PHYSICAL REVIEW B **104**, 144110 (2021)

Through the eyes of a descriptor: Constructing complete, invertible descriptions of atomic environments

Martin Uhrin[*]
Department of Energy Conversion and Storage, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark

(Received 20 April 2021; revised 5 October 2021; accepted 6 October 2021; published 26 October 2021)

In this work we apply methods for describing three-dimensional images to the problem of encoding atomic environments in a way that is invariant to rotations, translations, and permutations of the atoms and, crucially, can be decoded back into the original environment modulo global orientation without the need for training a model. From the point of view of decoding, the descriptor is optimally complete and can be extended to arbitrary order, allowing for a systematic convergence of the fidelity of the description. In experiments on molecules ranging from 3 to 29 atoms in size, we demonstrate that positions can be decoded with a 97% success rate and positions plus species with a 70% rate of success, rising to 95% if a second fingerprint is used. In all cases, consistent recovery is observed for molecules with 17 or fewer atoms. Additionally, we evaluate the descriptor's performance in predicting the energies and forces of bulk Ni, Cu, Li, Mo, Si, and Ge by means of a neural network model trained on DFT data. When comparing to six machine learning interaction potential methods that use various descriptors and regression schemes, our descriptor is found to be competitive, in several cases outperforming well established methods. The combined ability to both decode and make property predictions from a representation that does not need to be learned lays the foundations for a novel way of building generative models that are tasked with solving the inverse problem of predicting atomic arrangements that are statistically likely to have certain desired properties.

DOI: 10.1103/PhysRevB.104.144110

I. INTRODUCTION

Predicting the properties of collections of atoms forms a core pillar of many of the technological advances that have become a staple of modern life. Whether it be drug discovery, materials science, chemical synthesis, or chip fabrication, the ability to connect properties to the atomic scale invariably opens the door to accelerated discovery thanks to greater accuracy and transferability when compared to more coarse grained descriptions. While property prediction is now a mature scientific practice, the inverse problem of predicting atomic arrangements *given* particular target properties is still in its infancy (see [1] for a review of some common methods) and has the potential to have an even greater impact on society than property prediction alone.

Arguably, the most pervasive computational method for the discovery of novel materials and molecules is currently high-throughput screening, which typically involves calculating properties (and often stable atomic arrangements) for many candidates in the hopes of finding some that meet the desired criteria. Naturally this often incurs a high computational cost while simultaneously being limited to a predefined set of candidates. Modern machine learning methods, particularly generative models [2], offer the possibility of overcoming some of these limitations by learning patterns within large data sets and proposing novel candidates that are statistically likely to have the desired properties.

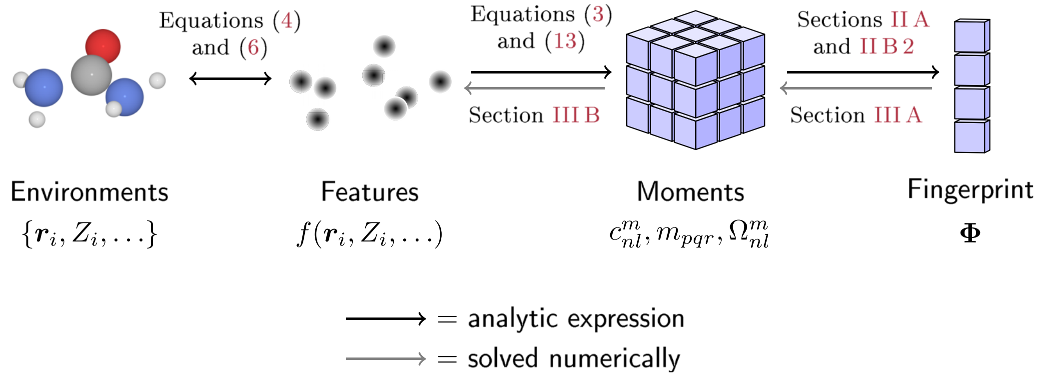
Of key importance when building any machine learning model is the representation of the atomic system upon which the learning occurs. This is reflected in the large volume literature dedicated to the topic (see, e.g., [3–13]). What is near universally recognized are the benefits of using representations that possess the same symmetries as the physical laws that govern the atomic interactions. Specifically, we seek descriptions that possess invariance or equivariance to permutation, translation, and rotation. In addition, for nonlinear machine learning methods such as neural networks, it is desirable to have compact descriptors that do not produce redundant information as this greatly improves training efficiency by keeping the dimensionality of the feature space as low as possible. Finally, for generative models, which rely on inverting latent representations to reproduce atomic structures, it is important that the descriptor be complete such that any set of atomic coordinates maps uniquely onto a point in latent space. It is this collective set of properties and, perhaps more significantly, the inversion procedure itself that are the focus of this work.

While the topic of symmetry invariant descriptions is now well established in the atomistic modeling community, its history goes back significantly further in the fields of image analysis and computer vision. For these communities, the fundamental objects of interests are pixels, voxels, or point clouds, however the methods developed map readily to atomic environments, often by simply placing delta or Gaussian functions on atomic sites. In this work we borrow and build on methods from three-dimensional (3D) image

---

[*] martin.uhrin.10@ucl.ac.uk

analysis to create a complete descriptor of atomic environments that can be inverted and highlight similarities and connections with existing methods in the atomistic modeling community.



The visual summary above shows a schematic of our method for constructing permutation, translation, and rotation invariant fingerprint vectors that can be systematically inverted to recreate the original atomic environment. References to specific portions of the paper are given to enable the reader to easily skip to the sections of most interest. Results from numerical experiments can be found in Sec. IV. In order to avoid ambiguity we will use the term *descriptor* to refer to the method and *fingerprint* or *fingerprint vector* (as this evokes the idea of a uniquely identifying artifact) to refer to the object that is produced by a descriptor.

II. MOMENT INVARIANTS AS (LOCAL) ATOMIC DESCRIPTORS

The moments of a function give a quantitative description of its shape and as such have a long history of use in the world of two-dimensional (2D) and 3D image analysis and computer vision. The first (raw) and second (central) moments of a function are familiar from probability distributions as the mean and variance respectively, while the nth moment of a real-valued continuous function $f(x)$ is given by

$$m_n = \int_{-\infty}^{\infty} x^n f(x) dx. \tag{1}$$

In this context n is typically called the moment's *order* and is equal to the degree of the polynomial onto which f is projected. More generally, moments (in certain contexts called expansion coefficients) can be defined as the projections of a square-integrable function onto a set of basis functions $m_n = \langle f|\Psi_n\rangle$. Many commonly used descriptors of atomic environments are, in fact, functions of moments which produce a fingerprint vector, denoted here as $\boldsymbol{\Phi}$.

For the purposes of faithfully encoding and, crucially, reconstructing atomic environments we are particularly interested in descriptors that have the following properties:

(1) Invariance to global translations and rotations, and permutations of atom labels.

(2) Orthonormality of the basis functions, i.e., $\langle\Psi_i|\Psi_j\rangle = \delta_{ij}$ [14]. Such basis functions lead to an optimal reconstruction of f in the sense that the contribution to the mean square error decreases with each successive moment order n [15].

(3) Completeness of the basis functions. If, for any function $f \in L^2$, the following condition is satisfied:

$$\lim_{n_{\max}\to\infty} \left\| f - \sum_{n=0}^{n_{\max}} m_n \Psi_n \right\|^2 = 0, \tag{2}$$

we say that the basis Ψ_n is complete as f can be approximated arbitrarily closely.

(4) A smooth, injective, invariants function $\boldsymbol{m} \mapsto \boldsymbol{\Phi}(\boldsymbol{m})$ that maps any two vectors of moments $[\boldsymbol{m} = (m_1, \ldots, m_{n_{\max}})]$ *not* related by a global rotation onto distinct fingerprint vectors.

In what follows the ultimate goal is to identify procedures for generating systematically improvable descriptions that satisfy, as much as possible, the aforementioned criteria and that can be used to both predict properties *and* to recreate atomic environments from a finite set of invariants. We call such descriptors moment invariants (local) atomic descriptors (MILAD).

A. Geometric moments

The extension of Eq. (1) to three dimensions is trivial; using three indices s, t, and u, the moments of a three-dimensional function $f(x_1, x_2, x_3)$ are given by

$$m_{stu} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^s x_2^t x_3^u f(x_1, x_2, x_3) dx_1 dx_2 dx_3, \tag{3}$$

where the order is now $s + t + u$. These are the, so-called, geometric moments and are widely used in 3D image analysis where the function $f(\boldsymbol{x})$, with $\boldsymbol{x} = (x_1, x_2, x_3)^T$, is typically a set of discrete voxels of varying intensity [16]. Importantly, one can prove [17] that for piecewise continuous $f(\boldsymbol{x})$ with compact support, all moment orders exist and that $f(\boldsymbol{x})$ is uniquely determined by the set of all moments to infinite order which are themselves uniquely determined by $f(\boldsymbol{x})$. While this is formally true only in the limit of infinite order, we will show later that it is possible to use a related set of orthogonal moments to recover atomic environments from finitely many moments and, indeed, rotation invariants thereof.

In order to use moments to represent atomic configurations, we must make a choice for the form of $f(\boldsymbol{x})$ and a convenient one is that of a sum (thereby inducing permutational invariance) of feature functions centered on each atom, perhaps the simplest of which is a set of N atom centered delta functions with weights w_i,

$$f(\boldsymbol{x}) = \sum_{i}^{N} w_i \delta(\boldsymbol{x} - \boldsymbol{r}_i), \tag{4}$$

where $\boldsymbol{r}_i$ is the position of the ith atom and the weights can either all be equal to one or be used to encode additional information, for example the atomic species. This choice of $f(\boldsymbol{x})$ is particularly appealing as the moments can be trivially computed as

$$m_{stu} = \sum_{i}^{N} w_i \boldsymbol{r}_{i,1}^{s} \boldsymbol{r}_{i,2}^{t} \boldsymbol{r}_{i,3}^{u}. \tag{5}$$

These moments m_{stu} can be grouped by order into tensors from which rotation invariants can be easily derived. This forms the basis of much early work in image analysis and more recently in the atomistic modeling community, most notably in the work of Shapeev [8] and Zaverkin and Kästner [18].

Another commonly used feature function (see, e.g., [5,19,20]) is that of a three-dimensional Gaussian, in which case $f(\boldsymbol{x})$ becomes

$$f(\boldsymbol{x}) = \sum_{i}^{N} \frac{w_i}{\sigma_i^3 \sqrt{(2\pi)^3}} \exp\left[-\frac{1}{2}\left(\frac{\boldsymbol{x} - \boldsymbol{r}_i}{\sigma_i}\right)^2\right], \tag{6}$$

where both w_i and the Gaussian widths σ_i can be used to encode information about atom i [21]. The analytic integrals $\iiint x_1^s x_2^t x_3^u f(x_1, x_2, x_3) dx_1 dx_2 dx_3$, and therefore the moments, can be found relatively easily [22] and are fast to compute.

The range of possible forms for $f(\boldsymbol{x})$ is infinite and the choice is somewhat arbitrary, and can therefore be guided by seeking forms that localize the atomic positions well and that have integrals that can be efficiently evaluated while providing enough degrees of freedom to encode the desired properties, e.g., atomic species, spins, chemical shieldings, or anything else relevant to the property being predicted.

Thus far, m_{stu} are permutationally invariant due to choosing $f(\boldsymbol{x})$ that are sums of feature functions. Translational invariance can be achieved by normalization that is by choosing a particular reference point to be the origin of our coordinate system. In the case of local atomic descriptors, this is typically taken to be the position of each atom in turn, around which a cutoff sphere is projected and all the atoms that lie in its interior are included in $f_i(\boldsymbol{x})$. This leads to a set of fingerprints, one for each atom, which can be used directly as inputs to a suitable fitting algorithm to predict local properties such as forces or chemical shifts [23]. Alternatively, the set of local fingerprints can be used together to predict global properties such as total energy or band gap, either by combining them to form a new descriptor or directly (e.g., as is done in Behler-Parinello neural network potentials [3]). For describing global environments such as entire molecules, nanoribbons, clusters, etc., the center of mass or average position can be used as the reference point or it can be chosen more specifically to fit the problem at hand.

The remaining invariance to tackle is that of rotation. The moments m_{stu} are not, themselves, rotationally invariant, however various methods have been developed all yielding polynomials of moments that are. Some of the earliest work on rotation invariants from moments in three dimensions was carried out by Sadjadi and Hall [24] however it was in the work of Suk and Flusser [25] that a general solution was first presented by extending the ideas of Cyganski and Orr [26] on 2D moment tensors to 3D.

Using Einstein notation, we start with the definition of the moment tensor

$$M^{i_1 i_2 \ldots i_k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^{i_1} x^{i_2} \cdots x^{i_k} f(x^1, x^2, x^3) dx^1 dx^2 dx^3, \tag{7}$$

which has k indices and each i_j can be either 1, 2, or 3 corresponding to the x_1, x_2, or x_3 coordinate, respectively. The total number of repetitions of each corresponds to the exponents s, t, and u in Eq. (3), giving the relation between $M^{i_1 i_2 \cdots i_k}$ and the geometric moments m_{stu}. For the case of rotations (as opposed to general affine transformations, which can also be symmetrized [27] but are more complicated and less relevant for atomic systems) the moment tensor can be treated as a Cartesian tensor, in which case the distinction between covariance and contravariance is lifted and $M_{i_1 i_2 \cdots i_k}$ transforms simply as

$$\hat{M}_{\alpha_1, \alpha_2, \ldots, \alpha_k} = Q_{\alpha_1 i_1} Q_{\alpha_2 i_2} \ldots Q_{\alpha_k i_k} M_{i_1, i_2, \ldots, i_k}, \tag{8}$$

for a given orthonormal rotation matrix Q_{ij}. It can be shown that the total contraction of a Cartesian tensor, e.g., M_{kk}, and total contractions of products of Cartesian tensors, e.g., $M_{kl}M_{kl}$, $M_{kl}M_{lm}M_{mk}$, etc., are all rotation invariants [16]. The first two Suk-Flusser invariants [28] expressed in terms of geometric moments are

$$\begin{aligned}
\Phi_1 &= M_{kk} = m_{200} + m_{020} + m_{002}, \\
\Phi_2 &= M_{kl}M_{kl} = m_{200}^2 + m_{020}^2 + m_{002}^2 \\
&\quad + 2m_{110}^2 + 2m_{101}^2 + 2m_{011}^2.
\end{aligned}$$

This result can be understood intuitively by considering the case of atom centered delta functions [Eq. (4)]. Using Eq. (5) we have that

$$\Phi_1 = \sum_{i=1}^{N} w_i \left(\boldsymbol{r}_{i,1}^2 + \boldsymbol{r}_{i,2}^2 + \boldsymbol{r}_{i,3}^2\right) = \sum_{i=1}^{N} w_i \boldsymbol{r}_i \cdot \boldsymbol{r}_i, \tag{9}$$

which is nothing more than the sum of the weighted dot products of each atom's position vector with itself, a quantity that is plainly rotationally invariant. We may also recognize this as the trace of the inertia tensor of a rigid body composed of point masses Tr($\mathbf{I}$), a quantity that is well known in engineering applications as the *first-principle invariant*. For the case of Φ_2 we have

$$\begin{aligned}
\Phi_2 = &\left(\sum_{i=1}^{N} w_i \boldsymbol{r}_{i,1}^2\right)^2 + \left(\sum_{i=1}^{N} w_i \boldsymbol{r}_{i,2}^2\right)^2 + \left(\sum_{i=1}^{N} w_i \boldsymbol{r}_{i,3}^2\right)^2 \\
&+ 2\left(\sum_{i=1}^{N} w_i \boldsymbol{r}_{i,1} \boldsymbol{r}_{i,2}\right)^2 + 2\left(\sum_{i=1}^{N} w_i \boldsymbol{r}_{i,1} \boldsymbol{r}_{i,3}\right)^2 \\
&+ 2\left(\sum_{i=1}^{N} w_i \boldsymbol{r}_{i,2} \boldsymbol{r}_{i,3}\right)^2,
\end{aligned}$$

which is the *second main invariant* Tr($\mathbf{I}^2$).

This procedure can be continued in a similar fashion to yield as many invariants as are needed to faithfully describe the atomic environments under investigation, however, care must be taken to avoid redundant invariants which can take several forms as discussed later. The procedure outlined thus far has much in common with that used to build invariants for moment tensor potentials (MTPs) [8,29] and Gaussian moment invariants [18], differing primarily in leaving the choice of $f(\boldsymbol{x})$ free.

B. Spherical harmonic based moments

While geometric moments are fast to calculate, their numerical properties make them less suitable for some applications, and other choices of basis function can yield better performance, particularly when it comes to reconstruction. Another commonly used set of functions are the spherical harmonics, which yield complex moments defined as

$$c_{nl}^{m} = \int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} R_{nl}(r)\overline{Y_l^m(\theta,\varphi)} f(r,\theta,\varphi) r^2 \sin\theta \, dr d\theta d\varphi, \tag{10}$$

where $R_{nl}(r)$ are a set of real-valued radial basis functions and $r^2 = x_1^2 + x_2^2 + x_3^2$, $\theta = \arccos(x_3/r)$, and $\varphi = \arctan(x_2/x_1)$. Here we use the definition

$$Y_l^m(\theta,\varphi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+1)!}} P_l^m[\cos(\theta)] e^{im\varphi}, \tag{11}$$

where P_l^m are the *associated Legendre functions*.

Spherical harmonics are particularly well suited to deriving rotation invariants and, as such, have been seen using 3D image analysis for some time. Pioneering work was carried out by Lo and Hon-Son [30] who used group-theoretical methods to find rotation invariants for 3D object identification and positioning. A similar approach was followed by Canterakis [31] and later Novotni and Klein [32] who both used Zernike polynomials [33] as radial basis functions to arrive at 3D rotation invariants. More recent work by Suk *et al.* [34] has demonstrated a systematic way of arriving at independent invariants from complex moments which we detail below.

The atomistic modeling, bioinformatics, and computational chemistry communities have also adopted such rotation invariants both for predicting properties and, in particular, learning potential energy surfaces (PESs). Some of the earliest examples can be found in work on molecules [19,35–42] while the last decade has seen widespread adoption of spherical harmonic based descriptors, often coupled with advanced machine learning techniques, to predict properties of solids as well. The well known SOAP descriptor [5] was one of the first and shares much in common with the invariants of Canterakis [31]. Recent work in the atomistic modeling community has also turned to the use of polynomials of moments very similar to those described here, showing a growing convergence between the directions taken by the 3D image analysis and atomistic modeling communities [43–47].

1. Radial basis

Perhaps the simplest radial basis is $R_n(r) = r^n$ which gives rise to a set of moments which are widely used in 2D and 3D image analysis (see, e.g., [48–51]). Using spherical harmonics in Cartesian form the, so called, *3D complex moments* are given by

$$c_{nl}^{m} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r^n Y_l^m(x_1,x_2,x_3) f(x_1,x_2,x_3) dx_1 dx_2 dx_3. \tag{12}$$

The first few spherical harmonics are

$$Y_0^0(x_1,x_2,x_3) = \frac{1}{2}\sqrt{\frac{1}{\pi}},$$

$$Y_1^0(x_1,x_2,x_3) = \frac{1}{2}\sqrt{\frac{3}{\pi}}\frac{x_3}{r},$$

$$Y_1^1(x_1,x_2,x_3) = -\frac{1}{2}\sqrt{\frac{3}{2\pi}}\frac{(x_1+ix_2)}{r},$$

$$Y_2^0(x_1,x_2,x_3) = \frac{1}{4}\sqrt{\frac{5}{\pi}}\frac{(3x_3^2-r^2)}{r^2},$$

$$Y_2^1(x_1,x_2,x_3) = -\frac{1}{2}\sqrt{\frac{15}{2\pi}}\frac{(x_1+ix_2)x_3}{r^2},$$

$$Y_2^2(x_1,x_2,x_3) = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\frac{(x_1+ix_2)^2}{r^2},$$

where $Y_l^{-m} = (-1)^m \overline{Y_l^m}$. When the basis consists of polynomials in (x_1, x_2, x_3) the moments provide a complete and independent description of f, we therefore make the restriction that $n - l$ be even and $l \leqslant n$, such as to cancel the factor of r^{-l} that comes from the spherical harmonics. This basis is, thus, a set of complex homogenous polynomials of degree n in (x_1, x_2, x_3).

While 3D complex moments are fast to calculate, they have some notable disadvantages compared to orthogonal moments. For one, they posses high dynamic range which can lead to greater numerical errors, in addition indirect methods must to be used to perform reconstruction (see, e.g., [52]).

Orthogonal radial functions have a region of orthogonality which we take to be $0 \leqslant r \leqslant 1$, this requires that atomic environments be scaled appropriately to fit within the unit sphere. In general, the corresponding *orthogonal moments* are given by

$$c_{nl}^{m} = \eta_{nlm}\int_0^{2\pi}\int_0^{\pi}\int_0^{1} R_{nl}(r)\overline{Y_l^m(\theta,\varphi)} \times f(r,\theta,\varphi) r^2 \sin\theta \, dr d\theta d\varphi, \tag{13}$$

where η_{nlm} is an normalization constant. Orthogonal basis functions make it trivial to reconstruct an approximation to the original function given a finite set of moments. The reconstruction up to order $n_{\max}$ is given by

$$\tilde{f}(r,\theta,\varphi) = \sum_{n}^{n_{\max}}\sum_{l=0}^{n}\sum_{m=-l}^{l} c_{nl}^{m} R_{nl}(r) Y_l^m(\theta,\phi). \tag{14}$$

This expansion minimizes the mean-square error to the original function and can be systematically converged making it ideally suited for the task of reconstruction.

2. Rotation invariants

Here we give a brief description of how to arrive at a set of rotation invariants to arbitrary correlation order, consisting of polynomials of complex moments which can subsequently be reduced to an independent set of rotation invariants. The flexibility to expand to arbitrary order is particularly appealing given the recent evidence that three- and four-body atom-centered features are insufficient to unambiguously describe all possible atomic environments [11]. We follow the procedure of Suk *et al.* [34] who use ideas from Lo and Hon-Son [30] and encourage the reader to refer to these sources for a more complete description.

Spherical harmonics transform as

$$Y_l^m(\mathbf{Q}^{-1}\boldsymbol{x}) = \sum_{m'=-l}^{l} D^l_{m'm}(\mathbf{Q}) Y_l^{m'}(\boldsymbol{x}), \tag{15}$$

when rotated by an arbitrary rotation matrix $\mathbf{Q}$ where $D^l_{m'm}$ are the, so-called, Wigner D functions. It is common to express the spherical harmonics as a vector $\boldsymbol{Y}_l(\boldsymbol{x}) = (Y_l^{-l}(\boldsymbol{x}), Y_l^{-l+1}(\boldsymbol{x}), \dots, Y_l^l(\boldsymbol{x}))^T$ and $D^l_{m'm}$ as $(2l+1)\times(2l+1)$ dimensional unitary matrices, called Wigner D matrices, that form an irreducible representations of the group of three-dimensional rotations SO(3). The elements of $\mathbf{D}^l$ are given by

$$D^l_{mm'} = \langle Y_l^m | \hat{R}(\alpha\beta\gamma) | Y_l^{m'} \rangle, \tag{16}$$

where $\hat{R}(\alpha\beta\gamma)$ is the rotation operator parametrized by three Euler angles.

For a given l, the corresponding spherical harmonics $\boldsymbol{Y}_l$ form a basis of the irreducible representation $\mathbf{D}^l$. Invariants to rotation can be found by identifying one-dimensional irreducible subspaces that transform according to $\mathbf{D}^0$. Y_0^0 is such an invariant, however spherical harmonics corresponding to irreducible representations with $l \neq 0$ are not. To find further invariants we can, for example, combine basis functions corresponding to representations $\mathbf{D}^{j_1}$ and $\mathbf{D}^{j_2}$ by taking their tensor product. Maschke's theorem tells us that the resulting representation can be expressed as the direct sum of irreducible representations

$$\mathbf{D}^{j_1} \otimes \mathbf{D}^{j_2} = \bigoplus_{l=|j_1-j_2|}^{j_1+j_2} \mathbf{D}^l, \tag{17}$$

where the right-hand side is a block diagonal matrix and the upper-left submatrix $\mathbf{D}^0$ corresponds to a scalar that is a rotation invariant. In general, the connection between basis functions $\varphi^i_{j_1}$ and $\varphi^i_{j_2}$ corresponding to the irreducible representations $\mathbf{D}^{j_1}$ and $\mathbf{D}^{j_2}$ and the basis functions Ψ^k_j of a particular $\mathbf{D}^l$ is

$$\Psi^k_l = \sum_{m=\max(-j_1, k-j_2)}^{\min(j_1, k+j_2)} \langle j_1, j_2, m, k-m | l, k \rangle \varphi^m_{j_1} \varphi^{k-m}_{j_2}, \tag{18}$$

where $\langle j_1, j_2, m, k-m | l, k \rangle$ are Clebsch-Gordan coefficients. As the radial basis functions do not couple to m, the tensor products of moment vectors $\boldsymbol{c}_{nl} = (c_{nl}^{-l}, c_{nl}^{-l+1}, \dots, c_{nl}^{l})^T$ are also related by Eq. (18). This led Lo and Hon-Son to define *composite complex moment forms*

$$c_n(l_1, l_2)^k_l = \sum_{m=\max(-l_1, k-l_2)}^{\min(l_1, k+l_2)} \langle l_1, l_2, m, k-m | l, k \rangle \, c^m_{nl_1} c^{k-m}_{nl_2}, \tag{19}$$

which can be used as a mathematical tool to calculate moments in the basis of representation $\mathbf{D}^l$. Modern equivariant neural networks [53–57] use similar ideas to ensure that vectorial or tensorial inputs are propagated through the network in a symmetry preserving way.

In the atomistic modeling community it is common to speak in terms of the *correlation order* of an invariant, labeled ν, which is given by the number of terms involved in a tensor products, while the signal processing community typically uses the term *degree* (or order) in reference to the polynomial degree in moments. The two are equivalent and here will use the former. Using Eq. (19) a set of rotation invariants can be defined as follows.

$\nu = 1$ *invariants* c^0_{n0} are invariants corresponding to $\mathbf{D}^0$, where n is even.

$\nu = 2$ *invariants* $c_n(l,l)^0_0$ are invariants corresponding to $(\mathbf{D}^l \otimes \mathbf{D}^l)^0$, given by

$$\sum_{m=-l}^{l} \langle l, l, m, -m | 0, 0 \rangle \, c^m_{nl} c^{-m}_{nl} = \frac{1}{\sqrt{2l+1}} \sum_{m=-l}^{l} (-1)^{l-m} c^m_{nl} c^{-m}_{nl}, \tag{20}$$

where $n - l$ must be even. This is often called the power spectrum and is used by a number of existing descriptors.

$\nu = 3$ *invariants* Further invariants can be arrived at by combining the bases of more representations, for example, $((\mathbf{D}^{l_1} \otimes \mathbf{D}^{l_2})^l \otimes \mathbf{D}^l)^0$ invariants are given by

$$c_{n_1}(l_1, l_2)_l c_{n_2} = \frac{1}{\sqrt{2l+1}} \sum_{k=-l}^{l} (-1)^{l-k} c_{n_1}(l_1, l_2)^k_l c^{-k}_{n_2 l}, \tag{21}$$

where $n_1 - l_1$, $n_1 - l_2$, and $n_2 - l$ are even. Swapping l_1 and l_2 produces the same invariant and therefore we take $l_2 \leqslant l_1$. This is often called the bispectrum.

$\nu = 4$ *invariants* Continuing in this fashion $((\mathbf{D}^{l_1} \otimes \mathbf{D}^{l_2})^l \otimes (\mathbf{D}^{l_3} \otimes \mathbf{D}^{l_4})^l)^0$ gives the invariants

$$c_{n_1}(l_1, l_2)_l c_{n_2}(l_3, l_4)_l = \frac{1}{\sqrt{2l+1}} \sum_{k=-l}^{l} (-1)^{l-k} c_{n_1}(l_1, l_2)^k_l c_{n_2}(l_3, l_4)^{-k}_l, \tag{22}$$

where $n_1 - l_1$, $n_1 - l_2$, $n_2 - l_3$, and $n_2 - l_4$ are even and to avoid duplicates we take $l_1 \leqslant n_1$, $l_1 - l \leqslant l_2 \leqslant l_1$, $l_3 \leqslant n_2$, and $l_3 - l \leqslant l_4 \leqslant l_3$.

This procedure may be extended to yield as many invariants as are needed however there will invariably be many that are dependent or identical to others which must then be reduced to an independent set as outlined below. The procedure of generating successively higher correlation order invariants is also at the core of the atomic cluster expansion (ACE) [43,45] scheme which uses linear regression to expand the energy of a collection of atoms in a basis of such invariants. Recent results [58,59] show very good performance both in terms of accuracy and computational efficiency.

C. Independent invariants

Be they based on geometric or complex moments, a set of rotation invariants can be arrived at by naively carrying out one of the procedures outlined above, however this will invariably lead to a set that is overcomplete. Certain types of dependency can be eliminated during generation, however, others have to be checked *post facto* which can have a high algorithmic complexity. Suk and Flusser [25] provide a useful analysis which begins with the following classification of redundancies: 1. *zero invariants*, 2. *identical invariants*, 3. *direct products* (where an invariant is a product of previously found invariants), 4. *linear combinations*, and 5. *polynomial dependencies.* In general, the first three can be eliminated relatively easily by brute force using a suitable symbolic mathematics library. Some practical methods for eliminating linear and polynomial dependencies are described below.

1. Linear dependencies

In general, linear combinations can be detected by calculating the column rank of the matrix representing the system of equations given by the invariants at each correlation order, where the columns consist of the moment coefficients of each invariant. If the number of columns exceeds the rank, then the number of invariants can be reduced by performing a singular value decomposition and discarding those with zero singular values. If a resulting matrix can be found that has full rank, then the invariants are linearly independent. While this procedure works, in principle, for any set of invariants it can suffer from problems of numerical instability, specifically when the polynomial coefficients are not integers.

An alternative approach for eliminating linear dependencies among invariants from geometric moments is to identify unique generating graphs from the networks corresponding to the tensor contractions [25]. The approach of using tensor networks to construct invariants is also used by Zaverkin and Kästner [18] for their Gaussian moments descriptor.

For invariants from spherical harmonics many linear dependencies can be eliminated by using standard results from the coupling of angular momenta. For example, as in Sec. II B 2, by restricting the l indices involved in a tensor product to be in sorted order, e.g., $l_1 \geqslant l_2 \geqslant \cdots \geqslant l_{\text{max}}$, a large number of dependencies are avoided. Furthermore, if any two or more two irreps involved in a tensor product share the same value of l, then the n indices may also be similarly sorted to avoid further dependencies. This scheme is employed as part of the construction of NICE Nigam *et al.* [47] descriptors. Finally, the work of Bachmayr *et al.* [44], which underpins the ACE descriptor, gives a numerically stable algorithm for generating a fully linearly independent set of invariants to arbitrary correlation order using a complete analysis of the permutational symmetries among invariants.

TABLE I. The number of invariants to translation and rotation from complex moments with the corresponding number of linearly and algebraically invariants up to a given l_{max} (here the condition that $l \leqslant n$ is assumed).

l_{max}	2	3	4	5	6	7
All	4	16	49	123	280	573
Linearly indep.	3	13	37	100	228	486
Algebraically indep.	3	13	28	49	77	113

2. Polynomial dependencies

A set of invariants with no polynomial dependencies is called independent or algebraically complete. General procedures to generate such a set are prohibitively expensive, even for relatively low correlation orders [60]. Nevertheless, a number of algorithms that try to reduce some or all such dependencies have been proposed [27,61,62], and while they are not guaranteed to produce a certifiably independent set, in practice it is sufficient to test their algebraic independence on a representative set of moments. Not all methods will produce the same independent invariants due to arbitrary choices that are made, but in all cases the total number of independent invariants to a given order should equal the number of moments minus the number of degrees of freedom that are being symmetrized.

Despite the algorithmic complexity this step is important for nonlinear regression schemes as the reduction in the number of invariants can be significant, greatly reducing the dimensionality of the feature space. Table I shows the number of invariants generated by the procedure in Sec. II B 2 up to $l_{\text{max}} = 7$ and the corresponding number of linearly and algebraically independent invariants.

In this work we use invariants from complex moments made independent by the method of Kostková *et al.* [27] which is based on earlier work [61,62], and can be summarized as follows. Given a set of n_k invariants $\{I_1, \dots, I_{n_k}\}$, if there are dependencies then we can write any invariant as some function of the others, e.g.,

$$I_1(c_{nl}^m) = f(I_2(c_{nl}^m), \dots, I_{n_k}(c_{nl}^m)). \tag{23}$$

Taking the derivative with respect to a moment, we get

$$\frac{\partial I_1(c_{nl}^m)}{\partial c_{nl}^m} = \frac{df(I_2(c_{nl}^m), \dots, I_{n_k}(c_{nl}^m))}{dc_{nl}^m} = \sum_{\beta=2}^{n_k} \frac{\partial f}{\partial I_\beta}\frac{\partial I_\beta}{\partial c_{nl}^m}, \tag{24}$$

which shows that the derivatives of the dependent invariants are, themselves, linearly dependent. This gives an intuitive way to understand the Jacobian criterion for algebraic independence which states that if the invariants are dependent, then, by rearranging Eq. (23), there must exist a function F such that

$$F\left(I_1, \ldots, I_{n_k}\right) = 0. \tag{25}$$

For convenience we sort the moments c_{nl}^m in lexicographic order such that we can use a single index $\alpha = (n, l, m)$, which runs from 1 to n_p. Using this notation, the derivatives of F are

$$\frac{\partial F\left(I_1, \ldots, I_{n_k}\right)}{\partial c_\alpha} = \sum_{\beta=1}^{n_k} \frac{\partial F\left(I_1, \ldots, I_{n_k}\right)}{\partial I_\beta} \frac{\partial I_\beta}{\partial c_\alpha} = 0, \tag{26}$$

where we have a known $n_k \times n_p$ Jacobian matrix $A_{\alpha\beta} = \frac{\partial I_\beta}{\partial c_\alpha}$ and an unknown coefficients vector $b_\beta = \frac{\partial F(I_1,\ldots,I_{n_k})}{\partial I_\beta}$ of length n_k. If all of the invariants are independent, then the only solution is $b_\beta = 0, \forall \beta$. Accordingly, if the Jacobian has full column rank, then the invariants are algebraically independent. Conversely, if the rank n_r is less than n_k, then there are only n_r independent invariants.

As the column rank of $A_{\alpha\beta}$ cannot be determined analytically this step must be performed numerically for several generic sets of moments and the maximum rank found is taken to be n_r. If n_r is less than n_k, then some criterion must be used to choose which invariants to keep. Kostková *et al.* follow the following procedure:

(1) Order the invariants first by correlation order and then by number of terms (this guarantees that any subset of invariants up to arbitrary $l_{\max}$ is also algebraically complete).

(2) Add one invariant at a time to the set of invariants.

(3) Check the new rank of $A_{\alpha\beta}$ by testing against the representative set of moments (in their case 5 were used) and if it increases by one, keep the invariant, otherwise discard it.

While this procedure is relatively computationally expensive, it need only be carried out once. It is important to either validate the tolerance of the method used to determine the rank or to use a library that supports exact arithmetic to ensure that the outcome is not affected by numerical issues.

The above procedure works for any set of invariants (irrespective of the basis functions used), while for spherical harmonic based invariants the procedure of Nigam *et al.* [47] offers an alternative that does not rely on numerical evaluation and therefore avoids the associated issues of choosing a tolerance and having to use a representative set of moments, however, their methods will not necessarily identify all polynomial dependencies.

3. Data driven invariants reduction

The invariants reduction analysis thus far focused solely on eliminating redundancies that do not lead to any information loss over all possible signals (i.e., sets of moments). There are, however, data driven methods for reducing the number of invariants that are based on features found in a particular data set and allow the descriptor to retain discriminative power among the entries of that set, potentially sacrificing universality. It may even be possible to use such techniques to find a reduced set of invariants that work well for all signals that are sums of atomic feature functions (with a corresponding loss of discriminative power for other square-integrable functions). As this work is focused on universal descriptors we will not detail any particular approach here but we refer the reader to [47,63] for two examples of such techniques.

D. This work

Our codebase contains support for delta and Gaussian feature functions as well as a set of 1185 linearly independent and 962 algebraically independent invariants from geometric moments up to 16th order. The code is fully modular and can be easily extended with new invariants, basis functions or feature functions.

For invariants based on spherical harmonics we use Zernike functions as the radial basis, however other choices have been proposed including Bessel [64] and Chebyshev [44] functions which may have different numerical characteristics. (See Goscinski *et al.* [63] for an excellent review of the performance of various radial basis functions.) The 3D Zernike polynomials are defined as

$$Z_{nl}^m(r, \theta, \varphi) = R_{nl}(r) Y_l^m(\theta, \varphi), \tag{27}$$

however, they are perhaps easiest to work with in Cartesian form,

$$Z_{nl}^m(\boldsymbol{x}) = \sum_{\nu=0}^{k} q_{kl}^\nu |\boldsymbol{x}|^{2\nu} e_l^m(\boldsymbol{x}), \tag{28}$$

where e_l^m are harmonic polynomials, $2k = n - l$, and the coefficients q_{kl}^ν are fixed by the orthonormality relation

$$\frac{3}{4\pi} \int_{|\boldsymbol{x}|\leqslant 1} Z_{nl}^m(\boldsymbol{x}) \overline{Z_{n'l'}^{m'}(\boldsymbol{x})} r^2 \sin\theta \, d\boldsymbol{x} = \delta_{nn'} \delta_{ll'} \delta^{mm'}, \tag{29}$$

the full derivation of which can be found in [31]. We calculate the Zernike moments $\Omega_{nl}^m := \frac{3}{4\pi} \langle f | Z_{nl}^m \rangle$ by means of a change of basis transformation from geometric moments according to the procedure outlined by Novotni and Klein [32]. Briefly, the Zernike moments are expressed as a linear combination of geometric moments m_{stu} [Eq. (3)],

$$\Omega_{nl}^m = \frac{3}{4\pi} \sum_{s+t+u \leqslant n} \overline{\chi_{nlm}^{stu}} m_{stu}, \tag{30}$$

where χ_{nlm}^{stu} are given by

$$\begin{aligned} \chi_{nlm}^{stu} =& c_l^m 2^{-m} \sum_{\nu=0}^{l} q_{kl}^\nu \sum_{\alpha=0}^{\nu} \binom{\nu}{\alpha} \sum_{\beta=0}^{\nu-\alpha} \binom{\nu-\alpha}{\beta} \sum_{u=0}^{m} (-1)^{m-u} \\ & \times \binom{m}{u} i^u \sum_{\mu=0}^{\lfloor \frac{l-m}{2} \rfloor} (-1)^\mu 2^{-2\mu} \binom{l}{\mu} \binom{l-\mu}{m+\mu} \sum_{\nu=0}^{\mu} \binom{\mu}{\nu}. \end{aligned} \tag{31}$$

A similar change of basis transformations for going from MTP geometric tensors to ACE spherical tensors can be found in Appendix A of [45]. The main advantage of starting with geometric moments is computational speed, particularly as χ^{stu}_{nlm} need only be calculated once and can then be cached or even stored offline for reuse.

From the Zernike moments we use an independent set of 117 rotation invariants up $n = 7$ generated by the procedure in Sec. II C 2. These invariants are brought together in a single fingerprint vector $\mathbf{\Phi}$. Derivatives are available for both positions and weights as these are important for reconstruction.

E. On completeness

The topic of completeness of atomic descriptors has received a lot of attention (see, e.g., [5,6,11,13,64] and it is, undoubtedly, important from the point of view that all regression schemes predicting properties of atomic systems make the assumption, explicitly or implicitly, that two atomic configurations that cannot be made to coincide by a combination of rigid body translations, rotations, and for certain properties reflections, will be mapped to different points in fingerprint space. Naturally, when inverting rotationally invariant fingerprints this assumption is equally important as, typically, we rely on the fact that up to global rotation there is a single atom density that explains a given fingerprint.

For the types of descriptors that are the focus of this work (those based on expanding an atom density in a chosen basis followed by computing rotation invariants from the resulting tensors) we can break our analysis of completeness down by considering each of the steps individually. If the output of one step is not unique with respect to the inputs, then, necessarily, the descriptor is not complete.

Provided no two atoms have identical coordinates, then the mapping of atoms onto localized feature functions is unique. This is true even for coincident atoms so long as fixed weights are used, however, with variable weights degeneracies can occur which would require additional fingerprints to resolve.

In Sec. I we deliberately restricted our focused to orthonormal basis functions as these are both complete and convenient for performing synthesis. However, a set of functions need not be orthogonal to be complete in the sense of being able to approximate any function in L^2 arbitrary closely. More formally, a basis can be said to be complete [65] over the interval (a, b) if there exists a series

$$\tilde{f}(\boldsymbol{x}) = \alpha_1\Psi_1(\boldsymbol{x}) + \alpha_2\Psi_2(\boldsymbol{x}) + \cdots + \alpha_{n_{\max}}\Psi_{n_{\max}}(\boldsymbol{x}) \tag{32}$$

such that for every $\epsilon > 0$

$$\int_a^b |f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})|^2 d\boldsymbol{x} < \epsilon. \tag{33}$$

Where the basis consists of polynomials, completeness is given by the Weierstrass approximation theorem which applies to both the geometric and Zernike moments used in this work. Furthermore, any linear transformation of a complete basis is, itself, complete whether or not the basis is orthogonal [66].

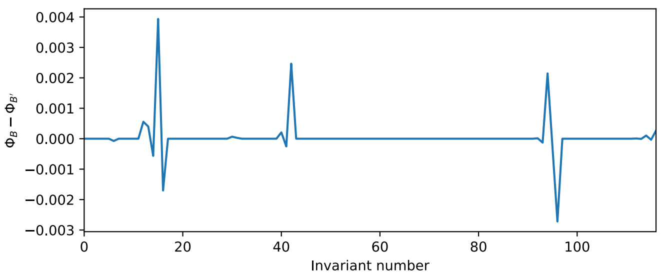


FIG. 1. Difference in fingerprint vector components between atomic arrangements (labeled B and B') that cannot be distinguished when using $\nu = 2$ and $\nu = 3$ invariants. All of the nonzero values correspond to $\nu = 4$ invariants.

Lastly, let us consider the completeness of the invariants themselves. A necessary condition for a set of invariants to uniquely map moments onto a point in fingerprint space is that there be at least as many algebraically independent invariants as the number of moments minus the number of symmetrized degrees of freedom. Working with a spherical harmonic expansion where $0 \leqslant n \leqslant n_{\max}$, $0 \leqslant l \leqslant l_{\max}$, and $-l \leqslant m \leqslant l$ there are $n_{\max} l_{\max}(l_{\max} + 2)$ moments. Bandeira *et al.* [67] have shown that with three or more radial function (i.e., $n_{\max} \geqslant 3$) it is sufficient to go up to $\nu = 3$ invariants to reach independence. While in this work we use the reduced basis

$$\sum_{n=0}^{n_{\max}} \sum_{l=0}^{n} \sum_{m=-l}^{l} c^m_{nl} R_{nl}(r) Y^m_l(\theta, \psi), \tag{34}$$

where $(n - l)$ must be even. This requires going up to $\nu = 4$ to arrive at enough invariants.

While having an algebraically complete set of invariants (or superset thereof) is necessary to distinguish all image functions, it is not sufficient as shown recently by Pozdnyakov *et al.* [11]. In this work they provide a pair of environments that cannot be distinguished by the bispectrum (i.e., $\nu = 3$) descriptor. In Fig. 1 we plot the difference in fingerprint vectors generated by the MILAD descriptor for these environments. The nonzero contributions come exclusively from 21 out of the 29 $\nu = 4$ invariants. The fact that MILAD can distinguish these environments is, however, no guarantee that two environments could not be found that require a higher correlation order to be distinguished.

It remains an open question as to whether there is some maximum bound on ν such that all atomic configuration could be distinguished or conversely if, in the general case, it is unbounded. As such, currently no formal statement about the injectivity of the map of moments to fingerprint used in this work can be made, however, results from Pozdnyakov *et al.* suggest that such examples are rare in practice. The recent work of Pozdnyakov *et al.* [68] gives a further in-depth analysis of some of the causes, and consequences, of incompleteness in rotationally invariant atomic descriptors.

Finally, assuming a certain minimum interatomic distance, the number of invariants needed to uniquely describe an atomic environment will naturally increase with the number of atoms N. By considering that there are $3N - 3$ degrees of freedom in a rotation symmetrized environment, we would expect to need at least this many invariants. In reality, the number needed is likely to be greater owing to the fact the the types of invariants discussed in this work are capable of describing functions that fall outside of the limited class of atom densities considered. One numerical way to probe this question more generally would be with the use of an autoencoder. By successively decreasing the number of artificial neurons in the bottleneck, one could look for a signature, discontinuous, increase in the decoding RMSD over a large set of examples environments. This will be explored as part of future work, however we briefly revisit this issue when discussing results from reconstruction experiments further below.

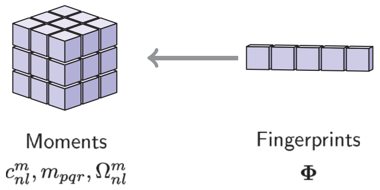
III. RECONSTRUCTION

The goal of reconstruction is to start with a set of moment invariants and general information about the set of feature functions used (function type, range of weights, etc.) and reproduce the original set of atomic positions, and optionally their atomic species. It is possible to start with semirandom configurations of atoms and perform a global optimisation of the residuals with respect to the known invariants (similar to random structure searching [69]), however this scales poorly as the number of atoms grows due to the large number of local minima encountered. On the other hand, performing a local minimization of the atomic coordinates with respect to known moments is significantly more reliable.

The problem of retrieving a set of moments starting from rotation invariants has commonalities with the more well known phase retrieval problem that lies at the heart of solving for crystal structures from their x-ray diffraction pattern. In this case, the symmetrization is over the translation group which manifests in only the intensities of the diffraction peaks being observed, preventing the structure from being solved directly by simply performing the inverse Fourier transform due to the missing phases. One of the most well known methods for phase retrieval is the Gerchberg-Saxton algorithm [70] which also has a similar, iterative optimization structure to that used here. The algorithm starts with random initial phases and proceeds by iteratively performing forward and backwards Fourier transforms, modifying the phases in-between to better match the observed diffraction pattern until the difference falls below a given threshold. The structure of the orientation retrieval problem being addressed here differs primarily by having to simultaneously solve for a hierarchy of many-point correlations (up to the maximum correlation order used), whereas phase retrieval typically deals with two point correlations (from the square of the Fourier transform).

Another closely related problem is that of multireference alignment, commonly used to solve the structure of single molecules in cryogenic electron microscopy (cryo-EM). Here an image is taken of many copies of the same molecule with different orientations from which the three-dimensional structure is determined. One of the more relevant algorithms from this community can be found in the work of Bandeira *et al.* [67]. The authors give a closed-form procedure to recover the spherical harmonic expansions coefficient from a set of $\nu = 2$ and $\nu = 3$ invariants. The algorithm is included in our codebase and has been tested on a number of atomic systems. The major shortcoming is that the procedure only works for, so called, generic signals and fails in the presence of symmetry. This is because symmetry causes some of the invariants to be trivial leading to a situation where a linear system that must be solved as part of the algorithm becomes underdetermined. By contrast, the numerical algorithm presented here relies on finding a least-squares solution which is biased by the prior knowledge we have about the class of signals making it more robust to such situations.

A. Moments from invariants



The invariants Φ_i form a latent space from which the corresponding moments can be recovered by solving the system of polynomials which couples the two. The system of equations is underdetermined by three equations corresponding to the missing orientation information; a degeneracy that will be resolved during reconstruction, however the new orientation will be uncorrelated with the original. The zeroth-order moment is itself a rotation invariant and encodes the total mass of the environment, in the case of geometric moments $m_{000} = \sum_i^{N_A} w_i$ while for Zernike moments $\Omega^0_{00} = \frac{3}{4\pi} m_{000}$. The invariant of first order moments

$$\Phi_1 = \frac{1}{\sqrt{3}}\left(2c^{-1}_{11}c^{1}_{11} - {c^{0}_{11}}^{2}\right)$$

encodes the distance of the center of mass from the origin, valid solutions for which can be found trivially. From here it is possible to solve the system of equations iteratively, where at each iteration we include all invariants containing moments up to a maximum angular frequency l' effectively building up a more and more detailed reconstruction of the original environment. At the end of each iteration, l' is incremented by one until we reach $l_{\max}$. This procedure can encounter local minima and we therefore make several attempts (typically up to two at each l') to find a solution until the root-mean-square deviation from the known invariants drops to below a given threshold, typically taken to

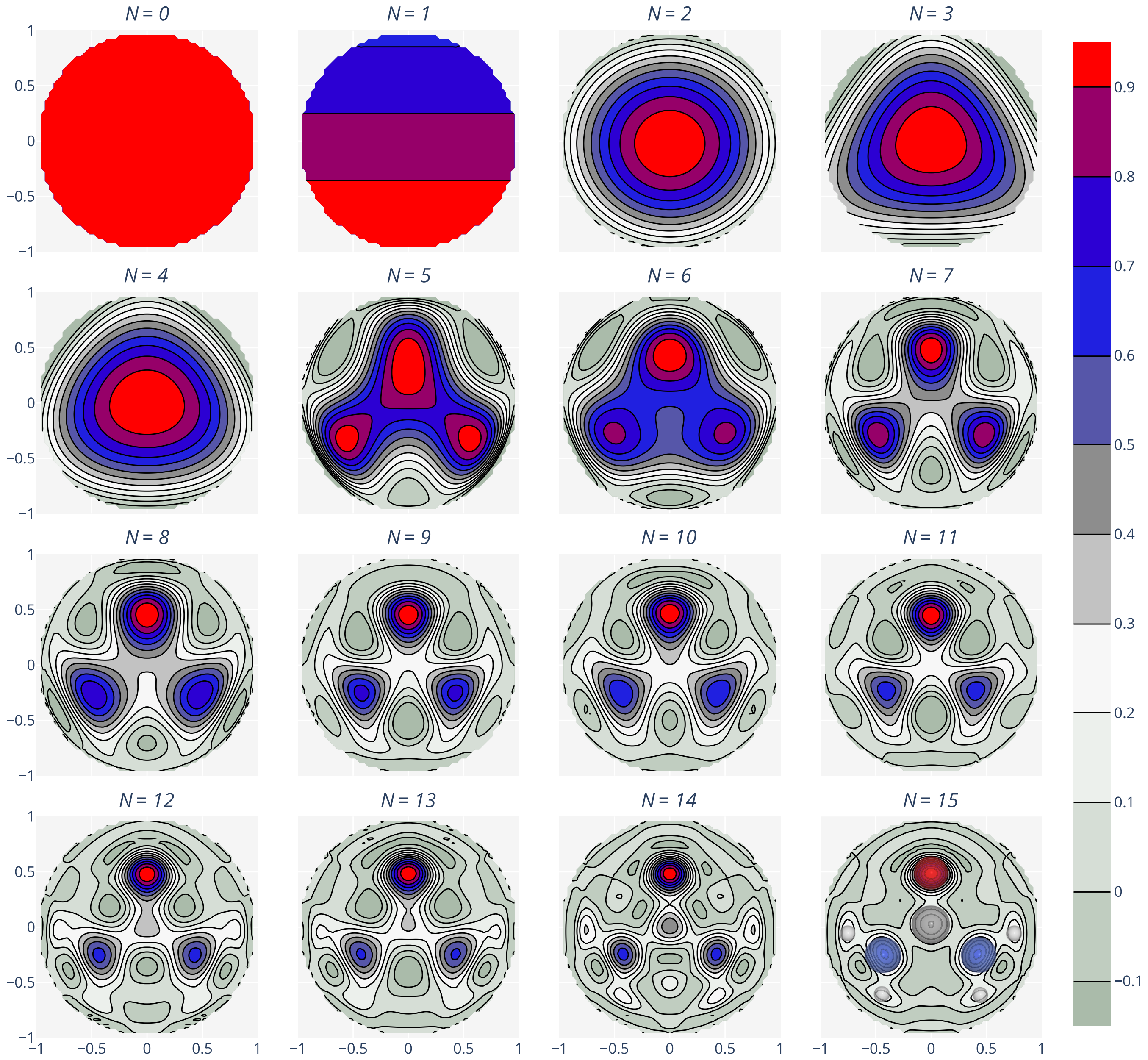


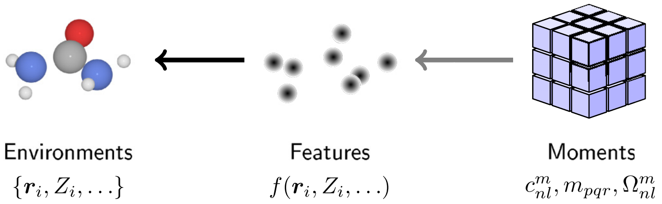
FIG. 2. Reconstructions from Zernike moments with increasing maximum expansion order N. The contour data is taken as a slice through a 3D grid corresponding to the plane of the molecule, a reduced opacity version of which is shown over $N = 15$. Atomic species have been mapped to the following weights of delta functions: $w_{\mathrm{H}} = 1.125$, $w_{\mathrm{C}} = 2.375$, $w_{\mathrm{N}} = 3.625$, $w_{\mathrm{O}} = 4.875$. The fractional color scale has been chosen to correspond to the standard CPK colors for each element.

be 10^{-5}. The orientation recovered is random, being determined by arbitrary choices made in picking a particular solution for the first and second order moments (or when solving for higher order moments if the original environment is highly symmetric). We use the Levenberg-Marquardt [71] least-squares solver implemented in SciPy [72] to perform this procedure, starting with an initial set of moments created from an environment of atoms placed randomly within the cutoff sphere and with species chosen randomly from the full set supported by the descriptor. However, we find that the solution is not particularly sensitive to the starting point and even random moments will often converge successfully. Algorithm 1 gives the pseudocode for the entire procedure and the python code is also available [73]. Once all of the moments are found, an approximation of the original function $f(\boldsymbol{r})$ can be reconstructed [Eq. (14)]. Figure 2 shows a urea molecule reconstructed using various values of maximum expansion order.

Algorithm 1. Numerical algorithm to solve for a set of moments from a given fingerprint. We use square brackets to indicate that only a subset of the array matching the condition is being used.

```
Inputs  : $\Phi_i$: fingerprint vector, $N_A$: num. atoms, $r_{cut}$: the descriptor cutoff radius, $r_{\min}$: minimum atoms
          separation, tol: the least-squares tolerance on the RMSD, max_retries: max retries for each $l$
Outputs: $c_{nl}^m$ the moments
# Make an initial guess and calculate the corresponding moments
positions ← RandomInitialConfiguration ($N_A$, $r_{\mathrm{cut}}$, $r_{\min}$);
$c_{nl}^m$ ← CalculateMoments (positions);
for $l' \leftarrow 1$ to $l_{max}$ do
    best_rmsd ← $\infty$;
    for 1 to max_retries do
        # Gradient descent w.r.t. moments
        moments, rmsd ← LstSq ($c_{nl}^m[l \leq l']$, $\Phi_i[l_{\mathrm{max}} \leq l']$);
        if rmsd < best_rmsd then
            # Retain this as the best minimum found so far
            best_rmsd ← rmsd;
            best_moments ← moments;
        if best_rmsd < tol then
            break;
        else
            $c_{nl}^m[l = l']$ ← Randomise ($c_{nl}^m[l = l']$);
    $c_{nl}^m[l \leq l']$ ← best_moments;
return $c_{nl}^m$;
```

B. Atoms from moments



Next, we are tasked with recovering the atomic positions and species from the moments. Once again this can be achieved by means of local minimization, this time between the atomic degrees of freedom and the moments themselves (for the sake of this procedure the features can be largely ignored as there is an analytic expression between them and the atomic coordinates).

If all of the weights are the same (i.e., all atoms are of the same specie), then the number of atoms can be determined directly from c_{00}^0, otherwise this moment simply acts as a constraint on the total weight of the delta (or Gaussian) functions in the environment, an alternative is to use two vectors of invariants, one encoding only positions and the other species, similar to the scheme proposed by Artrith *et al.* [74]. If no detailed information about the number of atoms of each specie is known, there are several ways to proceed: (1) attempt multiple optimizations with different numbers of atoms (that are consistent with c_{00}^0 and the allowed range of feature function weights) and keep the best fit or (2) use more atoms than necessary and merge overlapping atoms by summing their weights as part of post-processing.

The initial atomic positions for the optimisation are chosen using peak finding where we iteratively locate the highest density point in a 3D grid whose values are calculated according to Eq. (14). A cubic grid is used that contains the cutoff sphere, typically sampled using 31^3 points. Once an atom is found we

(1) subtract a single-atom signal from the density grid and

(2) zero out all grid values in the vicinity of the located atom to ensure the algorithm does not place any two atoms too close.

The procedure is summarized in algorithm 2 and an example is shown in Fig. 3.

Algorithm 2. Peak finding algorithm for determining discreet atom positions from the moments.

```
Inputs  : $c_{nl}^m$: moments, $N_A$: num. atoms, $r_{\mathsf{min}}$: minimum atoms separation, grid: the array of Cartesian
          coordinates to perform reconstruction over
Outputs: The atomic positions
positions ← list();
# Get values of function reconstruction on grid
values ← CalculateGrid (grid, $c_{nl}^m$);
for $i \leftarrow 1$ to $N_A$ do
    # Get grid values for isolated atom at
    # current maximum value
    point ← grid [argmax (values)];
    v ← CalculateGrid (CalculateMoments (point));
    # Remove contribution of this atom
    values ← values − v;
    # Zero grid values near this point
    # so no further atoms are placed near
    values [argwhere (grid - point < $r_{\mathsf{min}}$)] = 0;
    positions.append(point);
return positions;
```

Finally, to ensure that the optimisation avoids solutions that place atoms closer than is physically reasonable we add the following energy term to the cost function that biases it away from these configurations:

$$E(r_{ij}) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^n + k_1 r_{ij} + k_2\right] & \text{if } r_{ij} < r_{\min}, \\ 0 & \text{otherwise}, \end{cases} \tag{36}$$

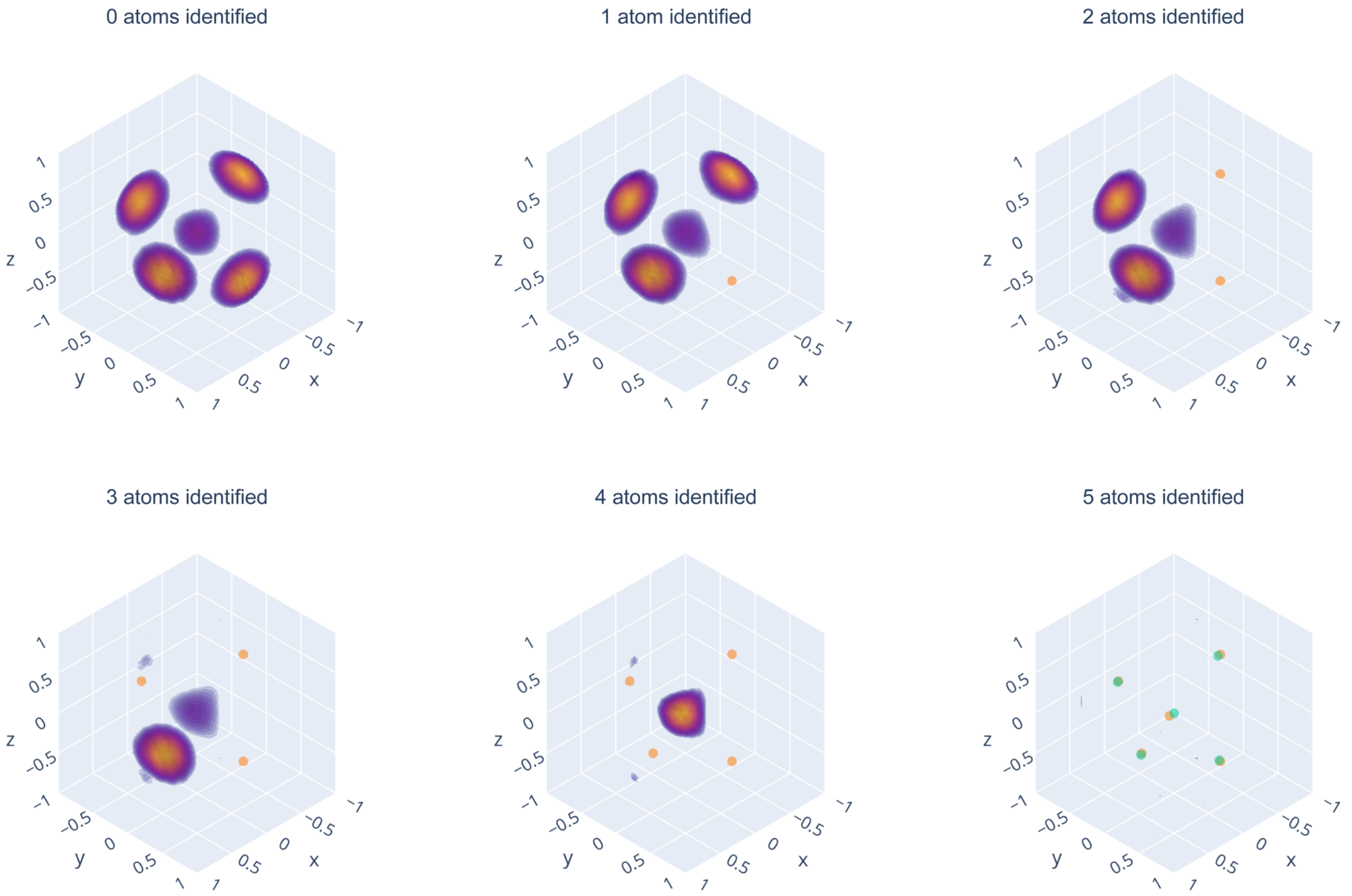

0 atoms identified
1 atom identified
2 atoms identified
3 atoms identified
4 atoms identified
5 atoms identified

FIG. 3. Peak finding from an approximation of the original environment $\tilde{f}$ based on Zernike moments. Atoms are located one by one whereupon they are replaced by an orange marker and the signal is subtracted for the next iteration. The final panel shows the original locations of the atoms as green markers.

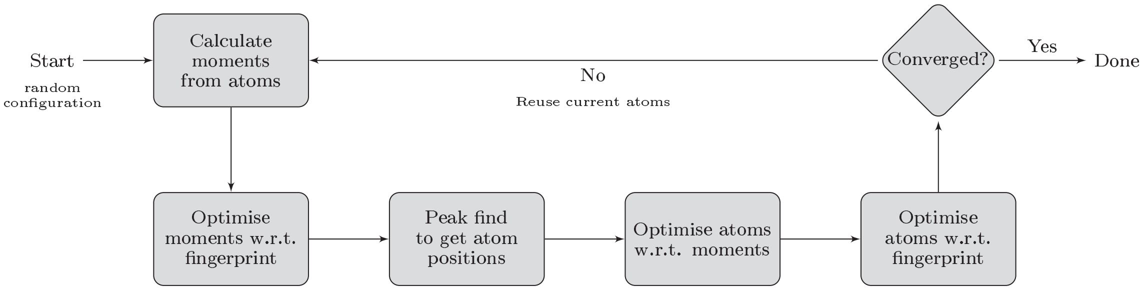


FIG. 4. Iterative scheme that finds atomic positions, and optionally species, by alternating between an optimization of the moments and an optimization of the atomic positions, both with respect to the known fingerprints.

where r_{ij} is the interatomic separation, k_1 and k_2 are constants set such that $E(r_{ij})$, and $\frac{dE}{dr_{ij}}$ both approach zero smoothly at the cutoff. We find that $\epsilon = 0.1, \sigma = 1, n = 2$, and $r_{\min} = 0.55$ Å work well and use these values throughout. The total loss function is thus

$$L\big(c^m_{nl}, c'^m_{nl}, \{\boldsymbol{r}_1, \dots, \boldsymbol{r}_{N_A}\}\big) = \mathrm{RMSD}\big(c'^m_{nl}, c^m_{nl}\big) + \sum_{ij,i\neq j}^{N_A} E(\|\boldsymbol{r_j} - \boldsymbol{r}_i\|), \qquad (37)$$

where c'^m_{nl} are the moments calculated from the atomic configuration at each step during the optimization.

C. Putting it all together

Even for the case of noise-free fingerprints the above procedures are not guaranteed to find the original atomic environment. The primary reason for this is that both optimizations (moments from invariants and atoms from moments) are performed over nonconvex functions and, as such, they can end up in the wrong minimum. This situation is easily detected as the gradients vanish while the RMSD remains high.

To overcome this we implement an algorithm according to the scheme shown in Fig. 4 which alternates between an optimization of the moments followed by a optimization of the corresponding atomic degrees of freedom. This is repeated a number of times until the RMSD drops below a preset threshold.

The algorithm presented here is designed to reconstruct a single atomic environment and makes the assumption that all atoms lie within the corresponding cutoff sphere. Outside this domain, the basis is no longer orthogonal, and therefore it is not strictly possible to reconstruct the atom density according to Eq. (14). Instead, the reconstruction of an atomic structure consisting of multiple overlapping cutoff spheres could be performed by simultaneously minimizing the loss over all environments in each step. However, this greatly complicates the problem as now the orientations and positions (if not fixed) of overlapping environments are coupled, potentially introducing additional minima to the optimization landscape. Furthermore, if the environments are not all identical, the problem is no longer permutationally invariant with respect to the assignment of fingerprints to each environment. For these reasons, we leave the reconstruction of multicenter configurations for future work.

IV. EXPERIMENTS

A. Reconstruction

To assess the quality of various reconstructions from moment invariants, we use a subset of the QM9 database [75,76], a database of small organic molecules. Three molecules of each size (in number of atoms) were chosen randomly, ranging from $N_A = 3$ to 29, with the exception of 3 atom molecules of which there are only two and 28 atom molecules of which there are none. This corresponds to a total test set of 77 molecules. The full set of QM9 IDs used can be found in the Supplemental Material [77]. We use weighted delta functions as features. If only position information is being used, then the weights are all fixed to 1, otherwise the atomic numbers of HCONF are mapped onto the continuous range 1 → 2 (i.e., $w_{\mathrm{H}} = 1.1$, $w_{\mathrm{C}} = 1.3$, $w_{\mathrm{O}} = 1.5$, $w_{\mathrm{N}} = 1.7$, $w_{\mathrm{F}} = 1.9$). In the latter case the weights are free to vary during optimization, and are mapped back onto the correct integers as part of post-processing. For example, a value falling in the interval [1, 1.2) would be mapped to hydrogen. The cutoff sphere is set at 5 Å such as to accommodate the largest molecule in the data set. Each molecule is positioned by finding the smallest bounding sphere containing all atoms using the miniball library [78] and translating the center of the sphere to the origin.

In the first instance, the quality of the reconstruction is judged by the root-mean square of the difference between the calculated fingerprint and that of the recovered atomic environment $\sqrt{\sum_i (\Phi_i - \Phi_i')^2 / N_\Phi}$. The validity of this as a measure of structural similarity is demonstrated in the Supplemental Material [77]. We note that MILAD fingerprints are unable to distinguish between chiral versions of a molecule and therefore a low RMSD can indicate that the original or a chiral counterpart have been recovered. A summary of the parameters used is tabulated below.

Total molecules	77
Species	HCONF
Feature function	delta
Weights	1 → 2
Cutoff	5Å
Invariants	Zernike (to $n_{\max} = 7$), 117 total
Reconstruction attempts	3

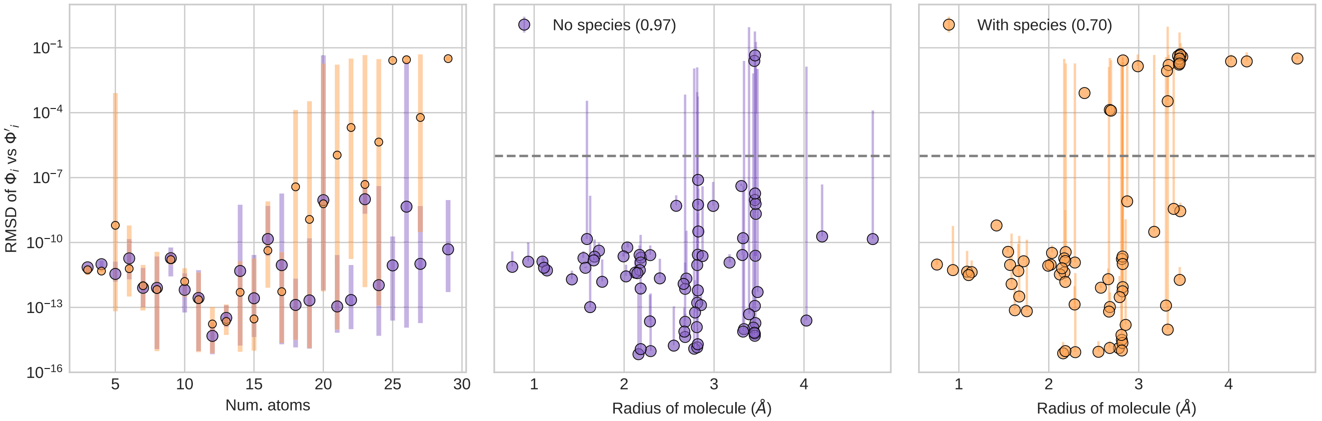


FIG. 5. Results showing the RMSD of the original fingerprints versus those calculated from the reconstructed molecules. The leftmost panel shows results aggregated by number of atoms. The right panels shows results for each molecule individually where the marker is placed at the lowest RMSD. Error bars indicate the min/max for that data point over several reconstruction attempts. A value of $\approx 10^{-6}$ (dashed line) represents a faithful reconstruction, the legend shows the proportion of markers falling below this threshold.

In each case the fingerprint is calculated and used to perform a reconstruction according to the procedure outlined in Sec. III C.

Figure 5 shows results from the reconstruction of atomic positions (orange) and atomic positions plus species (purple). For each molecule three reconstruction attempts were made, in each case starting from a random configuration with the correct number of atoms. A manual inspection of the recovered molecules indicates that an RMSD of $< 10^{-6}$ represents a visually indistinguishable reconstruction versus the original and we use this as the threshold of successful reconstruction. The recovery of atomic positions is generally highly reliable, achieving successful reconstruction for 97% of molecules and all molecules with a radius of less than ≈3.4 Å. Any larger than this and the effective resolution of the basis (see Fig. 2) is insufficient for consistent reconstruction on every attempt. The largest molecules that can be decoded have 29 atoms translating to 85 degrees of freedom after symmetrization, with consistent recovery up to 19 atoms or 54 degrees of freedom. These numbers provide some indication of the efficiency with which the fingerprint encodes the symmetrized atom density given that it consists of 117 invariants. Reconstruction of positions and species is less reliable, achieving an overall 70% success rate with consistent recovery up to ≈2.4 Å. In this case there is one more degree of freedom per atom meaning that the largest molecule decoded, with 27 atoms, has 105 degrees of freedom while the largest molecules that can be consistently decoded have 17 atoms and 65 degrees of freedom.

Next, to improve upon the recovery of species information we take the reconstructed atomic positions from the last experiment and perform a reconstruction of the weights with respect to fingerprints with species information. This represents a scenario where two fingerprint vectors are supplied, one containing position information only and the other augmented with species information, similar to using an additional color channel in an image.

Figure 6 shows the species recovery results with the original position recovery data replotted for easy comparison. This shows a notable improvement with 95% of structures now

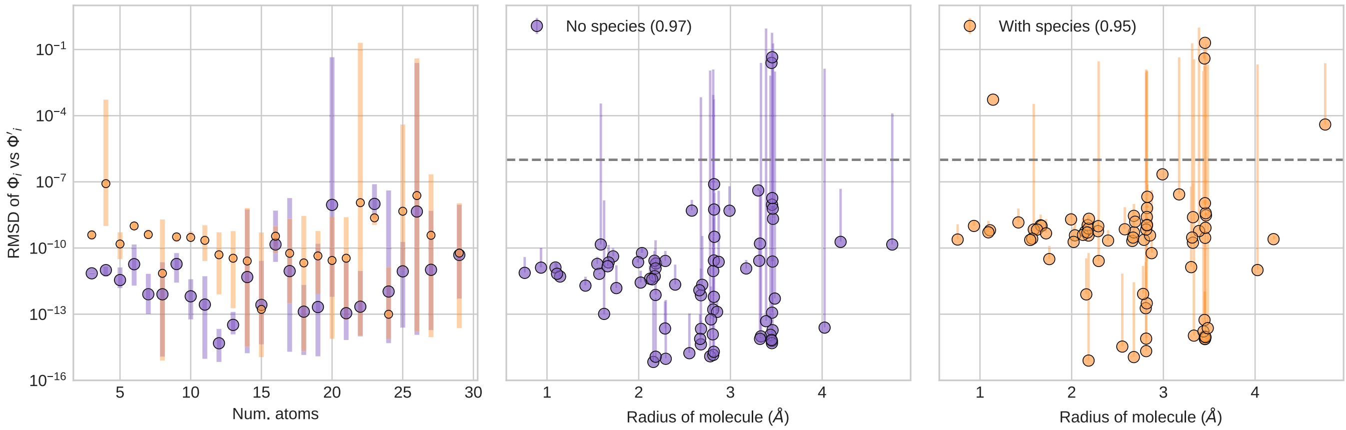
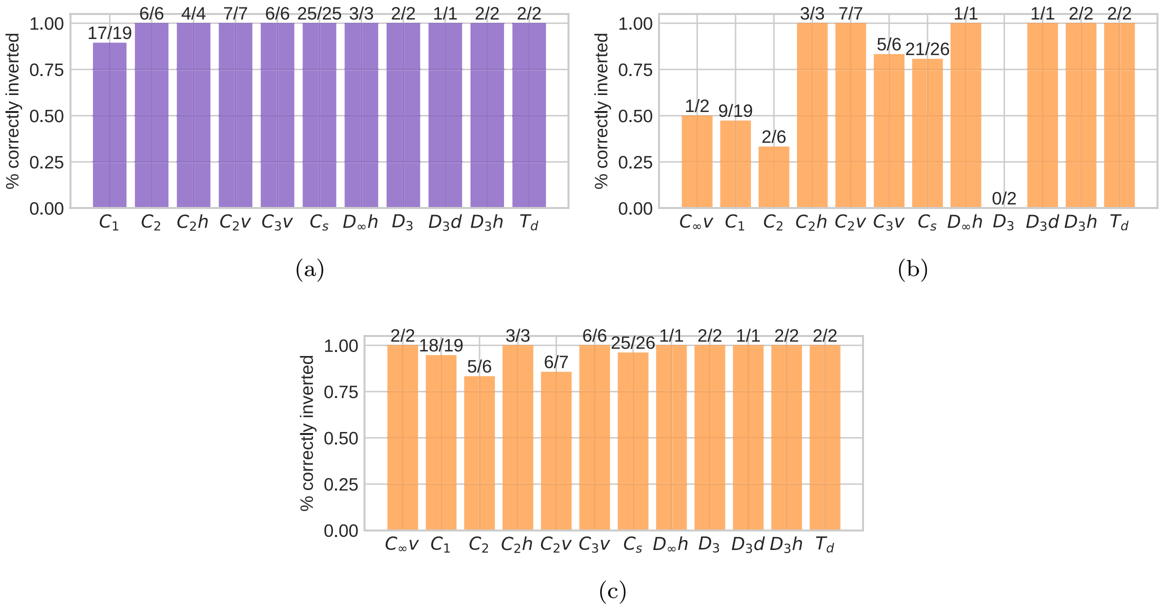


FIG. 6. A plot showing recovery of atomic species using a second fingerprint vector including species information. Results from decoding atomic positions only ("no species") are replotted from Fig. 5 for ease of comparison.



(a) (b)

(c)

FIG. 7. Breakdown by point group showing the proportion of molecules correctly inverted. Numbers above bars indicate the absolute number correctly inverted out of the total. (a) No species, (b) With species (single fingerprint), and (c) With species (two fingerprints).

being recovered correctly. With few exceptions, the structures whose positions were correctly recovered in the previous experiment now have their species correctly decoded.

Finally, Fig. 7 shows the reconstruction results broken down by the point group of the molecules. This analysis is important to verify that the inversion algorithm works correctly in the presence of symmetry which can lead to many invariants being zero. As can be seen, there is no discernible trend that would suggest that the algorithm cannot deal with symmetric molecules.

B. Neural network potential

To give an indication of the performance of MILAD for the prediction of properties we use a Behler-Parinello [79] like ANN from the AMP library [80] for the prediction of total energies and forces. As shown in Fig. 8 each atom-centered environment is mapped onto a fingerprint vector $\mathbf{\Phi}_i$ that is the input to an ANN consisting of a number of all-to-all connected linear layers. The local atomic contributions to the energy are then summed to give the total energy. We find the hyperbolic tangent activation function to offer a good combination of training speed and accuracy. The linear layers may be configured with or without a bias. Our testing suggests that it is preferable to have no bias but rather to rescale the model inputs and outputs to lie in the fixed range $(-1, 1)$ based on the range of values found in the training set. This also makes it easy to detect when extrapolation is taking place, as in such a case the input or output will fall outside of this range.

Forces are calculated via the chain rule using derivatives of the energy with respect to neural network weights and of the invariants with respect to atomic positions. To achieve smoothness of the fingerprint vector with respect to atoms entering and leaving the cutoff sphere, the following cosine cutoff function proposed by Behler [3] is used:

$$f(r_{ij}) = \begin{cases} \left[\cos\left(\frac{\pi r_{ij}}{r_{\mathrm{cut}}}\right) + 1\right]/2 & r_{ij} \leqslant r_{\mathrm{cut}}, \\ 0 & r_{ij} > r_{\mathrm{cut}}. \end{cases} \tag{38}$$

The value of this function is used as a prefactor to scale atomic feature weights.

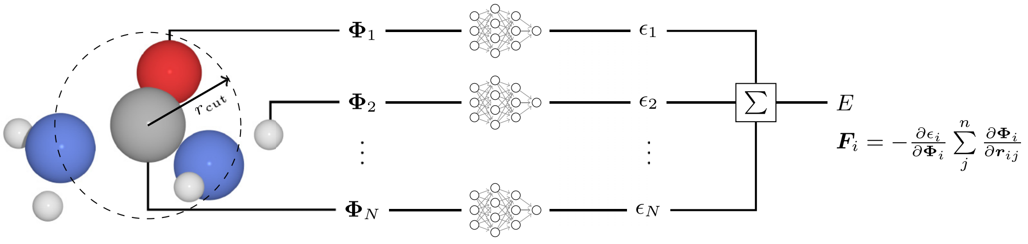


FIG. 8. The Behler-Parinello like ANN scheme used to fit total energies and local forces. A rotation invariant MILAD vector $\mathbf{\Phi}_i$ is created for each atomic environment by including neighbors up to the cutoff radius. Each $\mathbf{\Phi}_i$ is passed to a feed-forward neural network which outputs the corresponding contribution to the total energy ϵ_i. Finally, the total energy E is given by the sum of local contributions.

TABLE II. Model parameters for each of the systems studied. The number of neural network layers for the NNP models from [82] are also tabulated for comparison.

	Training point	Test points	r_{cut} (Å)	α Eq. (39)	Network layers	NNP network layers
Ni	263	31	3.8	0.04	32-32	24-24
Cu	262	31	3.9	0.04	24-24	8-8
Li	241	29	4.8	0.04	32-32	24-24
Mo	194	23	5.0	0.04	32-32	16-16
Si	214	25	4.7	0.04	32-32	24-24
Ge	228	25	5.1	0.05	32-32	24-24

1. Training

The following loss function is used to train the neural network:

$$L = \frac{1}{2}\sum_{k=1}^{M}\frac{1}{N_k}\left[(E_k - E'_k)^2 + \frac{\alpha}{3}\sum_{i=1}^{N_k}\sum_{j=1}^{3}\left(\boldsymbol{F}_{i,j} - \boldsymbol{F'}_{i,j}\right)^2\right], \tag{39}$$

where k labels each of the M systems in the training set, N_k is the number of atoms in the kth system, and primed quantities represent the known training values. During each training step predictions are made from which the loss function is calculated. Gradients with respect to the loss, obtained using back-propagation, are then used by the optimizer to update the network weights in an attempt to reduce the value of L. We use the BFGS optimizer throughout.

Given that artificial neural networks (ANNs) are universal approximators, care must be taken to avoid overfitting. Indeed, even random inputs can be successfully [81] learned, thus rendering any conclusions based on a training procedure where all of the data is used meaningless. To avoid this we keep back some of the data as an unseen test set and apply *early stopping*. This involves calculating the loss for both training and test sets and terminating training when the loss of the test set starts to increase. In general, we find this situation is much less likely to occur when training with both forces and energies.

2. The experiment

We use data from two recent studies [59,82] comparing various descriptors and regression schemes. The studies compared GAP [83,84], MTP [8], a neural network using Behler-Parinello symmetry functions as the descriptor (labeled NNP) [3,79], ACE [43], SNAP [85], and qSNAP [86]. The latter two both use hyperspherical harmonics as the basis where the radial component is mapped onto the surface of a four-dimensional sphere. SNAP and qSNAP both use bispectrum ($\nu = 3$) invariants, the former expressing the energy as a linear expansion while the latter includes quadratic terms.

The data set consists of elemental Ni, Cu, Li, Mo, Si, and Ge in a variety of configurations including the ground state crystal structures, strained structures, slabs, and *ab initio* molecular dynamics snapshots. The reference energies and forces were calculated using the VASP density functional theory package and the Perdew-Burke-Ernzerhof functional [87]. For each elemental system there are between 217 and 294 total configurations organized into a 9:1 split of training and test points.

Table II shows the settings used to configure the MILAD descriptor and the neural network for each system. Limited hyperparameter tuning was performed to determine the number of neural-network layers to use for each system starting with the same values as reported for the NNP model. In all cases we found that increasing the number of layers resulted in better accuracy, likely due to the fact that the MILAD fingerprint has more components then the number symmetry functions used by the NNP.

TABLE III. Root-mean-square errors in predicted energies and forces. Data for the three-body ACE descriptor are taken from Zeni *et al.* [59] while the rest come from Zuo *et al.* [82].

	Energy RMSD (meV/atom)						Force RMSD (eV/Å)					
	Ni	Cu	Li	Mo	Si	Ge	Ni	Cu	Li	Mo	Si	Ge
GAP	0.62	0.56	0.63	3.55	4.18	4.47	0.04	0.02	0.01	0.16	0.12	0.08
MTP	0.74	0.52	0.66	3.89	3.02	3.68	0.03	0.01	0.01	0.15	0.09	0.07
NNP	2.25	1.68	0.98	5.67	9.95	10.95	0.07	0.06	0.06	0.20	0.17	0.12
SNAP	1.17	0.87	1.13	9.06	8.06	10.96	0.08	0.08	0.04	0.37	0.34	0.29
qSNAP	1.04	1.16	0.85	3.96	6.28	10.55	0.07	0.05	0.04	0.33	0.29	0.20
Three-body ACE [59]	1.74	1.19	1.23	4.00	5.16	11.62	0.03	0.02	0.01	0.16	0.13	0.09
MILAD	1.39	0.96	0.64	5.79	5.65	5.47	0.08	0.07	0.02	0.36	0.19	0.14
Empirical potentials												
EAM	8.51	7.46	368.64	67.98	–	–	0.11	0.12	0.14	0.52	–	–
MEAM	23.04	10.49	–	36.42	111.67	–	0.33	0.24	–	0.22	0.40	–
Tersoff	–	–	–	–	202.37	550.72	–	–	–	–	0.74	1.36

Table III shows the results retabulated from [59,82] along with those obtained using MILAD. In most cases, the energy RMSDs achieved with MILAD on the test set lie in the range of ACE, NNP, SNAP, and qSNAP but, in most cases, higher than GAP and MTP. It is unclear at this stage whether this is due to the difference in regression scheme (GAP uses kernel regression while MTP uses linear regression) or the descriptors themselves. For example, Eq. (30) allows one to transform from the geometric moments used by MTP to Zernike moments used by MILAD suggesting that, at the same correlation order, they should share similar information content. However, there remain nontrivial differences in the way the radial basis is constructed. Specifically, MTP typically uses a data-driven approach with the basis being tailored to a particular data set rather than being universal. On the other hand, the training set size is relatively small for neural network regression which tends to be better suited to problems with large amounts of training data. Comparing to NNP, MILAD achieves lower RMSDs for most of the systems at the cost of requiring layers with more artificial neurons owing to the greater number of fingerprint components.

Looking to the forces the results are similar with MILAD RMSDs generally falling within in range achieved by NNP, SNAP, and qSNAP and higher than those achieved by MTP, GAP, and ACE. Our testing shows that this trend persists even when increasing the value of α to try and bias the optimizer to converge the forces preferentially over the energies. It is possible that by performing a grid search over the hyperparameters, as done by Zuo *et al.* [82], that the forces could be converged further still. An in depth investigation into this and other factors that affect the performance of MILAD for predicting properties will form part of future work.

V. CONCLUSION

In this work we have adapted a method for constructing complete, rotationally invariant, descriptions of finite energy functions from the image analysis community and applied it to the problem of describing atomic environments. The resulting invariants are algebraically complete, consisting of exactly three fewer equations than the number of moments due to the missing orientation degrees of freedom. This compactness makes MILAD fingerprints particularly well suited for use as inputs to neural networks where there is a significant training benefit to having a low-dimensional feature space.

The ability to invert MILAD fingerprints to recover atomic environments is particularly appealing and raises the possibility of building generative models that are not based on discontinuous description such as pixels or voxels [88], or linear representations such as SMILES strings [89]. Furthermore, many generative models depend on the use of an autoencoder that must be trained to find a latent space representation, and this would typically need to be retrained to be used for each new system. With our descriptor it is possible to encode a variable number of atoms and atomic species in a fixed length feature vector which is smooth with respect to its inputs. This enables alchemical models to be built that can be trivially extended to support greater numbers of species, the primary limitation being the effective resolution of the description which can be tuned by increasing the maximum order n_{max}. As the latent space formed by MILAD fingerprints can encode any $f \in L^2$, it remains overcomplete with respect to those f that correspond to sums of atomic feature functions. As a next step towards building a generative model, work is currently underway to further compress this latent space to find lower dimensional manifolds that correspond to valid atomic configurations. To be sure, this is an ambitious goal that raises many unsolved issues, it is nevertheless an avenue worth pursuing as it would eliminate the need for data augmentation and ensure that rotational symmetry is respected exactly.

The current set of invariants are limited to a maximum order of $n_{\text{max}} = 7$ (and therefore $l_{\text{max}} = 7$), which effectively limits their discriminative power. This is known to be particularly problematic for highly symmetric environments which require a high angular frequency to be described correctly (in such environments, low order moments are often zero) [90]. The limitation of $n_{\text{max}} = 7$ comes from numerical issues in correctly identifying independent invariants, as standard singular value decomposition methods operate on floating point numbers where a threshold for comparing numbers must be carefully chosen. This inevitably leads to errors as the number of invariants to be reduced increases. To overcome this, the method is currently being extended to higher order by using an exact arithmetic library and applying algorithms from the works of Bachmayr *et al.* [44] and Nigam *et al.* [47].

In summary, we have shown that MILAD fingerprints can be used for both high-fidelity reconstruction of atomic environments without the need for training and for accurately predicting the properties of atomic systems. This hints at a novel route to building generative models using existing machine learning tools with a view to enabling general-purpose inverse design of materials and molecules.

The codebase for calculating MILAD fingerprints, decoding them back into structures, training neural networks, and other related operations can be found at [91]. The notebooks that can be used to reproduce the experiments and most of the images found in this work are located at [92]. The QM9 data set used for reconstruction experiments can be found at Ramakrishnan *et al.* [76]. The data set for neural network potential fitting can be found at [93]. An additional neural network fitting experiment found in the Supplemental Material [77] uses data from Dragoni *et al.* [94–101].

ACKNOWLEDGMENTS

I would like to thank Tomáš Suk for an enlightening exchange about image moments, Nongnuch Artrith and Andrew Peterson for discussions on neural network based empirical potentials, Peter Bjørn Jørgensen for discussions about learning architectures and rotation invariants, Alexander Sougaard Tygesen for discussions about recovering atoms from moments, and Tess Smidt and Thomas Hardin for discussions on inverting invariants, particularly for suggesting recovering moments from invariants directly.

[1] B. Sanchez-Lengeling and A. Aspuru-Guzik, Science **361**, 360 (2018).

[2] A. Y. Ng and M. I. Jordan, in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, Series NIPS'01 (MIT Press, Cambridge, MA, USA, 2001), pp. 841–848.

[3] J. Behler, J. Chem. Phys. **134**, 074106 (2011).

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

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

[6] O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp, and A. Knoll, Int. J. Quantum Chem. **115**, 1084 (2015).

[7] G. Ferré, J. B. Maillet, and G. Stoltz, J. Chem. Phys. **143**, 104114 (2015).

[8] A. V. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).

[9] H. Huo and M. Rupp, arXiv:1704.06439.

[10] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, and M. Ceriotti, J. Chem. Phys. **148**, 241730 (2018).

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

[12] M. F. Langer, A. Goeßmann, and M. Rupp, arXiv:2003.12081.

[13] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Chem Rev. **121**, 9759 (2021).

[14] It is not strictly necessary to work with an orthogonal basis, but in this case the basis functions need to be linearly independent and invertible for direct synthesis to be possible.

[15] Note that for particular classes of image function there may still be differences in convergence rates for different choices of orthonormal basis functions.

[16] J. Flusser, T. Suk, and B. Zitová, *2D and 3D Image Analysis by Moments* (John Wiley & Sons, Ltd, Chichester, UK, 2016), pp. 1–529.

[17] T. H. Reiss, in *Recognizing Planar Objects Using Invariant Image Features*, Lecture Notes in Computer Science, edited by T. H. Reiss (Springer, Berlin, 1993), Vol. 676, Chap. 2.

[18] V. Zaverkin and J. Kästner, J. Chem. Theory Comput. **16**, 5410 (2020).

[19] I. Sommer, O. Muller, F. S. Domingues, O. Sander, J. Weickert, and T. Lengauer, Bioinformatics **23**, 3139 (2007).

[20] M. J. Willatt, F. Musil, and M. Ceriotti, J. Chem. Phys. **150**, 154110 (2019).

[21] In fact, σ can even be a covariance matrix, giving the atomic feature an asymmetric shape. Something that can be exploited to encode anisotropic properties such as spins or chemical shifts.

[22] The first few orders are given in the Supplemental Material [77], the MILAD source code contains explicit forms for up to order 16 inclusive.

[23] F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, and L. Emsley, Nat. Commun. **9**, 4501 (2018).

[24] F. A. Sadjadi and E. L. Hall, IEEE Trans. Pattern Analysis Machine Intelligence **PAMI-2**, 127 (1980).

[25] T. Suk and J. Flusser, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 6855 LNCS (2011), pp. 212–219.

[26] D. Cyganski and J. A. Orr, IEEE Trans. Pattern Analysis Machine Intelligence **PAMI-7**, 662 (1985).

[27] J. Kostková, T. Suk, and J. Flusser, IEEE Trans. Pattern Analysis Machine Intelligence **43**, 1140 (2021).

[28] Note that typically the Suk-Flusser invariants are given in terms of the central geometric moments and scale normalized by dividing by an appropriate power of the m_{000} moment which encodes the total "mass" of the system. For atomic systems, however, scale invariance is rarely desired.

[29] K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, J. Chem. Phys. **148**, 241727 (2018).

[30] C.-H. Lo and D. Hon-Son, IEEE Trans. Pattern Analysis Machine Intelligence **11**, 1053 (1989).

[31] N. Canterakis, in *11th Scandinavian Conference on Image Analysis* (Pattern Recognition Society of Denmark, 1999), pp. 85–93.

[32] M. Novotni and R. Klein, Computer-Aided Design **36**, 1047 (2004).

[33] v. F. Zernike, Physica **1**, 689 (1934).

[34] T. Suk, J. Flusser, and J. Boldyš, Pattern Recognition **48**, 3516 (2015).

[35] N. L. Max and E. D. Getzoff, IEEE Comput. Graphics Appl. **8**, 42 (1988).

[36] B. S. Duncan and A. J. Olson, Biopolymers **33**, 219 (1993).

[37] B. S. Duncan and A. J. Olson, Biopolymers **33**, 231 (1993).

[38] B. S. Duncan and A. J. Olson, J. Mol. Graphics **13**, 250 (1995).

[39] R. J. Morris, R. J. Najmanovich, A. Kahraman, and J. M. Thornton, Bioinformatics **21**, 2347 (2005).

[40] D. W. Ritchie and G. J. Kemp, J. Comput. Chem. **20**, 383 (1999).

[41] W. Cai, X. Shao, and B. Maigret, J. Mol. Graphics Model. **20**, 313 (2002).

[42] S. Grandison, C. Roberts, and R. J. Morris, J. Comput. Biol. **16**, 487 (2009).

[43] R. Drautz, Phys. Rev. B **99**, 014104 (2019).

[44] G. Dusson, M. Bachmayr, G. Csanyi, R. Drautz, S. Etter, C. van der Oord, and C. Ortner, arXiv:1911.03550.

[45] R. Drautz, Phys. Rev. B **102**, 024104 (2020).

[46] A. Grisafi, J. Nigam, and M. Ceriotti, Chem. Sci. **12**, 2078 (2021).

[47] J. Nigam, S. N. Pozdnyakov, and M. Ceriotti, J. Chem. Phys. **153**, 121101 (2020).

[48] Y. S. Abu-Mostafa and D. Psaltis, IEEE Trans. Pattern Analysis Machine Intelligence **PAMI-7**, 46 (1985); L. Vinet and A. Zhedanov, J. Phys. A: Math. Theor. **44**, 085201 (2011).

[49] D. Bhattacharya and S. Sinha, Pattern Recognition **30**, 1373 (1997).

[50] J. Flusser and T. Suk, IEEE Trans. Image Processing **15**, 3784 (2006).

[51] J. Flusser, T. Suk, and B. Zitová, *Moments and Moment Invariants in Pattern Recognition* (John Wiley & Sons, Ltd, Chichester, UK, 2009), pp. 1–296.

[52] F. Ghorbel, S. Derrode, R. Mezhoud, T. Bannour, and S. Dhahbi, Pattern Recognition Lett. **27**, 1361 (2006).

[53] N. Thomas, T. E. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, arXiv:1802.08219.

[54] R. Kondor, Z. Lin, and S. Trivedi, Adv. Neural Inform. Processing Syst. **31**, 10117 (2018).

[55] F. B. Fuchs, V. Fischer, D. E. Worrall, and M. Welling, arXiv:2006.10503.

[56] B. K. Miller, M. Geiger, T. E. Smidt, and F. Noé, arXiv:2008.08461.

[57] M. Geiger, T. E. Smidt, A. Musaelian, B. K. Miller, W. Boomsma, B. Dice, K. Lapchevskyi, M. Weiler, M. Tyszkiewicz, S. Batzner, M. Uhrin, J. Frellsen, N. Jung, S. Sanborn, J. Rackers, and M. Bailey, **0.3.5** (2021).

[58] Y. Lysogorskiy, C. van der Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, and R. Drautz, npj Comput. Mater. **7**, 97 (2021).

[59] C. Zeni, K. Rossi, A. Glielmo, and S. De Gironcoli, J. Chem. Phys. **154**, 224112 (2021).

[60] Z. Guo, N. Saxena, and A. Sinhababu, Leibniz Int. Proc. Inform. **102**, 101 (2018).

[61] M. Langbein and H. Hagen, A Generalization of Moment Invariants on 2D Vector Fields to Tensor Fields of Arbitrary Order and Dimension, in *Advances in Visual Computing*, edited by G. Bebis, R. Boyle, B. Parvin, D. Koracin, Y. Kuno, J. Wang, R. Pajarola, P. Lindstrom, A. Hinkenjann, M. L. Encarnação, C. Silva, and D. Coming (Springer Berlin Heidelberg, 2009), pp. 1151–1160.

[62] M. S. Hickman, J. Math. Imaging Vision **44**, 223 (2012).

[63] A. Goscinski, F. Musil, S. N. Pozdnyakov, and M. Ceriotti, J. Chem. Phys. **155**, 104106 (2021)

[64] E. Kocer, J. K. Mason, and H. Erturk, AIP Adv. **10**, 015021 (2020).

[65] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1989).

[66] P. R. Clement, SIAM Rev. **5**, 131 (1963).

[67] A. S. Bandeira, B. Blum-Smith, J. Kileel, A. Perry, J. Weed, and A. S. Wein, arXiv:1712.10163.

[68] S. N. Pozdnyakov, L. Zhang, C. Ortner, G. Csányi, and M. Ceriotti, arXiv:2109.11440.

[69] C. J. Pickard and R. J. Needs, J. Phys.: Condens. Matter **23**, 53201 (2011).

[70] R. W. Gerchberg and W. O. Saxton, Optik **35**, 237 (1972).

[71] J. J. Moré, *Numerical Analysis*, edited by G. A. Watson (Springer Berlin Heidelberg, 1978), pp. 105–116.

[72] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, Nat. Methods **17**, 261 (2020).

[73] https://github.com/muhrin/milad/blob/v0.3.0/milad/optimisers/moments_optimiser.py.

[74] N. Artrith, A. Urban, and G. Ceder, Phys. Rev. B **96**, 014112 (2017).

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

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

[77] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.104.144110 for definitions of associated Legendre functions used; four invariants added to set of 113 from Flusser-Suk; validation of fingerprint RMSD as measure of molecule similarity; QM9 IDs used in reconstruction experiments; additional neural network potential experiment on Fe.

[78] https://pypi.org/project/miniball/.

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

[80] A. Khorshidi and A. Peterson, Comput. Phys. Commun. **207**, 310 (2016).

[81] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, arXiv:1611.03530.

[82] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, J. Phys. Chem. A **124**, 731 (2020).

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

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

[85] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, J. Comput. Phys. **285**, 316 (2015).

[86] M. A. Wood and A. P. Thompson, J. Chem. Phys. **148** (2018).

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

[88] M. S. Jørgensen, H. L. Mortensen, S. A. Meldgaard, E. L. Kolsbjerg, T. L. Jacobsen, K. H. Sørensen, and B. Hammer, J. Chem. Phys. **151**, 054111 (2019).

[89] V. Nesterov, M. Wieser, and V. Roth, arXiv:2010.06477.

[90] T. Suk and J. Flusser, Symmetry **6**, 722 (2014).

[91] https://github.com/muhrin/milad.

[92] https://github.com/muhrin/milad-paper-2021.

[93] https://github.com/materialsvirtuallab/mlearn.

[94] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, Gaussian Approximation Potentials for iron from extended first-principles database (Data Download) (2017).

[95] J. Kromann, RMSD (2021).

[96] G. Landrum, RDKit: Open-source cheminformatics (2021).

[97] S. K. Lam, A. Pitrou, and S. Seibert, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC - LLVM '15* (ACM, New York, 2015), pp. 1–6.

[98] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry *et al.*, PeerJ Comput. Sci. **3**, e103 (2017).

[99] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. D'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), pp. 8024–8035.

[100] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, Phys. Rev. Mater. **2**, 013808 (2018).

[101] D. P. Kingma and J. Ba, AIP Conf. Proc. **1631**, 58 (2014).