizabilities of MD snapshots by extracting the displacements Δx_{ij} from it and applying eq 1. Figure 1a shows the performance of predictions with the LRM for the α_{xx} component in SiO_2 along a selected portion of an MD trajectory at 300 K, which we obtain using density functional theory (DFT) in VASP. Compared to the DFPT reference data for the same MD trajectory, the LRM accurately captures many of the α_{xx} oscillations.

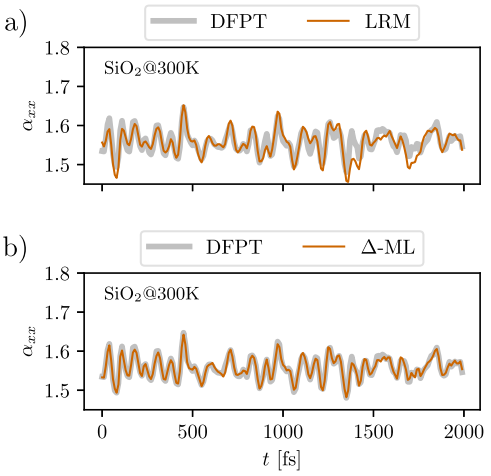


Figure 1. (a) LRM predictions of the α_{xx} component of the polarizability tensor in SiO_2 over a period of 2000 fs, compared to DFPT reference data. (b) Δ-ML predictions for the same data, obtained with $N_t = 50$ and $N_v = 10$. Note that the values shown here are polarizability per volume.

The LRM has several limitations: first, the polarizability can have a local extremum at the reference positions, in which case the derivatives, $\partial\alpha_{\alpha\beta}/\partial x_i$, vanish. This occurs when atomic motion along the respective axis is first-order Raman-inactive. Additionally, it is also possible that these derivatives are zero

https://doi.org/10.1021/acs.jpcc.4c00886
J. Phys. Chem. C 2024, 128, 6464−6470

for only certain components of $\boldsymbol{\alpha}$, which will be discussed below. In these cases, the LRM can only capture the constant term and the predicted polarizability time series for these components will therefore be approximately constant. These are important limitations of the present LRM, which, in principle, can be overcome in a straightforward way by accounting for further terms in the expansion of eq 1. Here, we focus on the fact that these limitations motivate an ML model as an additional second step that can capture higher-order displacement responses.

For the second step in our Δ-ML approach, we employ kernel-based methods.[48,49] The underlying idea is to use a descriptor that captures relevant aspects of atomic configurations and a kernel function that allows one to quantify similarities between different configurations along MD trajectories. These similarities can then be used to perform fitting using kernel ridge regression (KRR). For many physical quantities, descriptors based on overlap integrals are an appropriate choice. A common method which takes into account symmetries is smooth overlap of atomic positions (SOAP).[50−53] For fitting tensorial quantities such as $\boldsymbol{\alpha}$, the extended λ-SOAP approach is well-suited because it also takes into account tensorial covariance properties.[54] While the choice for λ-SOAP implies no difference for some tensor invariants such as the mean polarizability (see Section I.C in Supporting Information), it does provide a benefit for fitting the off-diagonal elements of $\boldsymbol{\alpha}$.

We perform the fitting procedure in KRR with λ-SOAP using the DScribe[55] and librascal[56] packages together with scikit-learn.[57] The training data for the ML model are obtained from DFPT calculations on a subset of MD trajectories. Note that the $6N + 1$ configurations that were required for the LRM can, in principle, be reused as training data for the ML model, which we did not attempt here to ease the comparison. Furthermore, a validation data set is selected from the MD trajectory in order to optimize ML hyperparameters with respect to validation error. For the sizes N_t and N_v of the training and validation set, respectively, we used a constant ratio $N_t/N_v = 5{:}1$. This ensures that the validation set is scaled in proportion whenever the number of training data points is increased. In the spirit of Δ-ML, LRM predictions are always subtracted from DFPT values before feeding them to ML, such that only differences are learnt. Further details on the fitting procedure can be found in Section I.F of the Supporting Information.

Figure 2 shows scatterplots of a polarizability time series for SiO_2 predicted with Δ-ML and an otherwise identical ML approach without the LRM (direct-ML), comparing both to reference DFPT calculations along an MD trajectory. The diagonal components of $\boldsymbol{\alpha}$ are already predicted well with direct-ML compared to the DFPT reference data. For this particular case, even the LRM alone achieved fairly accurate predictions (cf. Figure 1a), implying that the diagonal components are relatively easy to capture for SiO_2. However, the off-diagonal components are predicted far less accurately in direct-ML. Remarkably, we find that our Δ-ML method provides similarly accurate predictions for both the diagonal and off-diagonal components of $\boldsymbol{\alpha}$ (see Figures 2b and 1b). While the off-diagonal components are smaller in magnitude, it is still important to accurately capture their fluctuations in order to predict Raman spectra. Altogether, our findings suggest that the Δ-ML approach allows for more efficient learning of all components of $\boldsymbol{\alpha}$ because the LRM provides a good first approximation for dynamic fluctuations of this quantity.

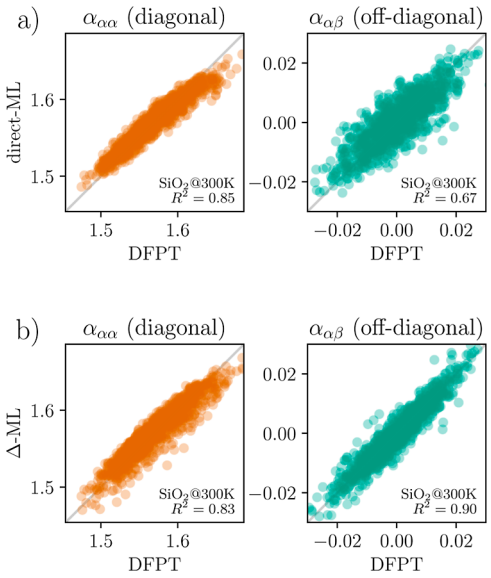


Figure 2. (a) Scatterplot comparing direct-ML predictions of diagonal (left) and off-diagonal (right) polarizability components in SiO_2 to DFPT reference data, obtained with $N_t = 20$ and $N_v = 4$. (b) Scatterplot comparing Δ-ML predictions to DFPT, obtained with the same training and prediction set. Note that the values shown here are polarizability per volume.

To assess the accuracy of our method, we compared it again to direct-ML for the case of SiO_2, using DFPT results as a reference and calculating the coefficient of determination, R^2 (see Section I.G in the Supporting Information for details). We focus on the tensor invariants a (mean polarizability) and γ^2 (anisotropy) of $\boldsymbol{\alpha}$ (see Section I.C of the Supporting Information), since these are directly relevant for Raman calculations and show R^2 for SiO_2 as a function of N_t in Figure 3. We find that for achieving R^2 close to 1 for both a and γ^2, Δ-ML requires N_t to be on the order of only 20, which outperforms direct-ML by at least a factor of 2. Thus, the LRM and ML methods encoded in Δ-ML complement each other well and may offer accurate predictions of $\boldsymbol{\alpha}$ with smaller training-set sizes.

The size of the training set required to obtain a good prediction performance is expected to strongly depend on the system. Therefore, we investigate the versatility of Δ-ML by computing R^2 for a broader range of physical systems that include extended solids and gas-phase molecules. The N_t values that are required to achieve good prediction performance for both polarizability components and Raman spectra are listed in Table 1 for each system. It should be noted that in many cases, an accurate prediction of the Raman spectrum requires much lower N_t than a prediction of the individual polarizability components. Therefore, the Δ-ML method can provide substantial computational savings if one is only interested in the quality of the spectra (see Section III of the Supporting Information for more details).

The solid AlN is an interesting, challenging example because of its known LO/TO splitting that makes the prediction of its dielectric and Raman properties a difficult problem.[17] Thus, AlN illustrates several complications that require an accurate computational methodology for the prediction of its Raman spectrum. For this case as well as for SiO_2, we find that Δ-ML

https://doi.org/10.1021/acs.jpcc.4c00886
J. Phys. Chem. C 2024, 128, 6464−6470

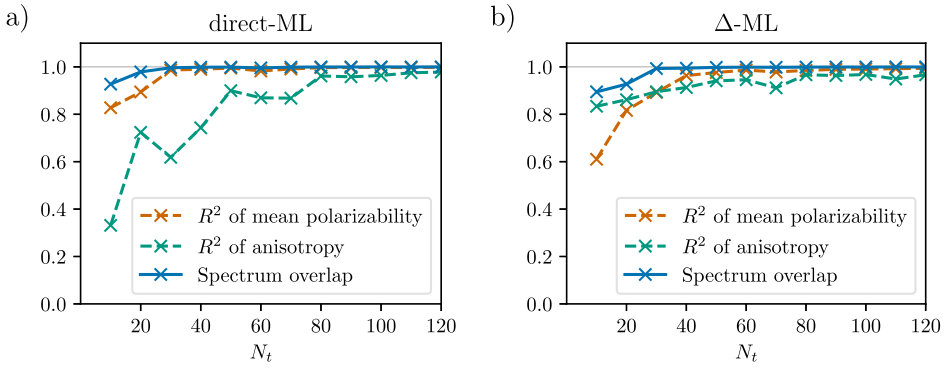


Figure 3. (a) Performance metrics for direct ML predictions in SiO_2 as a function of training-set size N_t and the (b) same performance metrics for Δ-ML predictions.

Table 1. Minimum Required Training-Set Sizes to Achieve $R^2 > 0.8$ in Both a (Mean Polarizability) and γ^2 (Anisotropy) of α for Different Systems[a]

	direct-ML	Δ-ML
SiO_2	50	20
AlN	360	280
Si	60	60
NaCl	440	440
H_2O	10	10
CH_4	20	10
CH_3OH	10	10

[a] N_t/N_v was kept constant at 5:1, N_t was increased in steps of 20 for AlN and NaCl and in steps of 10 for all other systems, and R^2 was evaluated on separate data sets of size 400.

achieves a significant reduction of the required N_t compared to direct-ML. The small gas-phase molecules (H_2O, CH_4, and CH_3OH) we consider here are found to require similarly low N_t ~10 in both approaches, further demonstrating the broad applicability of the Δ-ML method.

We consider two more extended systems, Si and NaCl, in order to demonstrate how the aforementioned inherent limitations of the chosen LRM are compensated by the proceeding ML step: the change of the diagonal components of $\boldsymbol{\alpha}$ with Δx_i is approximately an even function in Si (see Supporting Information), for which the simple LRM predicts a constant time series for the $\alpha_{\alpha\alpha}(\boldsymbol{x}(t))$ values that does not capture the pertinent temporal fluctuations in the system. NaCl is another challenging case because it is not first-order Raman-active and the LRM therefore merely predicts $\alpha_{\alpha\beta}(\boldsymbol{x}(t))$ = const. Table 1 shows that Δ-ML and direct-ML lie on par for these two cases; i.e., the proceeding ML can compensate for the lack of dynamical information in the underlying LRM. Hence, even when, for physical reasons, the choice of our specific LRM may not provide any benefit for learning components of $\boldsymbol{\alpha}$, it does not worsen prediction performance. Therefore, the Δ-ML approach can seamlessly integrate a simple physical model and ML procedure in order to capture temporal and spatial fluctuations of $\boldsymbol{\alpha}$. These findings suggest that Δ-ML is a promising approach for the dielectric predictions of dynamical systems. Furthermore, since the LRM is the simplest approximation to eq 1, it can still be extended in a straightforward way to further improve the prediction performance of Δ-ML if needed.

Close connections of the dielectric quantities discussed here to Raman spectra are established via a correlation-function analysis. Specifically, calculation of the Raman spectrum requires the Fourier-transformed VACFs of the tensor components of $\boldsymbol{\alpha}$, i.e.

$$\langle \dot{\alpha}_{\alpha\beta}(\tau)\dot{\alpha}_{\alpha\beta}(\tau + t)\rangle_\tau = \int_{-\infty}^{\infty} \dot{\alpha}_{\alpha\beta}(\tau)\dot{\alpha}_{\alpha\beta}(\tau + t)\mathrm{d}\tau \tag{2}$$

The terms a_τ and γ_τ^2 can then be computed from the VACFs,[20,58] see Section I.D in the Supporting Information for details. A spherically averaged Raman spectrum can then be obtained as

$$I(\omega) \propto \frac{(\omega_{\mathrm{in}} - \omega)^4}{\omega} \frac{1}{1 - \exp\left(-\frac{\hbar\omega}{k_B T}\right)} \frac{45a_\tau^2 + 7\gamma_\tau^2}{45} \tag{3}$$

Note that the frequency-dependent prefactor in eq 3 is not exact. Several other versions of this equation also exist, which can be derived based on different approximations for taking into account the quantum nature of the atomic motion.[59]

We applied our Δ-ML method to calculate MD-Raman spectra at 300 K for the test systems described above using full DFPT calculations as a reference. Figure 4 showcases two examples, and the other systems are discussed in Section IV of

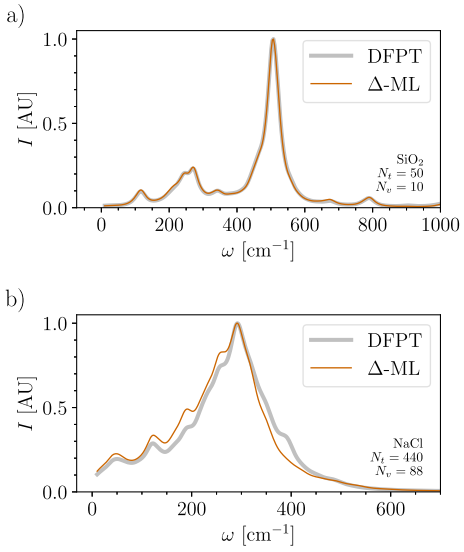


Figure 4. (a) Raman spectrum for SiO_2 computed from Δ-ML predictions, compared to the DFPT reference spectrum. (b) Raman spectrum for NaCl computed from Δ-ML predictions, compared to the DFPT reference spectrum.

https://doi.org/10.1021/acs.jpcc.4c00886
J. Phys. Chem. C 2024, 128, 6464−6470

the Supporting Information. While our focus is on the performance of the Δ-ML method, we point out that the spectra are also affected by other aspects of the computational setup, such as the choice of density functional approximation. Good agreement between Δ-ML and DFPT is obtained for SiO_2 as expected from our above findings and further quantified through the calculated cosine similarity, S_C, shown as a spectrum overlap in Figure 3 (see Section I.G in the Supporting Information). The case of NaCl is particularly interesting since, as we noted above, it is a first-order Raman-inactive material, which means that only higher-order effects contribute to the Raman spectrum. For this reason, the spectrum contains contributions from q-points other than Γ.[60] This makes MD-Raman calculations particularly challenging, requiring a large simulation cell in order to provide sufficient sampling of the q-space (see Section I.A in the Supporting Information for details). In addition, we find that the ML procedure required a relatively large N_t in this case. We speculate that N_t could be reduced via improvement of the first step in the Δ-ML procedure, for example, by including higher-order terms. With sufficient accuracy of terms a and γ, the Raman spectrum for NaCl shows a high overlap with the corresponding DFPT spectrum, with only some minor discrepancies remaining. Furthermore, both the ML and DFPT spectra also show good agreement with recent experimental data.[60] To the best of our knowledge, this is the first ML-based MD-Raman calculation for a higher-order Raman material.

It is interesting to note that for MD-Raman calculations, we observed a certain insensitivity on errors in the individual components of $\boldsymbol{\alpha}$. Specifically, relatively inaccurate predictions of tensor components can still result in a relatively accurate Raman spectrum (see spectrum overlap in Figure 3). In addition, errors at different points along $\boldsymbol{\alpha}(\boldsymbol{x}(t))$ are, to a good approximation, random and independent of each other. Thus, these errors tend to average out when computing autocorrelation functions. However, high accuracy of predictions of the Raman spectra is, of course, not always guaranteed. This can be shown even for the relatively simple gas-phase molecular systems by considering an extremely small training set with $N_t = 5$ and comparing the two ML procedures (see Section IV.C of the Supporting Information). Indeed, a direct-ML approach does not produce physically reasonable Raman spectra for all three molecules as expected from our findings (cf. Table 1). By contrast, the Δ-ML method still yields realistic spectra even with such small training data sets for these simpler molecules. While this is to be expected because of the additional information contained in the LRM, it again confirms the benefit of the Δ-ML method for predicting Raman spectra. Here, one is not restricted to using the LRM per se, and the Δ-ML approach provides room for further adaptation in future work. Possibilities include use of more advanced first-step models instead of the LRM, different descriptors in the ML step, and changes to the hyperparameter optimization scheme, such as employing cross validation.

CONCLUSIONS

In conclusion, we proposed a Δ-ML approach that unifies a physical model with symmetry-adapted ML for prediction of dielectric properties and Raman spectra and is applicable to a broad range of different systems. We focused on the polarizability tensor and chose a simple LRM as a starting point to describe the dynamic dielectric fluctuations, which are completed by an ML procedure in a second step. The Δ-ML method can perform better than an otherwise identical direct-ML approach with the same training-set size. This is especially true because the LRM step provides a benefit for predicting off-diagonal components of the polarizability tensor. Since the data points needed for parametrizing the LRM can be reused as ML training data, Δ-ML does not necessarily increase computational costs compared to direct-ML. We also investigated specific systems for which the LRM method provides no benefit by design and found that it does not deteriorate ML prediction performance for these cases. Our findings show that Δ-ML is a promising approach for the predictions of dielectric properties and Raman spectra of molecules and materials at finite temperature. It provides a way to reliably compute spectra that capture the full extent of atomic motions in molecules and materials without relying on the harmonic approximation at a reasonable computational cost. We speculate that the Δ-ML approach might also be useful for the calculation of other properties that require time-correlation functions such as infrared spectra or transport coefficients.

ASSOCIATED CONTENT

Data Availability Statement

The programs needed to perform the calculations presented in this work can be obtained upon reasonable request from the corresponding authors.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.4c00886.

Further computational details and additional results on displacement dependence of the polarizability in silicon and fitting performance and training-set size as well as additional Raman spectra (PDF)

AUTHOR INFORMATION

Corresponding Authors

Tomáš Bučko − *Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava SK-84215, Slovakia; Institute of Inorganic Chemistry, Slovak Academy of Sciences, Bratislava SK-84236, Slovakia;* orcid.org/0000-0002-5847-9478; Email: tomas.bucko@uniba.sk

David A. Egger − *Physics Department, TUM School of Natural Sciences, Technical University of Munich, Garching 85748, Germany;* orcid.org/0000-0001-8424-902X; Email: david.egger@tum.de

Authors

Manuel Grumet − *Physics Department, TUM School of Natural Sciences, Technical University of Munich, Garching 85748, Germany*

Clara von Scarpatetti − *Physics Department, TUM School of Natural Sciences, Technical University of Munich, Garching 85748, Germany;* orcid.org/0009-0004-3975-3779

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.4c00886

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acs.jpcc.4c00886
J. Phys. Chem. C 2024, 128, 6464−6470

■ ACKNOWLEDGMENTS

C.v.S. thanks Felix Schwarzfischer for helpful discussions. Funding provided by the Alexander von Humboldt-Foundation in the framework of the Sofja Kovalevskaja Award, endowed by the German Federal Ministry of Education and Research, by TUM solar in the context of the Bavarian Collaborative Research Project Solar Technologies Go Hybrid (SolTech), and by TU Munich—IAS, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763, is gratefully acknowledged. The authors further acknowledge the Gauss Centre for Supercomputing e.V. for funding this project by providing computing time through the John von Neumann Institute for Computing on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre. Part of the research was obtained using the computational resources procured in the national project National Competence Centre for high-performance computing within the Operational Programme Integrated Infrastructure (project code: 311070AKF2). T.B. acknowledges support from the Slovak Research and Development Agency under the contract no. APVV-20-0127 and the grant VEGA 1/0254/24 from the Ministry of Education Research, Development and Youth of the Slovak Republic.

■ REFERENCES

(1) Jones, R. R.; Hooper, D. C.; Zhang, L.; Wolverson, D.; Valev, V. K. Raman Techniques: Fundamentals and Frontiers. *Nanoscale Res. Lett.* **2019**, *14*, 231.
(2) Hartman, T.; Wondergem, C. S.; Kumar, N.; van den Berg, A.; Weckhuysen, B. M. Surface- and Tip-Enhanced Raman Spectroscopy in Catalysis. *J. Phys. Chem. Lett.* **2016**, *7*, 1570−1584.
(3) Loridant, S. Raman Spectroscopy as a Powerful Tool to Characterize Ceria-Based Catalysts. *Catal. Today* **2021**, *373*, 98−111.
(4) Hess, C. New Advances in Using Raman Spectroscopy for the Characterization of Catalysts and Catalytic Reactions. *Chem. Soc. Rev.* **2021**, *50*, 3519−3564.
(5) Yoo, R. M. S.; Yesudoss, D.; Johnson, D.; Djire, A. A Review on the Application of In-Situ Raman Spectroelectrochemistry to Understand the Mechanisms of Hydrogen Evolution Reaction. *ACS Catal.* **2023**, *13*, 10570−10601.
(6) Yaffe, O.; Guo, Y.; Tan, L. Z.; Egger, D. A.; Hull, T.; Stoumpos, C. C.; Zheng, F.; Heinz, T. F.; Kronik, L.; Kanatzidis, M. G.; Owen, J. S.; Rappe, A. M.; Pimenta, M. A.; Brus, L. E. Local Polar Fluctuations in Lead Halide Perovskite Crystals. *Phys. Rev. Lett.* **2017**, *118*, 136001.
(7) Ibaceta-Jaña, J.; Muydinov, R.; Rosado, P.; Mirhosseini, H.; Chugh, M.; Nazarenko, O.; Dirin, D. N.; Heinrich, D.; Wagner, M. R.; Kühne, T. D.; Szyszka, B.; Kovalenko, M. V.; Hoffmann, A. Vibrational Dynamics in Lead Halide Hybrid Perovskites Investigated by Raman Spectroscopy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 5604−5614.
(8) Huang, X.; Li, X.; Tao, Y.; Guo, S.; Gu, J.; Hong, H.; Yao, Y.; Guan, Y.; Gao, Y.; Li, C.; Lü, X.; Fu, Y. Understanding Electron−Phonon Interactions in 3D Lead Halide Perovskites from the Stereochemical Expression of 6s2 Lone Pairs. *J. Am. Chem. Soc.* **2022**, *144*, 12247−12260.
(9) Cohen, A.; Brenner, T. M.; Klarbring, J.; Sharma, R.; Fabini, D. H.; Korobko, R.; Nayak, P. K.; Hellman, O.; Yaffe, O. Diverging Expressions of Anharmonicity in Halide Perovskites. *Adv. Mater.* **2022**, *34*, 2107932.
(10) Yu, P.; Cardona, M. *Fundamentals of Semiconductors: Physics and Materials Properties*, 4th ed.; Springer: Berlin, Heidelberg, 2010.
(11) M Julien, C.; Mauger, A.; Julien, C. M.; Mauger, A. In Situ Raman Analyses of Electrode Materials for Li-ion Batteries. *AIMS Mater. Sci.* **2018**, *5*, 650−698.
(12) Famprikis, T.; Bouyanfif, H.; Canepa, P.; Zbiri, M.; Dawson, J. A.; Suard, E.; Fauth, F.; Playford, H. Y.; Dambournet, D.; Borkiewicz, O. J.; Courty, M.; Clemens, O.; Chotard, J.-N.; Islam, M. S.; Masquelier, C. Insights into the Rich Polymorphism of the Na^+ Ion Conductor Na3PS4 from the Perspective of Variable-Temperature Diffraction and Spectroscopy. *Chem. Mater.* **2021**, *33*, 5652−5667.
(13) Brenner, T. M.; Grumet, M.; Till, P.; Asher, M.; Zeier, W. G.; Egger, D. A.; Yaffe, O. Anharmonic Lattice Dynamics in Sodium Ion Conductors. *J. Phys. Chem. Lett.* **2022**, *13*, 5938−5945.
(14) Ángyán, J. G.; Dobson, J.; Jansen, G.; Gould, T. *London Dispersion Forces in Molecules, Solids and Nano- structures. An Introduction to Physical Models and Computational Methods*; The Royal Society of Chemistry: London, 2020; pp 346−347.
(15) Togo, A.; Tanaka, I. First Principles Phonon Calculations in Materials Science. *Scr. Mater.* **2015**, *108*, 1−5.
(16) Skelton, J. M.; Burton, L. A.; Jackson, A. J.; Oba, F.; Parker, S. C.; Walsh, A. Lattice Dynamics of the Tin Sulphides SnS_2, SnS and Sn_2S_3: Vibrational Spectra and Thermal Transport. *Phys. Chem. Chem. Phys.* **2017**, *19*, 12452−12465.
(17) Popov, M. N.; Spitaler, J.; Veerapandiyan, V. K.; Bousquet, E.; Hlinka, J.; Deluca, M. Raman Spectra of Fine-Grained Materials from First Principles. *Npj Comput. Mater.* **2020**, *6*, 121.
(18) Dove, M. T. *Introduction to Lattice Dynamics*; Cambridge University Press: Cambridge, 1993.
(19) Putrino, A.; Parrinello, M. Anharmonic Raman Spectra in High-Pressure Ice from Ab Initio Simulations. *Phys. Rev. Lett.* **2002**, *88*, 176401.
(20) Thomas, M.; Brehm, M.; Fligg, R.; Vöhringer, P.; Kirchner, B. Computing Vibrational Spectra from Ab Initio Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6608−6622.
(21) Ditler, E.; Luber, S. Vibrational Spectroscopy by Means of First-Principles Molecular Dynamics Simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, No. e1605.
(22) Baroni, S.; Resta, R. Ab Initio Calculation of the Macroscopic Dielectric Constant in Silicon. *Phys. Rev. B* **1986**, *33*, 7017−7021.
(23) Noé, F.; Tkatchenko, A.; Müller, K. R.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361−390.
(24) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference. *Phys. Rev. Lett.* **2019**, *122*, 225701.
(25) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K. R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142−10186.
(26) Berger, E.; Lv, Z.-P.; Komsa, H.-P. Raman Spectra of 2D Titanium Carbide MXene from Machine-Learning Force Field Molecular Dynamics. *J. Mater. Chem. C* **2023**, *11*, 1311−1319.
(27) Han, R.; Ketkaew, R.; Luber, S. A Concise Review on Recent Developments of Machine Learning for the Prediction of Vibrational Spectra. *J. Phys. Chem. A* **2022**, *126*, 801−812.
(28) Raimbault, N.; Grisafi, A.; Ceriotti, M.; Rossi, M. Using Gaussian Process Regression to Simulate the Vibrational Raman Spectra of Molecular Crystals. *New J. Phys.* **2019**, *21*, 105001.
(29) Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A.; Ceriotti, M. Accurate Molecular Polarizabilities with Coupled Cluster Theory and Machine Learning. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 3401−3406.
(30) Lewis, A. M.; Lazzaroni, P.; Rossi, M. Predicting the Electronic Density Response of Condensed-Phase Systems to Electric Field Perturbations. *J. Chem. Phys.* **2023**, *159*, 014103.
(31) Fang, M.; Tang, S.; Fan, Z.; Shi, Y.; Xu, N.; He, Y. A Study of Simulating Raman Spectra for Alkanes with a Machine Learning-Based Polarizability Model. *arXiv* **2023**, arXiv.2301.13490.
(32) Inoue, K.; Litman, Y.; Wilkins, D. M.; Nagata, Y.; Okuno, M. Is Unified Understanding of Vibrational Coupling of Water Possible? Hyper-Raman Measurement and Machine Learning Spectra. *J. Phys. Chem. Lett.* **2023**, *14*, 3063−3068.

https://doi.org/10.1021/acs.jpcc.4c00886
J. Phys. Chem. C 2024, 128, 6464−6470

(33) Litman, Y.; Lan, J.; Nagata, Y.; Wilkins, D. M. Fully First-Principles Surface Spectroscopy with Machine Learning. *J. Phys. Chem. Lett.* **2023**, *14*, 8175−8182.
(34) Sommers, G. M.; Calegari Andrade, M. F.; Zhang, L.; Wang, H.; Car, R. Raman Spectrum and Polarizability of Liquid Water from Deep Neural Networks. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10592−10602.
(35) Zhang, Y.; Ye, S.; Zhang, J.; Hu, C.; Jiang, J.; Jiang, B. Efficient and Accurate Simulations of Vibrational and Electronic Spectra with Symmetry-Preserving Neural Network Models for Tensorial Properties. *J. Phys. Chem. B* **2020**, *124*, 7284−7290.
(36) Tuan-Anh, T.; Zaleśny, R. Predictions of High-Order Electric Properties of Molecules: Can We Benefit from Machine Learning? *ACS Omega* **2020**, *5*, 5318−5325.
(37) Shang, H.; Wang, H. Anharmonic Raman Spectra Simulation of Crystals from Deep Neural Networks. *AIP Adv.* **2021**, *11*, 035105.
(38) Nguyen, V. H. A.; Lunghi, A. Predicting Tensorial Molecular Properties with Machine Learning Models. *Phys. Rev. B* **2022**, *105*, 165131.
(39) Zhang, Y.; Jiang, B. Universal Machine Learning for the Response of Atomistic Systems to External Fields. *Nat. Commun.* **2023**, *14*, 6424.
(40) Paul, A.; Ruffino, A.; Masiuk, S.; Spanier, J.; Grinberg, I. An Atomistic Model of Electronic Polarizability for Calculation of Raman Scattering from Large-Scale MD Simulations. *arXiv* **2023**, arXiv.2304.07536.
(41) Gandolfi, M.; Rognoni, A.; Aieta, C.; Conte, R.; Ceotto, M. Machine Learning for Vibrational Spectroscopy via Divide-and-Conquer Semiclassical Initial Value Representation Molecular Dynamics with Application to N-methylacetamide. *J. Chem. Phys.* **2020**, *153*, 204104.
(42) Ren, H.; Li, H.; Zhang, Q.; Liang, L.; Guo, W.; Huang, F.; Luo, Y.; Jiang, J. A Machine Learning Vibrational Spectroscopy Protocol for Spectrum Prediction and Spectrum-Based Structure Recognition. *Fundam. Res.* **2021**, *1*, 488−494.
(43) Petrusevich, E. F.; Bousquet, M. H. E.; Ośmiałowski, B.; Jacquemin, D.; Luis, J. M.; Zaleśny, R. Cost-Effective Simulations of Vibrationally-Resolved Absorption Spectra of Fluorophores with Machine-Learning-Based Inhomogeneous Broadening. *J. Chem. Theory Comput.* **2023**, *19*, 2304−2315.
(44) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.
(45) Gajdoš, M.; Hummer, K.; Kresse, G.; Furthmüller, J.; Bechstedt, F. Linear Optical Properties in the Projector-Augmented Wave Methodology. *Phys. Rev. B* **2006**, *73*, 045112.
(46) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(47) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(48) Karniadakis, G. E.; Kevrekidis, I. G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3*, 422−440.
(49) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **2021**, *121*, 10073−10141.
(50) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.
(51) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87*, 184115.
(52) Willatt, M. J.; Musil, F.; Ceriotti, M. Atom-Density Representations for Machine Learning. *J. Chem. Phys.* **2019**, *150*, 154110.
(53) Musil, F.; Grisafi, A.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. Physics-Inspired Structural Representations for Molecules and Materials. *Chem. Rev.* **2021**, *121*, 9759−9815.
(54) Grisafi, A.; Wilkins, D. M.; Csányi, G.; Ceriotti, M. Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. *Phys. Rev. Lett.* **2018**, *120*, 036002.
(55) Himanen, L.; Jäger, M. O.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of Descriptors for Machine Learning in Materials Science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
(56) Musil, F.; Veit, M.; Goscinski, A.; Fraux, G.; Willatt, M. J.; Stricker, M.; Junge, T.; Ceriotti, M. Efficient Implementation of Atom-Density Representations. *J. Chem. Phys.* **2021**, *154*, 114109.
(57) Pedregosa, F.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(58) Long, D. A. *The Raman Effect*; John Wiley & Sons, Ltd: Chichester, UK, 2002.
(59) Egorov, S. A.; Skinner, J. L. Semiclassical Approximations to Quantum Time Correlation Functions. *Chem. Phys. Lett.* **1998**, *293*, 469−476.
(60) Benshalom, N.; Reuveni, G.; Korobko, R.; Yaffe, O.; Hellman, O. Dielectric Response of Rock-Salt Crystals at Finite Temperatures from First Principles. *Phys. Rev. Mater.* **2022**, *6*, 033607.

https://doi.org/10.1021/acs.jpcc.4c00886
J. Phys. Chem. C 2024, 128, 6464−6470