The Journal of Chemical Physics

Operators in quantum machine learning: Response properties in chemical space EP

Cite as: J. Chem. Phys. **150**, 064105 (2019); https://doi.org/10.1063/1.5053562
Submitted: 23 August 2018 . Accepted: 19 November 2018 . Published Online: 12 February 2019

Anders S. Christensen, Felix A. Faber, and O. Anatole von Lilienfeld

COLLECTIONS

EP This paper was selected as an Editor's Pick

open access

View Online

Export Citation

CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Alchemical and structural distribution based representation for universal quantum machine learning
The Journal of Chemical Physics **148**, 241717 (2018); https://doi.org/10.1063/1.5020710

SchNet – A deep learning architecture for molecules and materials
The Journal of Chemical Physics **148**, 241722 (2018); https://doi.org/10.1063/1.5019779

Perspective: Computational chemistry software and its advancement as illustrated through three grand challenge cases for molecular science
The Journal of Chemical Physics **149**, 180901 (2018); https://doi.org/10.1063/1.5052551

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Operators in quantum machine learning: Response properties in chemical space

Cite as: J. Chem. Phys. 150, 064105 (2019); doi: 10.1063/1.5053562
Submitted: 23 August 2018 • Accepted: 19 November 2018 •
Published Online: 12 February 2019

Anders S. Christensen, Felix A. Faber, and O. Anatole von Lilienfeld[a)]

AFFILIATIONS

Department of Chemistry, University of Basel, Basel, Switzerland

[a)] **Electronic mail:** anatole.vonlilienfeld@unibas.ch

ABSTRACT

The role of response operators is well established in quantum mechanics. We investigate their use for universal quantum machine learning models of response properties in molecules. After introducing a theoretical basis, we present and discuss numerical evidence based on measuring the potential energy's response with respect to atomic displacement and to electric fields. Prediction errors for corresponding properties, atomic forces, and dipole moments improve in a systematic fashion with training set size and reach high accuracy for small training sets. Prediction of normal modes and infrared-spectra of some small molecules demonstrates the usefulness of this approach for chemistry.

I. INTRODUCTION

Time-independent electronic ground-state quantum properties can be expressed as expectation values of the electronic wave function and an operator, typically defined via the quantum-classical correspondence principle. The performance of supervised machine learning models of these quantum properties, a.k.a. quantum machine learning (QML),[1–4] can be conveniently assessed using learning curves that monitor the decay of the out-of-sample prediction error (deviation of the predicted properties from reference for query compounds not included in training) as a function of compound training set size N. Due to the leading prediction error decaying as a/N^b, log-log plots have become the recommended practice in the field with log(a) and b denoting the off-set and learning rate (or efficiency), respectively.[5–7] While, in principle, supervised ML models can be generated for any cause and effect relationship, it is the very philosophy of QML that representation and the kernel function (when using kernel ridge regression) are property independent[8,9] in the same way in which the electronic wave function and its Hamiltonian are property independent. However, there is a select and highly relevant set of quantum properties which can be understood as response properties, obtained through the use of response operators and perturbation theory. Common examples include derivatives of the energy with respect to the nuclear displacement or charge, an external electric field, an external magnetic field, or nuclear magnetic moments, and they can efficiently be accounted for within density functional theory.[10,11] We note that energy response properties also form the basis for conceptual density functional theory[12,13] as well as computational alchemy.[14–21] It has previously been observed that prediction errors of many conventional QML models of response properties can converge relatively slowly, even for QML models that are able to achieve remarkably high accuracy for energies.[2,8,22–24] In this paper, we investigate if the use of response operators is beneficial for deriving improved QML models that afford learning curves with lower off-sets and better learning rates.

Perhaps the most relevant quantum response property is the force exerted on each atom in the system, the first order energy derivative with respect to nuclear displacement.[25] Quite recently, tremendous efforts have been made to predict atomic forces accurately within QML models for the purpose of running *ab initio* quality molecular dynamics simulations at low computational cost.[26–38] Treating the force as the first derivative of the energy is tantamount to using the gradient operator, as commonly implemented in quantum chemistry packages. Doing so leads directly to energy conservation, a crucial property for most statistical mechanics applications, which has also been already obtained by others.[35,39] The use of response operators, however, has not yet been applied generally to generate QML models for other response properties.

Here, we extend the principle of using response operators to investigate the potential total energy and its response to a change in (i) atomic coordinates and (ii) an external electric field, i.e., the dipole moments. Other QML models capable of predicting dipole moments have already been published.[2,8,40–45] The work of Schütt *et al.* presented a neural network that is able to predict the dipole moment of the QM9 dataset[46,47] with very high accuracy[41] by training on the dipole moment vector itself. The other approaches rely on a charge model predicted from a neural network to estimate intensities in an infrared spectrum where the frequencies are obtained from a molecular dynamics simulation.[42,44] Similarly to Schütt *et al.*, we propose to learn the dipole moment by training on the quantum mechanical observable directly, but in contrast we train a model to describe the energy for which the dipole moment can be calculated as a response property by taking the derivative of the energy with respect to an external electric field. The modeling of highly accurate molecular potential energy surfaces has also been thoroughly investigated with several ML techniques, due to their important connection to infrared (IR) spectroscopy.[28,48–50] We show how our operator formalism can lead to ML potential energy surfaces that reproduce the vibrational normal modes of molecules across chemical space and even reproduces the IR spectrum of a molecule by using the relevant response operators with a suitable training set.

This paper is organized as follows: first we present the derivation for a kernel-based regression model capable of predicting response properties by letting the response operator act on the kernels. We then implement a representation that allows us to simultaneously train on properties that depend on both the external electric field and the internal degrees of freedom of the molecule. The hydrogen fluoride molecule is used as a toy model to demonstrate the principle. We benchmark the operator-based machine learning model on a number of existing data sets that account for forces, energies, and dipole moments across chemical space and show how our response model improves learning the dipole moment of molecules when compared to conventional kernel ridge regression models. Finally, we discuss how the model naturally couples force and energy predictions with dipole moment predictions and we show how the response model can directly predict properties related to second order derivatives, including mixed derivatives, such as infrared intensities, harmonic vibrational frequencies, and normal modes.

II. THEORY

A. Operator quantum machine learning (OQML)

Within kernel-based regression,[51–54] the total potential energy U^* of a query molecule C in its electronic ground-state can be decomposed into a sum of atomic energies, which are calculated using a basis of kernel functions

$$U^*_C = \sum_{I\in C} U^*_{\mathrm{local}}\left(q^*_I\right) = \sum_{I\in i}\sum_{J} k\left(q_J, q^*_I\right)\alpha_J, \tag{1}$$

where J runs over all atoms in the atomic environment in the basis, α_J is its regression weight, q_I is the representation of the Ith atom in the molecule, and here the asterisk denotes query atom.

Writing Eq. (1) in matrix form, we have

$$\mathbf{U} = \mathbf{K}\boldsymbol{\alpha}. \tag{2}$$

Note that in contrast to conventional Kernel Ridge Regression (KRR) and Gaussian Process Regression (GPR) based QML models,[9] this kernel matrix is not symmetric since first dimension is over the atoms used to build the basis and the second dimension has one entry for each observable, e.g., energies for molecules in the example in Eq. (2).

In this work, we approximate a response property ω, i.e., an observable which can be computed by applying a differential operator $\mathcal{O}$ acting on the energy U^*, defined in Eq. (1),

$$\boldsymbol{\omega} = \mathcal{O}[\mathbf{U}] = \mathcal{O}[\mathbf{K}]\boldsymbol{\alpha}. \tag{3}$$

The set of regression coefficients, $\boldsymbol{\alpha}$, can be obtained by minimizing the Lagrangian

$$\begin{aligned} J(\boldsymbol{\alpha}) &= \sum_{\gamma}\beta_\gamma \|\mathcal{O}_\gamma[\mathbf{U}^{\mathrm{ref}}] - \mathcal{O}_\gamma[\mathbf{K}\boldsymbol{\alpha}]\|^2_{L_2(\Omega_\gamma)} \\ &\equiv \sum_{\gamma}\beta_\gamma \int_{\Omega_\gamma}\left[\mathcal{O}_\gamma[\mathbf{U}^{\mathrm{ref}}] - \mathcal{O}_\gamma[\mathbf{K}\boldsymbol{\alpha}]\right]^{\mathrm{T}}\left[\mathcal{O}_\gamma[\mathbf{U}^{\mathrm{ref}}] - \mathcal{O}_\gamma[\mathbf{K}\boldsymbol{\alpha}]\right] \end{aligned} \tag{4}$$

with respect to $\boldsymbol{\alpha}$ over some training set of known values of $\mathcal{O}[\mathbf{U}^{\mathrm{ref}}]$. Ω_γ is the domain over which the corresponding operator should be minimized, e.g., all rotational degrees of freedom if the operator acts on a SO(3) group. γ denotes the specific perturbation of any order so that the model can be trained for multiple properties simultaneously, for example, energies, gradients, and dipole moments. β_γ is a weight, specific to each perturbation. In this work, β is set to 1 throughout. For simplicity, we pick Ω such that $\int_\Omega = 1$ for the remainder of this study. $\boldsymbol{\alpha}$ can be obtained, e.g., by solving the associated normal equations or using an orthogonal factorization such as a QR[55] or a singular-value decomposition (SVD). The corresponding normal equation (see the supplementary material for derivation) to this problem is given by

$$\boldsymbol{\alpha} = \left[\sum_{\gamma}\beta_\gamma\int_{\Omega_\gamma}\mathcal{O}_\gamma[\mathbf{K}]^{\mathrm{T}}\mathcal{O}_\gamma[\mathbf{K}]\right]^{-1}\left[\sum_{\gamma}\beta_\gamma\int_{\Omega_\gamma}\mathcal{O}_\gamma[\mathbf{U}^{\mathrm{ref}}]^{\mathrm{T}}\mathcal{O}_\gamma[\mathbf{K}]\right]. \tag{5}$$

However, solving the normal equations can be numerically unstable since it effectively squares the condition number, i.e., $\kappa(\mathbf{K}^{\mathrm{T}}\mathbf{K}) = (\kappa(\mathbf{K}))^2$.

For the practical implementation and the results discussed here, an SVD factorization has been used to solve Eq. (4), as it has several practical and efficient implementations. In contrast to the QR factorization, the SVD factorization is more numerically stable if $\mathbf{K}$ is rank-deficient, e.g., if **K** contains rows or columns that correspond to atoms or molecules that are identical or only differ by symmetry operations to which the representation is invariant.

In the case of under-determined equations, the SVD factorization is performed ignoring singular values smaller than a threshold, which can be treated as a hyperparameter similarly to regularization within ordinary KRR.

B. Operators

This section is dedicated to discussing some important response operators in quantum mechanics, defining the domain Ω over which the Lagrangian is to be minimized, and providing the corresponding solutions to the integrals in Eq. (5).

We define the response operator for some external parameter $\vec{\eta} = \{\eta_x, \eta_y, \eta_z\}$ which can be written as $\mathcal{O}_{\delta\vec{\eta}} \equiv \frac{\partial}{\partial\vec{\eta}}$. Applying such an operator would map the scalar field to a three dimensional vector field. All rotational degrees of freedom can then be integrated out with the following solutions. The solutions to the two integrals in Eq. (5), respectively, are thus

$$\int_{\Omega_{\delta\vec{\eta}}} \mathcal{O}_{\delta\vec{\eta}}[\mathbf{K}]^{\mathrm{T}}\mathcal{O}_{\delta\vec{\eta}}[\mathbf{K}] = \frac{1}{3}\sum_{\nu\in x,y,z}\left(\frac{\partial}{\partial\eta_\nu}\mathbf{K}\right)^{\mathrm{T}}\left(\frac{\partial}{\partial\eta_\nu}\mathbf{K}\right), \tag{6}$$

$$\int_{\Omega_{\delta\vec{\eta}}} \mathcal{O}_{\delta\vec{\eta}}[\mathbf{U}^{\mathrm{ref}}]^{\mathrm{T}}\mathcal{O}_{\delta\vec{\eta}}[\mathbf{K}] = \frac{1}{3}\sum_{\nu\in x,y,z}\left(\frac{\partial}{\partial\eta_\nu}\mathbf{U}^{\mathrm{ref}}\right)^{\mathrm{T}}\left(\frac{\partial}{\partial\eta_\nu}\mathbf{K}\right). \tag{7}$$

Correspondingly, this procedure can be used to solve the equations for the second order response operator, with respect to two different perturbations $\vec{\eta}$ and $\vec{\eta}'$,

$$\int_{\Omega_{\delta\vec{\eta}\delta\vec{\eta}'}} \mathcal{O}_{\delta\vec{\eta}\delta\vec{\eta}'}[\mathbf{K}]^{\mathrm{T}}\mathcal{O}_{\delta\vec{\eta}\delta\vec{\eta}'}[\mathbf{K}] = \frac{1}{9}\sum_{\nu,\nu'\in x,y,z}\left(\frac{\partial^2}{\partial\eta_\nu\partial\eta'_{\nu'}}\mathbf{K}\right)^{\mathrm{T}}\left(\frac{\partial^2}{\partial\eta_\nu\partial\eta'_{\nu'}}\mathbf{K}\right), \tag{8}$$

$$\int_{\Omega_{\delta\vec{\eta}\delta\vec{\eta}'}} \mathcal{O}_{\delta\vec{\eta}\delta\vec{\eta}'}[\mathbf{U}^{\mathrm{ref}}]^{\mathrm{T}}\mathcal{O}_{\delta\vec{\eta}\delta\vec{\eta}'}[\mathbf{K}] = \frac{1}{9}\sum_{\nu,\nu'\in x,y,z}\left(\frac{\partial^2}{\partial\eta_\nu\partial\eta'_{\nu'}}\mathbf{U}^{\mathrm{ref}}\right)^{\mathrm{T}}\left(\frac{\partial^2}{\partial\eta_\nu\partial\eta'_{\nu'}}\mathbf{K}\right). \tag{9}$$

A step-by-step derivation of these equations is given in the supplementary material. We note that the above equations are only true if the kernel is invariant with respect to rotations around θ and ϕ, which is true for the FCHL representation used in conjunction with a rotationally invariant kernel function, such as the Gaussian kernel.

Now we can explicitly write the matrix elements for the operators investigated within this study. The uppercase indices I, J, and K correspond to atomic centers, and the lowercase indices i, j, and k correspond to molecules.

The unperturbed kernel corresponds to the energy or identity operator acting on the kernel. The elements of the unperturbed kernel **K** are given as

$$(\mathbf{K})_{i\mathrm{J}} = \sum_{\mathrm{I}\in i} k\left(q_{\mathrm{J}}, q^*_{\mathrm{I}}\right). \tag{10}$$

The kernel elements that correspond to the force, i.e., minus the nuclear gradient operator acting on the kernel, are given by

$$-\frac{\partial}{\partial x^*_{\mathrm{I}}}(\mathbf{K})_{\mathrm{IJ}} = -\sum_{\mathrm{K}\in i}\frac{\partial k\left(q_{\mathrm{J}}, q^*_{\mathrm{K}}\right)}{\partial x^*_{\mathrm{I}}}, \quad \text{where} \quad I \in i. \tag{11}$$

The kernel elements that correspond to the response of the external electric field $\vec{\mathrm{E}}$ are given by

$$\frac{\partial}{\partial \mathrm{E}^*_\nu}(\mathbf{K})_{i_\nu \mathrm{J}} = \sum_{\mathrm{K}\in i}\frac{\partial k\left(q_{\mathrm{J}}, q^*_{\mathrm{K}}\right)}{\partial \mathrm{E}^*_\nu}, \quad \text{where} \quad \nu \in \{x, y, z\}. \tag{12}$$

Correspondingly, the nuclear Hessian kernel is given by

$$\frac{\partial^2}{\partial x^*_{\mathrm{I}'}\partial x^*_{\mathrm{I}}}(\mathbf{K})_{\mathrm{I'IJ}} = \sum_{\mathrm{K}\in i}\frac{\partial k\left(q_{\mathrm{J}}, q^*_{\mathrm{K}}\right)}{\partial x^*_{\mathrm{I}'}\partial x^*_{\mathrm{I}}}, \quad \text{where} \quad I', I \in i. \tag{13}$$

Finally, the kernel that yields the dipole derivatives necessary for the infrared intensities is written as the mixed second order derivative,

$$\frac{\partial^2}{\partial \mathrm{E}^*_\nu\partial x^*_{\mathrm{I}}}(\mathbf{K})_{i_\nu \mathrm{IJ}} = \sum_{\mathrm{K}\in i}\frac{\partial k\left(q_{\mathrm{J}}, q^*_{\mathrm{K}}\right)}{\partial \mathrm{E}^*_\nu\partial x^*_{\mathrm{I}}}, \quad \text{where } I \in i \text{ and } \nu \in \{x, y, z\}. \tag{14}$$

We are not aware of any other QML model which can account for these effects simultaneously.

C. Comparison to Gaussian process regression

In conventional GPR, the response properties (e.g., derivatives) of the learned function can be included in the training and the operators are enforced by adding a kernel for each operator of each learned function in the training set.[56] For example, including the nuclear gradient in addition to the energy will add one additional kernel function for each gradient component in the training set. The GPR kernel matrix that simultaneously incorporates the energy, u, and the gradient, g, is written as

$$\mathbf{K}^{\mathrm{GPR}} = \begin{bmatrix}\mathbf{K}^{u,u*} & \mathbf{K}^{u,g*}\\ \mathbf{K}^{g,u*} & \mathbf{K}^{g,g*}\end{bmatrix}, \tag{15}$$

where $\mathbf{K}^{u*,u}$ is the covariance between two molecules, i and j. For example, using a local decomposition, this is given by the following double sum:

$$\mathbf{K}^{u,u*}_{ij} = \sum_{\mathrm{I}\in i}\sum_{\mathrm{J}\in j} k\left(q_{\mathrm{J}}, q^*_{\mathrm{I}}\right). \tag{16}$$

Likewise, the first of the two blocks that contain only one derivative is given by

$$\mathbf{K}^{u,g*}_{i\mathrm{K}j} = \sum_{\mathrm{I}\in i}\sum_{\mathrm{J}\in j}\frac{\partial k\left(q_{\mathrm{J}}, q^*_{\mathrm{I}}\right)}{\partial x^*_{\mathrm{K}}} \tag{17}$$

and the second block is equal to the transpose. The last block that comprises the largest part of the full kernel matrix is the double derivative given by

$$\mathbf{K}^{g,g*}_{iKjL} = \sum_{I\in i}\sum_{J\in j} \frac{\partial k\left(q_J, q^*_I\right)}{\partial x_L \partial x^*_K}. \tag{18}$$

Thus, the memory requirement for a kernel for a training set with N molecules, each with M atoms is dominated by the 2nd derivative covariance kernel that scales as $\mathcal{O}\left(9N^2M^2\right)$. With numerical derivatives, a gradient is twice as expensive as the kernel itself and the 2nd derivative is four times as expensive. With these factors, the number of kernel evaluations of the 2nd derivative kernel scales as $\mathcal{O}\left(36N^2M^4\right)$.

Within our OQML formalism (Secs. II A and II B), we do not extend the basis by adding additional kernel functions, but we rather enforce the derivatives of the kernel elements in the regression. Note that OQML assigns only one α coefficient per atom, regardless of the dimensionality of the perturbation. This choice of basis has similarities to the sparsification introduced by Bartók and Csányi,[57] although the mathematical origins are different.

In practice, this means that the number of kernel function evaluations needed to train the model is reduced drastically.

The size of the kernel necessary to train our OQML model in Eq. (5) is $\mathcal{O}\left(N^2M^2\right)$, regardless of the perturbation. The number of kernel evaluations when the gradient is included in the training will scale as roughly $\mathcal{O}\left(6N^2M^3\right)$. For the examples in this work, memory requirements and training times are reduced by factors of ~10 and ~100, respectively, compared to conventional GPR with the same amount of training data.

In GPR, the training error will usually be close to zero since each additional label in the training set will be described by an additional basis kernel function. Since Eq. (5) uses a constant number of basis functions, the normal equation will describe an overdetermined set of equations, when the size of the perturbation exceeds the number of basis functions. For example, there are always more gradient components than the number of atoms in a molecule, while for molecules >3 atoms there are always more atoms than dipole moment components. The fact that the problem can become noticeable also means that training errors can become noticeable. Here, we found that in some cases they can even become as large as the test set error.

D. Representation

In this work, we extend the Faber-Christensen-Huang-Lilienfeld (FCHL) representation[23] to explicitly include the dependence on an externally applied electric field. This is crucial in order to learn dipole moments and other electric field-dependent properties. The FCHL representation consists of a set of M-body expansions $\mathcal{A}_M(I) = \{A_1(I), A_2(I), A_3(I), \ldots, A_M(I)\}$. The terms in the many-body expansion correspond to element type, interatomic distances, and interatomic angles, for the one-, two-, and three-body terms, up to order M, respectively.

It has previously been shown that the off-set in the learning curve is improved when the two- and three-body terms are multiplied by scaling factors such that features that contribute more to the learned property are weighted higher in the regression.[58] For energy learning, it was shown that $1/r^n$ and an Axilrod-Teller-Muto term[59,60] are suitable scaling factors for the FCHL two- and three-body terms, respectively.

In this paper, we extend the FCHL representation to include a dependence on the external electric field. Our modified FCHL* representation (denoted by an asterisk) compares the same features as the original formulation (i.e., element type and interatomic distances and angles), but an extra term is added to the scaling function to emulate the physics of the electric-field dependence of the representation and adjust the weighting accordingly. The new two-body scaling function (denoted by an asterisk) is given by

$$\xi_2^{*IJ} = \xi_2^{IJ} - \epsilon(\vec{\mu}_{IJ} \cdot \vec{E}), \tag{19}$$

where ξ_2^{IJ} is the $1/r^n$ scaling function in the original FCHL representation, $\vec{E}$ is the externally applied electric field, and $\vec{\mu}_{IJ}$ is a fictitious dipole arising from fictitious partial charges assigned to the atomic site of the atoms I and J, and ϵ is a scaling parameter that balances the two terms in the scaling function. This parameter was fitted *ad hoc* to ϵ = 0.005 Hartree^{-1} using toy models. The center-of-nuclear-charge convention is used to define the origin of the coordinate system. In practice, the fictitious partial charges are taken from the Gastieger charge model[61] as implemented in Open Babel.[62] However, we note that the exact values of the fictitious partial charges are unimportant and any partial charge model could likely be used. Note that the model does not learn these fictitious partial charges or does it use these as a proxy to learn the dipole moment. The model learns the scalar field of the energy, and the charges merely serve as dummy variables which enforce the right physical dependence of the kernel elements on the electric field.

The augmented three-body scaling function for an atom I interacting with the atoms J and K is similarly given by

$$\xi_3^{*IJK} = \xi_3^{IJK} - \epsilon(\vec{\mu}_{IJK} \cdot \vec{E}), \tag{20}$$

where ξ_3^{IJK} is the Axilrod-Teller-Muto scaling factor used to weight the three-body terms in the FCHL representation and $\vec{\mu}_{IJK}$ is the fictitious dipole arising from fictitious partial charges assigned to the atomic site of the atoms I, J, and K.

In the absence of an externally applied electric field, the FCHL* kernel elements are identical with the original FCHL kernel elements, but the derivative with respect to a perturbing field is now non-zero. We also note that this representation is "non-polarizable;" the second derivative of the representation with respect to the field is zero with a linear kernel. This could be amended, for example, by using on-site multipole moments with polarizability tensors, e.g., from a polarizable force field or a chemical-potential equalization charge model, rather than a static charge model.

III. RESULTS

A. Toy model for force learning

In this section, we demonstrate numerically the response of the kernel elements with respect to two very different kinds of perturbations, namely, (1) the nuclear coordinates and (2) an external electric field. The hydrogen fluoride molecule (H–F) is used as a toy model, and to show how including the vector quantities in the training improves learning.

We now show how the derivative of the kernel improves learning the potential energy of H–F. The MP2/aug-cc-pVTZ potential energy curve for the H-F molecule is used as training data. Upon selecting four training points (see Fig. 1), models were trained on these four points with and without the interatomic forces in the training set. Not training on forces using the FCHL representation with the default hyperparameters,[23] the resulting model poorly describes the dissociation curve; at the minimum-energy distances it even predicts a spurious transition state, and the energy decreases sharply when $d \rightarrow 0$. However, with forces included the potential energy surface is reproduced remarkably almost quantitatively, despite only four points being used to fit the model.

B. Toy model for electric field-dependent properties

Here we demonstrate the effect of including the dipole moment in addition to the energy in the training data. We now use a GPR model since our approach in Sec. II A would only contain two basis functions, while we are including up to four components, i.e., energy and dipole moment components. The toy model demonstrates the properties of the FCHL* representations which are fully transferable to the ML approach we present herein. We place a H–F molecule in an electric field of 0.001 a.u. which is rotated 360°, and the energy and dipole moment are calculated at each step of 1° at the MP2/aug-cc-pVTZ level of theory. We select just one point as a training set and train two GPR models: one with the MP2 energy and dipole moment components and the other with the MP2 energy but without the dipole moment. The energy predictions of these models as a function of the rotation of the field are displayed in Fig. 2. Without fitting to the dipole moment, the energy change due to the electric field is close to 0, only fluctuating by a bit of numerical noise from the fit. When the dipole moment is included, the curve is reproduced almost quantitatively with only a negligible deviation at the lowest energy point, presumably due to very small polarization effects and numerical noise.

This demonstrates how including a dipole-like dependence on the electric field in the representation is an efficient way to capture the underlying physics of the dipole moment into the kernel.

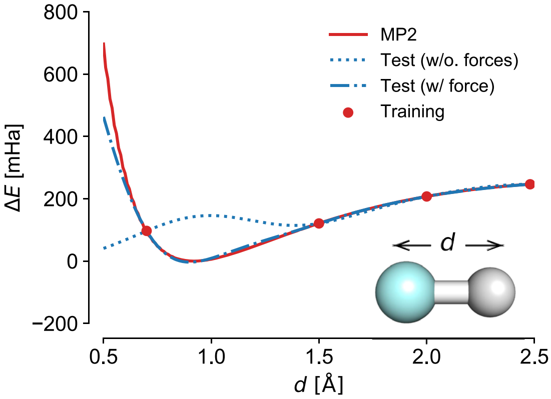

FIG. 1. The MP2/aug-cc-pVTZ potential energy surface of the hydrogen fluoride (H–F) molecule is displayed as a solid red line. Four training points (red dots) are selected, and two models are trained and used to predict the potential energy surface: one including the interatomic force in addition to the MP2 energy (blue, dashed-dotted) and the other using only the MP2 energy (blue, dotted).

C. Force and energy learning

Here we use the FCHL* representation within the presented OQML model to study two existing benchmark sets for learning forces and energies. The MD17 consists of molecular dynamics (MD) snapshots from MD trajectories of different molecules for which reference forces and energies are available.[35] We benchmark our models to seven molecules out of the MD17 dataset, namely, ethanol, salicylic acid, aspirin, malonaldehyde, toluene, naphthalene, and uracil. Similarly, the ISO17 consists of MD snapshots of isomers with the chemical formula $C_7O_2H_{10}$. The ISO17 additionally comes with two different test sets.[37,63] The first consists only of isomers with a connectivity that is present in the training set (“known”), and the other that contains only isomers with a connectivity that is not present in the training set (“unknown”). Briefly the two datasets benchmark the conformational freedoms and

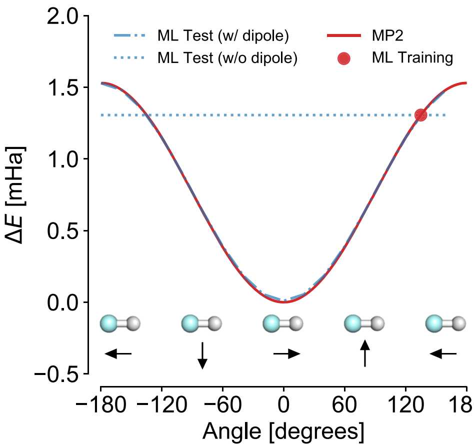

FIG. 2. A hydrogen fluoride (H–F) molecule is placed in an external electric field of 0.001 a.u., and the MP2/aug-cc-pVTZ energy is calculated as a function of the angle between the H–F molecule and the field vector, displayed as a red line. A single point is selected as a training set (red dot), and two models are trained and used to predict the energy in the electric field: one including the dipole moment of the molecule in addition to the MP2 energy (blue, dashed-dotted) and the other using only the MP2 energy (blue, dotted). The alignment between the field and the molecule is sketched at the bottom for clarity.

constitutional freedoms of molecules, respectively. Since there is no electric field applied to the molecules in these data sets, note that the FCHL* representation reduces to the original FCHL representation.[23]

Learning curves for the two datasets are displayed in Figs. 3 and 4. For reference, we compare FCHL* to the Gradient-Domain Machine Learning (GDML) method,[35] which is closely related to GPR with the inverse distance matrix as representation, and the SchNet neural network.[37] We note that a promising modification to GDML exists, sGDML, which shows higher accuracy compared to GDML for molecules that have atoms that are related by symmetry operations.[64] For the MD17 dataset, the out-of-sample MAE errors of the predicted energies are similar among FCHL*, GDML, and SchNet, with SchNet being slightly less accurate in most cases (see Fig. 3). FCHL* and SchNet perform best for ethanol and malonaldehyde, while GDML is best for salicylic acid and naphthalene. Uracil is best modeled by GDML, with relatively poor SchNet forces, and FCHL being in between. At this point, we remind the reader that the GDML approach is only applicable to a given system, while FCHL* and SchNet are capable of learning across chemical space. Note however, that a direct comparison between the different ML approaches is not possible. Ultimately, the OQML approach is different from SchNet and GDML, not only because of the use of operators, but also in the choice of representation.

Performance across constitutional space is tested on the constitutional isomers in the ISO17 dataset (Fig. 4). For the two test sets of "known" and "unknown" molecules in the ISO17, the FCHL* model displays a good learning rate, that is, qualitatively comparable to the SchNet model. Note that, here, the name "known" only implies that the isomers of the same constitution are known to the machine, but not the conformations in the test set. Unfortunately the learning curves between the FCHL*

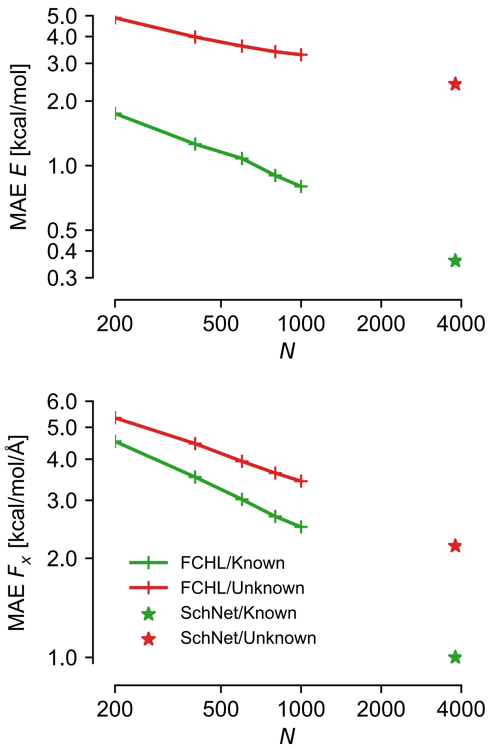

FIG. 4. The learning curves of our model for the ISO17 dataset, and in addition the accuracy for SchNet when using 4000 training samples is shown. The top panel shows the out-of-sample MAE energy prediction for a set of isomers known to the trained machine ("known") and for a set of unknown to the machine ("unknown"). The bottom panel shows the out-of-sample MAE force prediction for the same two sets. Note that "known" in this context only concerns whether the isomers are included in the training set or not. In both cases, only isomers with a conformation unknown to the machine are used as test data.

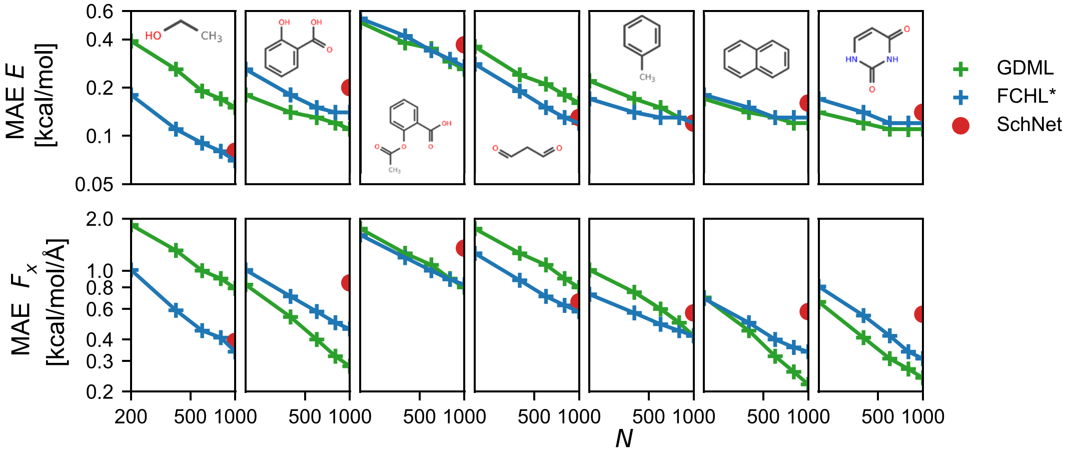

FIG. 3. The learning curves of our model for the MD17 dataset, for the seven molecules in the MD17 dataset (from left to right) ethanol, salicylic acid, aspirin, malonaldehyde, toluene, naphthalene, and uracil. The out-of-sample mean absolute error (MAE) energy prediction (E, top row) and MAE force component prediction (F_X, bottom row) are shown for the presented FCHL* (blue) model as well as for the GDML[35] (green) and SchNet (red) models.[37,63]

models and SchNet do not overlap, so the two models cannot be compared quantitatively here, but the out-of-sample accuracy seems comparable.

Overall, we find that our operator approach leads to forces with state-of-the-art accuracy, on par with two of the most accurate models already published in the literature.

D. Learning dipole moments of QM9

Prediction errors of machine learning models of dipole moments converge slowly for conventional QML models.[8,22,23] Here we demonstrate how including the underlying physics for the dipole moment into the representation improves the learning rate, as opposed to learning the dipole norm with conventional kernel ridge regression. We compare two approaches to learn the dipole moment norm of the molecules in QM9: (1) using the FCHL* representation with the OQML approach outlined in Sec. II A to fit the dipole moments as derivatives of the energy and (2) learning the dipole moment norm as a scalar using kernel ridge regression with the FCHL representation as done in our earlier paper.[23] The learning curves of the two models are displayed in Fig. 5. The MAE out-of-sample predicted dipole moment norm is decreased substantially with our new approach. For instance, training on 5000 random molecules, the out-of-sample MAE error is reduced by 54% (from 0.67 D to 0.31 D). We also note that not only is the learning curve offset lower when the dipole moment operator is used, compared to conventional KRR, but it is also substantially steeper. This demonstrates the strength of the approach of using the correct response operators in the kernel to learn the corresponding response properties.

E. Learning normal modes

In this section, we assess the ability of the methodology to predict vibrational normal modes of a number of organic molecules.

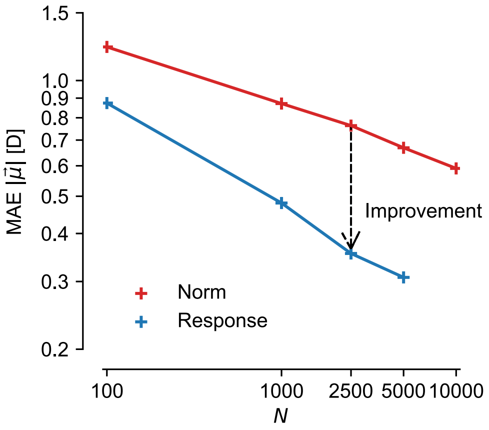

FIG. 5. The out-of-sample prediction error of the dipole norm as a function of the QM9 training data set size. The red curve corresponds to a conventional KRR model learning the scalar with the original FCHL representation, taken from Faber *et al.*[23] The blue curve shows the predictions from a machine trained on the energy and dipole moments of QM9 molecules, which in turn predicts the dipole vector from which the norm is calculated.

We randomly selected 83 molecules from the QM9 dataset with 9 heavy atoms. For each of these molecules, we create a minimal training set consisting of all sub-fragments of the molecules with up to 7 heavy atoms, following the methodology of Huang and von Lilienfeld.[58] Effectively this approach can be used to prove that the machine can extrapolate from the known properties of smaller molecules to predict the same properties for larger molecules.

For each of these generated fragments, a conformational search is performed using RDKit[65] and the unique conformers are minimized at the ωB97xD/6-31G(d) level of theory. From each of these minimized geometries, a number of distorted geometries are generated using normal-mode sampling[66] at the same level of theory. For each of the distorted geometries, a single-point energy and force evaluation is performed at the ωB97xD/6-31G(d) level of theory, and the forces and energies are saved. Using the sets of distorted fragment geometries for each of the 83 molecules, we train machines on forces and energies with increasing numbers of samples of each fragment in the sets.

In order to benchmark the performance of the trained machines, we set up the following test: a vibrational analysis is performed at the ωB97xD/6-31G(d) level of theory for each of the 83 molecules. Using the normal modes of the molecules obtained from the vibrational analysis, we generate scans of the potential energy surface along each normal mode. The scan consists of structures that are distorted from the equilibrium geometry along each of the normal modes in 10 steps along the positive and negative directions. The distortions along each normal mode are scaled using the force constants such that the energy of the geometry with the largest distortion along a normal mode is about 0.5 kcal/mol higher than that of the equilibrium geometry. For each of these potential energy scans along the normal modes, we let the trained machines to predict the potential energy and then we compare this to the QM energy. If the machine predicts a well-defined minimum within the 0.5 kcal/mol scan range, this is counted as a success, otherwise this is counted as a failure. As an example, we show predicted normal mode scans for the 15 normal modes with lowest frequency for a QM9 molecule ($C_6N_3H_7$, ID# 036682, SMILES string: `C1C2C3C4OCOC13C24`) in Fig. 6. The molecular structure and its corresponding atom-in-molecule fragments (am-ons) used for training are shown in Fig. 7.

In addition, we present predictions from machines trained on N ∈ {1, 2, 4, 8, 16, 32} distorted samples of each sub-fragment in the database. Data to reconstruct similar plots for all 83 molecules are available from Figshare at https://doi.org/10.6084/m9.figshare.6994445. For the machine trained on only N = 1 sample per fragment, a total of 11 normal modes do not have a well-defined minimum within the scan range. By increasing the training set to N = 2, the machine only predicts two normal modes with minima outside the scan range. At N = 4, all normal modes have a well-defined minimum inside the scan range, but when increasing to N = 8, two of the low normal modes that correspond to very non-local conformational changes are not identified correctly to lie within the scan range. Increasing

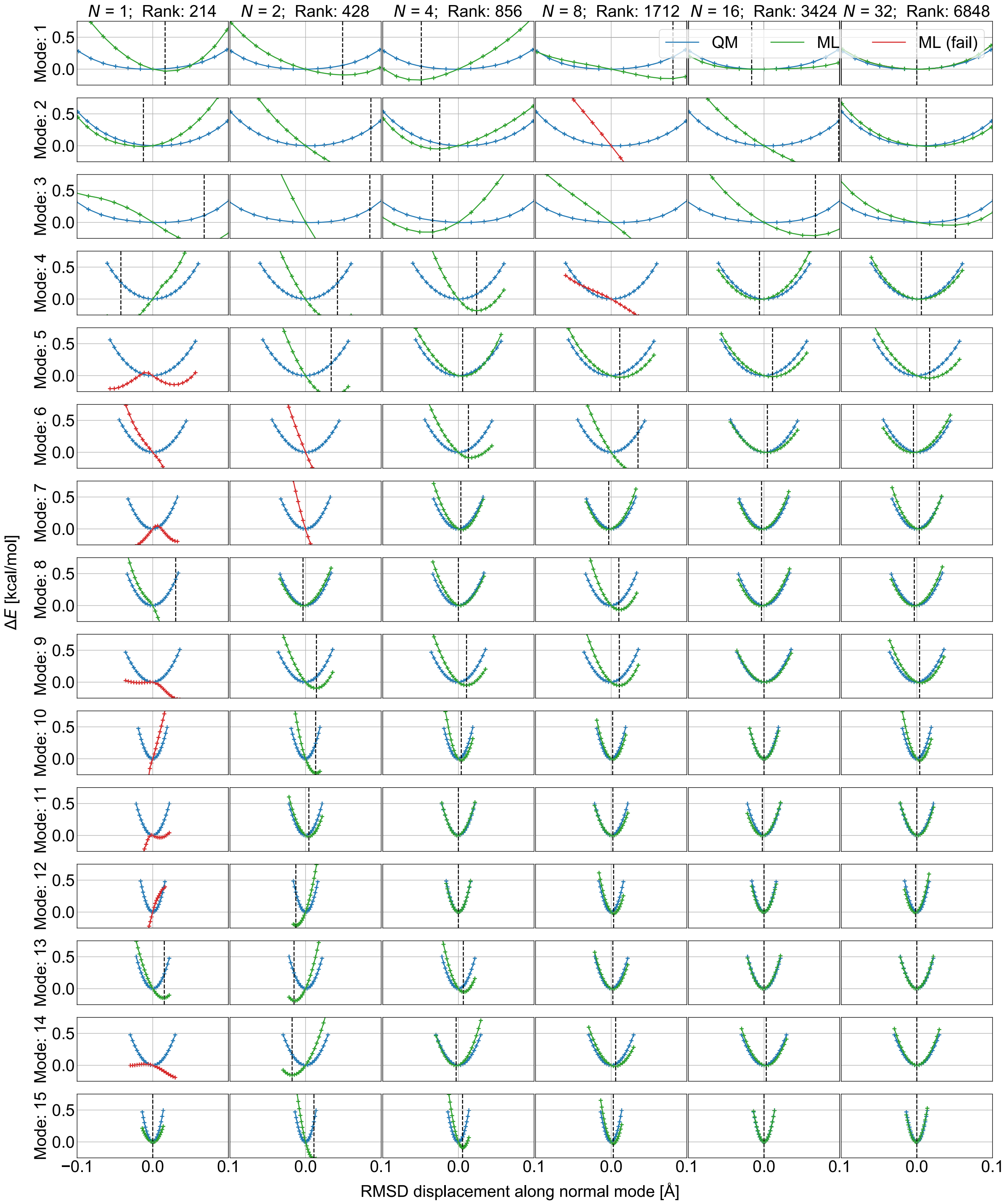

FIG. 6. ML predicted energy changes of $C_6N_3H_7$ as a function of distortion along each of the 15 normal modes with lowest frequency. The molecular structure and its corresponding atom-in-molecule fragments used for training are shown in the figure. Stiffer normal modes are easier to learn and therefore not shown. The complete result set is provided in the supplementary material. Each row and each column correspond to a normal mode and training set size N/maximum possible rank of the kernel matrix, respectively. N is the number of samples for each amon (i.e., sub-fragment). Displacements are scaled such that the maximum distortion energy is close to 0.5 kcal/mol. The X-axis displays the root-mean-square deviation (RMSD) in coordinates to the QM equilibrium geometry after the molecule has been displaced along that normal mode. The Y-axis is the energy difference to the equilibrium geometry, calculated with either QM (blue) or ML (green/red). The curves predicted from ML are displayed in green if there is a defined minimum within the scan range and red (fail) otherwise. The locations of the minima are marked by black vertical dashed lines.

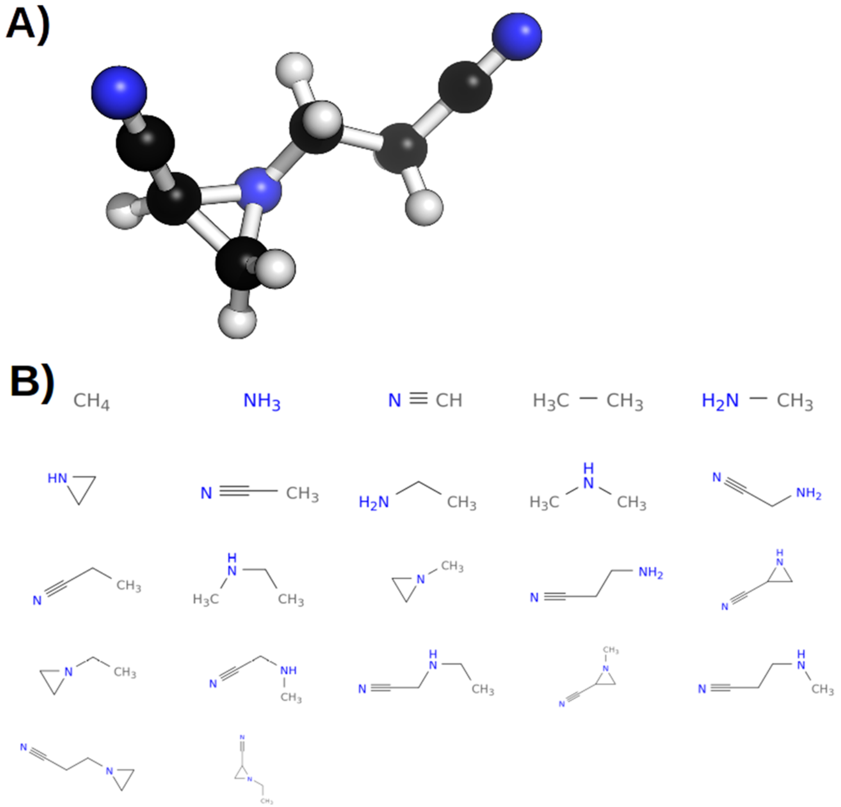

FIG. 7. Panel (a) displays the QM9 molecule with the ID# 036682 (SMILES string: `C1C2C3C4OCOC13C24`) for which normal modes are predicted in Fig. 8. Panel (b) displays the fragments identified using the method of Huang and von Lilienfeld,[58] which are used to generate the training set for the molecule.

again to N = 16 samples, the minima are well-defined again, and at N = 32, the QM potential energy curves are almost quantitatively reproduced.

We note that the higher normal modes, which mostly correspond to very local distortions such as a single hydrogen bond stretching, are almost always very well reproduced. By contrast, the lower normal modes, which often are more non-local in nature and correspond to very flat energy surfaces, require larger training set sizes to reproduce correctly.

Upon repeating the same test for all of the 83 QM9 molecules, we can plot the fraction of normal modes which are incorrectly described as a function of the training set size. Here, training set size is measured as the maximum possible rank of the kernel matrix, which corresponds to the number of regression coefficients and the number of atoms in the training set. This is plotted for all 83 molecules in Fig. 8 for the corresponding machines training on $N \in \{1, 2, 4, 8, 16, 32\}$ distorted samples of each sub-fragment. We note a trend that larger training sizes yield a smaller chance that the machine fails to identify a well-defined minimum close to the minimum in the reference geometry.

F. Infrared spectrum for dichloromethane

In order to demonstrate the utility of the above developments, we have combined them in order to learn and predict IR spectra. More specifically, a vibrational analysis is performed to get the harmonic frequencies and the IR intensities for the dichloromethane molecule. We note that although our methodology is transferable, the results of this exercise are very dependent on the training set. Thus we restrict this section to only one molecule and demonstrate that the methodology yields higher order derivatives, including mixed derivatives that systematically improve with the training set.

Models are trained on distorted geometries of the dicholoromethane molecule, for which MP2/def2-TZVP energies, forces, and dipole moments had been previously calculated. The training set consists of 100 distorted geometries that are generated by normal-mode sampling following the protocol of Smith *et al.*[66] Using the trained model, a standard vibrational analysis using the rigid-rotor harmonic-oscillator approximation is performed in a standard quantum chemistry package (Gaussian09)[67] via an interface to the QML code[68] which supplies the necessary energies and derivatives to the quantum chemistry program. First, the molecule is optimized on the machine learned potential energy surface by supplying the optimizer in the Gaussian program with the energies and nuclear gradients. Second, the vibrational analysis is performed by supplying the Gaussian program with the numerical nuclear Hessian and dipole derivatives.

As a reference, we compare the IR spectrum from the vibrational analysis on the potential energy surface of the machine learning model to the IR spectrum from a standard vibrational analysis at the MP2/def2-TZVP level.

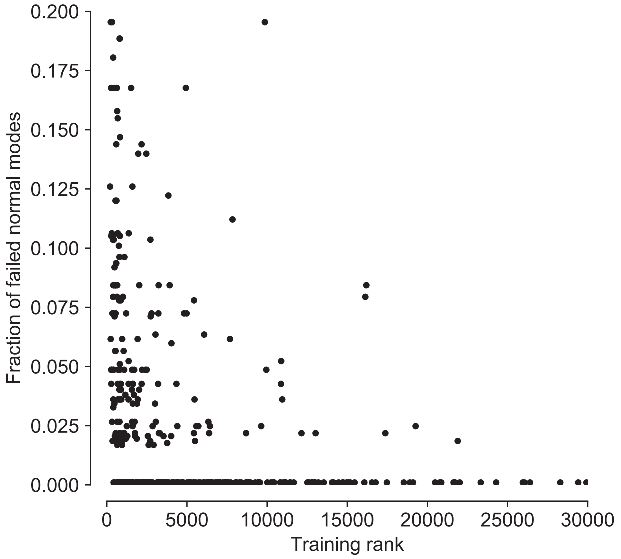

FIG. 8. Fraction of failed normal mode predictions for 83 QM9 molecules with 9 heavy atoms as a function of training set size. For each molecule, six machines are trained with increasing numbers of molecules in the training set. The X-axis shows the rank of the kernel matrix (i.e., the number of regression coefficients) for each training set used to train a model for a molecule. The Y-axis shows the fraction of modes for the same molecule, for which the machine predicts a well-defined minimum within a reasonable distance (see text) from the reference equilibrium geometry.

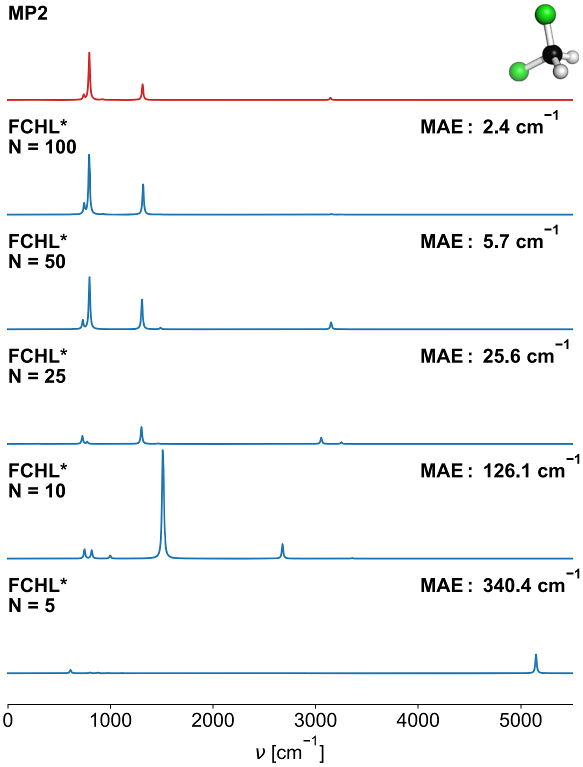

FIG. 9. The unscaled infrared spectrum of dichloromethane calculated via vibrational analysis. (Top/red) Calculated at the MP2/def2-TZVP level of theory; (bottom/blue) using QML to calculate the necessary derivatives of the energy with respect to the nuclear coordinate and the dipole moment. The spectra are convoluted using Lorentzian distributions[69] with a width of $\gamma = 8$ cm^{-1}.

Five models are trained on a decreasing number of samples (100, 50, 25, 10, and 5) of randomly selected configurations from the full 100 configuration training set. Then, a geometry optimization and a vibrational analysis are performed with each of the trained models. The resulting IR spectra for dichloromethane are displayed in Fig. 9. Qualitatively the FCHL* models reproduce the frequencies of the true MP2 reference with close agreement between the vibrational frequencies of the tallest peaks, even with as few as 10 training samples. The three most intense peaks in the spectrum are located at 743, 793, and 1318 cm^{-1} when using the largest training set (100 samples), compared to 740, 793, and 1315 cm^{-1}, respectively, for the reference MP2 spectrum. Training the model on only five randomly selected samples does not lead to a meaningful IR spectrum; however, already with ten instances, decent frequencies and underestimated intensities are obtained for the first two peaks. Learning the intensities via the dipole derivatives seems to be a harder task for the machine, compared to the peak locations, and the relative peak intensities are not qualitatively correct until N = 50 training samples.

We note that the dichloromethane molecule has 9 normal modes, and it is therefore expected that at the very least 9 samples would be necessary to have the minimally required sampling along all the possible normal modes. Further increasing the training set size to 25 and 50 samples improves the locations of the peaks to MAE vibrational frequencies of 25.6 and 5.7 cm^{-1}, respectively. At 100 training samples, the spectrum is almost at spectroscopic precision with an MAE of only 2.5 cm^{-1}.

This demonstrates the generality of the response operator-based machine learning model. The IR intensities correspond to a second order mixed derivative, indicating that the model accounts even for higher order effects after including only energy and first order derivatives. These results suggest that the systematic addition of higher order effects has the potential to improve the performance even further.

IV. METHODOLOGY

A. Used software

All energy, gradient, and dipole-moment calculations for the H-F molecule were performed in ORCA 4.0.1[70] at the MP2/aug-cc-pVTZ level of theory with no RI approximation and the `NoFrozenCore` keyword. The relaxed MP2 density was used to calculate the dipole moment as the correct derivative of the energy.

Since only the dipole norms are supplied with the QM9 dataset,[46,47] the dipole moment vectors of QM9 were recalculated using ORCA 4.0.1. To ensure consistency with the

B3LYP/6-31G(2df,p) method and basis set used in the original QM9 dataset, the `B3LYP/G` option was used for the B3LYP functional[71] and the 6-31G(2df,p) basis set was manually set up to the same contraction coefficients and exponents as used in the original calculations.

Energies, forces, and vibrational analyses for the QM9 molecules and fragments in Sec. III C were calculated at the ωB97xD/6-31G(d) level of theory using the Gaussian09 program.[67] The structures and the corresponding data can be found in comma-separated value (CSV) format from Figshare at https://doi.org/10.6084/m9.figshare.7000280.

The forces, energies, and dipole moments of the dichloromethane molecule were calculated at the MP2/def2-TZVP level of theory in the Gaussian09 program. The MP2 vibrational analysis was also carried out in Gaussian09. The vibrational analyses that employ machine learning were also carried in Gaussian09 via a Python interface to the machine learning code, and the keywords `freq=(numer,fourpoint,step=100)` were used to get the second derivatives. Our current implementation employs two-point numerical first derivatives, except for geometry optimizations for which it was necessary to use a five-point numerical derivative due to the sensitivity to numerical noise in the optimizer.

The reader can carry out machine learning with the presented algorithms, i.e., implemented kernel functions, efficient solvers, and the FCHL* representation. The necessary code is freely available from our open source machine learning toolkit QML[68] at http://github.com/qmlcode/qml.

B. Hyperparameters

All hyper parameters of the FCHL* representation were kept fixed to the same values as those found to be optimal in our previous paper,[23] and the only new parameter is the newly introduced ϵ = 0.0005 Hartree^{-1} parameter in the scaling functions. In all examples, a Gaussian kernel function is used with the kernel width set to σ = 0.64 and the cap for smallest singular values to keep in the SVD decomposition was set to 10^{-9} in units of the largest singular value. These parameters were not rigorously fitted to any dataset, so it is possible that more optimal values exist.

V. CONCLUSION

This paper explores a kernel-based supervised machine learning model that is capable of learning response properties by applying the corresponding response operator to the kernel function. Within this framework, we have extended the FCHL representation by a physically motivated response term for the application of an external electric field. Using the hydrogen fluoride molecule as a toy model, we have demonstrated how the machine learning model and representation can account for the right physics in simple systems with only a minimal number of training samples. Upon benchmarking the accuracy of our model for force and energy prediction on the MD17 and ISO17 dataset, our OQML model achieves state-of-the-art accuracy, on par or better than the GDML and SchNet models. For learning the dipole norm of the molecules in the QM9 dataset, using the operator formalism leads to an improvement of 54% compared to learning the same quantity as a scalar with the same representation. Finally, we allude to the possibility to obtain higher order derivatives, including mixed derivatives. This idea has been demonstrated by training a model on the energies, forces, and dipole moments for the dicholoromethane molecule. Using the resulting model, we have performed a vibrational analysis and presented the resulting infrared spectrum which systematically approaches the reference spectrum (calculated at the corresponding *ab initio* level of theory) as more training cases are being added.

Our results suggest that it is advantageous to learn response properties via the corresponding response operators. The OQML methodology presented here is, in principle, not limited to derivatives of the energy with respect to the nuclear positions or the external electric field. We envision extending the representation to account for a multitude of other properties, such as higher order response properties, including magnetic properties such as nuclear magnetic resonance (NMR) chemical shifts and spin-spin coupling constants or alchemical derivatives. Since the OQML formalism is not restricted to any choice of operator, it might also be possible to go beyond response operators. For instance, with the right representation, it should be possible to even learn more fundamental properties of molecules such as the electronic density or the kinetic energy.

SUPPLEMENTARY MATERIAL

See supplementary material for a detailed derivation of the formalism.

ACKNOWLEDGMENTS

The National Centre of Competence in Research (NCCR) Materials Revolution: Computational Design and Discovery of Novel Materials (MARVEL) of the Swiss National Science Foundation (SNSF) is acknowledged. This material is based upon the work supported by the Air Force Office of Scientific Research, Air Force Material Command, USAF under Grant No. FA9550-15-1-0026. Some calculations were performed at the sciCORE (http://scicore.unibas.ch/) scientific computing core facility at the University of Basel.

REFERENCES

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

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

[3] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, Sci. Rep. **3**, 2810 (2013).

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

[5] C. Cortes, L. D. Jackel, S. A. Solla, V. Vapnik, and J. S. Denker, "Learning curves: Asymptotic values and rate of convergence," in *Advances in Neural Information Processing Systems* 6, edited by J. D. Cowan, G. Tessuro, and J. Alspector (Morgan-Kaufmann, 1994), pp. 327–334.

[6] K. R. Müller, M. Finke, N. Murata, K. Schulten, and S. Amari, Neural Comput. **8**, 1085 (1996).

[7] O. A. von Lilienfeld, Angew. Chem., Int. Ed. **57**, 4164 (2018).

[8] R. Ramakrishnan and O. A. von Lilienfeld, CHIMIA Int. J. Chem. **69**, 182 (2015); e-print arXiv:1502.04563.

[9] R. Ramakrishnan and O. A. von Lilienfeld, "Machine learning, quantum chemistry, and chemical space," in *Reviews in Computational Chemistry* (John Wiley & Sons, Inc., 2017), Vol. 30, pp. 225–256.

[10] X. Gonze, Phys. Rev. A **52**, 1096 (1995).

[11] A. Putrino, D. Sebastiani, and M. Parrinello, J. Chem. Phys. **113**, 7102 (2000).

[12] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford Science Publications, 1989).

[13] P. Geerlings, F. D. Proft, and W. Langenaeker, Chem. Rev. **103**, 1793 (2003).

[14] O. A. von Lilienfeld, R. Lins, and U. Rothlisberger, Phys. Rev. Lett. **95**, 153002 (2005).

[15] O. A. von Lilienfeld, J. Chem. Phys. **131**, 164102 (2009).

[16] D. Sheppard, G. Henkelman, and O. A. von Lilienfeld, J. Chem. Phys. **133**, 084104 (2010).

[17] O. A. von Lilienfeld, Int. J. Quantum Chem. **113**, 1676 (2013).

[18] K. Y. S. Chang, S. Fias, R. Ramakrishnan, and O. A. von Lilienfeld, J. Chem. Phys. **144**, 174110 (2016).

[19] A. Solovyeva and O. A. von Lilienfeld, Phys. Chem. Chem. Phys. **18**, 31078 (2016).

[20] S. Fias, F. Heidar-Zadeh, P. Geerlings, and P. W. Ayers, Proc. Natl. Acad. Sci. U. S. A. **114**, 11633 (2017).

[21] K. S. Chang and O. A. von Lilienfeld, Phys. Rev. Mater. **2**, 073802 (2018).

[22] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, J. Chem. Theory Comput. **13**, 5255 (2017).

[23] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, J. Chem. Phys. **148**, 241717 (2018).

[24] W. Pronobis, A. Tkatchenko, and K.-R. Müller, J. Chem. Theory Comput. **14**, 2991 (2018).

[25] R. P. Feynman, Phys. Rev. **56**, 340 (1939).

[26] B. G. Sumpter and D. W. Noid, Chem. Phys. Lett. **192**, 455 (1992).

[27] S. Lorenz, A. Gross, and M. Scheffler, Chem. Phys. Lett. **395**, 210 (2004).

[28] S. Manzhos and T. Carrington, Jr., J. Chem. Phys. **125**, 084109 (2006).

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

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

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

[32] S. Manzhos, R. Dawes, and T. Carrington, Int. J. Quantum Chem. **115**, 1012 (2015).

[33] V. Botu and R. Ramprasad, Int. J. Quantum Chem. **115**, 1074 (2015).

[34] M. Rupp, R. Ramakrishnan, and O. A. von Lilienfeld, J. Phys. Chem. Lett. **6**, 3309 (2015); e-print arXiv:1505.00350.

[35] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Sci. Adv. **3**, e1603015 (2017).

[36] J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. **8**, 3192 (2017).

[37] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).

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

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

[40] B. Huang and O. A. von Lilienfeld, J. Chem. Phys. **145**, 161102 (2016).

[41] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Nat. Commun. **8**, 13890 (2017).

[42] A. E. Sifain, N. Lubbers, B. T. Nebgen, J. S. Smith, A. Y. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, and S. Tretiak, J. Chem. Phys. Lett. **9**, 4495 (2018).

[43] B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, and S. Tretiak, J. Chem. Theory Comput. **14**, 4687 (2018).

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

[45] K. T. Schütt, M. Gastegger, A. Tkatchenko, and K.-R. Müller, preprint arXiv:1806.10349 (2018).

[46] L. Ruddigkeit, R. van Deursen, L. Blum, and J.-L. Reymond, J. Chem. Inf. Model. **52**, 2864 (2012).

[47] R. Ramakrishnan, P. Dral, M. Rupp, and O. A. von Lilienfeld, Sci. Data **1**, 140022 (2014).

[48] S. Manzhos, X. Wang, R. Dawes, and T. Carrington, J. Phys. Chem. A **110**, 5295 (2006).

[49] S. Manzhos and T. Carrington, J. Chem. Phys. **129**, 224104 (2008).

[50] J. Cui and R. V. Krems, J. Phys. B: At., Mol. Opt. Phys. **49**, 224001 (2016).

[51] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, IEEE Trans. Neural Networks **12**, 181 (2001).

[52] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond* (MIT Press, 2002).

[53] V. Vovk, "Kernel ridge regression," in *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, edited by B. Schölkopf, Z. Luo, and V. Vovk (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), pp. 105–116.

[54] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Series in Statistics (Springer, New York, N.Y., 2001).

[55] G. Golub and C. Van Loan, *Matrix Computations* (Johns Hopkins University Press, 1996).

[56] C. E. Rasmussen and C. K. I. Williams, in *Gaussian Processes for Machine Learning*, edited by T. Dietterich (MIT Press, Cambridge, 2006), www.GaussianProcess.org.

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

[58] B. Huang and O. A. von Lilienfeld,"The 'DNA' of chemistry: Scalable quantum machine learning with 'amons,'" e-print arXiv:1707.04146.

[59] B. M. Axilrod and E. Teller, J. Chem. Phys. **11**, 299 (1943).

[60] Y. Muto, J. Phys. Math. Soc. Jpn. **17**, 629 (1943).

[61] J. Gasteiger and M. Marsili, Tetrahedron **36**, 3219 (1980).

[62] R. Guha, M. T. Howard, G. R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. K. Wegner, and E. Willighagen, J. Chem. Inf. Model. **46**, 991 (2006).

[63] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, A. Tkatchenko, and K.-R. Müller, preprint arXiv:1706.08566 (2018).

[64] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, e-print arXiv:1802.09238 (2018).

[65] See http://www.rdkit.org for RDKit, online, "RDKit: Open-source cheminformatics."

[66] J. S. Smith, O. Isayev, and A. E. Roitberg, Sci. Data **4**, 170193 (2017).

[67] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, GAUSSIAN 09, Revision D.01, Gaussian, Inc., 2009.

[68] A. S. Christensen, F. A. Faber, B. Huang, L. A. Bratholm, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Qml: A python toolkit for quantum machine learning, https://github.com/qmlcode/qml, 2017.

[69] K. Madanakrishna, N. Edith, S. Vincent, G. van der Rest, C. Duncan, and F. Gilles, Chem. - Eur. J. **23**, 8414 (2017).

[70] N. Frank, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1327 (2017).

[71] P. J. Stevens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1993).