EDGE ARTICLE

Check for updates

Multi-scale approach for the prediction of atomic scale properties†

Andrea Grisafi, [‡a] Jigyasa Nigam [‡abc] and Michele Ceriotti [★ab]

Electronic nearsightedness is one of the fundamental principles that governs the behavior of condensed matter and supports its description in terms of local entities such as chemical bonds. Locality also underlies the tremendous success of machine-learning schemes that predict quantum mechanical observables – such as the cohesive energy, the electron density, or a variety of response properties – as a sum of atom-centred contributions, based on a short-range representation of atomic environments. One of the main shortcomings of these approaches is their inability to capture physical effects ranging from electrostatic interactions to quantum delocalization, which have a long-range nature. Here we show how to build a multi-scale scheme that combines in the same framework local and non-local information, overcoming such limitations. We show that the simplest version of such features can be put in formal correspondence with a multipole expansion of permanent electrostatics. The data-driven nature of the model construction, however, makes this simple form suitable to tackle also different types of delocalized and collective effects. We present several examples that range from molecular physics to surface science and biophysics, demonstrating the ability of this multi-scale approach to model interactions driven by electrostatics, polarization and dispersion, as well as the cooperative behavior of dielectric response functions.

1 Introduction

The broad success of machine-learning approaches, used to predict atomic-scale properties bypassing the computational cost of first-principles calculations,[1–14] can be largely traced to the use of structural descriptors that are defined through localized atomic environments.[15–17] The assumption of locality is supported by the principle of nearsightedness of electronic matter first introduced by Walter Kohn,[18] which implies that far-field perturbations to the local properties of the system are usually screened, and exponentially-decaying. Locality has been long exploited to develop linear-scaling electronic-structure schemes,[19–23] and in the context of machine-learning methods it allows constructing models that are highly transferable and applicable to diverse problems as well as to complex, heterogeneous datasets.[24]

Structural descriptors that are built using only local information cannot, however, describe long-range interactions and non-local phenomena. In many contexts, particularly when describing homogeneous, bulk systems,[6] long-range tails can be incorporated in an effective way or approximated by increasing the range of the local environments.[25] On a fundamental level, however, the use of nearsighted representations undermines the reliability of machine-learning approaches whenever strong electrostatic and polarization effects guide the macroscopic behavior of the system. This is, for instance, the case when considering the electrostatic screening properties of water and electrolyte solutions,[26–31] the collective dispersion interactions that stabilize molecular crystals and biomolecules,[32,33] or the surface charge polarization of a metal electrode in response to an external electric field.[34–38] Several examples have been presented that demonstrate the shortcomings of local ML models in the presence of long-range physical effects.[39–41]

Global representations that incorporate information on the entire system exist,[42–45] but usually they reduce the transferability of the resulting model. In the context of modelling electronic potential energy surfaces, several strategies have been proposed to incorporate explicitly the physical effects that underlie long-range interactions. Approaches that use machine learning together with an explicit description of the electrons or the electron charge density[46–50] have the potential to also address this issue, but are considerably more cumbersome than ML schemes that use only the nuclear coordinates as inputs. Baselining the model with a cheaper electronic-structure method that incorporates electrostatic contributions,[2,10,51–53] fitting separately models for long-range contributions based on

aLaboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. E-mail: michele.ceriotti@epfl.ch
bNational Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
cIndian Institute of Space Science and Technology, Thiruvananthapuram 695547, India

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc04934d

‡ These authors contributed equally to this work.

physics-inspired functional forms,[54] or using free-energy perturbation to promote a short-range ML potential to full quantum chemical accuracy[14] are very effective, pragmatic approaches to circumvent the problem. Alternatively, one can directly machine-learn the atomic partial charges and multipoles that enter the definition of the electrostatic energy,[55–61] model the atomic polarizability that underlies dispersion interactions,[62] or atomic electronegativities that are then used to determine the partial charges of the system by minimizing its electrostatic energy.[63,64] The major shortcoming of these methods is that, on one side, they are highly system dependent and, on the other, they are limited to the prediction of energy-related properties, and to the specific physical interaction that they are designed to model. Some of the present Authors have recently proposed an alternative approach to incorporate non-local interactions into an atom-centred ML framework. Non-local information of the system is folded within local atomic environments thanks to the definition of smooth Coulomb-like potentials that are subsequently symmetrized according to the nature of the target property.[65] The resulting long-distance equivariant (LODE) representation is endowed with a long-range character while still being defined from the information sampled in a finite local neighbourhood of the atoms.

In this work, density and potential based descriptors are combined within a unified multi-scale representation. The resulting model can be formally related to an environment-dependent multipolar expansion of the electrostatic energy, but has sufficient flexibility to yield accurate predictions for a number of different kinds of interactions, and regression targets. We first consider, as an example, a dataset of organic dimers, partitioned into pairs that are representative of the possible interactions between charged, polar and apolar groups, demonstrating that the multi-scale LODE features can be used to describe permanent electrostatics, polarization and dispersion interactions with an accuracy that is only limited by the number of training points. We then show how our model is able to capture the mutual polarization between a water molecule and a metal slab of lithium. Finally, we reproduce the dipole polarizability of a dataset of poly-aminoacids, extrapolating the electric response of the system at increasing chain lengths.

2 Multi-scale equivariant representations

The general problem we intend to tackle is that of representing an atomic configuration in terms of a vector of features, in a way that is both complete and concise, and that reflects the structural and chemical insights that are most important in determining the target properties. Let us start by defining a density field associated with a structure A as the superposition of decorated atom-centred Gaussians

$$\langle a\mathbf{x}|A; \rho \rangle = \sum_{i \in A} \delta_{aa_i} g(\mathbf{x} - \mathbf{r}_i). \qquad (1)$$

In this expression, $\mathbf{r}_i$ indicates the position of atoms of A, and a_i labels their elemental nature. In analogy with the Dirac notation used in quantum mechanics, the bra–ket $\langle X|A;\text{rep}\rangle$ refers to a representation of the structure A, whose nature is described by a set of labels "rep" (such as ρ, or V further down), and whose entries are indexed by one or more variables X (such as a or $\mathbf{x}$). In the limit of a complete basis, $|A;\text{rep}\rangle$ is independent of the choice of $\langle X|$, and so the basis can be changed following analytical or numerical convenience. The notation and its usage are described in more detail in ref. 17.

From these smooth atomic densities, a Coulomb-like potential can be formally defined as a result of the integral operation

$$\langle a\mathbf{x}|A; V \rangle = \int d\mathbf{x}' \frac{\langle a\mathbf{x}'|A; \rho \rangle}{|\mathbf{x} - \mathbf{x}'|}. \qquad (2)$$

A schematic representation of this construction is reported in Fig. 1. One could build a general family of fields using a different integral transformation of the density, but here we focus on this $1/|\mathbf{x} - \mathbf{x}'|$ form which is well-suited to describe long-range interactions. The two primitive representations $|\rho\rangle$ and $|V\rangle$ can be individually symmetrized over the continuous translation group.[17] Imposing translational invariance on eqn (1) has the ultimate effect of centring the representation on the atoms i of the system, so that we can conveniently refer to the set of atom-centred densities§

$$\langle a\mathbf{x}|A; \rho_i \rangle = \sum_{j \in A} \delta_{aa_j} g(\mathbf{x} - \mathbf{r}_{ji}), \qquad (3)$$

where $\mathbf{r}_{ji} = \mathbf{r}_j - \mathbf{r}_i$ is the vector separating atoms i and j. As already shown in ref. 65, given that the Coulomb operator is translationally invariant, one can obtain an analogous result symmetrizing the tensor product $|\rho\rangle \otimes |V\rangle$, yielding a set of atom-centred potentials

$$\langle a\mathbf{x}|A; V \rangle_i = \int d\mathbf{x}' \frac{\sum_{j \in A} \delta_{aa_j} g(\mathbf{x}' - \mathbf{r}_{ji})}{|\mathbf{x} - \mathbf{x}'|}. \qquad (4)$$

The cost of the naive evaluation of $\langle a\mathbf{x}|A;V_i\rangle$ scales with the square of the number of particles in the system but a more favorable scaling can be obtained by applying one of the many schemes used to accelerate the solution of the Poisson equation in atomistic simulations.[66]

Either of eqn (3) or (4) contains information on the entire structure. Usually, however, the atom-centred density $|\rho_i\rangle$ is evaluated including only atoms within spherical environments of a given cutoff radius r_c. This truncation is not only a matter of practical convenience: the nearsightedness principle[18] indicates that molecular and materials properties are largely determined by local correlations, and increasing indefinitely r_c has been shown to reduce the accuracy of the model[24,67] because, in the absence of enormous amounts of uncorrelated training structures, the increase in model flexibility leads to overfitting. The fundamental intuition in the construction of the atom-density potential $|V_i\rangle$ is that, even if one evaluates it in a spherical neighbourhood of the central atom i, thereby avoiding an uncontrollable increase in the complexity of the model, it
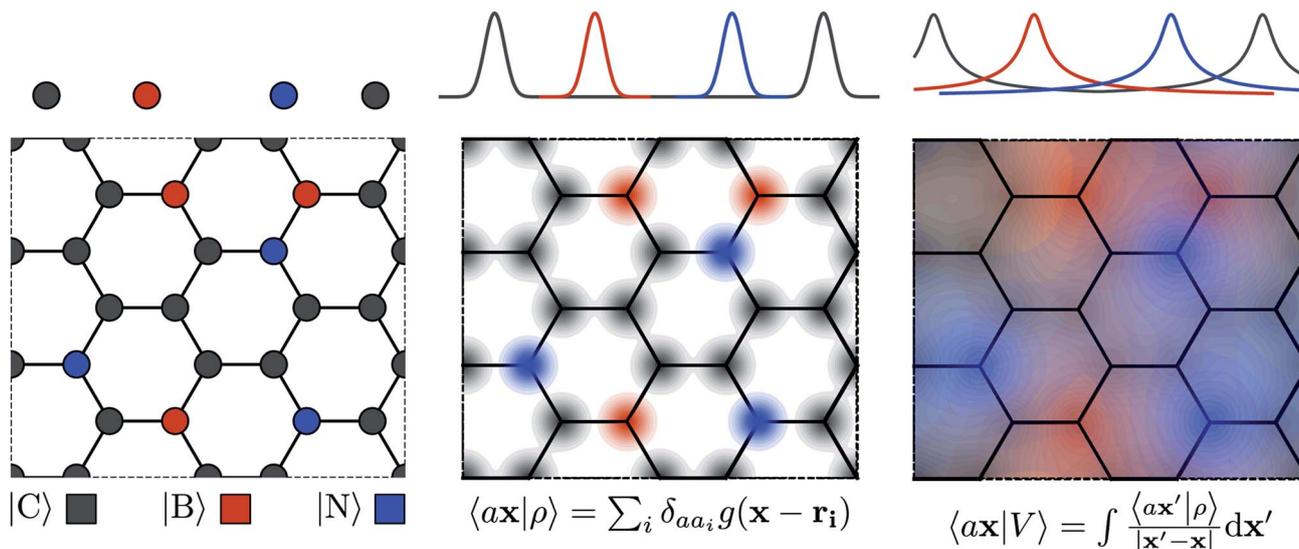
Fig. 1 Schematic representation of the construction of atomic–field representations. Top: 1-D atomic chain; bottom: hypothetical "doped graphene" 2-D system. Left: atoms are represented by their position in Cartesian coordinates. Middle: the structure is represented as an atom-density field, built as a superimposition of localized functions centred on each atom; each element is associated with a separate channel, represented by color coding. Right: the atom-density potential is formally equivalent to the electrostatic field generated by the decorated atom density; similarly to the density, each element is treated separately.

incorporates contributions from atoms that are very far away. The nature of $|V_i\rangle$ can be better understood by separating the near-field from the far-field potential in the definition of eqn (4), that is,

$$\langle a\mathbf{x}|V_i\rangle = \langle a\mathbf{x}|V_i^<\rangle + \langle a\mathbf{x}|V_i^>\rangle = \int \mathrm{d}\mathbf{x}' \frac{\langle a\mathbf{x}'|\rho_i^<\rangle}{|\mathbf{x}-\mathbf{x}'|} + \int \mathrm{d}\mathbf{x}' \frac{\langle a\mathbf{x}'|\rho_i^>\rangle}{|\mathbf{x}-\mathbf{x}'|} \tag{5}$$

where $\rho_i^<$ and $\rho_i^>$ are the atomic densities located inside and outside the i-th spherical environment. We omit the structure label A for convenience, as we will do often in what follows. The near-field term contributes information that is analogous to that included in $|\rho_i\rangle$. The far-field contribution instead determines the effect of the density beyond r_c, and the choice of the integral operator affects the asymptotic form of this effect, with $1/|\mathbf{x}-\mathbf{x}'|$ implying a Coulomb-like behavior.

Tensor products of the atom-centred density eqn (3) and potential (4) could be separately symmetrized over rotations and inversion, yielding respectively structural descriptors of short-range interatomic correlations, equivalent to SOAP-like representations,[16] or long-distance equivariants (LODE) features.[65] Here we introduce a more explicitly multi-scale family of representations, that couples $|\rho_i\rangle$ and $|V_i\rangle$ terms. Formally, one can obtain a symmetry-adapted ket that transforms like the irreducible representations of the O(3) group by computing the Haar integral over improper rotations¶ $\hat{S}$:

$$\left| \left\langle \rho_i^{\otimes\nu} \otimes V_i^{\otimes\nu'} \otimes \sigma \otimes \lambda\mu \right\rangle_{O(3)} \right\rangle = \int_{O(3)} \mathrm{d}\hat{S}\, \hat{S}|\sigma\rangle \otimes \hat{S}|\lambda\mu\rangle \otimes \underbrace{\hat{S}|\rho_i\rangle \otimes \dots \hat{S}|\rho_i\rangle}_{\nu \text{ times}} \otimes \underbrace{\hat{S}|V_i\rangle \otimes \dots \hat{S}|V_i\rangle}_{\nu' \text{ times}}, \tag{6}$$

which we indicate in what follows using the shorthand notation $\overline{|\rho_i^{\otimes\nu} \otimes V_i^{\otimes\nu'}; \sigma; \lambda\mu\rangle}$, omitting the $\sigma;\lambda\mu$ indices when considering invariant features ($\sigma = 1$, $\lambda = 0$). Within this construction, the ket $|\lambda\mu\rangle$ has the role of making the resulting features transform as a Y_λ^μ spherical harmonic,[39,68] while $|\sigma\rangle$ indicates the parity of the features under inversion.‖ eqn (6) might be intimidating, but it simply indicates a general, abstract recipe to combine different scalar fields that describe a local atomic environment in a way that is independent on the basis chosen to describe such fields. For instance, the ($\nu = 0$, $\nu' = 1$) invariant case can be readily evaluated by using a real-space basis to evaluate $\langle \mathbf{x}|V_i\rangle$

$$\langle x|\overline{V_i^{\otimes 1}}\rangle = \int \mathrm{d}\hat{R}\langle x\hat{\mathbf{x}}|\hat{R}|V_i\rangle = 2\pi \int \mathrm{d}\hat{\mathbf{x}}'\langle x\hat{\mathbf{x}}'|V_i\rangle, \tag{7}$$
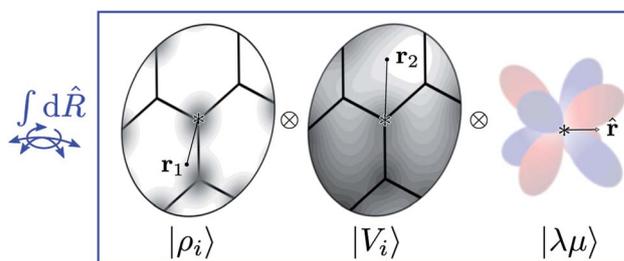

Fig. 2 A schematic representation of the Haar integral in eqn (6). Different representations of an environment centred on atom i are combined as tensor products (*i.e.* evaluated at different points, or on different basis functions) and averaged over all possible rotations of the system. Including also a set of spherical harmonics provides an absolute reference system and makes it possible to build ML models endowed with an equivariant behavior.

corresponding to a spherical average of the atom-centred potential. In the more general case, as sketched in Fig. 2, eqn (6) can be understood as the average of the product of atom-centered fields evaluated at $\nu + \nu'$ points.

In practical implementations, the abstract ket (6) can be computed by first expanding the atom-centred features (3) and (4) onto a discrete basis, and then evaluating the symmetrized ν-point correlation of the fields. A particularly clean, efficient, recursive formulation can be derived exploiting the fact that the equivariant features behave as angular momenta, and can then be combined using Clebsch–Gordan coefficients to build higher-order correlations.[69] In analytical derivations we use a partially-discretized basis, in which the radial contribution is kept as a continuous index, corresponding to

$$\langle arlm| = \int d\mathbf{x}\, \delta(r - x)\langle lm|\hat{\mathbf{x}}\rangle\langle a\mathbf{x}|, \qquad (8)$$

with $\langle \hat{\mathbf{x}}|lm\rangle \equiv Y_l^m(\hat{\mathbf{x}})$. Written in this basis, $\langle arlm|\rho_i\rangle$ expresses the decomposition of the density in independent angular momentum channels, evaluated at a distance r from the central atom. In practical implementations we use a basis of Gaussian type orbitals to also discretize the radial component.[68] This is the form that is usually chosen to write and compute the local SOAP features,[16] that can be evaluated as

$$\langle a_1 n_1; a_2 n_2; l|\overline{\rho_i^{\otimes 2}}\rangle \propto \sum_m \frac{(-1)^m}{\sqrt{2l+1}} \langle a_1 n_1 lm|\rho_i\rangle\langle a_2 n_2 l(-m)|\rho_i\rangle \qquad (9)$$

The nature of the representation, however, does not depend on such details. The basis-set independence is most clearly seen by considering the use of the equivariants in the context of a linear regression model. The value of a tensorial property $\mathbf{T}$ for a structure, expressed in its irreducible spherical components (ISC,[70] the combinations of the components of a Cartesian tensor that tranform under rotation as Y_λ^μ) and decomposed in atom-centred contributions, can be formulated as

$$T_\mu^{\sigma\lambda}(A, i) \approx \int dX \langle T; \sigma\lambda|X\rangle\big\langle X|A; \overline{\rho_i^{\otimes \nu} \otimes \rho_i^{\otimes \nu'}; \sigma; \lambda\mu}\big\rangle, \qquad (10)$$

where X indicates any complete basis that provides a concrete representation of the ket, and $\langle X|T;\sigma\lambda\rangle$ is the set of regression weights. One sees that (1) the regression model is invariant to a unitary transformation of the basis; (2) the equivariant nature of the model is associated with the $\lambda\mu$ indices of the ket, while the weights are invariant under symmetry operations. Linear models are especially useful to reveal the physical meaning of a representation: they allow to demonstrate the relation between short-range density correlations ($\nu' = 0$) and the body-order expansion of interatomic potentials,[17,71–73] and the relation between the first-order LODE($\nu = 0$, $\nu' = 1$) and fixed point-charge electrostatics. In the next section, we use this idea to show how the simplest multi-scale LODE($\nu = 1$, $\nu' = 1$) can be put in formal correspondence with the physics of multipole electrostatics.[74]

3 Features and models for long-range interactions

Even though neither the atom density $|\rho_i\rangle$ nor the associated potential field $|V_i\rangle$ correspond to physical quantities, the multi-scale combination of the two quantities in LODE($\nu = 1$, $\nu' = 1$) entails formal similarities with physics-based electrostatic models. This connection can be demonstrated both analytically and with numerical benchmarks.

3.1 Analytical connection with the multipole expansion

Consider a linear model to predict the ground-state electronic energy U of a system A. This corresponds to taking the scalar ($\lambda = 0$) and polar ($\sigma = 1$) limits within the prediction formula of eqn (10):

$$U(A) = \sum_{i=1}^N U_i(A) = \sum_{i=1}^N \int dX \langle U|X\rangle\langle X|A; \overline{\rho_i \otimes V_i}\rangle. \qquad (11)$$

We aim to prove that in the LODE(1,1) case, where the density and potential representations are both introduced to first order, this functional form can be used to model rigorously a multipolar expansion of the long-range contributions to U.

To see this, let us start by representing the energy prediction in terms of the partially-discretized basis of eqn (8). Upon symmetrization of the tensor product between ρ and V, and going to the coupled angular momentum basis,[69] one obtains a set of invariants that can be expressed using the basis $\langle X| \equiv \langle a_1 r_1; a_2 r_2; l|$

$$\langle a_1 r_1; a_2 r_2; l|\overline{\rho_i \otimes V_i}\rangle = \frac{1}{\sqrt{2l+1}} \sum_{|m| \le l} \langle a_1 r_1 lm|\rho_i^<\rangle\langle a_2 r_2 lm|V_i\rangle^*, \qquad (12)$$

where $\langle a_1 r_1 lm|\rho_i^<\rangle$ and $\langle a_2 r_2 lm|V_i\rangle$ indicate the spherical harmonics projections of the local density and the local potential fields respectively, and we omit the indication of the structure A, for brevity. Eqn (12) shows explicitly how $|\overline{\rho_i \otimes V_i}\rangle$ contains information on the correlation between the value of the atom density $|\rho_i\rangle$ and the potential $|V_i\rangle$, each evaluated at a given distance from the central atom. General symmetry considerations dictate how angular terms in the two correlations must be combined to obtain a rotationally invariant set of features, in clear analogy with the construction of the SOAP representation (9). By using in eqn (12) the splitting of the potential field in short and long-range parts, $|V_i\rangle = |V_i^<\rangle + |V_i^>\rangle$, we can partition the prediction for the atom-centred energy contribution in range separated terms, $U_i = U_i^< + U_i^>$. Focusing in particular on the long-range contribution, we can write explicitly eqn (11) as follows:

$$U_i^> = \sum_{l=0}^{l_{max}} \sum_{a_1 a_2} \int_0^{r_c} dr_1 \int_0^{r_c} dr_2 \langle U|a_1 r_1; a_2 r_2; l\rangle \times \sum_{|m| \le l} \langle a_1 r_1 lm|\rho_i^<\rangle$$
$$\times \langle a_2 r_2 lm|V_i^>\rangle^*. \qquad (13)$$

Here, $\langle U|a_1r_1;a_2r_2;l\rangle$ indicates the regression weights for the *total* potential energy, that also incorporate the $1/\sqrt{2l+1}$ factor in eqn (12). We are now interested in representing the spherical harmonic components of the potential in terms of the far-field contribution $\tilde{V_i}$ of eqn (5). Using the Laplace expansion of the Coulomb operator, we can rewrite $|\tilde{V_i^>}\rangle$ as:

$$\langle a_2r_2lm|V_i^>\rangle^* = \frac{4\pi}{2l+1}\int d\mathbf{x}\langle\rho_i^>|a_2\mathbf{x}\rangle\frac{r_2^l}{x^{l+1}}\langle\hat{\mathbf{x}}|lm\rangle$$
$$= \frac{4\pi}{2l+1}\int dr\langle\rho_i^>|a_2rlm\rangle\frac{r_2^l}{r^{l+1}}. \qquad (14)$$

Plugging this into eqn (13), one sees that the contribution to the energy coming from the far-field can be written as

$$U_i^> = \sum_{a_1a_2}\sum_{lm}\int_{r_c}^\infty dr\frac{1}{r^{l+1}}\langle\rho_i^>|a_2rlm\rangle\langle a_1a_2;lm|M_i^<(U)\rangle, \qquad (15)$$

where we introduce a set of spherical multipoles

$$\langle a_1a_2;lm|M_i^<(U)\rangle = \frac{4\pi}{2l+1}\int_0^{r_c}dr_2r_2^l\int_0^{r_c}dr_1\langle U|a_1r_1;a_2r_2;l\rangle$$
$$\times\langle a_1r_1lm|\rho_i^<\rangle \qquad (16)$$

that are determined indirectly as a consequence of the fitting of the energy, rather than by explicit calculations of the density distribution around the molecular components of the system.

Eqn (15) shares a striking resemblance with the expression for the interaction of a far-field charge density with the electrostatic potential generated by the near-field charge distribution.[75] As we shall see in what follows, this formal equivalence underpins the ability of $|\overline{\rho_i\otimes V_i}\rangle$ to model accurately several kinds of interactions. Crucially, however, ρ_i and V_i do not represent physical quantities, but are just a representation of the spatial arrangement of atoms. Atoms in the far-field respond in a way that depends only on their chemical nature, but the local multipoles are modulated in a highly flexible, non-trivial fashion by the distribution of atoms in the local environment. The form of eqn (16) also hints at how changing the representation would affect this derivation. Increasing the density order ν would allow for a more flexible, higher-body-order dependence of the local multipoles on the distribution of atoms in the vicinity of atom i, while increasing ν' would bring a more complicated dependency on the distribution of atoms in the far-field, leading to a linear regression limit that does not match formally the electrostatic multipole expansion and to an explicit coupling between $|\tilde{V_i^>}\rangle$ and $|\tilde{V_i^<}\rangle$. Changing the asymptotic form of the potential in eqn (2) could be used to incorporate a formal connection with dispersion-like, $1/r^6$ features. We want to stress that even in this form the model is not limited to describing the physics of permanent electrostatics. In fact, the coupling between the inner and outer atomic species (a_1 and a_2) carried by the definition of the regression weights makes it possible for the local multipoles to respond to species of the far-field distribution. We test the limits of this data-driven approach in Section 4.

3.2 A toy model for multipolar interactions

Before doing so, we want, however, to demonstrate numerically the mapping between these multi-scale LODE features and multipole models. We analyze quantitatively the behavior of a linear model based on $|\overline{\rho_i\otimes V_i}\rangle$ by observing its performance in representing the far-field interactions between an H_2O and a CO_2 molecule – since the interactions between the two molecules are essentially driven by permanent electrostatics. We build a dataset considering 33 non-degenerate reciprocal orientations between the two molecules, and learn the interaction over a range of distances between the centres of mass from 6.5 to 9 Å. We then extrapolate the predicted interaction profile in the asymptotic regime of $R > 9$ Å, verifying how the model converges towards the dissociated limit which is also included in the training set.

According to our construction, the cutoff value l_{max} chosen to define the angular resolution of the representation determines the number of multipoles that are included within the expansion of eqn (14). For example, taking the $l_{max} = 0$, $g \to \delta$, $r_c \to 0$ limits of eqn (15) leads to

$$U_i^> \approx \sum_{a_2}\int_{0_+}^\infty dr\frac{1}{r}\langle\delta_i^>|a_2r00\rangle\langle a_1a_2;00|M_i^>(U)\rangle, \qquad (17)$$

where we define the $g \to \delta$ limit of the atom density as

$$\langle a_2r|\delta_i^>\rangle = \sum_{j\neq i}\delta_{a_ja_2}\delta(r-|r_{ji}|), \qquad (18)$$

and we consider that the only atom inside the inner cutoff is the central atom, so the multipole coefficient depends only on the chemical nature of a_i and a_2. By evaluating explicitly the radial integral in eqn (17), one obtains an explicit sum over atom pairs

$$U_i^> \approx \sum_{j\neq i}\frac{\langle a_ia_j;00|M^<(U)\rangle}{r_{ji}}. \qquad (19)$$

If one interprets $\langle a_ia_j;00|M^<(U)\rangle$ as the product of the partial charges of the two species q_{a_i}, and q_{a_j}, this form is equivalent to a simple, Coulomb interaction energy between fixed point-charges. Including multipoles for $l > 0$ makes it possible to represent the anisotropy of the electrostatic interaction.

In Fig. 3 we report the results of the extrapolation for a given reciprocal orientation at increasing angular cutoffs l_{max}. We also compare different choices for the possible atomic centres that contribute to the energy prediction: in panel (a) we express the energy in terms of a single environment centred on the oxygen atom of the H_2O molecule; in panel (b) we use a single environment centred on the carbon atom of CO_2; in (c) we use multiple environments centred on each atom. This exercise probes the possibility of choosing between a model for the electrostatic energy that is based on the definition of molecular rather than atomic multipoles.[57,62] As one would expect from a classical interpretation of the long-range energy, the binding profile for the selected test configuration is ultimately driven by the interaction between the dipole moment of the water molecule and the quadrupole moment of CO_2. This is reflected in the sharp transition of the prediction accuracy when crossing
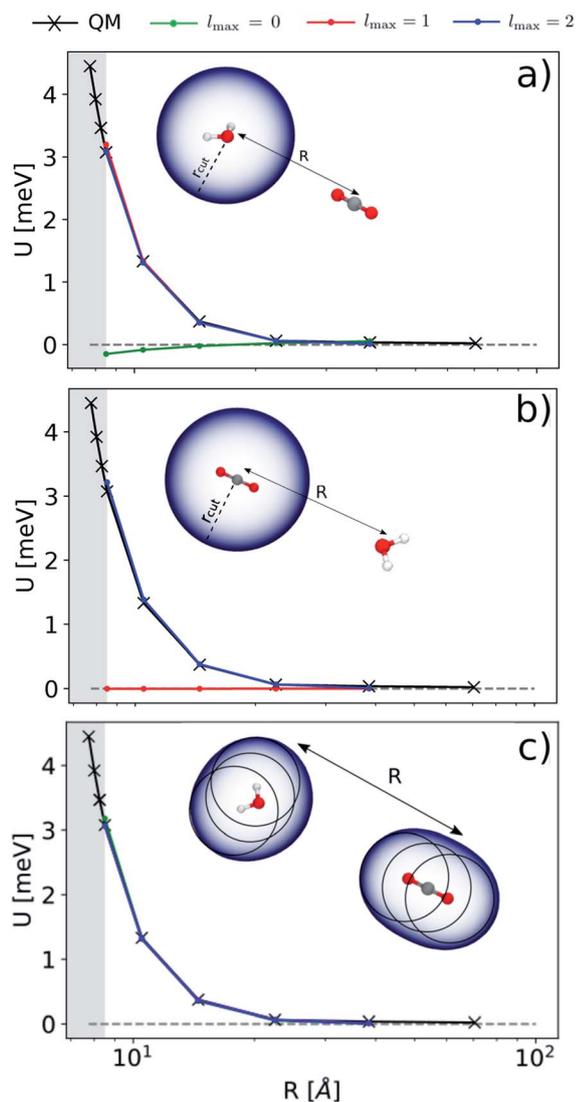
Fig. 3 Extrapolated interaction profiles for a given configuration of H_2O and CO_2 at different angular cutoff values l_{max}. Top, middle and bottom panels show the results of the asymptotic extrapolation when centring the representation on the oxygen atom of H_2O, the carbon atom of CO_2 and all the atoms of the system respectively.

a critical angular cutoff l_{max}. When centring the local environment on the water molecule (Fig. 3(a)), for instance, truncating the expansion at $l_{max} = 1$ is enough to reproduce the interaction between the dipolar potential of water and the CO_2 molecule. Conversely, when centring the representation on carbon dioxide (Fig. 3(b)), the H_2O density in the far-field has to interact with a CO_2 potential that is quadrupolar in nature, which requires an angular cutoff of at least $l_{max} = 2$. When centring the representation on all the atoms of the system (Fig. 3(c)), using an angular cutoff of $l_{max} = 0$ suffices to obtain qualitatively accurate interaction profiles.

The analogy between $|\overline{\rho_i \otimes V_i}\rangle$-based models and the multipole expansion raises the question of the relationship with ML electrostatic models based on atomic point charges. Traditional parametrized force fields as well as machine-learning potentials
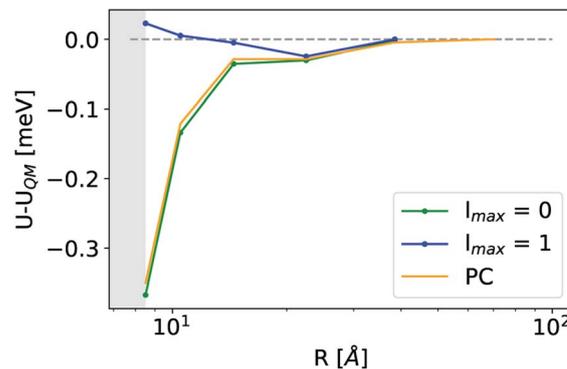

Fig. 4 Asymptotic prediction errors of a representative CO_2/H_2O configuration obtained at different levels of theory. Green and blue lines correspond to $l_{max} = 0$ and $l_{max} = 1$ LODE(1,1) models while the orange line refers to a fixed point–charges model.

that simply rely on representing the electrostatic energy of the system *via* a set of point-charges[56,63] generalize this form by making the atomic charges dependent on the local environment, and/or on overall charge neutrality conditions. Given that here we use rigid molecules, moving beyond the range of the local featurization, these ML schemes are well-approximated by a model based on fixed partial charges for C, H, O, and pairwise Coulomb interactions. Such a model yields binding curve profiles and overall accuracy similar to those of a LODE(1,1) model truncated at $l_{max} = 0$, which is consistent with the limiting case of eqn (19) (optimal charges correspond to $q_H = 0.24e$, $q_C = 0.96e$, $q_0 = -0.49e$, Fig. 4). The approach we take here is, instead, to increase the order of the expansion, and to use the additional flexibility to improve the accuracy of the model in a data-driven fashion, which allows to improve the accuracy further, particularly in the intermediate distance range.

On a conceptual level, the issue is to find the balance between a functional form that is flexible enough to describe arbitrary interactions, and one that maps naturally onto the physics of the interactions of interest. For this simple toy problem, increasing the expansion at $l_{max} = 1$ with an atomic multipole model achieves almost perfect predictions. However, a *too* general form is prone to overfitting and requires enormous amounts of training data: this is the case, for instance, one would encounter when increasing by brute force the cutoff of a local featurization.[65,67] The scattering transform[45] provides an entirely general framework that, similarly to the one we discuss here, aims at achieving a multi-scale description of interactions. The considerable improvement of its performance that is observed when applying feature selection[45,76,77] indicates a similar tendency to overfitting.

4 Results

The toy system we have discussed in the previous section reflects the behavior of the multi-scale LODE representation. In this section we present three applications to substantially more complicated systems to demonstrate that, even in their simplest

form, this family of features is suitable to address the complexity of challenging, real-life atomistic modelling problems, and physics well beyond that of permanent electrostatics. We compare multi-scale LODE models with those based on local machine-learning schemes, in particular with SOAP features.[16] We indicate these as the two-point density correlation $|\overline{\rho_i^{\otimes 2}}\rangle$ to stress that similar results are to be expected from any equivalent local featurization[17] such as atom-centered symmetry functions,[78] SNAP,[79] MTP,[80] ACE,[72] NICE.[69] We report errors in terms of the root mean square error (RMSE), or the percentage RMSE (RMSE%), which is expressed as a percentage of the standard deviation of the target properties.

4.1 Binding energies of organic dimers

We start by testing the ability of multi-scale LODE to describe different kinds of molecular interactions. To this end, we consider the interaction energy between 2291 pairs of organic molecules belonging to the BioFragment Database (BFDb).[81] For each dimer configuration, binding curves are generated by considering 12 rigid displacements in steps of 0.25 Å along the direction that joins the geometric centres of the two molecules. Then, unrelaxed binding energies are computed at the DFT/PBE0 level using the Tkatchenko–Scheffler self-consistent van der Waals method[82] as implemented in the FHI-aims package.[83] For each binding trajectory, we also include in the training set the dissociated limit of vanishing interaction energy, where the two monomers are infinitely far apart. The dataset so generated includes all the possible spectrum of interactions, spanning pure dispersion, induced polarization and permanent electrostatics. In order to better rationalize the learning capability of such a large variety of molecular interactions, we choose to partition the molecules in the dataset in three independent classes, namely, (1) molecules carrying a net charge, (2) neutral molecules that contain heteroatoms (N, O), and can therefore

exhibit a substantial polarity (3) neutral molecules containing only C and H, that are considered apolar and interacting mostly through dispersive interactions. Considering all the possible combinations of these kinds of molecules partitions the dimers into six classes, i.e., 184 charged–charged (CC), 267 charged-polar (CP), 210 charged-apolar (CA), 161 polar–polar (PP), 418 polar–apolar (PA) and 1051 apolar–apolar (AA) interactions. For each of the six classes, several, randomly selected binding curves are held out of the training set, to test the accuracy of our predictions. The remaining curves are used to fit one separate linear model for each class, using either local features (the SOAP power spectrum,[16] $|\overline{\rho_i^{\otimes 2}}\rangle$) or multi-scale LODE($\nu = 1$, $\nu' = 1$) features. In order to also assess the reliability of our predictions, we use a calibrated committee estimator[84] for the model uncertainty, which allows us to determine error bars for the binding curves. 8 random subselections of 80% of the total number of training configurations are considered to construct the committee model. The internal validation set is then defined by selecting the training structures that are absent from at least 25% of the committee members.

Fig. 5 shows characteristic interaction profiles for the six different classes of molecular pairs. The models use $r_c = 3$ Å environments centred on each atom. The configurations we report are those that exhibit median integrated errors within the test set of each class. The root mean square errors associated with the predictions over the entire test sets of each class are listed in Table 1. The results clearly show that while SOAP(2) is limited by the nearsightedness of the local environments, the LODE(1,1) multi-scale model is able to predict both the short and the long-range behaviour of the binding profiles on an equal footing. What is particularly remarkable is the fact that a simple, linear model can capture accurately different kinds of interactions, that occur on wildly different energy scales and asymptotic behavior: the typical binding energy of charged
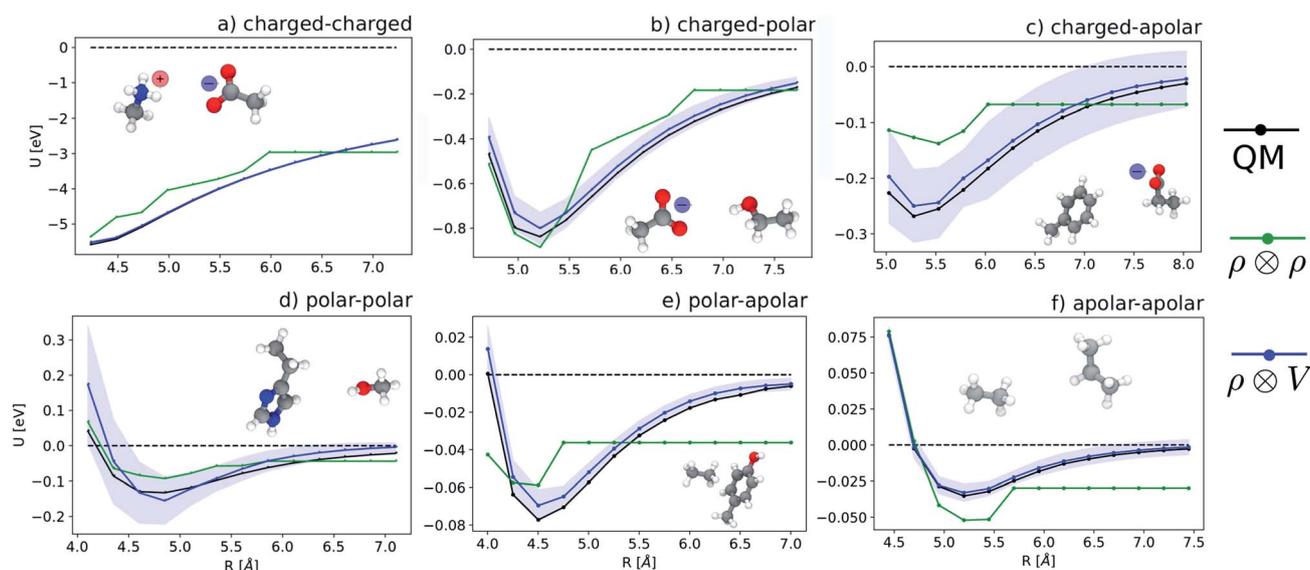

Fig. 5 Median-error binding curves for six different classes of intermolecular interactions. (black lines) Quantum-mechanical calculations. (green lines) Predictions of a $|\overline{\rho_i \otimes \rho_i}\rangle$ model. (blue lines) Predictions of a $|\overline{\rho_i \otimes V_i}\rangle$ model.

Table 1 Prediction performance expressed in terms of the RMSE over all the points of the binding curves, for the six classes of interactions and $\rho \otimes \rho$, $\rho \otimes V$ and $V \otimes V$ models. For each class we also indicate the number of training samples, and the characteristic energy scale, expressed in terms of the standard deviation of the energies in the test set

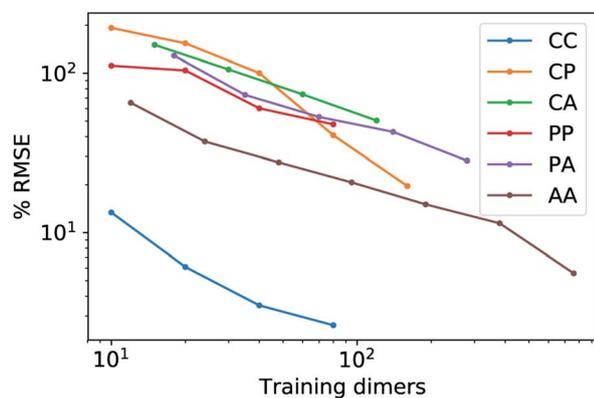|			RMSE/eV		
Class	n_{train}	STD/eV	$\rho \otimes \rho$	$\rho \otimes V$	$V \otimes V$
CC	100	1.86	0.72	0.049	0.058
CP	200	0.379	0.25	0.074	0.092
CA	150	0.083	0.056	0.041	0.034
PP	100	0.131	0.10	0.062	0.125
PA	350	0.046	0.032	0.013	0.021
AA	950	0.063	0.026	0.004	0.006

Fig. 6 Learning curves for the 6 classes of molecular interactions computed using the LODE(1,1) representation. The curves indicate that all interactions can be learned with comparable efficiency and that the accuracy of the model is limited by the small number of available reference structures. Interactions between charged molecules, that have a formal connection with the form of the multi-scale features, can be learned effectively with a small number of training samples.

dimers is of the order of several eV, and has a $1/r$ tail, while the typical interaction energy of two apolar molecules is of the order of a few 10 s of meV, and decays roughly as $1/r^6$.

A LODE($\nu' = 2$) model (*i.e.* based on $|\overline{V_i^{\otimes 2}}\rangle$ features) also allows to predict the binding curves beyond the 3 Å cutoff, but usually yields 50–100% larger errors than those observed with $|\overline{\rho_i \otimes V_i}\rangle$ – not only for charged molecules, but also for dimers that are dominated by dispersion interactions. The multi-scale nature of LODE($\nu = 1$, $\nu' = 1$) yields a better balance of short and long-range descriptions, and is sufficiently flexible to be adapted to the description of systems that are not dominated by permanent electrostatics, even though interactions between charged fragments are considerably easier to learn, in comparison to the others. We also observe that the uncertainty model works reliably as the predicted curves always fall within the estimated error bar. Larger uncertainties are found for interaction classes that have few representative samples in the training set, such as those associated with polar–polar molecular pairs (Fig. 5(d)). The learning curves, plotted in Fig. 6,

provide insights into the performance of LODE(1,1) for different kinds of interactions. CC dimers are learned with excellent relative accuracy – which is unsurprising given the formal connection with the multipole expansion. All other classes of interactions yield a relative accuracy for a given training set size which is an order of magnitude worse (with the exception of AA interactions, whose learning performance is intermediate). However, learning curves show no sign of saturation,[85] reflecting the fact that multi-scale features have sufficient flexibility to provide accurate predictions, but that the lack of a natural connection to the underlying physics would require a larger train set size. This is consistent with the considerations we made in the previous section based on the simple H_2O/CO_2 example.

4.2 Induced polarization on a metal surface

The previous example proves that linear $|\overline{\rho_i \otimes V_i}\rangle$ models capture a wide class of molecular interactions, ranging from pure dispersion to permanent electrostatics. Beyond molecular systems, however, a large number of phenomena occur in solid state physics that are driven by long-range effects, and involve more subtle, self-consistent interactions between far-away atoms. A particularly relevant example is represented by the induced macroscopic polarization that a metallic material undergoes in response to an external electric field, which underlies fundamentally and technologically important phenomena for surface science and nanostructures.[86–88] Physics-based modelling of these kinds of systems usually exploits the fact that, for a perfectly-conductive surface, the interaction is equivalent to that between the polar molecule and the mirror image, relative to the surface plane, of its charge distribution, with an additional inversion of polarity.[89] It would not appear at all obvious that our atom-centred framework, which does not include an explicit response of the far-field atom density to the local data-driven multipole, can capture the physics of a phenomenon associated with the polarization of electrons that are delocalized over the entire extension of the metallic solid.

To benchmark the performance of multi-scale LODE in this challenging scenario we consider the interaction of a slab of bcc lithium with a water molecule that is located at various distances from the (100)-surface. We start by selecting 81 water molecule configurations, differing in their internal geometry or in their spatial orientation relative to the surface. For each of these configurations, 31 rigid displacements are performed along the (100)-direction, spanning a range of distances between 0.5 Å and 8 Å from the lithium surface. Using this dataset we compute unrelaxed binding energies at the DFT/PBE level using the FHI-aims package.[83] We converge the slab size along the periodic xy-plane, minimizing the self-interaction between the periodic images of the water molecule, resulting in a 5 × 5 unit cell repetitions and a k-points sampling of 4 × 4 × 1 Å^{-1}. We set the slab extension along the non-periodic z-direction so that the Fermi energy is converged within 10 meV, resulting in a total of 13 layers. To remove the spurious interactions along the z-axis, we set a large vacuum space of roughly

80 Å in conjunction with a correction suitable to screen the dipolar potential.[90] Following these prescriptions, we obtain attractive potential profiles for all molecular geometries and orientation, consistently with the interaction between the dipolar field of the water molecule and the induced metal polarization.

For this example, we construct $|\rho_i\rangle$ and $|V_i\rangle$ representations within spherical environments of $r_c = 4$ Å with a Gaussian-density width of $\sigma = 0.3$ Å. The regression model is trained on 75 lithium–water binding curves while the remaining 6 are used for testing the accuracy of our predictions. Fig. 7 shows a comparison between a local $|\overline{\rho_i^{\otimes 2}}\rangle$ model and a multi-scale LODE $|\overline{\rho_i \otimes V_i}\rangle$ model in learning the interaction energy of the metal slab and the water molecule for one representative test trajectory (all test trajectories are reported in the ESI†). We observe that the local SOAP description is able to capture the short-range interactions but becomes increasingly ineffective as the water molecule moves outside the atomic environment, leading to an overall error of about 19 RMSE%. This is in sharp contrast to the performance of the $|\overline{\rho_i \otimes V_i}\rangle$ representation, which can capture both the effects of electrostatic induction at a large distance and the Pauli-like repulsion at short range with the same level of accuracy, halving the prediction error to about 9%. Learning curves are shown in the ESI.†

To further investigate what aspects of the physics of the molecule–surface interaction can be captured by the model, we perform a Mulliken population analysis on the reference DFT calculations, to extract the polarization vector of the water molecule in response to the interaction with the metal, *i.e.*, $\mathbf{P}^W = \boldsymbol{\mu}^W - \boldsymbol{\mu}_0^W$, where $\boldsymbol{\mu}^W$ and $\boldsymbol{\mu}_0^W$ are the dipole moment of the water molecule in the lithium-slab system and in vacuum respectively. Physically, the polarization $\mathbf{P}^W$ involves the response of water's electrons to the rearrangement of the electronic charge in the surface triggered by the dipolar field, and so it involves explicitly a back-reaction. Furthermore, the polarization shows both a (usually larger) component along the z-axis, and a tangential component in xy-plane. To account for the vectorial nature of $\mathbf{P}^W$, we take advantage of the tensorial extension of eqn (6). To single out the long-range nature of the
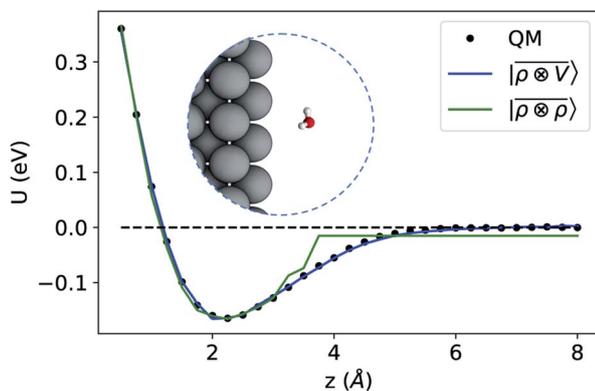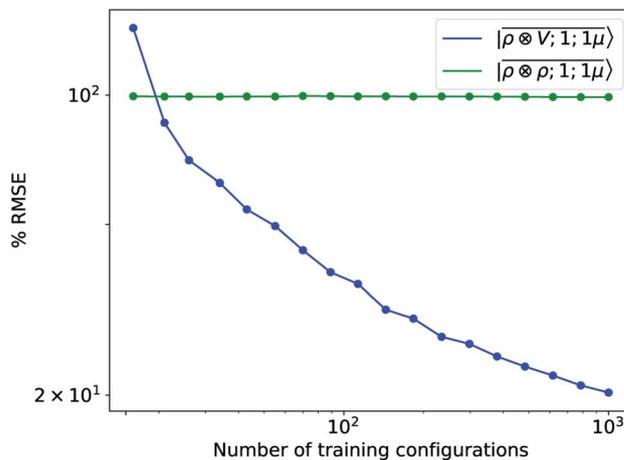

Fig. 8 Learning curves for the induced polarization of the water molecule due to interaction with image charges in the metal slab, computed only for separations greater than 4.5 Å. The error is computed as a fraction of the intrinsic variability of the test set of 215 configurations. Contrary to the local model (green), a linear $|\overline{\rho_i \otimes V_i}\rangle$ model (blue) can learn this self-consistent polarization, with no significant reduction of the learning rate up to 1000 training configurations.

polarization interaction, we restrict the regression of $\mathbf{P}^W$ to water configurations that are more than 4.5 Å far from the surface. Our dataset contains 1215 such configurations, out of which we randomly select 1000 for training, while the remaining 215 are retained for testing. Given that the training set contains no structures within the local descriptor cutoff, it comes as no surprise that a pure density-based tensor model $|\overline{\rho_i^{\otimes 2}; 1; 1\ \mu}\rangle$ entirely fails to learn the long-range polarization induced on the water molecule. Making use of the potential-based tensor model of eqn (10), in contrast, allows us to effectively learn the polarization vector $\mathbf{P}^W$, showing an error that decreases to ~20% RMSE at the maximum training set size available (Fig. 8). This example provides a compelling demonstration of the ability of $|\overline{\rho_i \otimes V_i}\rangle$ to build models of effects that go well-beyond permanent electrostatics.

4.3 Response functions of oligopeptides

As a final example, we consider the challenging task of predicting the polarizability of a dataset of poly-aminoacids. Dielectric response functions are strongly affected by long-range correlations because of the cooperative nature of the underlying physical mechanism. Poor transferability of local models between structures of different sizes has been observed for molecular dipole moments,[41] polarizability,[40] and the electronic dielectric constant of bulk water.[39] For this purpose, we use a training set composed of 27 428 conformers of single aminoacids and 370 dipeptides, testing the predictions of the model on a smaller test set containing 30 dipeptides, 20 tripeptides, 16 tetrapeptides and 10 pentapeptide configurations. Reference polarizability calculations are carried out with the Gaussian 16 quantum-chemistry code using the double-hybrid DFT functional PWPB95-D3 and the aug-cc-pVDZ basis set.[91]
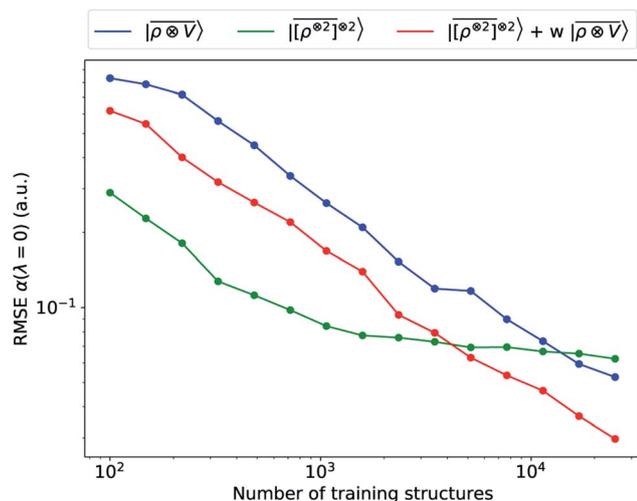

Fig. 7 Predicted binding curve of a test water–lithium configuration. (black dots) Reference DFT calculations; (green line) predictions of a $|\overline{\rho_i \otimes \rho_i}\rangle$ model; (blue line) $|\overline{\rho_i \otimes V_i}\rangle$ model.

Fig. 9 Learning curves for the $\lambda = 0$ component of the polarizability tensor of a database of polypeptide conformers. The green curve corresponds to the non-linear kernel which is equivalent to $|[\overline{\rho_i^{\otimes 2}}]^{\otimes 2}\rangle$, the blue curve to a linear kernel based on $|\overline{\rho_i \otimes V_i}\rangle$, and the red one to an optimal linear combination of the two.

Fig. 10 Absolute RMSE in learning the $\lambda = 0$ spherical tensor of polarizability of polypeptides as a function of the peptide length. The model was trained on 27 428 single-amino acids and 370 dipeptides. The error was computed on 30 dipeptides, 20 tripeptides, 16 tetrapeptides and 10 pentapeptides respectively.

We compute the multi-scale $|\overline{\rho_i \otimes V_i}\rangle$ features and their local counterparts using a Gaussian width of $\sigma = 0.3$ Å and a spherical environment cutoff of $r_c = 4$ Å. This data set is interesting, because it combines large structural variability with tens of thousands of distorted aminoacid configurations with longer-range interactions described by a few hundred dipeptide conformers.

We consider three models: a linear $|\overline{\rho_i \otimes V_i}\rangle$ multi-scale model; a square kernel model, that is equivalent to using a quadratic functional of the SOAP features, $|[\overline{\rho_i^{\otimes 2}}]^{\otimes 2}\rangle \equiv |\overline{\rho_i^{\otimes 2}}\rangle \otimes |\overline{\rho_i^{\otimes 2}}\rangle$, which incorporates 4 and 5-body correlations and enhance the many-body character of the representation at the local scale;[92] a weighted combination of the two. The learning curves for the trace ($\lambda = 0$) of the polarizability tensor, shown in Fig. 9, are very revealing of the behavior of these three models. The $|[\overline{\rho_i^{\otimes 2}}]^{\otimes 2}\rangle$ model, which disregards any non-local behavior beyond the atomic environment, is initially very efficient, but saturates to an error of 0.06 a.u. In contrast, equipped with non-local information, the $|\overline{\rho_i \otimes V_i}\rangle$ representation reduces the error of prediction to 0.05 a.u., but is initially much less effective. This is not due to the lack of higher-order local density correlations: a linear $|\overline{\rho_i^{\otimes 2}}\rangle$ model performs well, despite showing saturation due to its local nature (see discussion in the ESI†). We interpret the lackluster performance of the LODE model in the data-poor regime as an indication of the dominant role played by short-range effects in this diverse dataset, which can be learned more effectively by a nearsighted kernel, similarly to what observed in ref. 24, 67 and 93. Inspired by those works, we build a tunable kernel model based on a weighted sum of the local and the LODE kernels, that can be optimized to reflect the relative importance of the different ranges. We optimize the weight by cross-
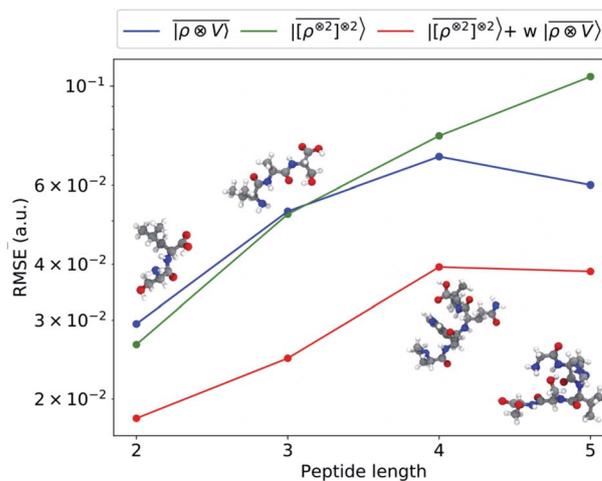
validation at the largest train size, obtaining a reduction of 50% of the test error, down to 0.028 a.u.

An analysis of the test error which separates the contributions from oligopeptides of different length, shown in Fig. 10, is consistent with this interpretation of the learning curves. All models show an error that increases with the size of the molecule because there are interactions that are just not described at the smaller train set size. However, the purely local model shows by far the worst extrapolative performance, while multi-scale models – in particular the one combining a non-linear local kernel and LODE features – show both a smaller overall error, and a saturation of the error for tetra and penta-peptides. This example illustrates the different approaches to achieve a multi-scale description of atomic-scale systems: the $|\overline{\rho_i \otimes V_i}\rangle$ features offer simplicity and physical interpretability, while a multi-kernel model makes it possible to optimize in a data-driven manner the balance between local and long-ranged correlations.

5 Conclusions

The lack of a description of long-range physical effects is one of the main limitations of otherwise greatly successful machine-learning schemes which are more and more often applied to model atomic-scale phenomena. We show how it is possible to construct a family of multi-scale equivariant features that combine the properties of well-established local ML schemes with the long-distance equivariant features that have been recently proposed by some of the authors. This multi-scale framework shows enticing formal correspondences with physically-meaningful interaction terms, such as multipole electrostatics. Still, the data driven nature of the construction allows the description of long-range interactions that do not fit this specific physical model. We show examples of how a multi-

scale LODE model can accurately predict interactions between different kinds of molecular dimers that include charged, polar and apolar compounds. Results are also very promising when it comes to modelling systems that clearly go beyond permanent electrostatics, such as a water molecule interacting with a metallic slab, and the dielectric response of oligopeptides. Our results also provide a glance at the remaining challenges in the description of long-range effects. The performance of the specific flavor of multi-scale features we discuss degrades progressively as the link of the target property with permanent electrostatics become less direct, even though this degradation translates in bigger data requirements rather than in an outright failure of the scheme. In the last example we present, that involves the tensorial response properties of oligopeptides, careful tuning of the relative importance of short and long range effects is necessary to achieve optimal transferability. Meanwhile, we have only scratched the surface of the more general idea that we introduce here. Different symmetry-adapted combinations of atom-centred fields can be computed with minor modifications of our scheme and could offer a strategy to further improvements. The combination of a physics-inspired formulation and data-driven flexibility that underlies this multi-scale LODE framework addresses one of the outstanding issues in atomistic machine-learning, and paves the way towards an even more pervasive use of statistical methods to support the computational investigation of molecules and condensed-phase systems.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

§ Strictly speaking, g in eqn (3) has twice the variance as that in (1), but we re-define the density function accordingly.

¶ Formally, improper rotations combine SO(3) rotation operators $\hat{R}$ and inversion $\hat{i}$, so $\int_{O(3)} d\hat{S} \equiv \sum_{k=0,1} \int_{SO(3)} d\hat{R}$ with $\hat{S} \equiv \hat{i}^k \hat{R}$.

‖ In particular, we consider $\sigma = 1$ if the learning target behaves as a polar tensor and $\sigma = -1$ if it mimics a pseudotensor under inversion symmetry.

1 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.

2 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, 136403.

3 R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler and M. Parrinello, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **81**, 100103.

4 H. Eshet, R. Z. Khaliullin, T. D. Kühne, J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2012, **108**, 115701.

5 G. C. Sosso, G. Miceli, S. Caravati, J. Behler and M. Bernasconi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **85**, 174103.

6 T. Morawietz, A. Singraber, C. Dellago and J. Behler, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 8368–8373.

7 V. L. Deringer and G. Csányi, *Phys. Rev. B*, 2017, **95**, 094203.

8 S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, *Nat. Commun.*, 2018, **9**, 3887.

9 K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.

10 M. Welborn, L. Cheng and T. F. Miller, *J. Chem. Theory Comput.*, 2018, **14**, 4772–4779.

11 F. A. Faber, A. S. Christensen, B. Huang and O. A. Von Lilienfeld, *J. Chem. Phys.*, 2018, **148**, 241717.

12 L. Zhang, J. Han, H. Wang, R. Car and E. Weinan, *Phys. Rev. Lett.*, 2018, **120**, 143001.

13 D. Dragoni, T. D. Daff, G. Csányi and N. Marzari, *Phys. Rev. Mater.*, 2018, **2**, 013808.

14 B. Cheng, E. A. Engel, J. Behler, C. Dellago and M. Ceriotti, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 1110–1115.

15 J. Behler, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17930–17955.

16 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.

17 M. J. Willatt, F. Musil and M. Ceriotti, *J. Chem. Phys.*, 2019, **150**, 154110.

18 E. Prodan and W. Kohn, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 11635–11638.

19 W. Yang, *Phys. Rev. Lett.*, 1991, **66**, 1438–1441.

20 G. Galli and M. Parrinello, *Phys. Rev. Lett.*, 1992, **69**, 3547–3550.

21 W. Kohn, *Phys. Rev. Lett.*, 1996, **76**, 3168–3171.

22 A. H. R. Palser and D. E. Manolopoulos, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 12704–12711.

23 S. Goedecker, *Rev. Mod. Phys.*, 1999, **71**, 1085–1123.

24 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, *Sci. Adv.*, 2017, **3**, e1701816.

25 S. Kondati Natarajan, T. Morawietz and J. Behler, *Phys. Chem. Chem. Phys.*, 2015, **17**, 8356–8371.

26 C. Zhang and G. Galli, *J. Chem. Phys.*, 2014, **141**, 084504.

27 A. M. Smith, A. A. Lee and S. Perkin, *J. Phys. Chem. Lett.*, 2016, **7**, 2157–2163.

28 Y. Chen, H. I. Okur, N. Gomopoulos, C. Macias-Romero, P. S. Cremer, P. B. Petersen, G. Tocci, D. M. Wilkins, C. Liang, M. Ceriotti and S. Roke, *Sci. Adv.*, 2016, **2**, e1501891.

29 A. P. Gaiduk and G. Galli, *J. Phys. Chem. Lett.*, 2017, **8**, 1496–1502.

30 L. Belloni, D. Borgis and M. Levesque, *J. Phys. Chem. Lett.*, 2018, **9**, 1985–1989.

31 F. Coupette, A. A. Lee and A. Härtel, *Phys. Rev. Lett.*, 2018, **121**, 075501.

32 A. M. Reilly and A. Tkatchenko, *Phys. Rev. Lett.*, 2014, **113**, 055701.

33 A. Ambrosetti, N. Ferri, R. A. DiStasio and A. Tkatchenko, *Science*, 2016, **351**, 1171–1176.

34 J. I. Siepmann and M. Sprik, *J. Chem. Phys.*, 1995, **102**, 511–524.

35 C. Merlet, C. Péan, B. Rotenberg, P. A. Madden, P. Simon and M. Salanne, *J. Phys. Chem. Lett.*, 2013, **4**, 264–268.

36 T. Dufils, G. Jeanmairet, B. Rotenberg, M. Sprik and M. Salanne, *Phys. Rev. Lett.*, 2019, **123**, 195501.

37 L. Scalfi, D. T. Limmer, A. Coretti, S. Bonella, P. A. Madden, M. Salanne and B. Rotenberg, *Phys. Chem. Chem. Phys.*, 2020, **22**, 10480–10489.

38 J. D. Elliott, A. Troisi and P. Carbone, *J. Chem. Theory Comput.*, 2020, **16**, 5253–5263.

39 A. Grisafi, D. M. Wilkins, G. Csányi and M. Ceriotti, *Phys. Rev. Lett.*, 2018, **120**, 036002.

40 D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio and M. Ceriotti, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 3401–3406.

41 M. Veit, D. M. Wilkins, Y. Yang, R. A. DiStasio and M. Ceriotti, *J. Chem. Phys.*, 2020, **153**, 024113.

42 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.

43 H. Huo and M. Rupp, 2017, arXiv:1704.06439.

44 M. Hirn, S. Mallat and N. Poilvert, *Multiscale Model. Simul.*, 2017, **15**, 827–863.

45 M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat and L. Thiry, *J. Chem. Phys.*, 2018, **148**, 241732.

46 J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller and K. Burke, *Phys. Rev. Lett.*, 2012, **108**, 253002.

47 F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke and K. R. Müller, *Nat. Commun.*, 2017, **8**, 872.

48 K. Mills, M. Spanner and I. Tamblyn, *Phys. Rev. A*, 2017, **96**, 042113.

49 A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf and M. Ceriotti, *ACS Cent. Sci.*, 2019, **5**, 57–64.

50 J. Hermann, Z. Schätzle and F. Noé, *Nat. Chem.*, 2020, **12**, 891–897.

51 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.

52 Z. Deng, C. Chen, X.-G. Li and S. P. Ong, *npj Comput. Mater.*, 2019, **5**, 75.

53 K. Rossi, V. Jurásková, R. Wischert, L. Garel, C. Corminboeuf and M. Ceriotti, *J. Chem. Theory Comput.*, 2020, **16**, 5139–5149.

54 M. Veit, S. K. Jain, S. Bonakala, I. Rudra, D. Hohl and G. Csányi, *J. Chem. Theory Comput.*, 2019, **15**, 2574–2586.

55 C. M. Handley and P. L. A. Popelier, *J. Chem. Theory Comput.*, 2009, **5**, 1474–1489.

56 N. Artrith, T. Morawietz and J. Behler, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 153101.

57 T. Bereau, D. Andrienko and O. A. Von Lilienfeld, *J. Chem. Theory Comput.*, 2015, **11**, 3225–3233.

58 P. Bleiziffer, K. Schaller and S. Riniker, *J. Chem. Inf. Model.*, 2018, **58**, 579–590.

59 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.*, 2018, **14**, 4687–4698.

60 K. Yao, J. E. Herr, D. Toth, R. Mckintyre and J. Parkhill, *Chem. Sci.*, 2018, **9**, 2261–2269.

61 L. Zhang, M. Chen, X. Wu, H. Wang and R. Car, *Phys. Rev. B*, 2020, **102**, 041121.

62 T. Bereau, R. A. DiStasio, A. Tkatchenko and O. A. von Lilienfeld, *J. Chem. Phys.*, 2018, **148**, 241706.

63 S. A. Ghasemi, A. Hofstetter, S. Saha and S. Goedecker, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **92**, 045131.

64 S. Faraji, S. A. Ghasemi, S. Rostami, R. Rasoulkhani, B. Schaefer, S. Goedecker and M. Amsler, *Phys. Rev. B*, 2017, **95**, 104105.

65 A. Grisafi and M. Ceriotti, *J. Chem. Phys.*, 2019, **151**, 204105.

66 Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror and D. E. Shaw, *J. Chem. Phys.*, 2005, **122**, 054101.

67 M. J. Willatt, F. Musil and M. Ceriotti, *Phys. Chem. Chem. Phys.*, 2018, **20**, 29661–29668.

68 A. Grisafi, D. M. Wilkins, M. J. Willatt and M. Ceriotti, *Machine Learning in Chemistry*, American Chemical Society, Washington, DC, 2019, vol. 1326, pp. 1–21.

69 J. Nigam, S. Pozdnyakov and M. Ceriotti, *J. Chem. Phys.*, 2020, **153**, 121101.

70 A. J. Stone, *Mol. Phys.*, 1975, **29**, 1461–1471.

71 A. Glielmo, C. Zeni and A. De Vita, *Phys. Rev. B*, 2018, **97**, 184307.

72 R. Drautz, *Phys. Rev. B*, 2019, **99**, 014104.

73 R. Jinnouchi, F. Karsai, C. Verdi, R. Asahi and G. Kresse, *J. Chem. Phys.*, 2020, **152**, 234102.

74 A. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, 1997.

75 D. J. Griffiths, *Introduction to electrodynamics*, Pearson, Boston, MA, 4th edn, 2013.

76 P. Sinz, M. W. Swift, X. Brumwell, J. Liu, K. J. Kim, Y. Qi and M. Hirn, *J. Chem. Phys.*, 2020, **153**, 084109.

77 X. Brumwell, P. Sinz, K. J. Kim, Y. Qi and M. Hirn, 2018, arXiv preprint arXiv:1812.02320.

78 J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.

79 A. Thompson, L. Swiler, C. Trott, S. Foiles and G. Tucker, *J. Comput. Phys.*, 2015, **285**, 316–330.

80 A. V. Shapeev, *Multiscale Model. Simul.*, 2016, **14**, 1153–1173.

81 L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz and C. D. Sherrill, *J. Chem. Phys.*, 2017, **147**, 161727.

82 A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.

83 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput. Phys. Commun.*, 2009, **180**, 2175–2196.

84 F. Musil, M. J. Willatt, M. A. Langovoy and M. Ceriotti, *J. Chem. Theory Comput.*, 2019, **15**, 906–915.

85 B. Huang and O. A. von Lilienfeld, *J. Chem. Phys.*, 2016, **145**, 161102.

86 R. A. Bell, M. C. Payne and A. A. Mostofi, *J. Chem. Phys.*, 2014, **141**, 164703.

87 Y. Litman, D. Donadio, M. Ceriotti and M. Rossi, *J. Chem. Phys.*, 2018, **148**, 102320.

88 D. Maksimov, C. Baldauf and M. Rossi, *Int. J. Quantum Chem.*, 2020, e26369.

89 M. W. Finnis, R. Kaschner, C. Kruse, J. Furthmuller and M. Scheffler, *J. Phys.: Condens. Matter*, 1995, **7**, 2001–2019.

90 J. Neugebauer and M. Scheffler, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1992, **46**, 16067–16080.

91 L. Mørch Folkmann Garner, MSc thesis, Department of Chemistry, University of Copenhagen, 2020.

92 S. N. Pozdnyakov, M. J. Willatt, A. P. Bartók, C. Ortner, G. Csányi and M. Ceriotti, *Phys. Rev. Lett.*, 2020, **125**, 166001.

93 F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti and L. Emsley, *Nat. Commun.*, 2018, **9**, 4501.