Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems

Andrea Grisafi,[1] David M. Wilkins,[1] Gábor Csányi,[2] and Michele Ceriotti[1]

[1]*Laboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*
[2]*Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge CB21PZ, United Kingdom*

Statistical learning methods show great promise in providing an accurate prediction of materials and molecular properties, while minimizing the need for computationally demanding electronic structure calculations. The accuracy and transferability of these models are increased significantly by encoding into the learning procedure the fundamental symmetries of rotational and permutational invariance of scalar properties. However, the prediction of tensorial properties requires that the model respects the appropriate geometric transformations, rather than invariance, when the reference frame is rotated. We introduce a formalism that extends existing schemes and makes it possible to perform machine learning of tensorial properties of arbitrary rank, and for general molecular geometries. To demonstrate it, we derive a tensor kernel adapted to rotational symmetry, which is the natural generalization of the smooth overlap of atomic positions kernel commonly used for the prediction of scalar properties at the atomic scale. The performance and generality of the approach is demonstrated by learning the instantaneous response to an external electric field of water oligomers of increasing complexity, from the isolated molecule to the condensed phase.

The past few years have seen a surge in applications of statistical learning approaches to the prediction of the properties of molecules and materials. Chemical and materials informatics approaches—in which large databases are mined to find correlations between structure and macroscopic properties—have become ubiquitous [1–7]. Furthermore, "machine-learning (ML) potentials" are increasingly used as surrogate models for demanding electronic structure calculations, and to obtain information on the stability and properties of a material as a function of the microscopic arrangement of its atoms [8–12]. For these approaches to be effective, it is crucial that the statistical learning algorithm and the mathematical representation of the atomic configurations respect the fundamental symmetries of the problem. For example, scalar properties should be invariant under rigid translations, rotations, or reflections of the atomic configurations, as well as permutations of the order of identical atoms. Methods that fulfill these requirements have demonstrated very promising performance for predicting scalar quantities such as electronic ground-state energies [8,13–16].

A complete description of molecular and condensed-phase systems, however, also requires the prediction of properties that are not scalars. The response of a material to mechanical, magnetic, or electric perturbations all require response coefficients that are tensorial in nature. The electrical response moments—the dipole moment $\boldsymbol{\mu}$, polarizability $\boldsymbol{\alpha}$, first hyperpolarizability $\boldsymbol{\beta}$, etc.—underlie in particular the modeling of experiments such as infrared [17], Raman [18–20], and second-harmonic spectroscopy [21–23]. No less importantly, they represent a fundamental ingredient to include many-body effects in atomistic simulations of a material through the development of polarizable force fields [24–29].

Gaussian process regression (GPR) is a commonly used machine-learning technique, which is formally equivalent to kernel-ridge regression [30,31], and is built upon the definition of a kernel function $k(\mathcal{X}, \mathcal{X}')$ that encodes the similarity between two configurations $\mathcal{X}$ and $\mathcal{X}'$ [31–33]. In order to guarantee that predicted properties respect the relevant physical symmetries, the kernel function must obey corresponding transformation rules. For instance, when predicting a scalar, $k(\mathcal{X}, \mathcal{X}')$ should be invariant to rotations of the two configurations. The extension to tensorial quantities is not straightforward. As discussed recently for the case of the learning of vectorial properties such as forces [34], the regression framework must be designed so that the predicted properties are covariant with respect to symmetry operations applied to the system. Under certain conditions, suitable strategies can be used to bypass the problem: for example, in the presence of relatively rigid molecular units (e.g., in water) it is possible to define a local reference frame, so that response tensors can be learned by comparing mutually aligned molecules [35,36]. However, this approach is not generally applicable to flexible or dissociable molecular systems. A learning algorithm that handles symmetries in a more general and mathematically rigorous fashion is required.

In this Letter, we introduce a GPR framework that explicitly includes the rotational symmetry of tensorial properties of arbitrary order, generalizing an earlier framework designed for the kernel-ridge regression of forces [34], and can treat molecular or condensed-phase systems of arbitrary complexity. As a practical implementation, we

define a family of kernels that are based on the smooth overlap of atomic positions (SOAP) kernels of Ref. [13], which we modify to account for the covariance of the tensorial property.

The objective of any regression framework is to predict a property y for a configuration $\mathcal{X}$, based on a set of reference inputs $\{\mathcal{X}_I\}$ for which the property values y_I have been determined already. In the case of GPR [33], the prediction is written as a linear combination of kernel functions $k(\mathcal{X}, \mathcal{X}')$ that are used to quantify the similarity between the trial configuration and the references:

$$y(\mathcal{X}) = \sum_I w_I k(\mathcal{X}, \mathcal{X}_I). \tag{1}$$

The weights can be determined by solving a linear problem $\mathbf{w} = (\mathbf{K} + \eta^2 \mathbf{1})^{-1} \mathbf{y}$, where $K_{IJ} = k(\mathcal{X}_I, \mathcal{X}_J)$ and $\mathbf{y}$ contains the values of the target property for the $\{\mathcal{X}_I\}$. The regularization η can be interpreted as the expected error of the fit, due both to any intrinsic noise in the target data and the limitations of the model representation.

Consider now the case of a tensorial property $\mathbf{T}$. We will label the components of the tensor using a compact notation T_μ, where μ indicates, for example, a set of Cartesian axes $\mu \equiv (\alpha\beta\ldots)$. Within a Bayesian interpretation, the kernel k represents a measure of correlations between the values of the tensorial property associated with the configurations $(\mathcal{X}, \mathcal{X}')$. In particular, we can write

$$k_{\mu\nu}(\mathcal{X}, \mathcal{X}') \equiv \langle T_\mu(\mathcal{X}); T_\nu^\dagger(\mathcal{X}') \rangle, \tag{2}$$

where $\langle A; B \rangle$ indicates the covariance between A and B. In this formalism the learning algorithm is expected to simultaneously take into account all the components of $\mathbf{T}$. Equation (2) represents a block of a full kernel matrix, which can be built by merging the portions associated with each pair of configurations. The complete matrix is Hermitian, so that for each block $k_{\mu\nu}(\mathcal{X}, \mathcal{X}') = k_{\nu\mu}^*(\mathcal{X}', \mathcal{X})$.

When a generalized symmetry operation $\hat{S}$ is applied to one configuration $\mathcal{X}$ of the system, the corresponding tensorial property transforms as $T_\mu(\hat{S}\mathcal{X}) = \sum_{\mu'} S_{\mu\mu'} T_{\mu'}(\mathcal{X})$. Given two independent symmetry operations $\hat{S}$ and $\hat{S}'$ acting on the two configurations, it follows from Eq. (2) that $\mathbf{k}$ must satisfy the following transformation rule:

$$k_{\mu\nu}(\hat{S}\mathcal{X}, \hat{S}'\mathcal{X}') = \sum_{\mu'\nu'} S_{\mu\mu'} S'_{\nu\nu'} k_{\mu'\nu'}(\mathcal{X}, \mathcal{X}'). \tag{3}$$

This is the generalization of the covariance conditions introduced in Ref. [34] for the special case of learning vectors. Similarly to that case, one can then verify that a kernel that satisfies Eq. (3) can be obtained starting from a scalar kernel $\kappa(\mathcal{X}, \mathcal{X}')$, by averaging over the matrix that represents the symmetry operation $\hat{S}$:

$$k_{\mu\nu}(\mathcal{X}, \mathcal{X}') = \int d\hat{S} S_{\mu\nu} \kappa(\mathcal{X}, \hat{S}\mathcal{X}'). \tag{4}$$

The scalar kernel κ only needs to be independent of the absolute reference frame, i.e., $\kappa(\hat{S}\mathcal{X}, \hat{S}\mathcal{X}') = \kappa(\mathcal{X}, \mathcal{X}')$, but not of the relative orientation of the two configurations.

In the case of a Cartesian tensor $T_{\alpha\beta\ldots}$ of rank r, a full hierarchy of Cartesian kernels can be built by combining r orthogonal rotation matrices, i.e., $S_{(\alpha\beta\ldots)(\alpha'\beta'\ldots)} = R_{\alpha\alpha'} R_{\beta\beta'} \cdots$ in Eq. (4), generating a kernel with blocks of size $3^r \times 3^r$. However, this strategy is unnecessarily complicated. The actual dimensionality of the problem can be significantly reduced by a unitary transformation that transforms the tensor into a block-diagonal form, its irreducible spherical tensor (IST) representation $\mathbf{T}$: $\{\mathbf{T}^\lambda\}$. Each λ identifies an orthogonal subspace of dimension $2\lambda + 1$, according to SO(3) algebra [37]. Depending on the rank and the symmetries of the tensor, the decomposition contains a different number of elements, which in any case correspond to diagonal blocks of size smaller than $2r + 1$. Performing a decomposition into the IST components makes the statistical learning faster and more transparent, since each tensorial component $\mathbf{T}^\lambda$ can now be independently learned as a vector of dimension $2\lambda + 1$. What is more, in the spherical basis, the covariance conditions of Eqs. (3) and (4) can be reformulated by using the fact that each spherical component $\mathbf{T}^\lambda$ of a covariant tensor follows the same transformation rules as the corresponding vector spherical harmonics $\mathbf{Y}^\lambda$ [37]. It follows that if the kernel is required to encode rotational symmetry in three dimensions, the generalized transformation matrix $S_{\mu\nu}$ of Eq. (4) is given by the Wigner matrix $\mathbf{D}^\lambda$ associated with the active rotations $\hat{R}$ of the system [38].

As a practical implementation of Eq. (4), we consider the case where $\kappa(\mathcal{X}, \mathcal{X}')$ is given by the overlap between Gaussian smoothed atom densities,

$$\kappa(\mathcal{X}, \mathcal{X}') = \left| \int \rho(\mathbf{r})\rho'(\mathbf{r}) d\mathbf{r} \right|^2, \tag{5}$$

where $\rho(\mathbf{r}) = \sum_{\mathbf{x} \in \mathcal{X}} g_\sigma(\mathbf{r} - \mathbf{x})$ is a sum over the atoms making up the environment $\mathcal{X}$ and $g_\sigma(\mathbf{r} - \mathbf{x})$ is a Gaussian of width σ centered on $\mathbf{x}$. The range of the kernel can be tuned by introducing a cutoff function that zeros out the contribution from atoms that lie farther than a given distance r_c from the central atom. With this choice of $\kappa(\mathcal{X}, \mathcal{X}')$, the matrix kernel $\mathbf{k}^\lambda(\mathcal{X}, \mathcal{X}')$ associated with a given IST component is

$$\mathbf{k}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} \mathbf{D}^\lambda(\hat{R}) \left| \int \rho(\mathbf{r})\rho'(\hat{R}\mathbf{r}) d\mathbf{r} \right|^2. \tag{6}$$

As shown in the Supplemental Material (SM) [39], when an angular decomposition of the atom-centered Gaussian densities is applied [49], this integration can be performed
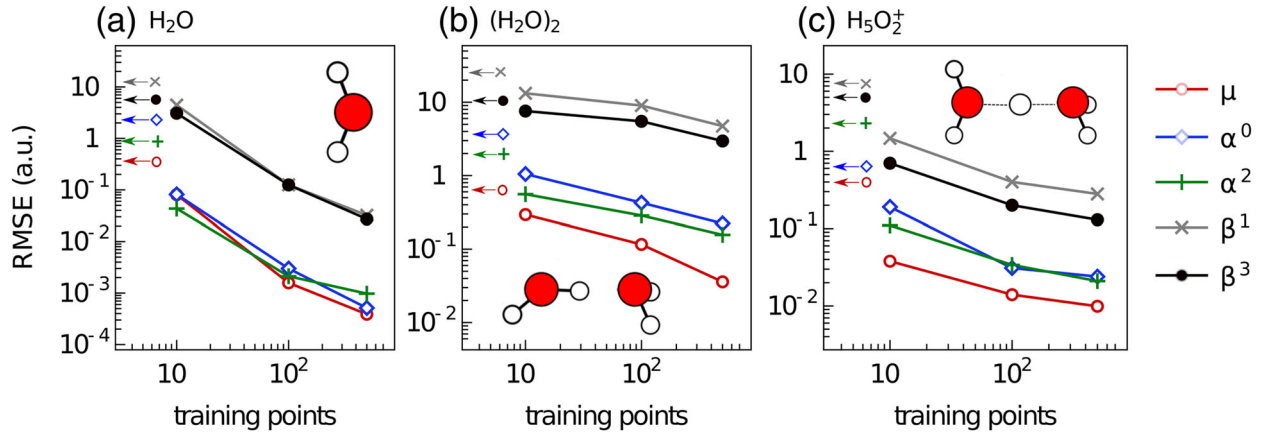
FIG. 1. Learning curves of the IST components of dipole $\boldsymbol{\mu}$ ($\lambda = 1$), polarizability $\boldsymbol{\alpha}$ ($\lambda = 0, 2$), and hyperpolarizability $\boldsymbol{\beta}$ ($\lambda = 1, 3$) for water monomer (left), water dimer (center), and Zundel cation (right). For all cases the testing data set consists of 500 independent configurations. Arrows with symbols indicate the intrinsic standard deviation of the testing data set. The kernel has been computed with an environment cutoff of 4 Å for the monomer and $H_5O_2^+$, and 5 Å for the water dimer.

analytically. The $\lambda = 0$ case recovers the scalar SOAP kernel of Ref. [13], which has been demonstrated to be very effective for the statistical learning of scalar properties of materials and molecules [50–55]. As detailed in the SM [39], such a "λ-SOAP" hierarchy of tensorial kernels can be recast as an inner product of $(2\lambda + 1)$-size vectors $\boldsymbol{P}_{nn'll'}^\lambda$:

$$k_{\mu\nu}^\lambda(\mathcal{X}, \mathcal{X}') = \sum_{nn'll'} P_{nn'll'}^{\lambda\mu}(\mathcal{X}) P_{nn'll'}^{\lambda\nu\star}(\mathcal{X}'), \qquad (7)$$

where the contraction indices n, n' and l, l' run, respectively, over the basis sets of the radial and the angular expansion of atomic densities [56]. Each $\boldsymbol{P}_{nn'll'}^\lambda$ represents a symmetry-adapted fingerprint associated with the individual configurations $\mathcal{X}$, generalizing the SOAP power spectrum of Ref. [13].

The advantage of this formulation, which builds on a mathematically rigorous treatment of SO(3) group symmetry, is that it can be applied seamlessly to molecules as well as to systems undergoing chemical reactions or to condensed phases of matter. Discrete symmetries can also be included straightforwardly, as they correspond to a discrete-sum equivalent of Eq. (4). For instance, an O(3) kernel with inversion symmetry can be computed as

$$\boldsymbol{k}_{\mathrm{O}(3)}^\lambda(\mathcal{X}, \mathcal{X}') = \frac{1}{2}[\boldsymbol{k}^\lambda(\mathcal{X}, \mathcal{X}') + (-1)^\lambda \boldsymbol{k}^\lambda(\mathcal{X}, \hat{i}\mathcal{X}')], \quad (8)$$

where $\hat{i}$ denotes inversion and $\boldsymbol{k}^\lambda$ is an SO(3) kernel. [57].

As a demonstration of the general applicability of the framework, we now show the performance of our symmetry-adapted GPR algorithm (SA-GPR) in predicting the static polarizability series of neutral and charged water oligomers, as well as the instantaneous dielectric response tensor of liquid water configurations. Details on the training sets and the kernel hyperparameters used in each case are provided in the SM [39]. As a first example, we consider the polarizability series of flexible and arbitrarily oriented

water molecules in vacuum. The dipole moment $\boldsymbol{\mu}$, polarizability $\boldsymbol{\alpha}$, and first hyperpolarizability $\boldsymbol{\beta}$ were computed with high-end quantum chemical methods for 1000 configurations. Because of the symmetry with respect to permutations of Cartesian indices—which is implied by the definition of response tensors as the derivatives of the electronic energy with an applied electric field—$\boldsymbol{\alpha}$ corresponds to an irreducible representation involving the $\lambda = 0$ and $\lambda = 2$ spherical components only, while $\boldsymbol{\beta}$ has an IST decomposition containing $\lambda = 1$ and $\lambda = 3$. Figure 1(a) shows the learning curves (i.e., the test error as a function of the number of training structures included) for all the IST components. Without explicitly using information on the orientation of water molecules, the SA-GPR framework can easily achieve a root mean square error (RMSE) below 5% for all components with only 100 training points.

A natural approach to extend the λ-SOAP framework to complex molecules, and eventually to condensed phases, involves decomposing the overall properties of the system into atom-centered components. It is straightforward to see that an atom-centered decomposition is equivalent to the learning of the system's properties using a single kernel that is built by breaking down each configuration into multiple environments, and defining the kernel to be used in the GPR as the sum of all possible local similarities between two configurations [55],

$$\boldsymbol{K}^\lambda(\mathcal{X}, \mathcal{X}') = \frac{1}{NN'} \sum_{i=1}^N \sum_{j=1}^{N'} \boldsymbol{k}^\lambda(\mathcal{X}_i, \mathcal{X}_j'), \qquad (9)$$

with $\mathcal{X}_i$ representing the ith environment of the configuration $\mathcal{X}$. $\boldsymbol{k}^\lambda(\mathcal{X}_i, \mathcal{X}_j')$ is the tensorial kernel that compares the ith local environment of the $\mathcal{X}$ configuration with the jth local environment of the $\mathcal{X}'$ configuration.

Considering a water dimer as an example, we take the two O atoms as centers of the environments (so that

$N = 2$), and allow all of the surrounding atoms (H and O) to contribute to the smoothed atom density. The extension of this formalism to multiple chemical species involves a generalization of the scalar kernel of Eq. (5), discussed in Ref. [15]. With 500 training samples, both the isotropic and anisotropic components of the dimer polarizability can be learned with a RMSE below 10% of the intrinsic variance [see Fig. 1(b)]. It is worth stressing that although we use the dimer responses as learning targets, the additive kernel implies a decomposition in (environment-corrected) monomer responses. Equation (9) allows us to write, e.g.,

$$\boldsymbol{\alpha}(\mathcal{X}) = \frac{1}{N}\sum_i \boldsymbol{\alpha}(\mathcal{X}_i), \qquad (10)$$

where $\boldsymbol{\alpha}(\mathcal{X}_i) = \sum_{J,j} w_J N_J^{-1} \boldsymbol{k}(\mathcal{X}_i, \mathcal{X}_j^J)$ is the contribution of the ith environment to the dimer polarizability. As shown in the SM, when the two molecules are far apart the monomer polarizabilities predicted using Eq. (10) converge to the values computed separately for the two monomers [39]. Thus, the discrepancy observed when the molecular separation is small can be seen as the two-body correction to the dielectric response function of individual monomers.

As the next step, we consider the case of the Zundel cation $H_2O_5^+$. Being both charged and chemically active, this system would be difficult to describe in terms of separate molecular contributions. Figure 1(c) compares the learning curves for the moduli of the spherical components of $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$, obtained using a spherical cutoff of 4 Å around each oxygen atom. Note that although each environment encompasses the entire molecule, learning with atom-centered environments implies enforcing the covariance condition at the level of O atoms, which better captures the physics of the problem. The errors for all components are well below 5% with 500 training samples, showing that λ-SOAP kernels are well suited to extend the SA-GPR method to systems which are intrinsically not separable into smaller molecular units.

In order to test the robustness and generality of our scheme, we consider the prediction of the dielectric response tensor $\boldsymbol{\varepsilon}$ of instantaneous configurations of condensed-phase water. The reference data set has been collected by computing $\boldsymbol{\varepsilon}$ through the modern theory of polarization [58] within density functional theory, for 1000 different snapshots of a 32-molecule path integral simulation [59] of room-temperature q-TIP4P/f water [60] (see SM for further details [39]). Figure 2 shows how an O-centered, $r_c = 4$ Å, λ-SOAP kernel allows us to learn directly both the isotropic and anisotropic components of $\boldsymbol{\varepsilon}$ with a RMSE well below 0.01 a.u. with just 500 training samples. As we discuss in the SM [39], training is much more effective if performed on the molecular polarizability as obtained from the Clausius-Mossotti (CM) relation, $\boldsymbol{\alpha} = (\boldsymbol{\varepsilon} - 1)(\boldsymbol{\varepsilon} + 2)^{-1}V$. This underscores the importance of reducing the impact of nonlocal effects—which appear in the definition of $\boldsymbol{\varepsilon}$ through the volume term and through
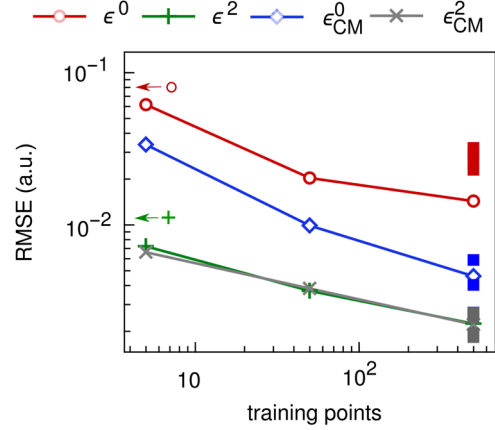

FIG. 2. Learning curves of the IST components of water dielectric response tensors $\boldsymbol{\varepsilon}$, through direct learning (red and green lines) and indirect learning going through the CM relation. The testing data set consists of 500 independent configurations. Arrows indicate the intrinsic standard deviation of the testing samples. Filled square points show the predictions for 5 ice Ih structures using the ML model trained on liquid water.

macroscopic field effects—when applying a machine-learning strategy that is based on an atom-centered decomposition. Indeed, similar performance can be obtained by learning $\boldsymbol{\varepsilon}$ if r_c is increased to 5 Å, so that information on the volume of the simulation is captured by the kernel (see SM [39]). In Fig. 2, we also show the errors for predicting the dielectric constant of 5 proton-disordered configurations of ice Ih [61] using the model trained on liquid water. Direct predictions of ε are less accurate than what is seen for the liquid. When going through the local CM response, however, the accuracy becomes comparable, underscoring the transferability of the model, and the ease with which it can be applied to solids.

The SA-GPR framework we introduced in this work provides a generally applicable strategy to perform kernel-based machine learning of tensorial properties, fully incorporating their rotational symmetries. Extensions to other discrete or continuous symmetries (e.g., to cylindrical geometries, or translational invariances) is straightforward. Building on the existing SOAP kernel between atomic environments, we obtain a hierarchy of λ-SOAP kernels which can be used to predict the electric response tensors of systems of increasing complexity, from isolated molecules to liquids and solids. Being able to apply statistical learning to tensors opens the way to the prediction of anisotropic materials properties: elastic and magnetic response, NMR chemical shifts, etc. Machine learning of molecular electric responses, which we used here as an example, makes it possible to improve the computation of linear and nonlinear optical spectra, as well as to design more accurate polarizable force fields for complex systems that cannot be described well in terms of rigid molecular entities. Another application with immense potential is related to the calculation of the building blocks of electronic-structure

methods, such as the ground-state charge density, or the matrix elements of Hamiltonians written in an atom-centered basis. Learning the Hamiltonian would allow one to obtain "tight-binding-like" schemes free of an explicit parametrization, which can match the accuracy of higher levels of electronic-structure theory when computing properties such as electronic bands. Statistical learning methods are finding applications across all fields of science and technology. This Letter provides an example of how making these approaches consistent with the fundamental physical symmetries of the problem at hand is crucial to realize their potential.

Source code and data availability.—A PYTHON code containing a rudimentary implementation of the λ-SOAP SA-GPR framework can be found in Ref. [62]. The repository also contains the data we used for training and testing.

[1] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, and A. Aspuru-Guzik, J. Phys. Chem. Lett. **2**, 2241 (2011).

[2] G. L. W. Hart, S. Curtarolo, T. B. Massalski, and O. Levy, Phys. Rev. X **3**, 041035 (2013).

[3] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, JOM **65**, 1501 (2013).

[4] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richard, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 011002 (2013).

[5] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, Comput. Mater. Sci. **111**, 218 (2016).

[6] C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. Hart, M. B. Nardelli, and S. Curtaralo, Comput. Mater. Sci. **108**, 233 (2015).

[7] L. Ward and C. Wolverton, Curr. Opin. Solid State Mater. Sci. **21**, 167 (2017).

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

[9] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[10] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K. R. Müller, and O. Anatole Von Lilienfeld, New J. Phys. **15**, 095003 (2013).

[11] Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. **114**, 096405 (2015).

[12] F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, Phys. Rev. Lett. **117**, 135502 (2016).

[13] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[14] M. Rupp, Int. J. Quantum Chem. **115**, 1058 (2015).

[15] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18**, 13754 (2016).

[16] G. Ferré, T. Haut, and K. Barros, J. Chem. Phys. **146**, 114107 (2017).

[17] F. Perakis, L. D. Marco, A. Shalit, F. Tang, Z. R. Kann, T. D. Kühne, R. Torre, M. Bonn, and Y. Nagata, Chem. Rev. **116**, 7590 (2016).

[18] X. Zhang, Q.-H. Tan, J.-B. Wu, W. Shi, and P.-H. Tan, Nanoscale **8**, 6435 (2016).

[19] R. J. Heaton and P. A. Madden, Mol. Phys. **106**, 1703 (2008).

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

[21] R. W. Boyd, *Nonlinear Optics* (Academic Press, New York, 2008).

[22] S. Roke and G. Gonella, Annu. Rev. Phys. Chem. **63**, 353 (2012).

[23] G. Tocci, C. Liang, D. M. Wilkins, S. Roke, and M. Ceriotti, J. Phys. Chem. Lett. **7**, 4311 (2016).

[24] A. Morita, J. Comput. Chem. **23**, 1466 (2002).

[25] G. S. Fanourgakis and S. S. Xantheas, J. Chem. Phys. **128**, 074506 (2008).

[26] P. Cieplak, F.-Y. Dupradeau, Y. Duan, and J. Wang, J. Phys. Condens. Matter **21**, 333102 (2009).

[27] P. E. M. Lopes, B. Roux, and A. D. MacKerell, Theor. Chem. Acc. **124**, 11 (2009).

[28] C. M. Baker, Comput. Mol. Sci. **5**, 241 (2015).

[29] G. A. Cisneros, K. T. Wikfeldt, L. Ojame, J. Lu, Y. Xu, H. Torabifard, A. P. Bartk, G. Csnyi, V. Molinero, and F. Paesani, Chem. Rev. **116**, 7501 (2016).

[30] B. Schölkopf, A. Smola, and K.-R. Müller, Neural Comput. **10**, 1299 (1998).

[31] C. K. I. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).

[32] M. Cuturi, arXiv:0911.5367v2.

[33] A. P. Bartók and G. Csányi, Int. J. Quantum Chem. **115**, 1051 (2015).

[34] A. Glielmo, P. Sollich, and A. De Vita, Phys. Rev. B **95**, 214302 (2017).

[35] T. Bereau, D. Andrienko, and O. A. von Lilienfeld, J. Chem. Theory Comput. **11**, 3225 (2015).

[36] C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke, and M. Ceriotti, Phys. Rev. B **96**, 041407 (2017).

[37] U. Weinert, Arch. Ration. Mech. Anal. **74**, 165 (1980).

[38] A. M. Morrison and A. G. Parker, Aust. J. Phys. **40**, 465 (1987).

[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.120.036002 for mathematical details of the derivation of the λ-SOAP framework and for computational details, which also includes Refs. [40–48].

[40] A. J. Stone, Mol. Phys. **29**, 1461 (1975).

[41] M. A. Blanco, M. Flórez, and M. Bermejo, J. Mol. Struct. **419**, 19 (1997).

[42] T. Regge, Il Nuovo Cimento **10**, 544 (1958).

[43] G. R. Medders, V. Babin, and F. Paesani, J. Chem. Theory Comput. **10**, 2906 (2014).

[44] K. Aidas *et al.*, Comput. Mol. Sci. **4**, 269 (2014).

[45] X. Huang, B. J. Braams, and J. M. Bowman, J. Chem. Phys. **122**, 044308 (2005).

[46] P. Umari and A. Pasquarello, Phys. Rev. Lett. **89,** 157602 (2002).

[47] I. Souza, J. Ĩniguez, and D. Vanderbilt, Phys. Rev. Lett. **89,** 117602 (2002).

[48] K. Urano and M. Inoue, J. Chem. Phys. **66,** 791 (1977).

[49] K. Kaufmann and W. Baumeister, J. Phys. B **22,** 1 (1989).

[50] W. J. Szlachta, A. P. Bartók, and G. Csányi, Phys. Rev. B **90,** 104108 (2014).

[51] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18,** 13754 (2016).

[52] S. De, F. Musil, T. Ingram, C. Baldauf, and M. Ceriotti, J. Cheminf. **9,** 6 (2017).

[53] M. J. Cliffe, A. P. Bartók, R. N. Kerber, C. P. Grey, G. Csányi, and A. L. Goodwin, Phys. Rev. B **95,** 224108 (2017).

[54] V. L. Deringer and G. Csányi, Phys. Rev. B **95,** 094203 (2017).

[55] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. Kermode, G. Csányi, and M. Ceriotti, Sci. Adv. **3,** e1701816 (2017).

[56] Note that, in agreement with the covariance condition of Eq. (3), the fingerprint vectors $\boldsymbol{P}^{\lambda}_{nn'll'}$ transform like the λth vector spherical harmonics when the configuration $\mathcal{X}$ is rotated.

[57] In this work, the overlap between atomic densities is raised to an even power in the definition of the scalar base kernel [Eq. (5)], which means that inversion symmetry is taken into account automatically when learning invariant properties. An overlap kernel raised to an odd power would instead be necessary when learning chiral properties: [13,15].

[58] R. Resta, J. Phys. Condens. Matter **22,** 123201 (2010).

[59] M. Ceriotti, J. More, and D. E. Manolopoulos, Comput. Phys. Commun. **185,** 1019 (2014).

[60] S. Habershon, T. E. Markland, and D. E. Manolopoulos, J. Chem. Phys. **131,** 024501 (2009).

[61] J. A. Hayward and J. R. Reimers, J. Chem. Phys. **106,** 1518 (1997).

[62] https://github.com/cosmo-epfl/SA-GPR.