Tensorial Properties via the Neuroevolution Potential Framework: Fast Simulation of Infrared and Raman Spectra

JCTC Journal of Chemical Theory and Computation

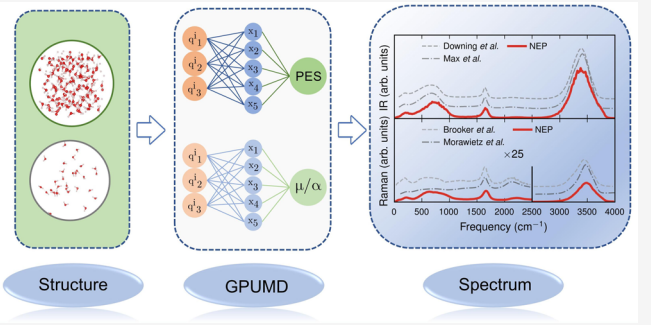
pubs.acs.org/JCTC Article

Tensorial Properties via the Neuroevolution Potential Framework: Fast Simulation of Infrared and Raman Spectra

Nan Xu, Petter Rosander, Christian Schäfer, Eric Lindgren, Nicklas Österbacka, Mandi Fang, Wei Chen, Yi He,* Zheyong Fan,* and Paul Erhart*

Cite This: https://doi.org/10.1021/acs.jctc.3c01343 Read Online

ABSTRACT: Infrared and Raman spectroscopy are widely used for the characterization of gases, liquids, and solids, as the spectra contain a wealth of information concerning, in particular, the dynamics of these systems. Atomic scale simulations can be used to predict such spectra but are often severely limited due to high computational cost or the need for strong approximations that limit the application range and reliability. Here, we introduce a machine learning (ML) accelerated approach that addresses these shortcomings and provides a significant performance boost in terms of data and computational efficiency compared with earlier ML schemes. To this end, we generalize the neuroevolution potential approach to enable the prediction of rank one and two tensors to obtain the tensorial neuroevolution potential (TNEP) scheme. We apply the resulting framework to construct models for the dipole moment, polarizability, and susceptibility of molecules, liquids, and solids and show that our approach compares favorably with several ML models from the literature with respect to accuracy and computational efficiency. Finally, we demonstrate the application of the TNEP approach to the prediction of infrared and Raman spectra of liquid water, a molecule ($PTAF^-$), and a prototypical perovskite with strong anharmonicity ($BaZrO_3$). The TNEP approach is implemented in the free and open source software package GPUMD, which makes this methodology readily available to the scientific community.



1. INTRODUCTION

Infrared (IR) and Raman spectroscopy are widely used techniques for the nondestructive characterization of the dynamics and to some extent chemistry of materials spanning the entire range from the gas phase to condensed matter.[1−3] Over the years, various theoretical approaches have been developed for simulating IR and Raman spectra, including in particular methods based on *ab initio* molecular dynamics (MD) simulations.[4−8] While these approaches are capable of reproducing experimental IR and Raman spectra of gases, liquids, and solids,[5,7−9] they are severely limited with respect to the system sizes and time scales attainable for two main reasons:[5,10] First, *ab initio* MD simulations rely on computationally demanding electronic structure calculations that scale strongly with system size in order to obtain energy and forces at every time step. Second, similarly expensive calculations of dipole moment ($\boldsymbol{\mu}$), polarizability ($\boldsymbol{\alpha}$), or electric susceptibility ($\boldsymbol{\chi}$) are required for at least many thousands of configurations to achieve numerical convergence of the underlying correlation functions.[5]

MD simulations can be accelerated by using classical force fields[11−13] or empirical interatomic potentials,[14,15] which approximate the potential energy surface (PES) with physically motivated yet constrained functions and few fitted parameters. The accuracy of such approaches for general materials is, however, often limited, negatively affecting the prediction of IR and Raman spectra.[16] Machine learning (ML) potentials are well suited to address this challenge as they bridge between the accuracy of quantum mechanical methods and the computational efficiency of classical force fields or empirical interatomic potentials.[17−21] The power of this approach, in particular for capturing vibrational properties of materials, has been shown repeatedly, see, e.g., refs.[22−26]

The calculation of $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, or $\boldsymbol{\chi}$ can be accelerated using parametric models in a similar fashion. Considering only static charges, the dipole moment is given by $\boldsymbol{\mu} = \sum_{i=1}^{N} Q_i \boldsymbol{r}_i$, where Q_i and $\boldsymbol{r}_i$ are the charge and position of atom i. Many classical force fields[11−13] assign fixed charges to atoms and thereby provide a convenient approach for calculating $\boldsymbol{\mu}$. Such fixed-charge models neglect, however, polarization effects, which can lead to large errors.[27] While this situation can in principle be

Received: December 8, 2023
Revised: March 19, 2024
Accepted: April 1, 2024

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

ameliorated by fluctuating-charge models,[28,29] the latter tend to lack robustness and can be difficult to generalize.[10,30]

Both $\boldsymbol{\alpha}$ and $\boldsymbol{\chi}$ describe the dielectric response to an applied electric field. For $\boldsymbol{\alpha}$ or $\boldsymbol{\chi}$, the bond polarizability model is one of the most frequently used parametric ones and has for example been applied to alkanes[31,32] and zeolites[33] as well as carbon nanotubes.[34] However, this simple model often suffers from unsatisfactory transferability when used in different environments.[35] POLI2VS[36] and MB-pol[37] are two other parametric models that can be used for predicting $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ but are limited to molecular systems such as water.[10]

The successful applications of ML potentials have inspired the development of ML dipole, polarizability, and susceptibility models.[22,38−41] For $\boldsymbol{\mu}$, a rank-1 tensor, both partial-charge and the partial-dipole ML models have been developed.[30] The objective of the partial-charge models is to assign proper partial charges for atoms in order to fit the total dipole moment.[22,30,42] Here, one concern is the balance between the fitting quality of $\boldsymbol{\mu}$ and the reproducibility of total charges.[22,30] By contrast partial-dipole models such as symmetry-adapted Gaussian process regression (SA-GPR),[38] tensorial embedded-atom neural network (T-EANN),[39] and deep potential (DP)[40] treat $\boldsymbol{\mu}$ as a sum of vectors[30,38] that can be determined from atom-centered chemical environments.

While this approach works for $\boldsymbol{\mu}$, which is a rank-1 tensor, it does not transfer to the construction of ML models for $\boldsymbol{\alpha}$ or $\boldsymbol{\chi}$, which are rank-2 tensors. This has motivated the pioneering development of the SA-GPR method for tensorial properties[38] as well as later the T-EANN[39,43] and DP models.[10]

The combination of ML potentials with ML models for $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, or $\boldsymbol{\chi}$ enables the simulations of IR and Raman spectra. This approach has been used to predict, e.g., the IR spectra of methanol, *n*-alkanes, and a peptide,[22] IR and Raman spectra of liquid water,[10,20,39,44] or the Raman spectra of various solid materials.[45] While these earlier studies have established the usefulness of ML models for predicting IR and Raman spectra, there is still ample room for improvement of current models for $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, or $\boldsymbol{\chi}$ in terms of computational and data efficiency[30,39] as well as the accessibility of these techniques in order to lower the threshold for the widespread adoption of such approaches.

This situation motivates the present work, in which we introduce accurate as well as computationally and data efficient ML models for rank-1 and rank-2 tensors based on the NEP framework.[21,46,47] We demonstrate the efficacy and efficiency of the resulting TNEP approach by training models for $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\chi}$ and combining these with models for the PES to predict IR and Raman spectra for a molecule ($PTAF^-$), a liquid (water), and a solid ($BaZrO_3$; Figure 1). We make this methodology available via the GPUMD package,[47] enabling comprehensive simulations of high-quality IR and Raman spectra with limited user effort.

2. METHODOLOGY

2.1. NEP Models for the PES. Since the ML models for $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ that we introduce below are based on the NEP framework for modeling PESs,[21,46,47] we first provide a brief review of the latter. Originally NEPs are ML potentials that model the high-dimensional PES of finite or extended systems, in the spirit of the neural network potential model proposed by Behler and Parrinello.[48] In this formalism, the total energy of the system is given by the sum of atomic site energies $U = \sum_i U_i$. The site energy U_i for a given atom i depends on the local environment of the atom, which is represented by an

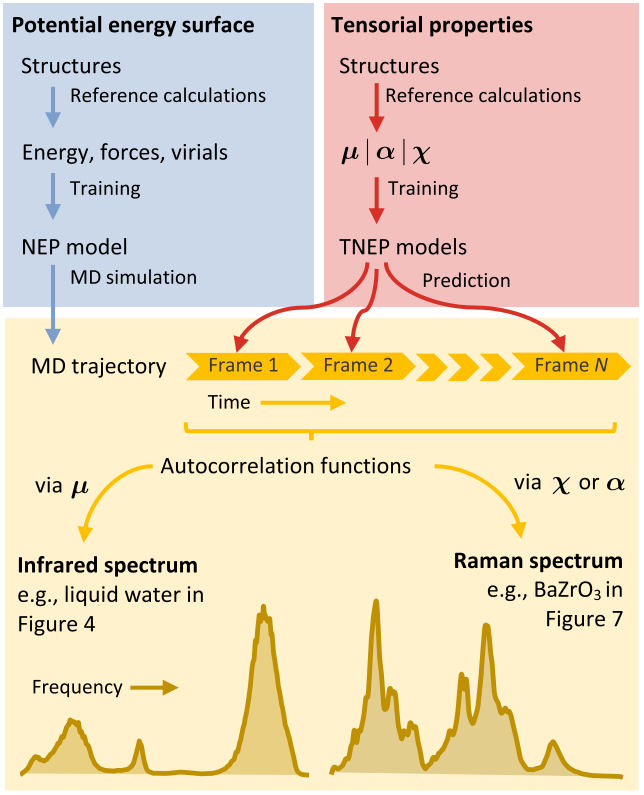


Figure 1. Workflow for simulations of IR and Raman spectra using NEP models for the PES and TNEP models for the dipole moment $\boldsymbol{\mu}$, the polarizability $\boldsymbol{\alpha}$, or the susceptibility $\boldsymbol{\chi}$.

abstract vector q_i^ν with a number of components indexed by ν. The function mapping from the descriptor to the site energy is represented by a feed-forward neural network (also known as a multilayer perceptron) with typically a single hidden layer. The input layer of the neural network is thus the descriptor vector, and the output layer consists of a single node whose value is the site energy U_i of the considered atom i, which can be formally expressed as

$$U_i = U_i(q_i^\nu) \tag{1}$$

From the energy, we can derive the rank-2 virial tensor that serves as the foundation for the dipole and polarizability models developed in the present work. For a given structure with N atoms, the virial tensor can be expressed as[47]

$$W^{\upsilon\nu} = -\sum_i^N \sum_{j\neq i} r_{ij}^\upsilon \frac{\partial U_i}{\partial r_{ij}^\nu} \tag{2}$$

where r_{ij}^υ is the υ-component of the vector $\boldsymbol{r}_{ij} \equiv \boldsymbol{r}_j - \boldsymbol{r}_i$, and $\boldsymbol{r}_i$ is the position of atom i. We refer to the term $\partial U_i/\partial r_{ij}^\nu$ as the partial force, explicit expressions for which have been presented in the original works developing the NEP approach.[21,47]

2.2. TNEP Rank-1 Tensor Models. To develop an ML model for predicting $\boldsymbol{\mu}$, we first note that it is a rank-1 tensor commonly expressed as a vector, in contrast to the energy, which is a rank-0 tensor (i.e., a scalar). The partial force in eq 2 is a vector, but the summation of it over the whole structure would be zero as a result of Newton's third law. To obtain a vector representation that does not vanish for a general structure, we note that the quantity defined in eq 2 is a rank-2

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

tensor that can adopt both positive and negative values (as it is the virial tensor in the context of PES models). We can thus obtain an expression for a vector quantity by contracting this rank-2 tensor with a vector. A natural choice for the vector to be contracted is $\boldsymbol{r}_{ij}$, which yields the following expression for rank-1 tensors such as the dipole moment

$$\boldsymbol{\mu} = -\sum_{i}^{N}\sum_{j\neq i}\boldsymbol{r}_{ij}\cdot\left(\boldsymbol{r}_{ij}\otimes\frac{\partial U_i}{\partial \boldsymbol{r}_{ij}}\right) = -\sum_{i}^{N}\sum_{j\neq i} r_{ij}^2\left(\frac{\partial U_i}{\partial \boldsymbol{r}_{ij}}\right) \tag{3}$$

where $r_{ij}^2 = \boldsymbol{r}_{ij}\cdot\boldsymbol{r}_{ij}$ is the distance squared between atoms i and j. We note that U_i here should have the dimension of charge instead of energy. Crucially this goes to show that the NEP formalism for PESs can be directly used to construct an ML model for rank-1 tensors such as the dipole moment. Below we refer to eq 3 as the TNEP dipole model.

2.3. TNEP Rank-2 Tensor Models. To develop ML models for predicting $\boldsymbol{\alpha}$ or $\boldsymbol{\chi}$, we first note that these are rank-2 tensors. Clearly, the quantity defined in eq 2 is an ideal candidate. However, using only eq 2 to represent $\boldsymbol{\alpha}$ or $\boldsymbol{\chi}$ does not lead to high regression accuracy because the diagonal terms of $\boldsymbol{\alpha}$ or $\boldsymbol{\chi}$ are usually much larger than the off-diagonal ones. We therefore represent $\boldsymbol{\alpha}$ (and equivalently $\boldsymbol{\chi}$) as a combination of eqs 1 and 2 as follows

$$\alpha^{\upsilon\nu} = \sum_{i}^{N} U_i\delta^{\upsilon\nu} - \sum_{i}^{N}\sum_{j\neq i} r_{ij}^{\upsilon}\frac{\partial U_i}{\partial r_{ij}^{\nu}} \tag{4}$$

where $\delta^{\upsilon\nu}$ is the Kronecker delta. Note that both the first and second terms on the right-hand side contribute to the diagonal elements of $\alpha^{\upsilon\nu}$, but only the second term contributes to the off-diagonal elements. U_i here has the dimension of polarizability instead of energy, yet the entire NEP formalism can be reused. Below we refer to eq 4 as the TNEP polarizability or susceptibility model.

2.4. Loss Functions. The NEP approach is named after the underlying ML model (a neural network) and the separable natural evolution strategy used as the training algorithm.[49] The latter is a principled real-valued black-box optimization method that is very well suited for training the weight and bias parameters in the neural network, of which there are typically a few thousand. The optimization is driven by the minimization of a loss function that is given by the weighted sum of the root-mean-square error (RMSE) of physical quantities as well as $\mathcal{L}_1$ and $\mathcal{L}_2$ regularization terms. For the construction of PES models, the physical quantities included in the loss function are the energies, forces, and virial tensors of the structures in the training set

$$L(\boldsymbol{z}) = \lambda_e\Delta U(\boldsymbol{z}) + \lambda_f\Delta F(\boldsymbol{z}) + \lambda_v\Delta W(\boldsymbol{z}) + \text{regularization terms} \tag{5}$$

where $\Delta U(\boldsymbol{z})$, $\Delta F(\boldsymbol{z})$, and $\Delta W(\boldsymbol{z})$ are the RMSEs of energies, forces, and virials calculated using a set of trainable parameters $\boldsymbol{z}$, and λ_e, λ_f, and λ_v are the corresponding relative weights. Explicit expressions for the regularization terms can be found in ref 47. For the construction of dipole TNEP models, the loss function is defined in terms of the RMSE of the dipole $\Delta\mu(\boldsymbol{z})$

$$L(\boldsymbol{z}) = \Delta\mu(\boldsymbol{z}) + \text{regularization terms} \tag{6}$$

For the construction of polarizability TNEP models, the loss function is defined in terms of the RMSE of the polarizability $\Delta\alpha(\boldsymbol{z})$

$$L(\boldsymbol{z}) = \Delta\alpha(\boldsymbol{z}) + \text{regularization terms} \tag{7}$$

2.5. Dielectric Response. It is instructive to recall some relations that describe the response of finite systems (such as molecules) and extended systems (such as solids and liquids) to an applied electric field.

If a molecule is subjected to an electric field $\boldsymbol{E}$, the resulting displacement of nuclei and electrons induces a dipole, which is given by[50]

$$\boldsymbol{\mu}_{\text{ind}} = \boldsymbol{\alpha}\boldsymbol{E}$$

where $\boldsymbol{\alpha}$ is the *molecular polarizability*.

For an extended system such as a solid or a liquid, one considers equivalently the dipole moment per unit volume, i.e., the polarization

$$\boldsymbol{P} = \epsilon_0\boldsymbol{\chi}\boldsymbol{E}$$

where $\boldsymbol{\chi}$ is the *electric susceptibility*. In the context of bulk liquids, the latter has also been referred to as the bulk polarizability. For clarity in the following, we use the term polarizability only to refer to molecular polarizability. There are different conventions for expressing $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\chi}$ leading to different units (Sect. S7). Here, we use e·bohr for $\boldsymbol{\mu}$ and bohr^3 for $\boldsymbol{\alpha}$, whereas $\boldsymbol{\chi}$ is unitless.

We note that under certain conditions, one can approximately connect the molecular polarizability and the electric susceptibility via the Clausius-Mossotti relation, which is based on a mean-field treatment of local field effects (see Sect. S8 in the Supporting Information).

2.6. The IR Intensity. The IR absorption cross section is given by[50]

$$\sigma(\omega) = \frac{4\pi^2}{\hbar cn}\omega(1 - e^{-\beta\hbar\omega})M(\omega) \tag{8}$$

where n is the refractive index of the material, c is the speed of light, $\beta = 1/k_BT$, and $M(\omega)$ is the absorption line shape given by the Fourier transform of the autocorrelation function (ACF) of the (total) dipole moment $\boldsymbol{\mu}$

$$M(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\langle(\hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{\mu}(0))(\hat{\boldsymbol{\epsilon}}\cdot\boldsymbol{\mu}(t))\rangle e^{-i\omega t}dt$$

where $\langle\cdots\rangle$ indicates the average over time origins, and $\hat{\boldsymbol{\epsilon}}$ is the polarization of the light.[50] For an isotropic sample, the time correlation should be averaged over the three directions, i.e., the line shape reduces to one-third of the trace of the dipole time correlation. Since the line shape is sampled classically, we make a classical approximation for the prefactor by expanding the Boltzmann factor to first order, which gives

$$\sigma(\omega) \propto \omega^2 M(\omega) \tag{9}$$

2.7. The Raman Intensity. The differential Raman cross-section for Stokes scattering is given by[50−52]

$$\frac{\partial^2\sigma}{\partial\omega_{\text{out}}\partial\Omega} = \left(\frac{\omega_{\text{in}} - \omega}{c}\right)^4\sum_{\gamma\delta\mu\nu}\hat{n}_\gamma\hat{n}_\mu L_{\gamma\delta\mu\nu}(\omega)\hat{\epsilon}_\delta\hat{\epsilon}_\nu \tag{10}$$

where $\hat{n}$ is the polarization of observed light, $\hat{\boldsymbol{\epsilon}}$ is the polarization of the incoming light, and Ω is a solid angle. Here, it is assumed that the frequency of the incoming light ω_{in} is significantly larger than the Raman shift ω and significantly smaller than the band gap, i.e., far from any electronic excitations. $L(\omega)$ is the Raman line shape given by the Fourier transform of the time-dependent polarizability $\boldsymbol{\alpha}(t)$ (finite

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

systems) or susceptibility $\chi(t)$ (extended systems), e.g., in the case of the former

$$L_{\gamma\delta\mu\nu}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\langle\alpha_{\gamma\delta}(0)\alpha_{\mu\nu}(t)\rangle e^{-i\omega t}dt \quad (11)$$

Note that the elements of the polarizability (or susceptibility) tensor are selected by the polarization of the incoming and outgoing light as indicated in eq 10. Polarized Raman measurements can be directly related to eq 10 by combinations of the Raman line shape $L(\omega)$. One can also calculate an average spectrum for isotropic samples.[50] The polarizability tensor (and equivalently the susceptibility tensor) can also be written as $\boldsymbol{\alpha} = \gamma\boldsymbol{I} + \boldsymbol{\beta}$ where $\gamma = \mathrm{Tr}(\boldsymbol{\alpha})/3$, and $\boldsymbol{\beta}$ is a traceless tensor to obtain the isotropic (polarized) and anisotropic (depolarized) spectrum. This leads to the decomposition

$$L_{\mathrm{iso}}(\omega) \propto \int_{-\infty}^{\infty}\langle\gamma(0)\gamma(t)\rangle e^{-i\omega t}dt$$
$$L_{\mathrm{aniso}}(\omega) \propto \int_{-\infty}^{\infty}\langle\mathrm{Tr}[\boldsymbol{\beta}(0)\boldsymbol{\beta}(t)]\rangle e^{-i\omega t}dt \quad (12)$$

The electric susceptibility (Sect. 2.5) can be separated into an electronic and an ionic contribution

$$\chi = \chi_{\mathrm{ion}}(\omega) + \chi_{\mathrm{e}}(\omega)$$

where the general frequency dependence of these terms is emphasized. For the prediction of Raman spectra, we only need to consider the electronic contribution $\chi_{\mathrm{e}}(\omega)$. Furthermore, we limit ourselves to nonresonant Raman spectroscopy. This means that we require the electric susceptibility in the ion-clamped static limit, i.e., $\chi_{\mathrm{e}}(0)$, and do not have to consider the frequency dependence of $\chi_{\mathrm{e}}(\omega)$, which arises from electronic transitions.

2.8. Workflow for Simulations of IR and Raman Spectra. By combining a NEP model for the PES with TNEP models for dipole, polarizability, or susceptibility, one obtains a simple yet general workflow for the computation of IR and Raman spectra (Figure 1). Starting from a NEP PES model, large-scale MD simulations are performed to sample the PES via the GPUMD package, typically for a few hundred picoseconds. TNEP dipole, polarizability, or susceptibility models are then employed to predict $\boldsymbol{\mu}(t)$, $\boldsymbol{\alpha}(t)$, or $\chi(t)$ along the trajectory. Finally, IR or Raman spectra are obtained via Fourier transformation of the respective ACFs via eq 9 or 10.

3. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the TNEP dipole, polarizability, and susceptibility models in comparison with models from the literature with respect to both regression accuracy and computational speed. The comparison includes the molecules H_2O, $(H_2O)_2$, and $H_5O_2^+$ (the Zundel cation), as well as a set of configurations representing liquid water. Structures with dipole, polarizability, and/or susceptibility data were retrieved from the repository maintained by the developers of the SA-GPR models[38,53] (see Sect. S1 in the Supporting Information for details). The data set for each of these systems comprises 1000 configurations, half of which were use for training, while the other half were used for validation. The hyperparameters used in the training of the TNEP models are presented in Tables S1 and S2. In the case of the SA-GPR method, the results for liquid water were computed using a publicly available model,[54] while the models for the molecules were trained by us (see Sect. S3 for details). In the case of the T-EANN method, we only use those data available in the literature.[39]

3.1. Dipole Moment. The TNEP dipole models can achieve very high precision when predicting $\boldsymbol{\mu}$ for both molecules and liquid water with very low RMSEs (Table 1) and coefficients of determination (R^2) very close to one (Figure S2).

Table 1. RMSEs (in e·bohr) and RRMSEs (unitless) for $\boldsymbol{\mu}$ for the Validation Sets Using NEP Rank-1 Tensor Models[a]

System	RMSE	RRMSE
H_2O	2×10^{-4}	0.069%
$(H_2O)_2$	105×10^{-4}	1.681%
$H_5O_2^+$	14×10^{-4}	0.371%
liquid water	17×10^{-4}	0.852%

[a]For liquid water, the dipole moment is given per water molecule.

As a further, more intuitive measure, one can also consider the root-mean-square-error relative to standard deviation (RRMSE),[39] defined as the RMSE divided by the standard deviation of the reference data (Figure 2a). For the water

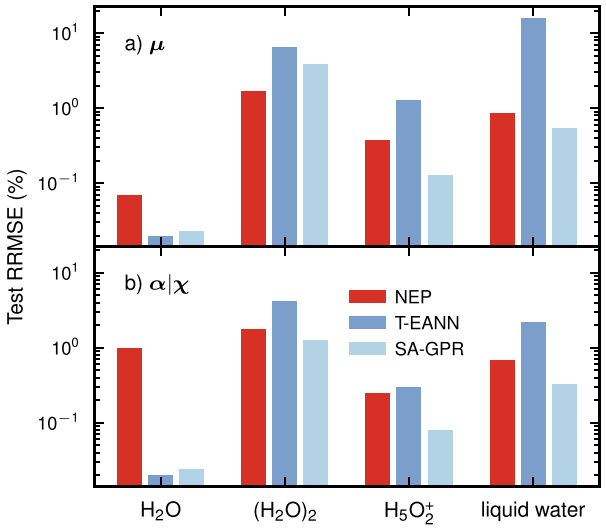


Figure 2. RRMSEs for the validation sets according to TNEP, T-EANN, and SA-GPR models for water systems for (a) $\boldsymbol{\mu}$ as well as (b) $\boldsymbol{\alpha}$ and χ/ρ. Validation RRMSEs for liquid water from T-EANN[39] were reported for the averaged molecular polarizability obtained via the Clausius-Mossotti relation (S8). The validation RRMSEs for χ/ρ should be somewhat higher than that for the averaged molecular polarizability (also see Table S5).

monomer (H_2O), all three methods yield extremely small RRMSEs below 0.1%. For the other three systems, including liquid water, the TNEP and SA-GPR models achieve comparable accuracy, while the T-EANN models perform systematically worse. This behavior is particularly pronounced for liquid water and might arise since the T-EANN model uses the positions relative to the center of mass as input, which are not well-defined in periodic systems.[55,56]

Neutral Molecules. The $\boldsymbol{\mu}$ of neutral molecules such as H_2O or $(H_2O)_2$ is uniquely defined. In the TNEP approach, $\boldsymbol{\mu}$ is calculated by summing over atomic contributions which, in contrast to, e.g., the T-EANN approach, does not require

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

choosing a reference point. Therefore, the TNEP dipole models are naturally suitable for neutral molecules.

In this context, we note that we also trained and validated a model for the QM7B data set containing thousands of neutral organic molecules,[57,58] for which we make similar observations (Sect. S4). The TNEP model yields a very low RMSE for the validation set of 1.80×10^{-3} e·bohr atom^{-1} and a very high R^2 score for the validation set of about 0.998.

Charged Molecules. The $\boldsymbol{\mu}$ of charged molecules is nonunique and depends on the choice of the reference point.[4,59] For charged molecules, one should therefore employ the relative permanent dipole $\boldsymbol{\mu}_r$ defined with respect to the center of mass when training TNEP dipole models. The reference $\boldsymbol{\mu}$ values in the $H_5O_2^+$ data set[38,53] have already been transformed to $\boldsymbol{\mu}_r$. Therefore, the absolute dipole moment of $H_5O_2^+$ including the movement of the center of mass should be $\boldsymbol{\mu} = \boldsymbol{\mu}_r + e \cdot \boldsymbol{r}_{COM}$. The same procedure was applied to the $PTAF^-$ molecule below (Sect. 4.3).

Periodic Systems. Traditional methods for calculating $\boldsymbol{\mu}$ cannot be applied to periodic systems since the position operator is not uniquely defined.[56,60] This issue is overcome via the modern theory of polarization,[38,60,61] which provides a rigorous definition for the polarization of periodic systems and established a methodology for calculating $\boldsymbol{\mu}$. It was therefore used in the present work to obtain $\boldsymbol{\mu}$ for periodic systems including water (Sect. S1) and α-Fe_2O_3 (Sect. S5). The TNEP model for α-Fe_2O_3 yields a very high R^2 score for the validation set close to one.

3.2. Polarizability and Susceptibility.

The RMSEs for the diagonal and off-diagonal elements of $\boldsymbol{\alpha}$ of (H_2O), $(H_2O)_2$, and $H_5O_2^+$ are quite small (Table 2), indicating the high accuracy of the TNEP polarizability model. The coefficients of determination are larger than 0.98 mirroring this trend (Figure S7 and Figure S8). For liquid water, we consider χ/ρ, which has the unit of polarizability per atom. The RMSEs for χ/ρ are on the same order of magnitude as the RMSEs for $\boldsymbol{\alpha}$ (Table 2).

Table 2. RMSEs (in bohr^3) and RRMSEs (unitless) for $\boldsymbol{\alpha}$ (Molecules) and χ/ρ (Liquid Water) for the Validation Sets Using TNEP Rank-2 Tensor Models[a]

	diagonal elements		off-diagonal elements	
System	RMSE	RRMSE	RMSE	RRMSE
H_2O	85×10^{-3}	5.89%	4×10^{-3}	1.22%
$(H_2O)_2$	227×10^{-3}	8.82%	137×10^{-3}	12.59%
$H_5O_2^+$	23×10^{-3}	1.20%	17×10^{-3}	1.06%
liquid water	54×10^{-3}	16.28%	37×10^{-3}	20.38%

[a]For liquid water, χ/ρ is given per water molecule.

The NEP models achieve an accuracy that is comparable to that of the T-EANN and SA-GPR models for the polarizability of $(H_2O)_2$ and $H_5O_2^+$ as well as the susceptibility of liquid water (Figure 2b). While the performance for the water monomer H_2O is worse, the TNEP model still yields a validation RRMSE of less than 1%.

As a further test, we constructed a TNEP polarizability model for the QM7B data set (Sect. S4). The RMSE values for the validation set are 4.64×10^{-2} bohr^3 atom^{-1} and 2.58×10^{-2} bohr^3 atom^{-1} for the diagonal and off-diagonal elements of $\boldsymbol{\alpha}$, respectively. For comparison, Wilkins et al.[62] reported a higher RMSE value of 5.50×10^{-2} bohr^3 atom^{-1} over both the diagonal and off-diagonal elements of $\boldsymbol{\alpha}$ using an SA-GPR model.

3.3. Computational Speed.

It is now instructive to evaluate the computational performance of TNEP models in comparison with publicly available SA-GPR models.[53,54] To this end, we consider liquid water systems with varying numbers of atoms. Starting from a cell containing 96 atoms, larger samples with up to 69984 atoms were created by replication.

The SA-GPR models can be run only serially on a central processing unit (CPU). In contrast, the TNEP model can be run on a CPU using `NEP_CPU`,[63] e.g., via the interface provided by the CALORINE package,[64] or on a graphics processing unit (GPU) by using the GPUMD package. The SA-GPR and TNEP (CPU) models were tested on a server containing two Intel XEON Platinum 8275CL processors with a system memory of 256 GB, while the TNEP (GPU) models were tested on a heterogeneous server containing two Intel XEON Gold 6148 processors and an Nvidia GeForce RTX 4090 card with a graphics memory of 24 GB.

The comparisons show that for system sizes $\gtrsim$1000 atoms the TNEP CPU models are at least 1 order of magnitude faster than the SA-GPR models on CPUs for both dipole and polarizability (Figure 3). On CPUs, the TNEP models exhibit nearly perfect weak scaling over the system sizes considered here. In contrast, the SA-GPR models show a notable decrease in speed as the system size increases. Running the TNEP models on GPUs enables an additional speedup of an order of magnitude or more. For very small systems, the GPU implementation is limited by IO. In addition, we note that GPUMD allows one to evaluate TNEP models on-the-fly during MD simulations for prediction of tensorial properties with a small impact on simulation speed (Sect. S10).

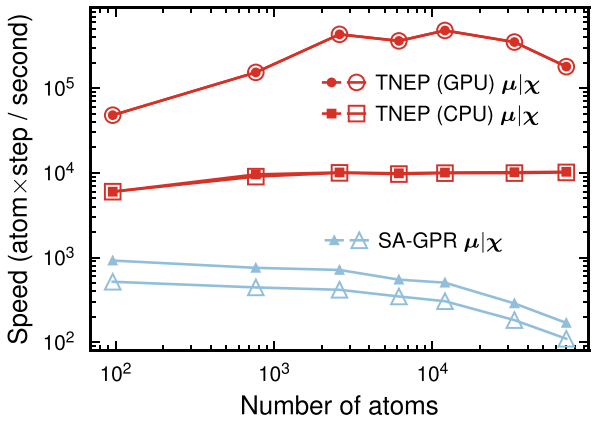


Figure 3. Comparison of computational speed of SA-GPR and TNEP models for the dipole ($\boldsymbol{\mu}$) and susceptibility (χ) of liquid water. Here, the SA-GPR results were obtained using the `TENSOAP-FAST` implementation.[54]

4. APPLICATIONS

Having established the accuracy and computational performance of the TNEP approach by comparison with reference data sets, we now demonstrate the application of NEP and TNEP models in combination for predicting the IR and Raman spectra of molecules, liquids, and solids. To this end, we employ the correlation function approach outlined above (Sect. 2.8 and Figure 1).

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

4.1. IR Spectrum of Water. First, we developed a NEP PES model for liquid water using energy, atomic forces, and virial data from density functional theory (DFT) calculations (Sect. S2).

Next, a system of 216 water molecules was equilibrated in the NPT ensemble for 100 ps using the trained PES model at 298 K and 1 bar, followed by a further equilibration run in the NVT ensemble for another 100 ps. Three production runs were carried out in the NVE ensemble for a duration of 200 ps. A time step of 0.5 fs was used throughout. We note that quantum effects can be actually rather pronounced in water as has been shown by path integral MD simulations in, e.g., refs.[65−67] Here, however, we decided to carry out classical MD simulations in order to enable a one-to-one comparison with the results of earlier studies.

The time dependence of the dipole ($\boldsymbol{\mu}(t)$) was computed for the production trajectories with a spacing of 1 fs using the TNEP dipole model for liquid water described above (Sect. 3.1). The IR spectrum was then obtained by Fourier transform of the dipole moment ACF via eq 9. The final IR spectrum was obtained by averaging the IR spectra from the production runs.

For comparison, we also ran a 200 ps MD simulation with the TIP3P force field[68] via the CP2K software package,[69] where the TIP3P force field uses charges of −0.834 e and +0.417 e for oxygen and hydrogen, respectively.

The NEP-TNEP method yields an IR spectrum that is in very good agreement with experimental data[70,71] over the entire frequency range from 0 to 4000 cm^{-1} (Figure 4a). This includes the hydrogen-bond stretching band[10] between 160 and 250 cm^{-1}, the libration band[10] from 400 to 800 cm^{-1} associated with hindered molecule rotations,[37] and the bending modes[37,72] at about 1650 cm^{-1} as well as the OH stretching band[37,72] from 2800 to 4000 cm^{-1}. The NEP and TNEP models for PES and $\boldsymbol{\mu}$ in conjunction with the underlying exchange-correlation functional thus succeed in capturing the entire range stretching from soft intermolecular to stiff intramolecular modes. This performance is also observed for the DP model (Figure 4a).

By comparison, classical models produce rather large errors for the location of several features in the IR spectrum of water. MD simulations with classical force fields[68,73] such as TIP3P (Figure 4a) and SPC/E tend to predict a blue-shifting of the bending modes by roughly 100 to 200 cm^{-1}. A similar tendency was also observed for the POLI2VS model.[36] The results from the MB-pol model on the other hand exhibit a blue-shift of the OH stretching band by about 50 cm^{-1}.[37]

The width of the OH stretching band has been proven to be quite difficult to predict due to the anharmonicity of the OH stretch mode.[37] The NEP-TNEP approach yields a value of 380 cm^{-1} for the full width at half-maximum of this band, which is in good agreement with experimental estimates of about 350 cm^{-1} from Downing et al.[70] Both NEP-TNEP and DP predictions exhibit a slight high-frequency tail for this band, which is not visible in the experimental spectra. This small difference could originate from the strongly constrained and appropriately normed (SCAN) functional[74] that was used for generating the PES training data[10,75] and/or the absence of quantum effects in the (classical) MD simulations.[10,37]

4.2. Raman Spectra of Water. To obtain the Raman spectra of liquid water, we sampled the time dependence of $\chi(t)$ using the TNEP susceptibility model and subsequently computed the ACFs for the same trajectories used for the prediction of the IR spectra. The full spectrum given by eq 10 and averaged over the available trajectories was then split into isotropic (polarized) and anisotropic (depolarized) contributions via eq 12.

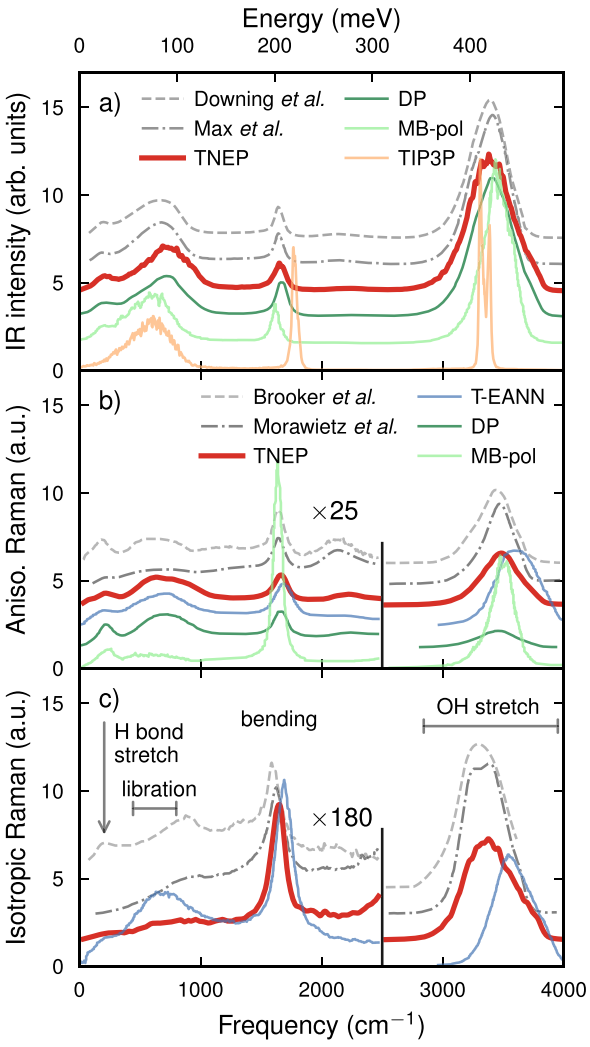


Figure 4. Comparison of (a) infrared as well as (b) anisotropic (depolarized) and (c) isotropic (polarized) Raman spectra of water at ambient conditions from simulations and experiment. Experimental data from Downing et al.,[70] Max et al.,[71] Brooker et al.,[76] and Morawietz et al.[77] Simulated spectra from T-EANN,[39] MB-pol,[37] and DP[10,44] models were adapted from the literature. In (a) and (b), the spectra were normalized by the integral between 80 and 2500 cm^{-1}, while in (c), they were normalized by the integral between 1000 and 2500 cm^{-1}.

The **anisotropic spectrum** predicted by the NEP-TNEP approach is overall in very good agreement with experimental data (Figure 4b).[76,77] The locations of the peaks and relative intensities of the stretching, bending, and librational modes in the simulated anisotropic Raman spectra are all well produced. It is noteworthy that in the low frequency region below approximately 1000 cm^{-1}, the variation between the experimental spectra is larger than the variation between the ML models and the experimental data. This could be related to difficulties associated with processing the experimental raw data in this frequency region.

The T-EANN and DP models yield results similar to those of the NEP-TNEP approach in the region up to about 1900

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

cm^{-1}. On the other hand, all ML models underestimate the intensity of the association band between 1900 and 2500 cm^{-1}, which is arising from the combination of librational and bending modes.[37,77] Here, the NEP-TNEP prediction is actually still the one that comes closest to the experimental spectra.

The broad high-frequency peak above 3000 cm^{-1}, which is associated with the OH stretch mode, is notably blue-shifted and broadened for the T-EANN model, while the DP model strongly underestimates the intensity of this peak. In contrast, the NEP-TNEP combination predicts this feature in good agreement with the experimental data.

Finally, the parametric MB-pol model yields the worst agreement with experiment, for example, strongly overestimating the intensity of the bending band while underestimating the libration band.

With regard to the **isotropic Raman spectrum** (Figure 4c), one should first note the variation among the experimental data. In particular, in the region below 1000 cm^{-1}, the resulting uncertainty is comparable or even larger than the deviation between the NEP-TNEP prediction and the experimental data, while the position of the libration band predicted by T-EANN appears red-shifted. With regard to the higher frequency region, both NEP-TNEP and T-EANN reproduce the bending band well. In the case of NEP-TNEP, this also applies for the OH stretch band, whereas in the case of T-EANN, a blue-shift can be observed similar to that of the anisotropic spectrum (Figure 4b).

4.3. IR Spectrum of PTAF$^-$. The NEP-TNEP method for predicting IR spectra can be easily adopted for other molecular systems as long as the underlying observables to be learned are available. Naturally, this includes the molecular configurations along a chemical reaction such that experimentally observable spectral changes can be connected to metastable complexes. One such complex is PTAF$^-$ (see inset in Figure 5), the intermediate reaction minimum in the deprotection reaction 1-phenyl-2-trimethylsilylacetylene (PTA) with tetra-n-butylammonium fluoride.[78−81]

To train NEP and TNEP models, we obtained PES and $\boldsymbol{\mu}$ data for a set of 20170 structures via DFT calculations using the ORCA code,[82] the PBE functional,[83] and a def2-TZVP basis set[84] while enforcing tight convergence of the self-consistent field cycles. Subsequently, MD simulations at various temperatures were performed in the NVE ensemble using a time step of 0.1 fs for 1 ns, during which $\boldsymbol{\mu}(t)$ was recorded with a time resolution of 0.5 fs.

The IR spectra obtained via the analysis of the ACF of $\boldsymbol{\mu}$ show a pronounced temperature dependence in particular of the line widths (Figure 5). The molecule supports several soft modes with frequencies in the region below 250 cm^{-1}, which are associated with the bending of and rotation about the ethynyl linker. These modes in particular lead to strong mode coupling (i.e., anharmonicity), which underlies the changes in line width and the redistribution of the dipole strength across the spectrum. Here, the computational efficiency of the NEP-TNEP implementation in GPUMD was crucial to resolve these features, as it enabled sampling on the nanosecond time scale, which would be prohibitive for DFT-MD simulations and computationally very expensive for a CPU implementation.

4.4. Raman Spectra of $BaZrO_3$. $BaZrO_3$ is a perovskite that is being investigated, e.g., as a proton conductor for applications in fuel cells. It has also been the subject of various fundamental studies, as it is a prototypical antiferroelectric perovskite.[85−87] It features soft and strongly temperature-dependent phonon modes,[88,89] which have been carefully analyzed with Raman spectroscopy,[90] rendering $BaZrO_3$ an ideal application for the present approach.

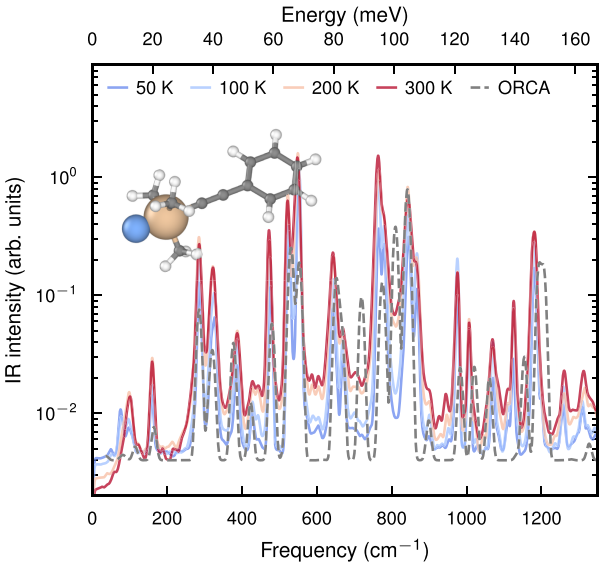


Figure 5. IR spectra for the metastable PTAF$^-$ complex (see the inset) at various temperatures. The gray dashed line represents the broadened integrated absorption coefficients of the harmonic spectrum obtained directly from DFT calculations. The overall agreement is good considering the lack of anharmonic corrections (intensity and vibrational frequencies) and temperature sensitivity of the spectrum obtained from DFT calculations.

For benchmarking, we constructed models for χ using both the TNEP and SA-GPR approaches. The reference data set comprised cubic and tetragonal supercells with up to 40 atoms. The training structures were taken from MD simulations at different temperatures and pressures, generated using a NEP PES model constructed in an earlier study.[89] In total, the reference data set contained 940 structures. 140 structures were randomly placed in a holdout set for validation, while training sets were compiled by the shuffle-split method (random selection with replacement) with 200 to 800 structures and five data sets per training set size.

A comparison of models generated using different choices for the size of the neural network as well as the descriptor demonstrates that viable models can be obtained for a wide range of parameters and that even small models with as few as 1500 or so parameters can yield very good results (Figure S12). Yet fine-tuning of these parameters as well as the regularization parameters (Figure S13) allows one to maximize model performance.

The convergence of RMSEs and R^2 scores with training set size is similar for TNEP and SA-GPR with slightly better performance for TNEP (Figure 6). In both cases, training sets of about 400 structures already yield very good models, demonstrating the data efficiency of these approaches. This behavior has also been observed in the construction of models for amino acids.[91]

Next MD simulations were carried out using 12 × 12 × 12 supercells (8640 atoms) and a time step of 1 fs using the NEP model for the PES. Following equilibration at 300 K and 0 GPa in the NPT ensemble, the time-dependent susceptibility $\chi(t)$

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

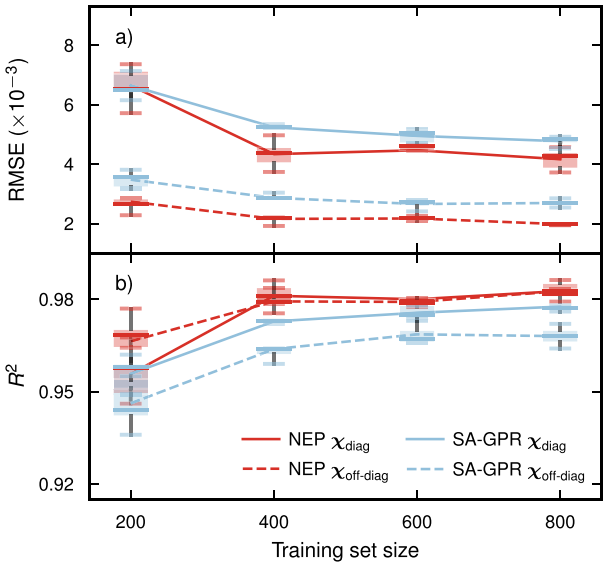


Figure 6. Variation of (a) RMSEs and (b) R^2 scores with training set size for TNEP and SA-GPR models based on five training sets per size generated by shuffle-split. In the case of TNEP, we used $N_{neu} = 20$, $n^R_{max} = n^A_{max} = 4$, and $\lambda_1 = \lambda_2 = 2 \times 10^{-3}$ (compare Figures S12 and S13).

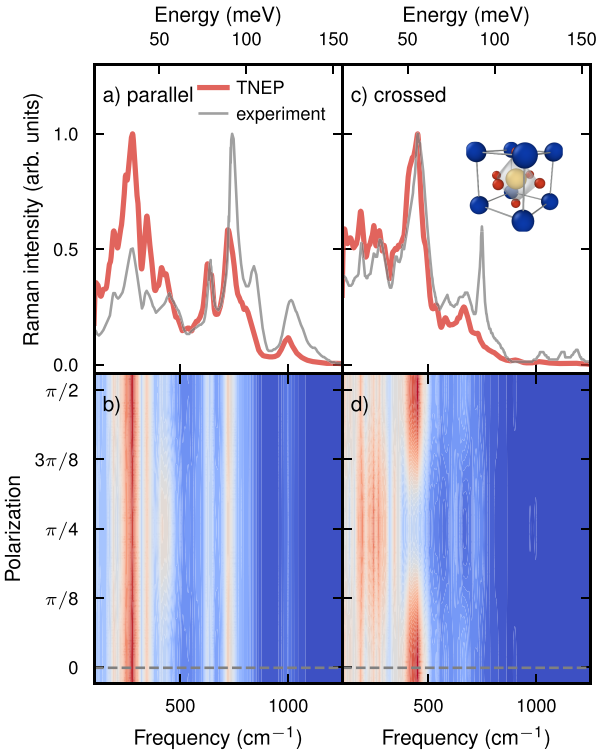


Figure 7. Raman spectra of $BaZrO_3$ for (a,b) parallel and (c,d) crossed polarization from simulations using a combination of NEP and TNEP models (red lines) as well as experiment (gray lines).[90] The spectra shown in (a,c) have been predicted for the nominal alignments used in the experimental measurements. The corresponding polarizations are indicated by the dashed horizontal lines in (b,d).

was recorded for 500 ps using a time resolution of 5 fs. For production, we used a TNEP model for χ trained against the full data set, but we found that models based on at least approximately 400 structures yield results that are practically indistinguishable within the statistical uncertainty. The Raman line shape was subsequently obtained via the ACF of χ according to eq 11. We then computed the Raman spectra for parallel (Figure 7a,b) and crossed polarization (Figure 7c,d), which in Porto notation correspond to $Z(XX)\,\overline{Z}$ and $Z(XY)\overline{Z}$, respectively, where X and Y are arbitrary crystal axes. The final spectra were obtained by averaging over 20 independent MD trajectories.

The results are overall in very good agreement with experiment, especially considering the very strong anharmonicity of this material and the strong temperature dependence of the vibrational spectrum.[88] The main difference with respect to the position of the peaks is a slight red-shift in the predicted spectra in the region above 600 cm^{-1}. This overly soft response can be attributed to the underlying exchange-correlation functional (vdW-DF-cx, refs.[92,93]), which the NEP model truthfully reproduces. One can also observe an inversion in the intensity of the low- and high energy features. This effect is almost certainly due to the classical sampling used here. It is rather common to correct for quantum effects in IR and *first order* Raman spectra by including a factor similar to the prefactor in eq 8. In the case of $BaZrO_3$ the room-temperature Raman spectrum arises, however, due to *second-order* scattering, i.e., due to combinations of modes. In that case, the application of the commonly used correction factor is no longer valid. Here, we omit such corrections entirely.

The Raman spectra depend on the crystal orientation with respect to the excitation laser. The present approach allows one to readily map out this dependence via eqs 10 and 11 (Figure 7b,d). While we are unaware of experimental measurements of the polarization dependence for $BaZrO_3$, we note that such experiments have been carried out for, e.g., NaCl.[94] As demonstrated in the previous study, such measurements can provide valuable additional information.

5. CONCLUSIONS

In this contribution, we introduced an extension of the NEP approach to tensors, resulting in the TNEP scheme. This was achieved by constructing expressions for rank-1 and rank-2 tensors based on the expression for the virial, which is a rank-2 tensor that arises naturally from derivatives of the energy (a rank-0 tensor) with respect to the atomic distances. This approach, which can be extended to tensors of higher rank, thus allows one to easily construct models that are equivariant.

We demonstrated the accuracy of this approach and its computational efficiency by constructing models for the dipole moment $\boldsymbol{\mu}$, the molecular polarizability $\boldsymbol{\alpha}$, and the electric susceptibility χ for several molecules, a liquid, and two crystalline materials. In particular, the computational speed of the current method and its implementation in the GPUMD package provide a significant advantage in terms of both the time scales and system sizes that can be sampled.

Finally, we applied the approach to predict IR and Raman spectra of liquid water, the molecule $PTAF^-$, and the perovskite $BaZrO_3$ in very good agreement with available experimental data, illustrating the range of systems that can be readily addressed by using the TNEP methodology introduced here.

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Data Availability Statement

The source code and documentation for GPUMD are available at https://github.com/brucefan1983/GPUMD and https://gpumd.org, respectively. The source code and documentation for CALORINE are available at https://gitlab.com/materials-modeling/calorine and https://calorine.materialsmodeling.org, respectively. NEP models and data are available via Zenodo via 10.5281/zenodo.10257363 (water and $BaZrO_3$), 10.5281/zenodo.10255268 ($PTAF^-$), and 10.5281/zenodo.8337182 ($BaZrO_3$).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c01343.

Training process of SA-GPR models for water systems; calculations of dipole moment for liquid water and α-Fe_2O_3; notes on the Clausius-Mossotti relation and the units of polarizability and electric susceptibility; parity plots of the TNEP predicted dipole moment, diagonal and off-diagonal elements of the polarizability tensor versus the *ab initio* references for water systems; demonstration of rotational invariance; parameters used in the training of NEP and TNEP models; dependence of TNEP model performance for $BaZrO_3$ on model and training hyperparameters; timing of on-the-fly model evaluation during MD simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Yi He − *Institute of Zhejiang University-Quzhou, Quzhou 324000, P. R. China; College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, P. R. China; Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0002-8807-0892; Email: yihezj@zju.edu.cn

Zheyong Fan − *College of Physical Science and Technology, Bohai University, Jinzhou 121013, P. R. China;* orcid.org/0000-0002-2253-8210; Email: brucenju@gmail.com

Paul Erhart − *Department of Physics, Chalmers University of Technology, SE-41296 Gothenburg, Sweden;* orcid.org/0000-0002-2516-6061; Email: erhart@chalmers.se

Authors

Nan Xu − *Institute of Zhejiang University-Quzhou, Quzhou 324000, P. R. China; College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, P. R. China;* orcid.org/0000-0003-3986-6883

Petter Rosander − *Department of Physics, Chalmers University of Technology, SE-41296 Gothenburg, Sweden*

Christian Schäfer − *Department of Physics, Chalmers University of Technology, SE-41296 Gothenburg, Sweden;* orcid.org/0000-0002-8557-733X

Eric Lindgren − *Department of Physics, Chalmers University of Technology, SE-41296 Gothenburg, Sweden;* orcid.org/0000-0002-8549-6839

Nicklas Österbacka − *Department of Physics, Chalmers University of Technology, SE-41296 Gothenburg, Sweden;* orcid.org/0000-0002-6043-4607

Mandi Fang − *Institute of Zhejiang University-Quzhou, Quzhou 324000, P. R. China; College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, P. R. China;* orcid.org/0009-0006-8273-3508

Wei Chen − *State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, P. R. China;* orcid.org/0000-0002-0165-4896

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c01343

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

N.X. is grateful for the financial support provided by the Startup Funds of the Institute of Zhejiang University-Quzhou. P.R., C.S., E.L., N.Ö., and P.E. acknowledge funding from the Swedish Research Council (Nos. 2020-04935 and 2021-05072), the European Union under the Marie Skłodowska-Curie program (No. 101065117), and the Swedish Foundation for Strategic Research via the SwedNESS graduate school (GSn15-0008) as well as computational resources provided by the National Academic Infrastructure for Supercomputing in Sweden at NSC, PDC, and C3SE partially funded by the Swedish Research Council through grant agreement No. 2022-06725. Y.H. acknowledges funding from the National Key Research and Development Program of China (No. 2022YFE0106100) and the National Natural Science Foundation of China (Nos. 22178299, 51933009). W.C. acknowledges funding from the National Natural Science Foundation of China (No. 22373104).

■ REFERENCES

(1) Schrader, B. Vibrational Spectroscopy of Different Classes and States of Compounds. In *Infrared and Raman spectroscopy: Methods and applications*; Wiley-VCH: 1995.

(2) Dendisová, M.; Jeništová, A.; Parchaňská-Kokaislová, A.; Matějka, P.; Prokopec, V.; Švecová, M. The use of infrared spectroscopic techniques to characterize nanomaterials and nanostructures: A review. *Anal. Chim. Acta* **2018**, *1031*, 1−14.

(3) Das, R. S.; Agrawal, Y. Raman spectroscopy: Recent advancements, techniques and applications. *Vib. Spectrosc.* **2011**, *57*, 163−176.

(4) 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.

(5) Silvestrelli, P. L.; Bernasconi, M.; Parrinello, M. Ab initio infrared spectrum of liquid water. *Chem. Phys. Lett.* **1997**, *277*, 478−482.

(6) Putrino, A.; Parrinello, M. Anharmonic Raman spectra in high-pressure ice from ab initio simulations. *Phys. Rev. Lett.* **2002**, *88*, 176401.

(7) Aprà, E.; Bhattarai, A.; Crampton, K. T.; Bylaska, E. J.; Govind, N.; Hess, W. P.; El-Khoury, P. Z. Time domain simulations of single molecule Raman scattering. *J. Phys. Chem. A* **2018**, *122*, 7437−7442.

(8) Luber, S.; Iannuzzi, M.; Hutter, J. Raman spectra from ab initio molecular dynamics and its application to liquid S-methyloxirane. *J. Chem. Phys.* **2014**, *141*, 094503.

(9) Kou, Z.; Hashemi, A.; Puska, M. J.; Krasheninnikov, A. V.; Komsa, H.-P. Simulating Raman spectra by combining first-principles and empirical potential approaches with application to defective MoS_2. *NPJ. Comput. Mater.* **2020**, *6*, 59.

(10) Sommers, G. M.; Andrade, M. F. C.; 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.

(11) MacKerell, A. D., Jr.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257−265.
(12) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(13) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(14) Foiles, S. M.; Baskes, M. I.; Daw, M. S. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Phys. Rev. B* **1986**, *33*, 7983−7991.
(15) Tersoff, J. Empirical Interatomic Potential for Carbon, with Applications to Amorphous Carbon. *Phys. Rev. Lett.* **1988**, *61*, 2879−2882.
(16) Henschel, H.; Andersson, A. T.; Jespers, W.; Mehdi Ghahremanpour, M.; Van der Spoel, D. Theoretical infrared spectra: quantitative similarity measures and force fields. *J. Chem. Theory Comput.* **2020**, *16*, 3307−3315.
(17) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
(18) 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.
(19) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R.; E, W. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. *Adv. Neural Inf. Process. Syst.*; 2018; Vol. *2018*, pp 4436−4446.
(20) Zhang, Y.; Xia, J.; Jiang, B. Physically motivated recursively embedded atom neural networks: Incorporating local completeness and nonlocality. *Phys. Rev. Lett.* **2021**, *127*, 156002.
(21) Fan, Z.; Zeng, Z.; Zhang, C.; Wang, Y.; Song, K.; Dong, H.; Chen, Y.; Ala-Nissila, T. Neuroevolution machine learning potentials: Combining high accuracy and low cost in atomistic simulations and application to heat transport. *Phys. Rev. B* **2021**, *104*, 104309.
(22) Gastegger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **2017**, *8*, 6924−6935.
(23) Chmiela, S.; Sauceda, H. E.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **2019**, *240*, 38−45.
(24) Kwac, K.; Freedman, H.; Cho, M. Machine learning approach for describing water OH stretch vibrations. *J. Chem. Theory Comput.* **2021**, *17*, 6353−6365.
(25) Shanavas Rasheeda, D.; Martín Santa Daría, A.; Schröder, B.; Mátyus, E.; Behler, J. High-dimensional neural network potentials for accurate vibrational frequencies: the formic acid dimer benchmark. *Phys. Chem. Chem. Phys.* **2022**, *24*, 29381−29392.
(26) Fransson, E.; Rosander, P.; Eriksson, F.; Rahm, J. M.; Tadano, T.; Erhart, P. Limits of the phonon quasi-particle picture at the cubic-to-tetragonal phase transition in halide perovskites. *Commun. Phys.* **2023**, *6*, 173.
(27) Leontyev, I.; Stuchebrukhov, A. Accounting for electronic polarization in non-polarizable force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2613−2626.
(28) Thaunay, F.; Jana, C.; Clavaguéra, C.; Ohanessian, G. Strategy for modeling the infrared spectra of ion-containing water drops. *J. Phys. Chem. A* **2018**, *122*, 832−842.
(29) Giovannini, T.; Grazioli, L.; Ambrosetti, M.; Cappelli, C. Calculation of IR spectra with a fully polarizable QM/MM approach based on fluctuating charges and fluctuating dipoles. *J. Chem. Theory Comput.* **2019**, *15*, 5495−5507.
(30) Veit, M.; Wilkins, D. M.; Yang, Y.; DiStasio, J.; Robert, A.; Ceriotti, M. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J. Chem. Phys.* **2020**, *153*, 024113.
(31) Smirnov, K. S.; Bougeard, D. Quantum-chemical derivation of electro-optical parameters for alkanes. *J. Raman Spectrosc.* **2006**, *37*, 100−107.
(32) Chen, Q.; Milner, S. T. Predicting Raman spectra of condensed polymer phases from MD simulations. *Macromolecules* **2017**, *50*, 9773−9787.
(33) Bornhauser, P.; Bougeard, D. Intensities of the vibrational spectra of siliceous zeolites by molecular dynamics calculations. IIRaman spectra. *J. Raman Spectrosc.* **2001**, *32*, 279−285.
(34) Rahmani, A.; Sauvajol, J.-L.; Cambedouzou, J.; Benoit, C. Raman-active modes in finite and infinite double-walled carbon nanotubes. *Phys. Rev. B* **2005**, *71*, 125402.
(35) Bougeard, D.; Smirnov, K. S. Calculation of off-resonance Raman scattering intensities with parametric models. *J. Raman Spectrosc.* **2009**, *40*, 1704−1719.
(36) Hasegawa, T.; Tanimura, Y. A polarizable water model for intramolecular and intermolecular vibrational spectroscopies. *J. Phys. Chem. B* **2011**, *115*, 5545−5553.
(37) Medders, G. R.; Paesani, F. Infrared and Raman spectroscopy of liquid water through first-principles many-body molecular dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1145−1154.
(38) 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.
(39) 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.
(40) Zhang, L.; Chen, M.; Wu, X.; Wang, H.; E, W.; Car, R. Deep neural network for the dielectric response of insulators. *Phys. Rev. B* **2020**, *102*, 041121.
(41) 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.
(42) Beckmann, R.; Brieuc, F.; Schran, C.; Marx, D. Infrared spectra at coupled cluster accuracy from neural network representations. *J. Chem. Theory Comput.* **2022**, *18*, 5492−5501.
(43) Feng, C.; Xi, J.; Zhang, Y.; Jiang, B.; Zhou, Y. Accurate and interpretable dipole interaction model-based machine learning for molecular polarizability. *J. Chem. Theory Comput.* **2023**, *19*, 1207−1217.
(44) Zhang, C.; Tang, F.; Chen, M.; Xu, J.; Zhang, L.; Qiu, D. Y.; Perdew, J. P.; Klein, M. L.; Wu, X. Modeling liquid Water by climbing up Jacob's ladder in density functional theory facilitated by using deep neural network potentials. *J. Phys. Chem. B* **2021**, *125*, 11444−11456.
(45) Berger, E.; Komsa, H.-P. Polarizability Models for Simulations of Finite Temperature Raman Spectra from Machine Learning Molecular Dynamics. *arXiv*. 2023, arXiv:2310.13310. https://arxiv.org/abs/2310.13310 (accessed 2024-04-03).
(46) Fan, Z. Improving the accuracy of the neuroevolution machine learning potential for multi-component systems. *J. Phys.: Condens. Matter* **2022**, *34*, 125902.
(47) Fan, Z.; Wang, Y.; Ying, P.; Song, K.; Wang, J.; Wang, Y.; Zeng, Z.; Xu, K.; Lindgren, E.; Rahm, J. M.; Gabourie, A. J.; Liu, J.; Dong, H.; Wu, J.; Chen, Y.; Zhong, Z.; Sun, J.; Erhart, P.; Su, Y.; Ala-Nissila, T. GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations. *J. Chem. Phys.* **2022**, *157*, 114801.
(48) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
(49) Schaul, T.; Glasmachers, T.; Schmidhuber, J. High dimensions and heavy tails for natural evolution strategies. *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*; New York, NY, USA, 2011; pp845−852.
(50) McQuarrie, D. A. *Statistical Mechanics*; Harper's chemistry series; Harper Collins: New York, 1976.

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(51) Cardona, M. *Light Scattering in Solids I*; Topics in Applied Physics; Springer: Berlin, Heidelberg, 1983; Vol. *8*.
(52) Born, M.; Huang, K. *Dynamical theory of crystal lattices*; International series of monographs on physics; Clarendon Press: 1954.
(53) TENSOAP. https://github.com/dilkins/TENSOAP (accessed 2023-08-12).
(54) TENSOAP-FAST. https://github.com/dilkins/TENSOAP-FAST/tree/master/example/models/bulk-water (accessed 2023-08-12).
(55) Makov, G.; Payne, M. C. Periodic boundary conditions in ab initio calculations. *Phys. Rev. B* **1995**, *51*, 4014−4022.
(56) Ditler, E.; Mattiat, J.; Luber, S. The position operator problem in periodic calculations with an emphasis on theoretical spectroscopy. *Phys. Chem. Chem. Phys.* **2023**, *25*, 14672−14685.
(57) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
(58) Yang, Y.; Lao, K. U.; Wilkins, D. M.; Grisafi, A.; Ceriotti, M.; DiStasio, R. A. Quantum mechanical static dipole polarizabilities in the QM7b and AlphaML showcase databases. *Sci. Data* **2019**, *6*, 152.
(59) Piela, L. *Ideas of Quantum Chemistry*; Elsevier: Amsterdam, 2007; pp 615−680.
(60) Spaldin, N. A. A beginner's guide to the modern theory of polarization. *J. Solid State Chem.* **2012**, *195*, 2−10.
(61) Krishnamoorthy, A.; Nomura, K.-i.; Baradwaj, N.; Shimamura, K.; Rajak, P.; Mishra, A.; Fukushima, S.; Shimojo, F.; Kalia, R.; Nakano, A.; Vashishta, P. Dielectric constant of liquid water determined with neural network quantum molecular dynamics. *Phys. Rev. Lett.* **2021**, *126*, 216403.
(62) 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.
(63) NEP_CPU. https://github.com/brucefan1983/NEP_CPU (accessed 2023-08-12).
(64) (a) Lindgren, E.; Rahm, J. M.; Fransson, E.; Eriksson, F.; Österbacka, N.; Fan, Z.; Erhart, P. calorine: A Python package for constructing and sampling neuroevolution potential models. *Journal of Open Source Software* **2024**, *9*, 6264. (b) https://calorine.materialsmodeling.org/; DOI: 10.5281/zenodo.7919206 (accessed 2024-03-19).
(65) Kapil, V.; Wilkins, D. M.; Lan, J.; Ceriotti, M. Inexpensive Modeling of Quantum Dynamics Using Path Integral Generalized Langevin Equation Thermostats. *J. Chem. Phys.* **2020**, *152*, 124104.
(66) Shepherd, S.; Lan, J.; Wilkins, D. M.; Kapil, V. Efficient Quantum Vibrational Spectroscopy of Water with High-Order Path Integrals: From Bulk to Interfaces. *The Journal of Physical Chemistry Letters* **2021**, *12*, 9108−9114.
(67) Kapil, V.; Kovács, D. P.; Csányi, G.; Michaelides, A. First-Principles Spectroscopy of Aqueous Interfaces Using Machine-Learned Electronic and Quantum Nuclear Effects. *Faraday Discuss.* **2024**, *249*, 50.
(68) Praprotnik, M.; Janežič, D. Molecular dynamics integration and molecular vibrational theory. III. The infrared spectrum of water. *J. Chem. Phys.* **2005**, *122*, 174103.
(69) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J.; et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.
(70) Downing, H. D.; Williams, D. Optical constants of water in the infrared. *J. Geophys. Res.* **1975**, *80*, 1656−1661.
(71) Max, J.-J.; Chapados, C. Isotope effects in liquid water by infrared spectroscopy. III. H_2O and D_2O spectra from 6000 to 0 cm^{-1}. *J. Chem. Phys.* **2009**, *131*, 184505.
(72) Rognoni, A.; Conte, R.; Ceotto, M. How many water molecules are needed to solvate one? Chem. *Sci.* **2021**, *12*, 2060−2064.
(73) Praprotnik, M.; Janežič, D.; Mavri, J. Temperature dependence of water vibrational spectrum: A molecular dynamics simulation study. *J. Phys. Chem. A* **2004**, *108*, 11056−11062.
(74) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.
(75) Xu, J.; Chen, M.; Zhang, C.; Wu, X. First-principles study of the infrared spectrum in liquid water from a systematically improved description of H-bond network. *Phys. Rev. B* **2019**, *99*, 205123.
(76) Brooker, M. H.; Hancock, G.; Rice, B. C.; Shapter, J. Raman frequency and intensity studies of liquid H_2O, $H_2{}^{18}O$ and D_2O. *J. Raman Spectrosc.* **1989**, *20*, 683−694.
(77) Morawietz, T.; Marsalek, O.; Pattenaude, S. R.; Streacker, L. M.; Ben-Amotz, D.; Markland, T. E. The interplay of structure and dynamics in the Raman spectrum of liquid water over the full frequency and temperature range. *J. Phys. Chem. Lett.* **2018**, *9*, 851−857.
(78) Chintareddy, V. R.; Wadhwa, K.; Verkade, J. G. Tetrabutylammonium fluoride (TBAF)-catalyzed addition of substituted trialkylsilylalkynes to aldehydes, ketones, and trifluoromethyl ketones. *J. Org. Chem.* **2011**, *76*, 4482−4488.
(79) Thomas, A.; George, J.; Shalabney, A.; Dryzhakov, M.; Varma, S. J.; Moran, J.; Chervy, T.; Zhong, X.; Devaux, E.; Genet, C.; Hutchison, J. A.; Ebbesen, T. W. Ground-state chemical reactivity under vibrational coupling to the vacuum electromagnetic field. *Angew. Chem., Int. Ed.* **2016**, *128*, 11634−11638.
(80) Schäfer, C.; Flick, J.; Ronca, E.; Narang, P.; Rubio, A. Shining light on the microscopic resonant mechanism responsible for cavity-mediated chemical reactivity. *Nat. Commun.* **2022**, *13*, 7817.
(81) Schäfer, C.; Fojt, J.; Lindgren, E.; Erhart, P. Machine Learning for Polaritonic Chemistry: Accessing Chemical Kinetics. *J. Am. Chem. Soc.* **2024**, *146*, 5402−5413.
(82) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, 224108.
(83) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(84) Weigend, F.; Baldes, A. Segmented contracted basis sets for one-and two-component Dirac−Fock effective core potentials. *J. Chem. Phys.* **2010**, *133*, 174102.
(85) Akbarzadeh, A. R.; Kornev, I.; Malibert, C.; Bellaiche, L.; Kiat, J. M. Combined theoretical and experimental study of the low-temperature properties of $BaZrO_3$. *Phys. Rev. B* **2005**, *72*, 205104.
(86) Knight, K. S. Low-temperature thermophysical and crystallographic properties of $BaZrO_3$ perovskite. *J. Mater. Sci.* **2020**, *55*, 6417−6428.
(87) Perrichon, A.; Jedvik Granhed, E.; Romanelli, G.; Piovano, A.; Lindman, A.; Hyldgaard, P.; Wahnström, G.; Karlsson, M. Unraveling the Ground-State Structure of $BaZrO_3$ by Neutron Scattering Experiments and First-Principles Calculations. *Chem. Mater.* **2020**, *32*, 2824−2835.
(88) Rosander, P.; Fransson, E.; Milesi-Brault, C.; Toulouse, C.; Bourdarot, F.; Piovano, A.; Bossak, A.; Guennou, M.; Wahnström, G. Anharmonicity of the antiferrodistortive soft mode in barium zirconate $BaZrO_3$. *Phys. Rev. B* **2023**, *108*, 014309.
(89) Fransson, E.; Rosander, P.; Erhart, P.; Wahnström, G. Understanding correlations in $BaZrO_3$: Structure and dynamics on the nano-scale. *Chem. Mater.* **2024**, *36*, 514−523.
(90) Toulouse, C.; Amoroso, D.; Xin, C.; Veber, P.; Hatnean, M. C.; Balakrishnan, G.; Maglione, M.; Ghosez, P.; Kreisel, J.; Guennou, M. Lattice dynamics and Raman spectrum of $BaZrO_3$ single crystals. *Phys. Rev. B* **2019**, *100*, 134102.
(91) Berger, E.; Niemelä, J.; Lampela, O.; Juffer, A. H.; Komsa, H.-P. Raman spectra of amino acids and peptides from machine learning

https://doi.org/10.1021/acs.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

polarizabilities. *arXiv*. 2024, arXiv:2401.14808. https://arxiv.org/abs/2401.14808 (accessed 2024-04-03).
(92) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.
(93) Berland, K.; Hyldgaard, P. Exchange functional that tests the robustness of the plasmon description of the van der Waals density functional. *Phys. Rev. B* **2014**, *89*, 035412.
(94) 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.jctc.3c01343
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX