Tensorial properties via the neuroevolution potential framework: Fast simulation of infrared and Raman spectra

Nan Xu,[1, 2] Petter Rosander,[3] Christian Schäfer,[3] Eric Lindgren,[3] Nicklas Österbacka,[3]
Mandi Fang,[1, 2] Wei Chen,[4] Yi He,[1, 2, 5, *] Zheyong Fan,[6, †] and Paul Erhart[3, ‡]
[1] *Institute of Zhejiang University-Quzhou, Quzhou 324000, P. R. China*
[2] *College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, P. R. China*
[3] *Department of Physics, Chalmers University of Technology, SE-41296, Gothenburg, Sweden*
[4] *State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing, P. R. China*
[5] *Department of Chemical Engineering, University of Washington, Seattle, WA 98195, USA*
[6] *College of Physical Science and Technology, Bohai University, Jinzhou 121013, P. R. China*

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 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 to 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.

I. INTRODUCTION

Infrared (IR) and Raman spectroscopy are widely used techniques for the non-destructive 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]: Firstly, *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. Secondly, similarly expensive calculations of dipole moment ($\boldsymbol{\mu}$), polarizability ($\boldsymbol{\alpha}$) or electric susceptibility ($\boldsymbol{\chi}$) are required for at least many thousand 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 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 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

---

[*] yihezj@zju.edu.cn
[†] brucenju@gmail.com
[‡] erhart@chalmers.se

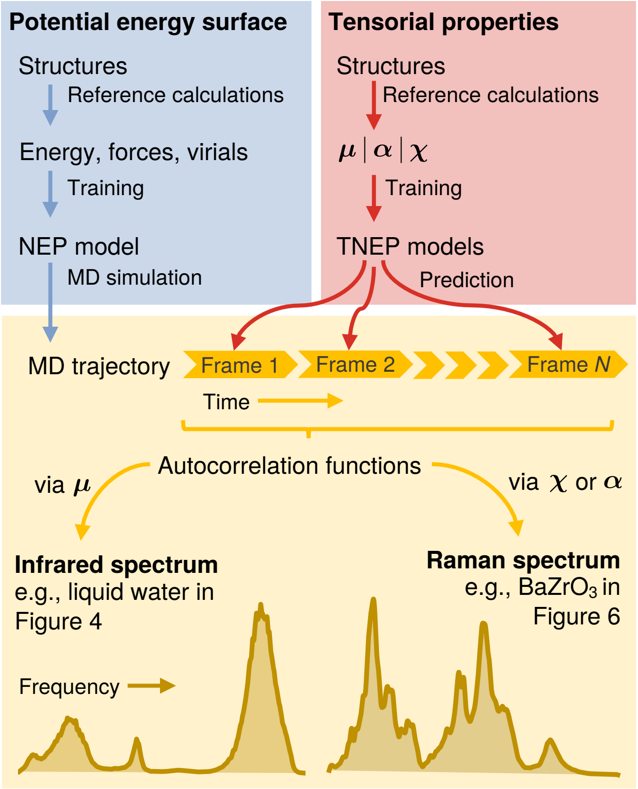

FIG. 1. Workflow for simulations of IR and Raman spectra using NEP models for the PES and TNEP models for the dipole moment $\boldsymbol{\mu}$, the polarizability $\boldsymbol{\alpha}$ or the susceptibility $\boldsymbol{\chi}$.

[31, 32], 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 ($BaZrO_3$; Fig. 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.

II. METHODOLOGY

A. 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 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

$$U_i = U_i(q_i^\nu). \qquad (1)$$

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 errors (RMSEs) 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. The full expressions for these quantities can be found in Ref. 47. Here, we only present the expression for the 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].

B. TNEP rank-1 tensor models

To develop a 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 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

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

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 a ML model for rank-1 tensors such as the dipole moment. Below we refer to Eq. (3) as the TNEP dipole model.

C. 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 term 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.

D. Dielectric response

It is instructive to recall some relations that describe the response of finite (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 the molecular polarizability. There are different conventions for expressing $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\chi}$ leading to different units (Sect. S7). Here, we use $\mathrm{e}\cdot\mathrm{Å}$ for $\boldsymbol{\mu}$ and $\mathrm{Å}^3$ 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).

E. The IR intensity

The IR absorption cross section is given by [50]

$$\sigma(\omega) = \frac{4\pi^2}{\hbar c n}\omega\left(1 - e^{-\beta\hbar\omega}\right)M(\omega),$$

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

$$M\left(\omega\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\langle \boldsymbol{\mu}\left(0\right)\cdot\boldsymbol{\mu}\left(t\right)\right\rangle e^{-i\omega t}\mathrm{d}t,$$

where $\langle\cdots\rangle$ indicates the average over time origins. Since the lineshape is sampled classically, we make a classical approximation for the prefactor by expanding the Boltzmann factor to first order, which gives

$$\sigma(\omega) \propto \omega^2 M(\omega). \tag{5}$$

F. The Raman intensity

The differential Raman cross-section is given by [50, 51]

$$\frac{\partial^2\sigma}{\partial\omega_\mathrm{out}\partial\Omega} = \left(\frac{\omega_\mathrm{in}\pm\omega}{2\pi}\right)^4 L_{ijkl}(\omega), \tag{6}$$

where 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. $L_{ijkl}(\omega)$ is the Raman lineshape given by the Fourier transform of the time-dependent polarizability $\boldsymbol{\alpha}(t)$ (finite system) or susceptibility $\boldsymbol{\chi}(t)$ (extended systems), e.g., in the case of the former

$$L_{ijkl}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\langle \alpha_{ij}\left(0\right)\alpha_{kl}\left(t\right)\right\rangle e^{-i\omega t}\mathrm{d}t. \tag{7}$$

The elements of the polarizability (or susceptibility) tensor are selected by components of the polarization of the incoming and outgoing light as

$$\alpha_{ij} = \hat{\boldsymbol{\epsilon}}_i^\mathrm{in}\cdot\boldsymbol{\alpha}\cdot\hat{\boldsymbol{\epsilon}}_j^\mathrm{out}. \tag{8}$$

This allows one to split the spectrum into an isotropic and an anisotropic part. To this end, the polarizability tensor (and equivalently the susceptibility tensor) can be rewritten as $\boldsymbol{\alpha} = \gamma\boldsymbol{I} + \boldsymbol{\beta}$ where $\gamma = \mathrm{Tr}(\boldsymbol{\alpha})/3$ and $\boldsymbol{\beta}$ is a traceless tensor. This leads to the decomposition

$$\begin{aligned} L_\mathrm{iso}(\omega) &\propto \int_{-\infty}^{\infty}\langle\gamma\left(0\right)\gamma\left(t\right)\rangle e^{-i\omega t}\mathrm{d}t \\ L_\mathrm{aniso}(\omega) &\propto \int_{-\infty}^{\infty}\langle\mathrm{Tr}[\boldsymbol{\beta}\left(0\right)\boldsymbol{\beta}\left(t\right)]\rangle e^{-i\omega t}\mathrm{d}t. \end{aligned} \tag{9}$$

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

$$\boldsymbol{\chi} = \boldsymbol{\chi}_\mathrm{ion}(\omega) + \boldsymbol{\chi}_\mathrm{e}(\omega),$$

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

G. 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 (Fig. 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 $\boldsymbol{\chi}(t)$ along the trajectory. Finally, IR or Raman spectra are obtained via Fourier transformation of the respective ACFs via Eqs. (5) or (6).

III. PERFORMANCE EVALUATION

In this section, we evaluate the performance of 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 Zündel 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, 52] (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 [53] 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].

A. Dipole moment

The TNEP dipole models can achieve very high precision when predicting $\boldsymbol{\mu}$ for both molecules and liquid

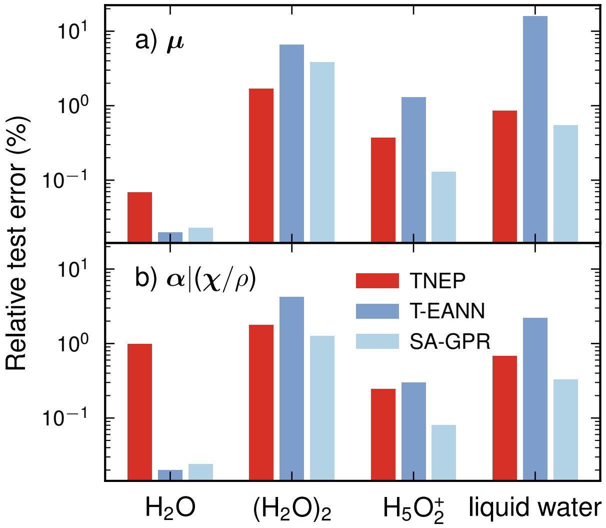

FIG. 2. Relative test errors for the validation sets according to TNEP, T-EANN, and SA-GPR models for water systems for (a) μ as well as (b) α and χ/ρ. The relative test error for liquid water from T-EANN [39] was reported for the averaged molecular polarizability obtained via the Clausius-Mossotti relation (Sect. S8). The relative test for χ/ρ should be somewhat higher than that for the averaged molecular polarizability.

TABLE I. RMSEs in $e \cdot Å$/atom for μ for the validation sets using NEP rank-1 tensor models.

System	RMSE
H_2O	3.8×10^{-5}
$(H_2O)_2$	92.8×10^{-5}
$H_5O_2{}^+$	10.9×10^{-5}
liquid water	29.3×10^{-5}

water with very low RMSEs (Table I) and correlation coefficients very close to one (Fig. S1).

As a further, more intuitive measure, one can also consider the relative error [39], defined as the RMSE divided by the standard deviation of the reference data (Fig. 2a). For the water monomer (H_2O) all three methods yield extremely small relative errors 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 [54, 55].

Neutral molecules. The μ of neutral molecules such as H_2O or $(H_2O)_2$ is uniquely defined. In the TNEP approach μ is calculated by summing over atomic contributions which, by contrast to, e.g., the T-EANN approach, does not require 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 [56, 57], for which we make similar observations (Sect. S4). The TNEP model yields a very low RMSE for the validation set of $9.51 \times 10^{-4}\, e \cdot Å$/atom.

Charged molecules. The μ of charged molecules is non-unique and depends on the choice of the reference point [4, 58]. For charged molecules, one should therefore employ the relative permanent dipole μ_r defined with respect to the center of mass, when training TNEP dipole models. The reference μ in the $H_5O_2{}^+$ data set [38, 52] have already been transformed to μ_r. Therefore, the absolute dipole moment of $H_5O_2{}^+$ including the movement of the center of mass should then be $\mu = \mu_r + e \cdot r_{\mathrm{COM}}$. The same procedure has been applied to the PTAF^- molecule below (Sect. IV C).

Periodic systems. Traditional methods for calculating μ cannot be applied to periodic systems since the position operator is not uniquely defined [55, 59]. This issue is overcome via the modern theory of polarization [38, 59, 60], which provides a rigorous definition for the polarization of periodic systems and established a methodology for calculating μ. It was therefore used in the present work to obtain μ for periodic systems including water (Sect. S1) and α-Fe_2O_3 (Sect. S5).

B. Polarizability and susceptibility

The RMSEs for the diagonal and off-diagonal elements of α of (H_2O), $(H_2O)_2$ and $H_5O_2{}^+$ are quite small (Table II), indicating the high accuracy of the TNEP polarizability model. The coefficients of determination are larger than 0.98 mirroring this trend (Fig. S6 and Fig. S7). For liquid water, we consider χ/ρ, which has the same unit as the polarizability per atom. The RMSEs for χ/ρ are on the same order of magnitude as the RMSEs for α (Table II).

The NEP models achieve an accuracy that is comparable to 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 (Fig. 2b). While the performance for the water monomer H_2O is worse, the TNEP model still yields a relative testing errors 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 $6.88 \times 10^{-3}\, Å^3\, \mathrm{atom}^{-1}$ and $3.28 \times 10^{-3}\, Å^3\, \mathrm{atom}^{-1}$ for the diagonal and off-diagonal elements of α, respectively. For comparison, Wilkins *et al.* [61] reported a higher RMSE value of $8.15 \times 10^{-3}\, Å^3\, \mathrm{atom}^{-1}$ over both the diagonal and off-diagonal elements of α using a SA-GPR model.

TABLE II. RMSEs in $\text{Å}^3\,\text{atom}^{-1}$ for $\boldsymbol{\alpha}$ (molecules) and $\boldsymbol{\chi}/\rho$ (liquid water) for the validation sets using TNEP rank-2 tensor models.

System	diagonal	off-diagonal
H_2O	42.0×10^{-4}	1.9×10^{-4}
$(H_2O)_2$	56.1×10^{-4}	33.8×10^{-4}
$H_5O_2{}^+$	4.8×10^{-4}	3.6×10^{-4}
liquid water	26.6×10^{-4}	18.2×10^{-4}

C. Computational speed

It is now instructive to evaluate the computational performance of TNEP models in comparison with publicly available SA-GPR models [52, 53]. 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 69 984 atoms were created by replication.

The SA-GPR models can only be run serially on a central processing unit (CPU). In contrast, the TNEP model can be run on CPUs using `NEP_CPU` [62], e.g., via the interface provided by the CALORINE package [63], or on graphics processing units (GPUs) 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 the TNEP CPU models are at least one order of magnitude faster than the SA-GPR models on CPUs for both dipole and polarizability (Fig. 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 speed up by an order of magnitude or more. For very small systems the GPU implementation is limited by IO.

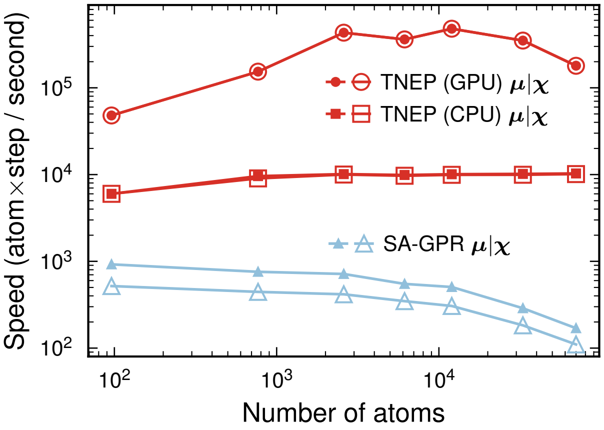

FIG. 3. Comparison of computational speed of SA-GPR and TNEP models for dipole ($\boldsymbol{\mu}$) and susceptibility ($\boldsymbol{\chi}$) of liquid water. Here, the SA-GPR results were obtained using the `TENSOAP-FAST` implementation [53].

IV. 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 IR and Raman spectra of molecules, liquids, and solids. To this end, we employ the correlation function approach outlined above (Sect. II G and Fig. 1).

A. IR spectrum of water

Firstly, 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.

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. III A). The IR spectrum was then obtained by Fourier transforming the dipole moment ACF via Eq. (5). 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 [64] via the CP2K software package [65], 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 [66, 67] over the entire frequency range from 0 to 4000 cm^{-1} (Fig. 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], the bending modes [37, 68] at about 1650 cm^{-1} as well as the OH stretching band [37, 68] 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 the soft intermolecular to the stiff intramolecular modes. This performance is also observed for the DP model (Fig. 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 [64, 69] such as TIP3P (Fig. 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's experiment [66]. 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 [70] that was used for generating the PES training data [10, 71] and/or the absence of quantum effects in the (classical) MD simulations [10, 37].

B. Raman spectra of water

To obtain the Raman spectra of liquid water we sampled the time dependence of $\boldsymbol{\chi}(t)$ 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. (6) and averaged over the available trajectories was then split into isotropic and anisotropic contributions via Eq. (9).

The **anisotropic spectrum** predicted by the NEP-TNEP approach is overall in very good agreement with experimental data (Fig. 4b) [72, 73]. The locations of 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 similar results as the NEP-TNEP approach in the region up to about 1900 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, 73]. 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.

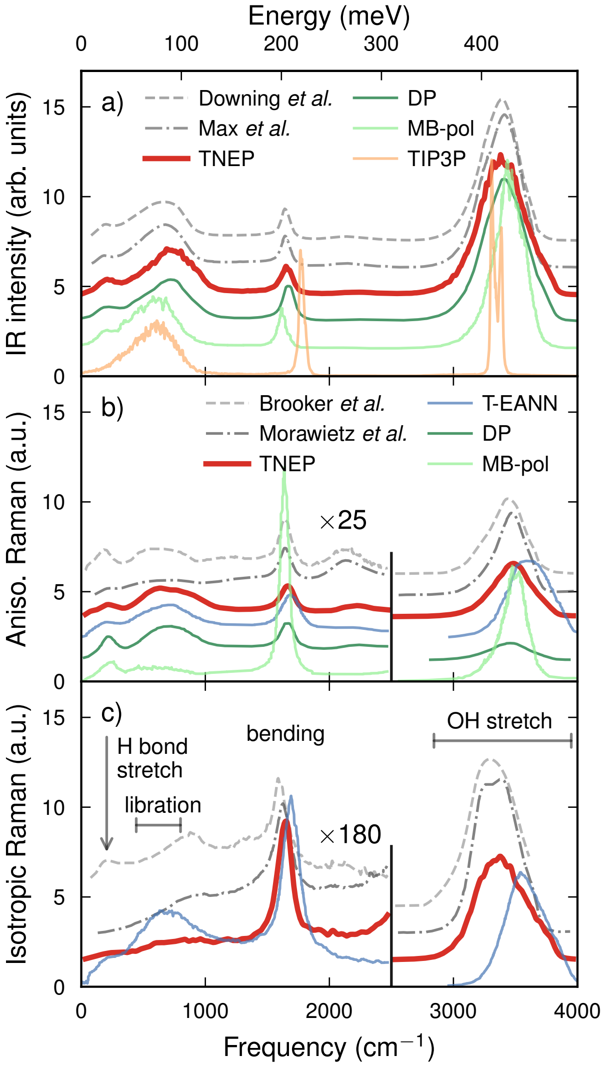

FIG. 4. Comparison of (a) infrared as well as (b) anisotropic and (c) isotropic Raman spectra of water at ambient conditions from simulations and experiment. Experimental data from Downing *et al.* [66], Max *et al.* [67], Brooker *et al.* [72], and Morawietz *et al.* [73]. Simulated spectra from T-EANN [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}.

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** (Fig. 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 redshifted. 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 blueshift can be observed similar to the anisotropic spectrum (Fig. 4b).

C. 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 Fig. 5), the intermediate reaction minimum in the S$_N$2 deprotection reaction 1-phenyl-2-trimethylsilylacetylene (PTA) with tetra-n-butylammonium fluoride [74–76].

To train NEP and TNEP models, we obtained PES and $\boldsymbol{\mu}$ data for a set of 20 170 structures via DFT calculations using the ORCA code [77], the PBE functional [78], and a def2-TZVP basis set [79] while enforcing tight convergence of the self-consistent field cycles. Subsequently, MD simulations at various temperatures were performed in the NVE ensemble using a timestep of 0.1 fs for 1 ns, during which $\boldsymbol{\mu}(t)$ was recorded with a time resolution of 0.5 fs.

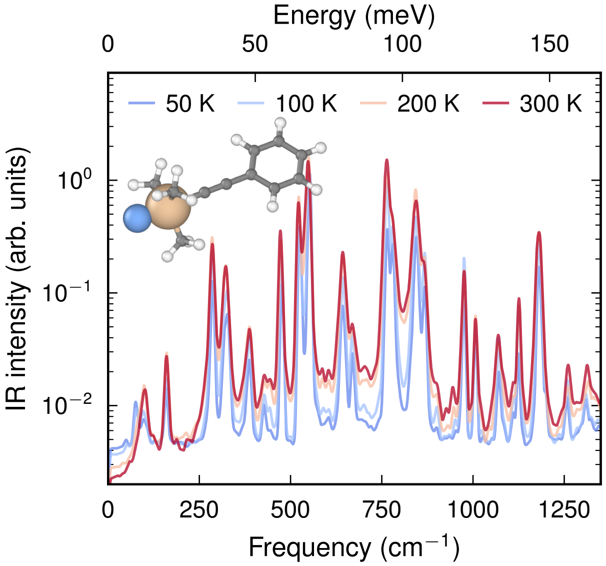

FIG. 5. IR spectra for the metastable PTAF$^-$ complex (see inset) at various temperatures.

The IR spectra obtained via the analysis of the ACF of $\boldsymbol{\mu}$ show a pronounced temperature dependence in particular of the linewidths (Fig. 5). The molecule supports several soft modes with frequencies in the region below 250 cm^{-1}, which are associated with 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 linewidth 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 a DFT-MD simulations and computationally very expensive for a CPU implementation.

D. 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 [81–83]. It features soft and strongly temperature-dependent phonon modes [84, 85], which have been carefully analyzed with Raman spectroscopy [80], rendering BaZrO$_3$

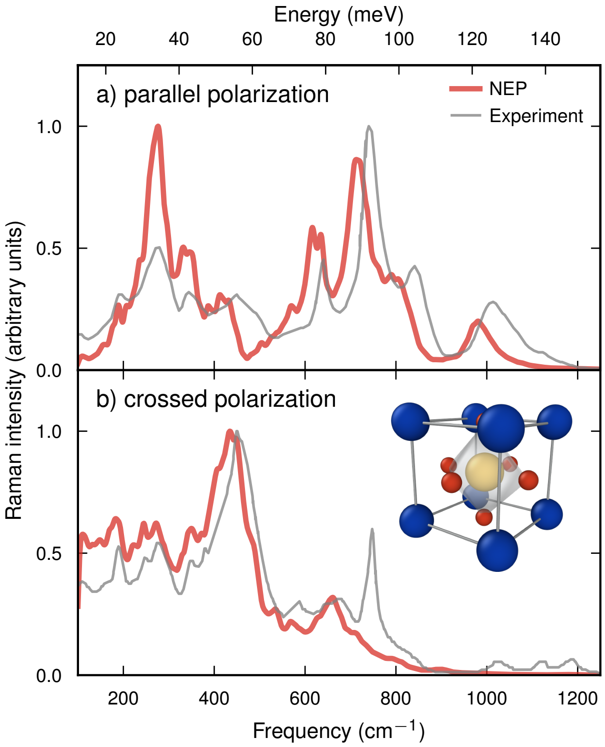

FIG. 6. Raman spectra of BaZrO$_3$ for (a) parallel and (b) crossed polarization from simulations using a combination of NEP and TNEP models as well as experiment [80].

an ideal application for the present approach.

A TNEP model for $\boldsymbol{\chi}$ was trained using 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 [85]. In total the reference data set contained 940 structures.

MD simulations were carried out using $12 \times 12 \times 12$ supercells (8640 atoms) and a timestep of 1 fs. Following equilibration at 300 K and 0 GPa in the NPT ensemble, the time-dependent susceptibility $\boldsymbol{\chi}(t)$ was recorded for 500 ps using a time resolution of 5 fs. The Raman lineshape was subsequently obtained via the ACF of $\boldsymbol{\chi}$ according to Eq. (7). We then computed the parallel and crossed Raman spectra, which in Porto notation correspond to $Z(XX)\bar{Z}$ ($xxyy$ in Eq. (7)) and $Z(XY)\bar{Z}$ ($xyxy$ in Eq. (7)), respectively. 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 [84]. The main difference is a slight red-shift in the predicted spectra in the region above $600\,\mathrm{cm}^{-1}$. This overly soft response can be attributed to the underlying exchange-correlation functional (vdW-DF-cx, Refs. 86, 87), which the NEP model truthfully reproduces.

V. CONCLUSIONS

In this contribution, we have 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 $\boldsymbol{\chi}$ for several molecules, a liquid as well as two crystalline materials. In particular, the computational speed of the current method and its implementation in the GPUMD package provide a significant advantage both in terms of 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 $\mathrm{BaZrO_3}$ in very good agreement with available experimental data, illustrating the range of systems that can be readily addressed using the TNEP methodology introduced here.

ACKNOWLEDGEMENTS

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 (Nos. 2022YFE0106100), and the National Natural Science Foundation of China (Nos. 22178299, 51933009). W. C. acknowledges funding from the National Natural Science Foundation of China (Nos. 22373104).

SUPPLEMENTARY INFORMATION

Supplemental material. Training process of SA-GPR models for water systems; calculations of dipole moment for liquid water and α-$\mathrm{Fe_2O_3}$; 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 polarizability versus the *ab initio* references for water systems; parameters used in the training of NEP and TNEP models.

Data availability. 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 `https://doi.org/10.5281/zenodo.10257363` (water and $\mathrm{BaZrO_3}$), `https://doi.org/10.5281/zenodo.10255268` (PTAF$^-$), and `https://doi.org/10.5281/zenodo.8337182` ($\mathrm{BaZrO_3}$).

---

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

[2] M. Dendisová, A. Jeništová, A. Parchaňská-Kokaislová, P. Matějka, V. Prokopec, and M. Švecová, Anal. Chim. Acta **1031**, 1 (2018).

[3] R. S. Das and Y. Agrawal, Vib. Spectrosc. **57**, 163 (2011).

[4] M. Thomas, M. Brehm, R. Fligg, P. Vöhringer, and

B. Kirchner, Phys. Chem. Chem. Phys. **15**, 6608 (2013).

[5] P. L. Silvestrelli, M. Bernasconi, and M. Parrinello, Chem. Phys. Lett. **277**, 478 (1997).

[6] A. Putrino and M. Parrinello, Phys. Rev. Lett. **88**, 176401 (2002).

[7] E. Aprà, A. Bhattarai, K. T. Crampton, E. J. Bylaska, N. Govind, W. P. Hess, and P. Z. El-Khoury, J. Phys. Chem. A **122**, 7437 (2018).

[8] S. Luber, M. Iannuzzi, and J. Hutter, J. Chem. Phys. **141**, 094503 (2014).

[9] Z. Kou, A. Hashemi, M. J. Puska, A. V. Krasheninnikov, and H.-P. Komsa, NPJ Comput. Mater. **6**, 59 (2020).

[10] G. M. Sommers, M. F. C. Andrade, L. Zhang, H. Wang, and R. Car, Phys. Chem. Chem. Phys. **22**, 10592 (2020).

[11] A. D. MacKerell Jr., N. Banavali, and N. Foloppe, Biopolymers **56**, 257 (2000).

[12] W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).

[13] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[14] S. M. Foiles, M. I. Baskes, and M. S. Daw, Phys. Rev. B **33**, 7983 (1986).

[15] J. Tersoff, Phys. Rev. Lett. **61**, 2879 (1988).

[16] H. Henschel, A. T. Andersson, W. Jespers, M. Mehdi Ghahremanpour, and D. Van der Spoel, J. Chem. Theory Comput. **16**, 3307 (2020).

[17] J. Behler, J. Chem. Phys. **145**, 170901 (2016).

[18] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Chem. Rev. **121**, 10142 (2021).

[19] L. Zhang, J. Han, H. Wang, W. A. Saidi, R. Car, and W. E, Adv. Neural Inf. Process. Syst. **2018**, 4436 (2018).

[20] Y. Zhang, J. Xia, and B. Jiang, Phys. Rev. Lett. **127**, 156002 (2021).

[21] Z. Fan, Z. Zeng, C. Zhang, Y. Wang, K. Song, H. Dong, Y. Chen, and T. Ala-Nissila, Phys. Rev. B **104**, 104309 (2021).

[22] M. Gastegger, J. Behler, and P. Marquetand, Chem. Sci. **8**, 6924 (2017).

[23] S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, Comput. Phys. Commun. **240**, 38 (2019).

[24] K. Kwac, H. Freedman, and M. Cho, J. Chem. Theory Comput. **17**, 6353 (2021).

[25] D. Shanavas Rasheeda, A. Martín Santa Daría, B. Schröder, E. Mátyus, and J. Behler, Phys. Chem. Chem. Phys. **24**, 29381 (2022).

[26] E. Fransson, P. Rosander, F. Eriksson, J. M. Rahm, T. Tadano, and P. Erhart, Commun. Phys **6**, 173 (2023).

[27] I. Leontyev and A. Stuchebrukhov, Phys. Chem. Chem. Phys. **13**, 2613 (2011).

[28] F. Thaunay, C. Jana, C. Clavaguéra, and G. Ohanessian, J. Phys. Chem. A **122**, 832 (2018).

[29] T. Giovannini, L. Grazioli, M. Ambrosetti, and C. Cappelli, J. Chem. Theory Comput. **15**, 5495 (2019).

[30] M. Veit, D. M. Wilkins, Y. Yang, J. DiStasio, Robert A., and M. Ceriotti, J. Chem. Phys. **153**, 024113 (2020).

[31] K. S. Smirnov and D. Bougeard, J. Raman Spectrosc. **37**, 100 (2006).

[32] Q. Chen and S. T. Milner, Macromolecules **50**, 9773 (2017).

[33] P. Bornhauser and D. Bougeard, J. Raman Spectrosc. **32**, 279 (2001).

[34] A. Rahmani, J.-L. Sauvajol, J. Cambedouzou, and C. Benoit, Phys. Rev. B **71**, 125402 (2005).

[35] D. Bougeard and K. S. Smirnov, J. Raman Spectrosc. **40**, 1704 (2009).

[36] T. Hasegawa and Y. Tanimura, J. Phys. Chem. B **115**, 5545 (2011).

[37] G. R. Medders and F. Paesani, J. Chem. Theory Comput. **11**, 1145 (2015).

[38] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, Phys. Rev. Lett. **120**, 036002 (2018).

[39] Y. Zhang, S. Ye, J. Zhang, C. Hu, J. Jiang, and B. Jiang, J. Phys. Chem. B **124**, 7284 (2020).

[40] L. Zhang, M. Chen, X. Wu, H. Wang, W. E, and R. Car, Phys. Rev. B **102**, 041121 (2020).

[41] E. Berger, Z.-P. Lv, and H.-P. Komsa, J. Mater. Chem. C **11**, 1311 (2023).

[42] R. Beckmann, F. Brieuc, C. Schran, and D. Marx, J. Chem. Theory Comput. **18**, 5492 (2022).

[43] C. Feng, J. Xi, Y. Zhang, B. Jiang, and Y. Zhou, J. Chem. Theory Comput. **19**, 1207 (2023).

[44] C. Zhang, F. Tang, M. Chen, J. Xu, L. Zhang, D. Y. Qiu, J. P. Perdew, M. L. Klein, and X. Wu, J. Phys. Chem. B **125**, 11444 (2021).

[45] E. Berger and H.-P. Komsa, Polarizability Models for Simulations of Finite Temperature Raman Spectra from Machine Learning Molecular Dynamics (2023), arxiv:2310.13310 [cond-mat, physics:physics].

[46] Z. Fan, J. Phys. Condens. Matter **34**, 125902 (2022).

[47] Z. Fan, Y. Wang, P. Ying, K. Song, J. Wang, Y. Wang, Z. Zeng, K. Xu, E. Lindgren, J. M. Rahm, A. J. Gabourie, J. Liu, H. Dong, J. Wu, Y. Chen, Z. Zhong, J. Sun, P. Erhart, Y. Su, and T. Ala-Nissila, J. Chem. Phys. **157**, 114801 (2022).

[48] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[49] T. Schaul, T. Glasmachers, and J. Schmidhuber, in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11 (Association for Computing Machinery, New York, NY, USA, 2011) p. 845–852.

[50] D. A. McQuarrie, *Statistical Mechanics*, Harper's chemistry series (Harper Collins, New York, 1976).

[51] M. Cardona, *Light Scattering in Solids I*, Topics in Applied Physics, Vol. 8 (Springer, Berlin, Heidelberg, 1983).

[52] `TENSOAP` (accessed: 12-08-2023), `https://github.com/dilkins/TENSOAP`.

[53] `TENSOAP-FAST` (accessed: 12-08-2023), `https://github.com/dilkins/TENSOAP-FAST/tree/master/example/models/bulk-water`.

[54] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[55] E. Ditler, J. Mattiat, and S. Luber, Phys. Chem. Chem. Phys. **25**, 14672 (2023).

[56] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).

[57] Y. Yang, K. U. Lao, D. M. Wilkins, A. Grisafi, M. Ceriotti, and R. A. DiStasio, Sci. Data **6**, 152 (2019).

[58] in *Ideas of Quantum Chemistry*, edited by L. Piela (Elsevier, Amsterdam, 2007) pp. 615–680.

[59] N. A. Spaldin, J. Solid State Chem. **195**, 2 (2012).

[60] A. Krishnamoorthy, K.-i. Nomura, N. Baradwaj, K. Shimamura, P. Rajak, A. Mishra, S. Fukushima, F. Shimojo, R. Kalia, A. Nakano, and P. Vashishta, Phys. Rev. Lett.

126, 216403 (2021).

[61] D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, and M. Ceriotti, Proc. Natl. Acad. Sci. U.S.A. **116**, 3401 (2019).

[62] `NEP_CPU` (accessed: 12-08-2023), `https://github.com/brucefan1983/NEP_CPU`.

[63] CALORINE (accessed: 12-08-2023), `https://calorine.materialsmodeling.org/`; DOI: 10.5281/zenodo.7919206.

[64] M. Praprotnik and D. Janežič, J. Chem. Phys. **122**, 174103 (2005).

[65] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, J. Chem. Phys. **152**, 194103 (2020).

[66] H. D. Downing and D. Williams, J. Geophys. Res. **80**, 1656 (1975).

[67] J.-J. Max and C. Chapados, J. Chem. Phys. **131**, 184505 (2009).

[68] A. Rognoni, R. Conte, and M. Ceotto, Chem. Sci. **12**, 2060 (2021).

[69] M. Praprotnik, D. Janežič, and J. Mavri, J. Phys. Chem. A **108**, 11056 (2004).

[70] J. Sun, A. Ruzsinszky, and J. P. Perdew, Phys. Rev. Lett. **115**, 036402 (2015).

[71] J. Xu, M. Chen, C. Zhang, and X. Wu, Phys. Rev. B **99**, 205123 (2019).

[72] M. H. Brooker, G. Hancock, B. C. Rice, and J. Shapter, J. Raman Spectrosc. **20**, 683 (1989).

[73] T. Morawietz, O. Marsalek, S. R. Pattenaude, L. M. Streacker, D. Ben-Amotz, and T. E. Markland, J. Phys. Chem. Lett. **9**, 851 (2018).

[74] V. R. Chintareddy, K. Wadhwa, and J. G. Verkade, J. Org. Chem **76**, 4482 (2011).

[75] A. Thomas, J. George, A. Shalabney, M. Dryzhakov, S. J. Varma, J. Moran, T. Chervy, X. Zhong, E. Devaux, C. Genet, J. A. Hutchison, and T. W. Ebbesen, Angew. Chem. Int. Ed. **128**, 11634 (2016).

[76] C. Schäfer, J. Flick, E. Ronca, P. Narang, and A. Rubio, Nat. Commun. **13**, 7817 (2022).

[77] F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, J. Chem. Phys. **152**, 224108 (2020).

[78] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[79] F. Weigend and A. Baldes, J. Chem. Phys. **133** (2010).

[80] C. Toulouse, D. Amoroso, C. Xin, P. Veber, M. C. Hatnean, G. Balakrishnan, M. Maglione, P. Ghosez, J. Kreisel, and M. Guennou, Phys. Rev. B **100**, 134102 (2019).

[81] A. R. Akbarzadeh, I. Kornev, C. Malibert, L. Bellaiche, and J. M. Kiat, Physical Review B **72**, 205104 (2005).

[82] K. S. Knight, Journal of Materials Science **55**, 6417 (2020).

[83] A. Perrichon, E. Jedvik Granhed, G. Romanelli, A. Piovano, A. Lindman, P. Hyldgaard, G. Wahnström, and M. Karlsson, Chemistry of Materials **32**, 2824 (2020).

[84] P. Rosander, E. Fransson, C. Milesi-Brault, C. Toulouse, F. Bourdarot, A. Piovano, A. Bossak, M. Guennou, and G. Wahnström, Phys. Rev. B **108**, 014309 (2023).

[85] E. Fransson, P. Rosander, P. Erhart, and G. Wahnström, Chemistry of Materials **accepted**, 10.48550/arXiv.2310.05565 (2023).

[86] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).

[87] K. Berland and P. Hyldgaard, Physical Review B **89**, 035412 (2014).

Supporting Information:

Tensorial properties via the neuroevolution potential framework: Fast simulation of infrared and Raman spectra

Nan Xu[1,2], Petter Rosander[3], Christian Schäfer[3], Eric Lindgren[3], Nicklas Österbacka[3], Mandi Fang[1,2], Wei Chen[4], Yi He[1,2,5,*], Zheyong Fan[6,†], and Paul Erhart[3,‡]

[1] *Institute of Zhejiang University-Quzhou, Quzhou 324000, P. R. China*
[2] *College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, P. R. China*
[3] *Department of Physics, Chalmers University of Technology, SE-41296, Gothenburg, Sweden*
[4] *State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing, P. R. China*
[5] *Department of Chemical Engineering, University of Washington, Seattle, WA 98195, USA*
[6] *College of Physical Science and Technology, Bohai University, Jinzhou 121013, P. R. China*
[*] *yihezj@zju.edu.cn*
[†] *brucenju@gmail.com*
[‡] *erhart@chalmers.se*

December 8, 2023

Contents

S1 Dipole, polarizability, and susceptibility data for water systems 2

S2 Training of NEP PES model for water 2

S3 Training of SA-GPR models for water molecules 2

S4 Training of TNEP models for QM7b data set 2

S5 Training of TNEP dipole model for α-Fe_2O_3 2

S6 Training of TNEP susceptibility model for $BaZrO_3$ 3

S7 Note on units of polarizability and susceptibility 3

S8 Note on the Clausius-Mossotti relation 4

S1 Dipole, polarizability, and susceptibility data for water systems

Data for the dipole moment ($\boldsymbol{\mu}$), the polarizability ($\boldsymbol{\alpha}$), and the electronic part of the electric susceptibility ($\boldsymbol{\chi}_e$) for molecular water species as well as liquid water were retrieved from a public repository.[1,2] The $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ data for the molecules H_2O, $(H_2O)_2$, and $H_5O_2{}^+$ provided in the repository were calculated at the coupled cluster singles and doubles (CCSD) level of theory[3,4] with the d-aug-cc-pVTZ[5,6] basis set. The data for liquid water in the repository were generated using the Perdew–Burke–Ernzerhof (PBE)[7] functional and ultra-soft pseudo-potentials (USPPs).[8] The repository contains data for the relative dielectric permittivity $\boldsymbol{\varepsilon}_r$. Here, the latter was converted to the electric susceptibility $\boldsymbol{\chi}_e = \boldsymbol{\varepsilon}_r - \mathbf{1}$ for training.

S2 Training of NEP PES model for water

We used the potential energy surface (PES) data set for liquid water from Ref. 9,10 to train a neuroevolution potential (NEP) PES model to be used for molecular dynamics (MD) simulations. The data set contains 1888 structures in total, for which energy, forces, and virials have been computed using density functional theory (DFT) calculations. The strongly constrained and appropriately normed (SCAN)[11] functional and the projector augmented wave (PAW) method[12] (with hard setups) were used. All data were randomly divided into training and validation data sets with a ratio of 4:1. The hyperparameters used in the training of the NEP model are tabulated in Table S3. The root-mean-square errors (RMSEs) for energy, forces, and virials converged after 3×10^5 generations of training, and the predicted energies, forces, and virials closely match the DFT reference data, as shown in Fig. S8. We also performed an NPT (isothermal-isobaric) MD simulation of liquid water (64 molecules) using the NEP PES model at 330 K and 1 bar. The radial distribution functions (RDFs) for O–O and O–H pairs extracted from these simulations agree well with *ab initio* molecular dynamics (AIMD) simulation in the literature[13] (Fig. S9).

S3 Training of SA-GPR models for water molecules

Since we did not find publically accessible models for H_2O, $(H_2O)_2$ and $H_5O_2{}^+$, we trained new symmetry-adapted Gaussian process regression (SA-GPR) models for $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ using the same data sets as those used for training the tensorial neuroevolution potential (TNEP) models (Sect. S1). The default hyperparameters were used in the training. The RMSEs of $\boldsymbol{\mu}$ for the validation sets of H_2O, $(H_2O)_2$, and $H_5O_2{}^+$ were calculated to be 0.0001, 0.0242, and 0.0005 a.u., close to the values of 0.0004, 0.036, and 0.001 a.u. that we estimated from the literature.[2] This consistence indicates that the SA-GPR models trained here have achieved the accuracy of those reported in literature.

S4 Training of TNEP models for QM7b data set

We also consider the QM7b data set[14] that comprises 7211 small organic molecules with up to six elements (H, C, N, O, S, Cl). The reference $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ data[15] were calculated at the CCSD level of theory using the d-aug-cc-pVDZ basis set[5,6]. 70% of the data were use for training, while the remaining data were used for validation. The parameters in the training of the TNEP dipole and polarizability models are shown in Table S1 and Table S2. Both the dipole and polarizability models achieve very high precision for the QM7B data set (Fig. S2).

S5 Training of TNEP dipole model for α-Fe_2O_3

We also consider yet another crystalline system in α-Fe_2O_3. The primitive structure of α-Fe_2O_3 reported by Mendili *et al.*[16] was used here, which is defined by three vectors: $\boldsymbol{a} = (5.03, 0.00, 0.00)$, $\boldsymbol{b} = (-2.515, 4.356, 0.00)$, and $\boldsymbol{c} = (0.00, 0.00, 13.75)$, measured in units of Å. A $2 \times 2 \times 1$ supercell was created and used as the initial structure for AIMD simulations. The simulation system contained 120 atoms. The AIMD simulations were performed using a time step of 0.5 fs and the CP2K software package[17]. The forces on the atoms were evaluated using the DZVP-GTH-PADE basis set and GTH-PADE pseudopotentials[18,19]. To maintain the temperature at 300 K, a Nosé-Hoover thermostat was employed[20]. During the simulation, a magnetic moment of $5\,\mu_B$ was assigned to each iron atom, where μ_B represents the Bohr magneton. The simulation was run for a total of 2000 time steps, and snapshots were saved with an interval of 1 fs. In total, 1000 structures were sampled and used for post-processing. The $\boldsymbol{\mu}$ data of the 1000 structures were calculated using the Berry phase formulation,[21] where the origin of the coordinate system was used as the reference point. The calculated values vary continuously with the simulation time, indicating that all configurations are on the same branches of the Berry phase and no shifts are required for calculating $\boldsymbol{\mu}$ (Fig. S5a). The 1000 data points were randomly divided into a training data set and a testing data set with a ratio of 7:3. The hyperparameters used for training the TNEP dipole model are tabulated in Table S1. The $\boldsymbol{\mu}$ values predicted by the TNEP model are consistent with the DFT reference values (Fig. S5b). The coefficient of determination (R^2) is greater than 0.999.

S6 Training of TNEP susceptibility model for $BaZrO_3$

The training data points were taken from MD snapshots at various temperatures and pressures obtained using a previously published NEP model for the $BaZrO_3$ PES.[22] Both the cubic and tetragonal phases were included, with supercell sizes varying from the primitive cell to 4 × 2 × 1 repetitions. The final dataset consisted of 940 structures, for which the relative permittivity $\boldsymbol{\varepsilon}_r$ was obtained from DFT calculations using the PAW formalism[12,23] as implemented in the Vienna Ab-initio Simulation Package[24,25]. The van-der-Waals density functional with consistent exchange (vdW-DF-cx) was used to describe the exchange-correlation energy contribution[26,27]. A plane-wave energy cutoff of of 510 eV was used along with Gaussian smearing with a width of 0.01 eV. Projection operators were evaluated in reciprocal space and an additional support grid was used for evaluation of augmentation charges for increased accuracy.

The hyperparameters used for training the TNEP models for $BaZrO_3$ are tabulated in Table S2. The relative permittivity was converted to electric susceptibility via the relation $\boldsymbol{\chi}_e = \boldsymbol{\varepsilon}_r - \mathbf{1}$ for training. Five-fold leave-one-out cross validation was carried out, where the dataset was split into five equal parts. Five separate models were then trained, each using one of these splits as a test set. These models were trained until their test set RMSE stopped improving, which occurred after 6×10^5 generations. Their mean coefficient of determination (R^2) was 0.950(2) for the diagonal and 0.984(2) for the off-diagonal elements. These quantities serve as accuracy estimates for the final model, which was trained on the entire dataset. Training of this model was considered converged after 6×10^5 generations based on the cross validation result, after which the TNEP model predicted values for the susceptibility consistent with DFT reference values (Fig. S10).

S7 Note on units of polarizability and susceptibility

For a single molecule, the *molecular polarizability* $\boldsymbol{\alpha}$ connects the induced dipole moment $\boldsymbol{\mu}_{\text{ind}}$ (charge times distance per molecule) to the electric field $\boldsymbol{E}$ (potential per distance),

$$\boldsymbol{\mu}_{\text{ind}} = \boldsymbol{\alpha}\boldsymbol{E}. \tag{S1}$$

For a bulk material, the *electric susceptibility* $\boldsymbol{\chi}$ connects the polarization (or dipole density) $\boldsymbol{P}$ (charge times distance per volume) to the electric field $\boldsymbol{E}$ (potential per distance),

$$\boldsymbol{P} = \epsilon_0\boldsymbol{\chi}\boldsymbol{E}. \tag{S2}$$

It is now instructive to consider the different units involved in the expressions above. Here, we explicitly included the unit mol, which is strictly part of the SI system but often left out. This is done in order to emphasize the transition from molecular quantities ($\boldsymbol{\alpha}$) to bulk quantities ($\epsilon_0\boldsymbol{\chi}$). This occurs, for example, in the Clausius-Mossotti relation, where $\boldsymbol{\alpha}$ is related to $\epsilon_0\boldsymbol{\chi}$ by scaling with the number volume density ρ. Here, we see that $\boldsymbol{\alpha}$ has units of $\mathrm{F\,m^2/mol}$ whereas $\epsilon_0\boldsymbol{\chi}$ has units of $\mathrm{F/m}$ and thus units of $\boldsymbol{\alpha}$ per volume. The following table summarizes the SI units of the quantities as defined above.

Quantity	SI unit
$\boldsymbol{\mu}_{\mathrm{ind}}$	$\mathrm{C\,m/mol}$
$\boldsymbol{E}$	$\mathrm{V/m}$
$\boldsymbol{\alpha}$	$\mathrm{C\,m^2/(V\,mol)} = \mathrm{F\,m^2/mol}$
$\boldsymbol{P}$	$\mathrm{C/m^2}$
$\boldsymbol{\epsilon}_0$	$\mathrm{F/m}$
$\boldsymbol{\chi}$	1
ρ	$\mathrm{mol/m^3}$

We note that there are various ways in which the relations Eqs. (S1) and (S2) are written in the literature. One can for example subsume ϵ_0 into $\boldsymbol{\chi}$, which emphasizes the symmetry with $\boldsymbol{\alpha}$. One can also choose to express $\boldsymbol{\alpha}$ in units of ϵ_0, i.e., $\boldsymbol{\alpha} = \epsilon_0\boldsymbol{\alpha}'$, in which case $\boldsymbol{\alpha}'$ has units of volume.

S8 Note on the Clausius-Mossotti relation

For non-polar liquids or gases, the Clausius-Mossotti relation[28] can be used to approximate the relation between the average molecular polarizability $\alpha = \mathrm{Tr}(\boldsymbol{\alpha})/3$ and the average electric susceptibility $\chi = \mathrm{Tr}(\boldsymbol{\chi})/3$. It is based on a mean-field approximation to account for local field effect and is given by

$$\frac{\chi_e}{\chi_e + 3} = \frac{\rho\alpha}{3\epsilon_0},$$

where ρ is the number volume density. The quantities are expressed in SI units. The Clausius-Mossotti can also be written in terms of the relative permittivity ε_r, which is related to the electric susceptibility via $\varepsilon_r = 1 + \chi$.

Note that all quantities here are isotropic and therefore represented by scalars. Here, α is the molecular polarizability, ρ is the number volume density, ϵ_0 is the permittivity of vacuum, and ε_r is the relative permittivity.

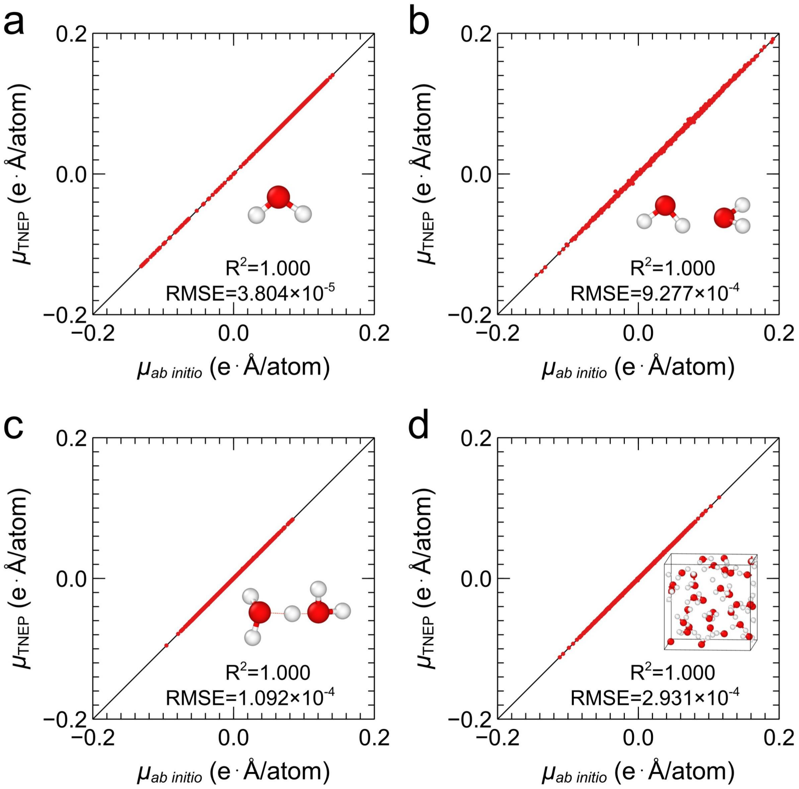

Figure S1: TNEP predicted dipole moment compared to *ab initio* data for the validation sets of (a) H_2O, (b) $(H_2O)_2$, (c) $H_5O_2{}^+$, and (d) liquid water.

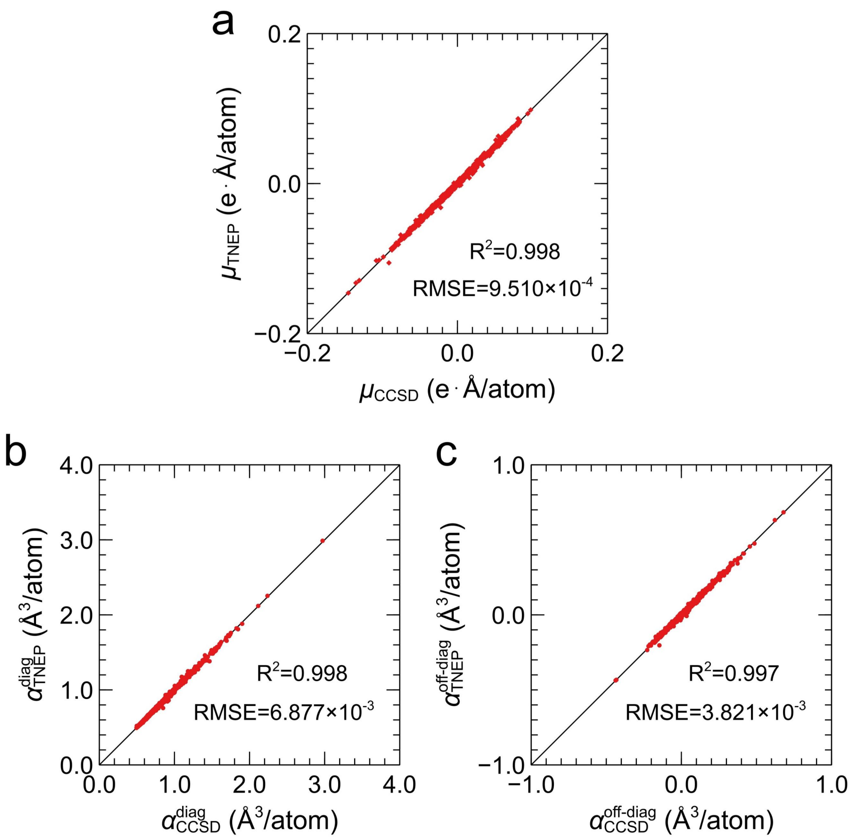

Figure S2: TNEP predictions as compared to CCSD reference data of (a) dipole moment, (b) diagonal elements of the polarizability, and (c) off-diagonal elements of the polarizability for the validation set of the QM7b set.

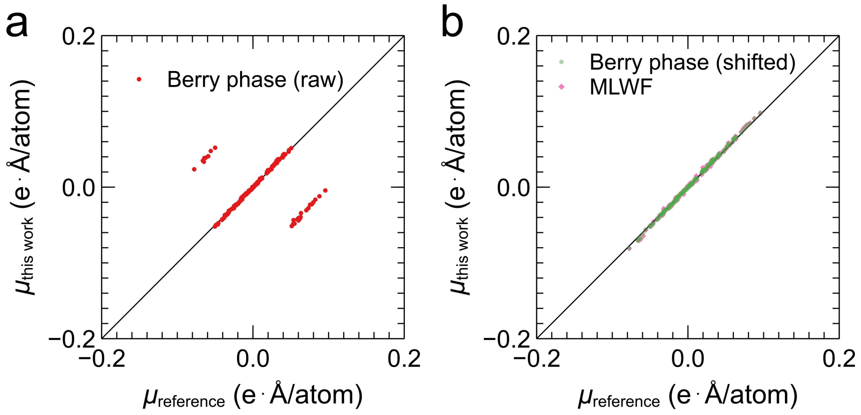

Figure S3: Dipole moments calculated in this work compared to reference data[1,2] for 50 liquid water structures that were evaluated in this work.

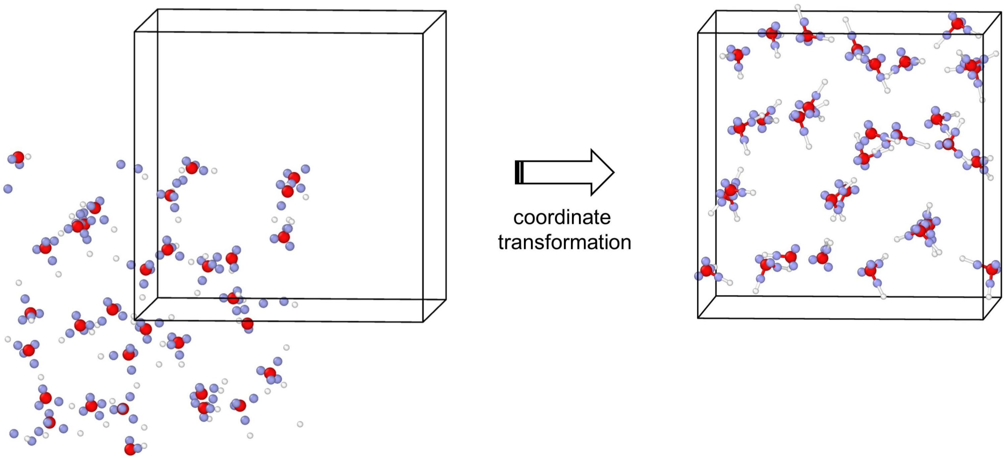

Figure S4: Schematic plot of coordinate transformation of water molecules and their Wannier centers before calculating the total dipole moment. Red, white, and slate-blue balls represent oxygen atoms, hydrogen atoms, and Wannier centers, respectively.

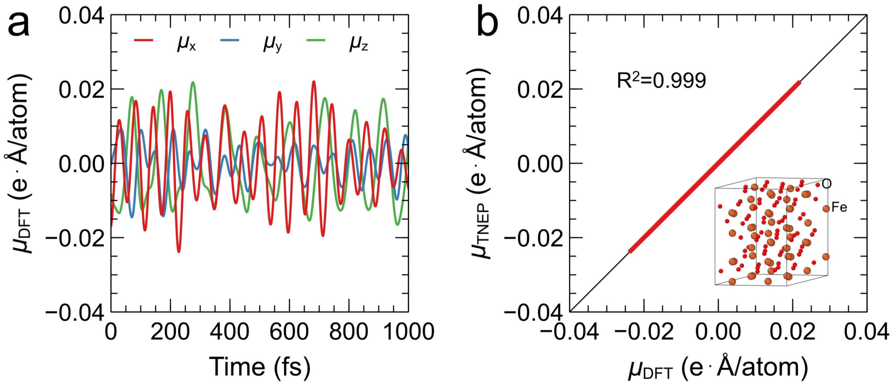

Figure S5: (a) Dipole moment of α-Fe_2O_3 calculated by DFT as a function of simulation time. (b) The comparison between the TNEP predictions and DFT values of dipole moment for the testing data set of α-Fe_2O_3.

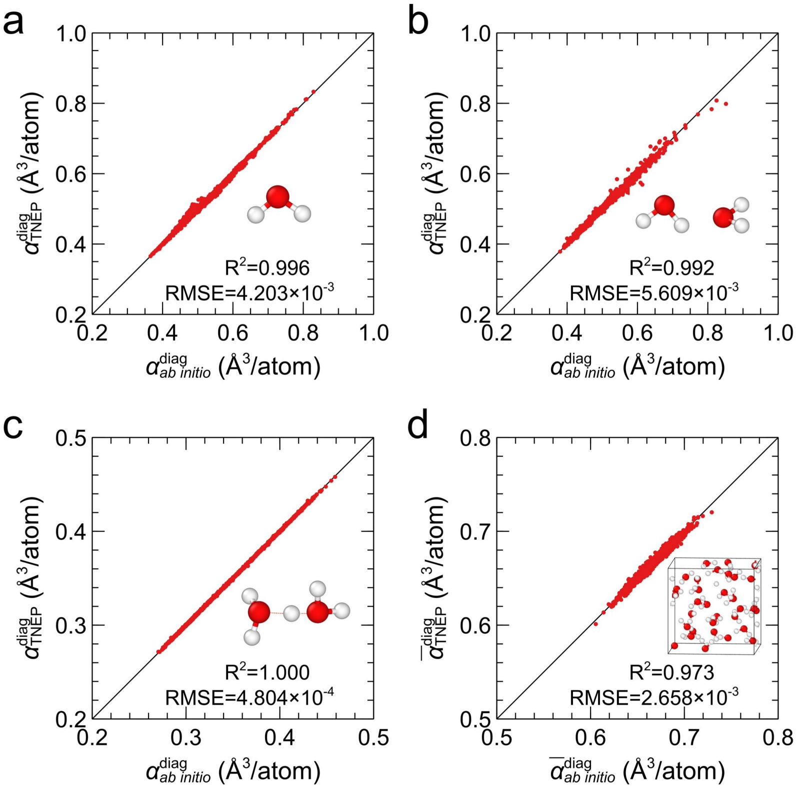

Figure S6: TNEP predicted diagonal polarizability as compared to *ab initio* data for the validation sets of (a) H_2O, (b) $(H_2O)_2$, (c) $H_5O_2{}^+$, and (d) liquid water. For liquid water we show the effective polarizability given by $\bar{\boldsymbol{\alpha}} = \boldsymbol{\chi}/\rho$, where ρ in the number volume density.

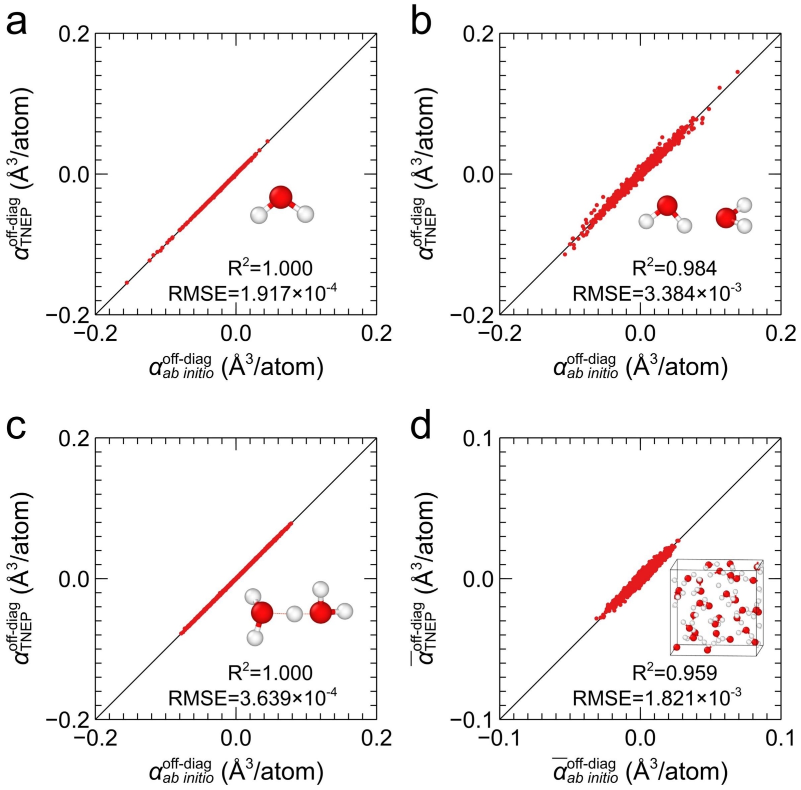

Figure S7: TNEP predicted off-diagonal polarizability compared to the *ab initio* values for the validation sets of (a) H_2O, (b) $(H_2O)_2$, (c) $H_5O_2{}^+$, and (d) liquid water. For liquid water we show the effective polarizability given by $\bar{\boldsymbol{\alpha}} = \boldsymbol{\chi}/\rho$, where ρ in the number volume density.

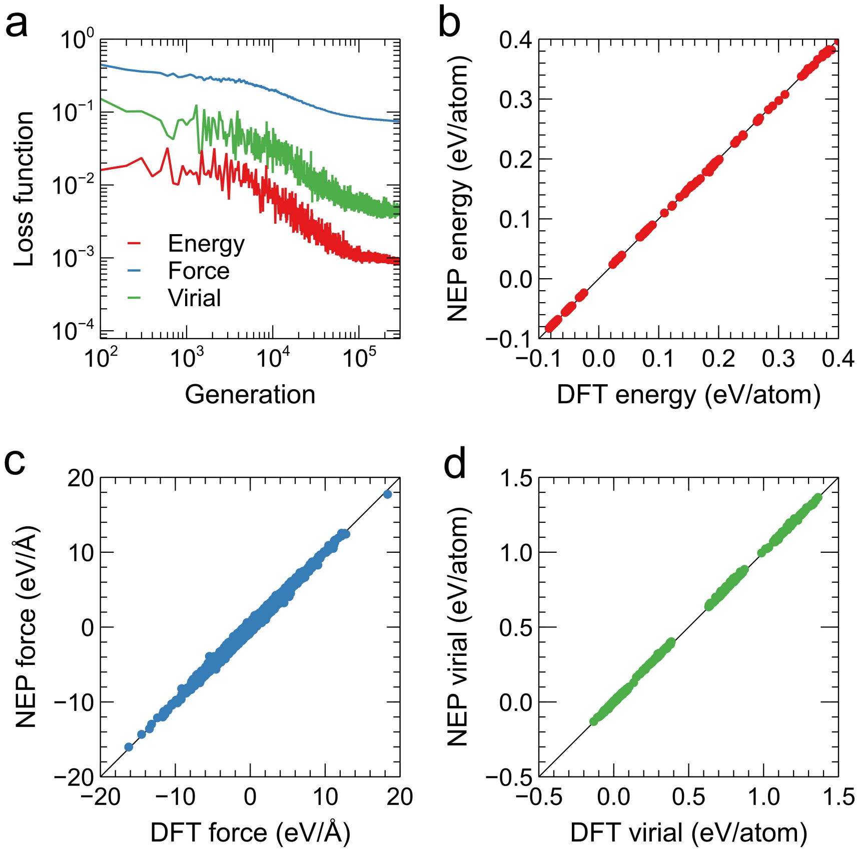

Figure S8: (a) RMSEs of energy, force, and virial for the validation set as a function of the number of generations. (b–d) Comparison between TNEP predictions and DFT reference values of energies, forces, and virials for the validation set of liquid water.

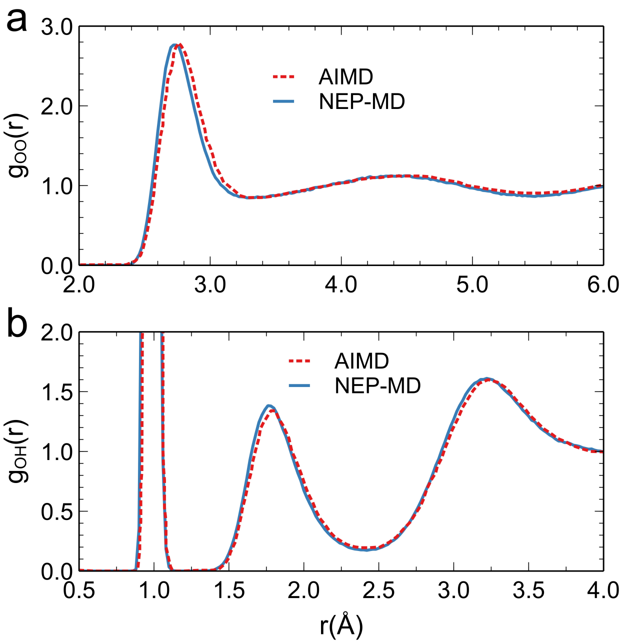

Figure S9: RDFs for (a) O–O and (b) O–H pairs extracted from MD simulations based on the NEP PES model constructed in this work and AIMD simulations[13] at 330 K and 1 bar.

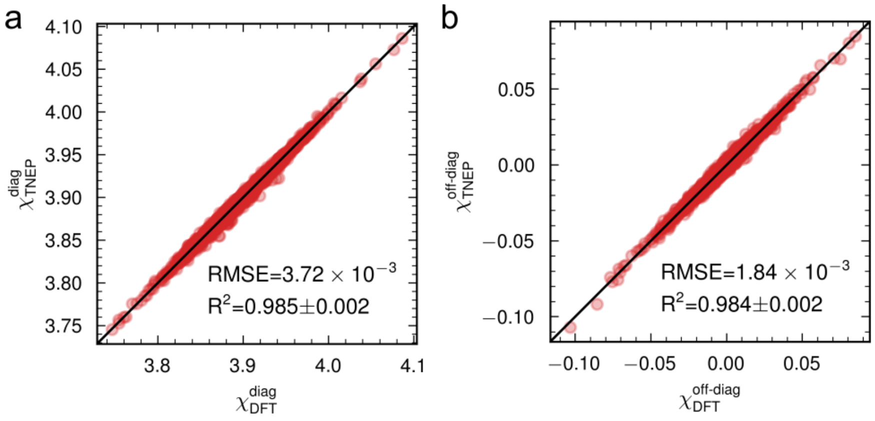

Figure S10: Comparison between TNEP predictions and DFT reference values for the (a) diagonal and (b) off-diagonal elements of the susceptibility $\boldsymbol{\chi}$ of $BaZrO_3$. The R^2 and RMSEs represent the means from five-fold hold-one-out cross-validation.

Table S1: Hyperparameters used in training of TNEP models for the dipole moment ($\boldsymbol{\mu}$) for H_2O, $(H_2O)_2$, $H_5O_2{}^+$, liquid water, organic molecules (QM7B set), and α-Fe_2O_3. Here, r_c^{R} (r_c^{A}) is the cutoff radius for the radial (angular) components of the descriptor, $n_{\mathrm{max}}^{\mathrm{R}}$ ($n_{\mathrm{max}}^{\mathrm{A}}$) is the Chebyshev polynomial expansion order for the radial (angular) components, l_{max}^{3b} (l_{max}^{4b}, l_{max}^{5b}) is the Legendre polynomial expansion order for the three-body (four-body, five-body) terms angular components, $N_{\mathrm{bas}}^{\mathrm{R}}$ ($N_{\mathrm{bas}}^{\mathrm{A}}$) is the number of basis functions that are used to build the radial (angular) descriptor functions, N_{neu} is the number of neurons in the hidden layer of the neural network, λ_1 (λ_2) is the $\mathcal{L}_1$ ($\mathcal{L}_2$) regularization parameter, N_{pop} is the population size in the natural evolution strategy algorithm, N_{bat} is the size of each batch used during training, and N_{gen} is the maximum number of generations to be evolved.

Parameter	H_2O	$(H_2O)_2$	$H_5O_2{}^+$	liquid water	QM7B set	α-Fe_2O_3
r_c^{R} (Å)	6	6	6	6	6	6
r_c^{A} (Å)	4	4	4	4	4	4
$n_{\mathrm{max}}^{\mathrm{R}}$	6	6	6	6	6	6
$n_{\mathrm{max}}^{\mathrm{A}}$	6	6	6	6	6	6
l_{max}^{3b}	4	4	4	4	4	4
l_{max}^{4b}	2	2	2	2	2	2
l_{max}^{5b}	1	1	1	1	1	1
$N_{\mathrm{bas}}^{\mathrm{R}}$	10	10	10	10	10	10
$N_{\mathrm{bas}}^{\mathrm{A}}$	10	10	10	10	10	10
N_{neu}	10	10	10	10	30	10
λ_1	0.00005	0.0008	0.0012	0.0005	0.001	0.0001
λ_2	0.00005	0.0008	0.0012	0.0005	0.001	0.0001
N_{batch}	full-batch	full-batch	full-batch	full-batch	full-batch	full-batch
N_{pop}	80	80	80	80	80	80
N_{gen}	2×10^5	2×10^5	2×10^5	2×10^5	4×10^5	2×10^5

Table S2: Hyperparameters used in training of TNEP polarizability and susceptibility models for H_2O, $(H_2O)_2$, $H_5O_2{}^+$, liquid water, and $BaZrO_3$. Compared to the TNEP dipole model, an additional parameter λ_s should be set, which stands for the relative weight between the off-diagonal elements and diagonal elements of rank-2 tensors in the construction of loss functions.

Parameter	H_2O	$(H_2O)_2$	$H_5O_2{}^+$	liquid water	QM7B set	$BaZrO_3$
r_c^{R} (Å)	6	6	6	6	6	6
r_c^{A} (Å)	4	4	4	4	4	4
$n_{\mathrm{max}}^{\mathrm{R}}$	6	6	6	6	6	4
$n_{\mathrm{max}}^{\mathrm{A}}$	6	6	6	6	6	4
l_{max}^{3b}	4	4	4	4	4	4
l_{max}^{4b}	2	2	2	2	2	0
l_{max}^{5b}	1	1	1	1	1	0
$N_{\mathrm{bas}}^{\mathrm{R}}$	10	10	10	10	10	12
$N_{\mathrm{bas}}^{\mathrm{A}}$	10	10	10	10	10	12
N_{neu}	10	10	10	10	30	20
λ_1	0.008	0.02	0.002	0.001	0.03	-1 (adaptive adjustment)
λ_2	0.008	0.02	0.002	0.001	0.03	-1 (adaptive adjustment)
N_{batch}	full-batch	full-batch	full-batch	full-batch	full-batch	full-batch
N_{pop}	80	80	80	80	80	50
N_{gen}	2×10^5	2×10^5	2×10^5	2×10^5	4×10^5	6×10^5
λ_s	10	1	1	1	1	1

Table S3: Hyperparameters used in training a NEP PES model for MD simulations of water.

Parameter	Liquid water
r_c^{R} (Å)	6
r_c^{A} (Å)	4
$n_{\mathrm{max}}^{\mathrm{R}}$	9
$n_{\mathrm{max}}^{\mathrm{A}}$	7
l_{max}^{3b}	4
l_{max}^{4b}	2
l_{max}^{5b}	0
$N_{\mathrm{bas}}^{\mathrm{R}}$	9
$N_{\mathrm{bas}}^{\mathrm{A}}$	7
N_{neu}	100
λ_1	-1 (adaptive adjustment)
λ_2	-1 (adaptive adjustment)
N_{batch}	750
N_{pop}	50
N_{gen}	3×10^5

Table S4: Hyperparameters used in training a NEP models for the prediction of infrared spectra for PTAF$^-$.

Parameter	PTAF$^-$ PES	PTAF$^-$ $\boldsymbol{\mu}$
r_c^{R} (Å)	8	8
r_c^{A} (Å)	4	6
$n_{\max}^{\mathrm{R}}$	8	15
$n_{\max}^{\mathrm{A}}$	6	8
$l_{\max}^{3b}$	4	4
$l_{\max}^{4b}$	0	2
$l_{\max}^{5b}$	0	0
$N_{\mathrm{bas}}^{\mathrm{R}}$	8	12
$N_{\mathrm{bas}}^{\mathrm{A}}$	8	12
N_{neu}	40	80
λ_1	0.1	-1 (adaptive adjustment)
λ_2	0.1	-1 (adaptive adjustment)
λ_e	1	1
λ_f	3	1
λ_v	0	0.1
N_{batch}	1×10^5	5×10^5
N_{pop}	50	50
N_{gen}	2×10^5	5×10^5

Supplemental References

[1] URL `https://github.com/dilkins/TENSOAP`.

[2] Andrea Grisafi, David M. Wilkins, Gábor Csányi, and Michele Ceriotti. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.*, 120(3):036002, 2018. doi: 10.1103/PhysRevLett.120.036002.

[3] Hendrik J. Monkhorst. Calculation of properties with the coupled-cluster method. *Int. J. Quantum Chem.*, 12(S11):421–432, 1977. doi: https://doi.org/10.1002/qua.560120850. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560120850`.

[4] Henrik Koch and Poul Jørgensen. Coupled cluster response functions. *J. Chem. Phys.*, 93(5): 3333–3344, 1990. doi: 10.1063/1.458814. URL `https://doi.org/10.1063/1.458814`.

[5] David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. iv. calculation of static electrical response properties. *J. Chem. Phys.*, 100(4):2975–2988, 1994. doi: 10.1063/1.466439. URL `https://doi.org/10.1063/1.466439`.

[6] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989. ISSN 0021-9606. doi: 10.1063/1.456153. URL `https://aip.scitation.org/doi/10.1063/1.456153`.

[7] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996. doi: 10.1103/PhysRevLett.77.3865. URL `https://link.aps.org/doi/10.1103/PhysRevLett.77.3865`.

[8] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41:7892–7895, Apr 1990. doi: 10.1103/PhysRevB.41.7892. URL `https://link.aps.org/doi/10.1103/PhysRevB.41.7892`.

[9] Ke Xu, Yongchao Hao, Ting Liang, Penghua Ying, Jianbin Xu, Jianyang Wu, and Zheyong Fan. Accurate prediction of heat conductivity of water by a neuroevolution potential. *J. Chem. Phys.*, 158(20), 05 2023. ISSN 0021-9606. doi: 10.1063/5.0147039. URL `https://doi.org/10.1063/5.0147039`.

[10] Linfeng Zhang, Han Wang, Roberto Car, and Weinan E. Phase diagram of a deep potential water model. *Phys. Rev. Lett.*, 126:236001, Jun 2021. doi: 10.1103/PhysRevLett.126.236001. URL `https://link.aps.org/doi/10.1103/PhysRevLett.126.236001`.

[11] Kerwin Hui and Jeng-Da Chai. Scan-based hybrid and double-hybrid density functionals from models without fitted parameters. *J. Chem. Phys.*, 144(4):044114, 2016. doi: 10.1063/1.4940734. URL `https://doi.org/10.1063/1.4940734`.

[12] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, Dec 1994. doi: 10.1103/PhysRevB.50.17953. URL `https://link.aps.org/doi/10.1103/PhysRevB.50.17953`.

[13] Mohan Chen, Hsin-Yu Ko, Richard C. Remsing, Marcos F. Calegari Andrade, Biswajit Santra, Zhaoru Sun, Annabella Selloni, Roberto Car, Michael L. Klein, John P. Perdew, and Xifan Wu. Ab initio theory and modeling of water. *Proc. Natl. Acad. Sci. U.S.A.*, 114(41):10846–10851, 2017. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1712499114`.

[14] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 108:058301, Jan 2012. doi: 10.1103/PhysRevLett.108.058301. URL `https://link.aps.org/doi/10.1103/PhysRevLett.108.058301`.

[15] Yang Yang, Ka Un Lao, David M. Wilkins, Andrea Grisafi, Michele Ceriotti, and Robert A. DiStasio. Quantum mechanical static dipole polarizabilities in the qm7b and alphaml showcase databases. *Sci. Data*, 6(1):152, 2019. doi: 10.1038/s41597-019-0157-8. URL `https://doi.org/10.1038/s41597-019-0157-8`.

[16] Yassine El Mendili, Jean-François Bardeau, Nirina Randrianantoandro, Fabien Grasset, and Jean-Marc Greneche. Insights into the mechanism related to the phase transition from γ-Fe_2O_3 to α-Fe_2O_3 nanoparticles induced by thermal treatment and laser irradiation. *J. Phys. Chem. C*, 116(44):23785–23792, 2012.

[17] Thomas D. Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V. Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z. Khaliullin, Ole Schütt, Florian Schiffmann, Dorothea Golze, Jan Wilhelm, Sergey Chulkov, Mohammad Hossein Bani-Hashemian, Valéry Weber, Urban Borštnik, Mathieu Taillefumier, Alice Shoshana Jakobovits, Alfio Lazzaro, Hans Pabst, Tiziano Müller, Robert Schade, Manuel Guidon, Samuel Andermatt, Nico Holmberg, Gregory K. Schenter, Anna Hehn, Augustin Bussy, Fabian Belleflamme, Gloria Tabacchi, Andreas Glöß, Michael Lass, Iain Bethune, Christopher J. Mundy, Christian Plessl, Matt Watkins, Joost VandeVondele, Matthias Krack, and Jürg Hutter. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.*, 152(19), 05 2020. ISSN 0021-9606. doi: 10.1063/5.0007045. URL `https://doi.org/10.1063/5.0007045`. 194103.

[18] By Gerald Lippert, Jurg Hutter Parrinello, and Michele. A hybrid gaussian and plane wave density functional scheme. *Mol. Phys.*, 92(3):477–488, 1997.

[19] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103–128, 2005. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2004.12.014. URL `https://www.sciencedirect.com/science/article/pii/S0010465505000615`.

[20] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97(4):2635–2643, 08 1992. ISSN 0021-9606. doi: 10.1063/1.463940. URL `https://doi.org/10.1063/1.463940`.

[21] Nicola A. Spaldin. A beginner's guide to the modern theory of polarization. *J. Solid State Chem.*, 195:2–10, 2012. ISSN 0022-4596. doi: https://doi.org/10.1016/j.jssc.2012.05.010. URL `https://www.sciencedirect.com/science/article/pii/S0022459612003234`.

[22] Erik Fransson, Petter Rosander, Paul Erhart, and Göran Wahnström. Understanding correlations in $BaZrO_3$: Structure and dynamics on the nano-scale. *Chemistry of Materials*, accepted, 7 2023. doi: 10.48550/arXiv.2310.05565.

[23] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758–1775, 1999. doi: 10.1103/PhysRevB.59.1758.

[24] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47: 558–561, Jan 1993. doi: 10.1103/PhysRevB.47.558.

[25] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 1996. doi: 10.1016/0927-0256(96)00008-0.

[26] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der waals density functional for general geometries. *Phys. Rev. Lett.*, 92:246401, 2004. doi: 10.1103/PhysRevLett.92.246401.

[27] Kristian Berland and Per Hyldgaard. Exchange functional that tests the robustness of the plasmon description of the van der waals density functional. *Physical Review B*, 89:035412, Jan 2014. doi: 10.1103/PhysRevB.89.035412.

[28] Kaoru Urano and Masayoshi Inoue. Clausius–Mossotti formula for anisotropic dielectrics. *J. Chem. Phys*, 66(2):791–794, 08 2008. doi: 10.1063/1.433957.