

Tensor-Reduced Atomic Density Representations

James P. Darby^{1,2,*}, Dávid P. Kovács^{2,*}, Ilyes Batatia,^{2,3} Miguel A. Caro⁴, Gus L. W. Hart⁵,
Christoph Ortner,⁶ and Gábor Csányi²

¹Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry, CV4 7AL, United Kingdom

²Engineering Laboratory, University of Cambridge, Cambridge, CB2 1PZ, United Kingdom

³ENS Paris-Saclay, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

⁴Department of Electrical Engineering and Automation, Aalto University, FIN-02150 Espoo, Finland

⁵Department of Physics and Astronomy, Brigham Young University, Provo, Utah, 84602, USA

⁶Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, British Columbia, Canada V6T 1Z2



(Received 6 October 2022; accepted 18 April 2023; published 13 July 2023)

Density-based representations of atomic environments that are invariant under Euclidean symmetries have become a widely used tool in the machine learning of interatomic potentials, broader data-driven atomistic modeling, and the visualization and analysis of material datasets. The standard mechanism used to incorporate chemical element information is to create separate densities for each element and form tensor products between them. This leads to a steep scaling in the size of the representation as the number of elements increases. Graph neural networks, which do not explicitly use density representations, escape this scaling by mapping the chemical element information into a fixed dimensional space in a learnable way. By exploiting symmetry, we recast this approach as tensor factorization of the standard neighbour-density-based descriptors and, using a new notation, identify connections to existing compression algorithms. In doing so, we form compact tensor-reduced representation of the local atomic environment whose size does not depend on the number of chemical elements, is systematically convergible, and therefore remains applicable to a wide range of data analysis and regression tasks.

DOI: [10.1103/PhysRevLett.131.028001](https://doi.org/10.1103/PhysRevLett.131.028001)

Over the past decade, machine learning methods for studying atomistic systems have become widely adopted [1–3]. Most of these methods utilize representations of local atomic environments that are invariant under relevant symmetries—typically rotations, reflections, translations and permutations of equivalent atoms [4,5]. Enforcing these symmetries allows for greater data efficiency during model training and ensures that predictions are made in a physically consistent manner. There are many different ways of constructing such representations which are broadly split into two categories: (i) descriptors based on internal coordinates, such as the Behler-Parrinello atom-centered symmetry functions [6], and (ii) density-based descriptors such as smooth overlap of atomic positions (SOAP) [7] or the bispectrum [8,9], which employ a symmetrized expansion of ν correlations of the atomic neighborhood density ($\nu = 2$ for SOAP and $\nu = 3$ for the bispectrum). A major drawback of all these representations is that their size increases dramatically with the number of

chemical elements S in the system. For instance, the number of features in the linearly complete atomic cluster expansion (ACE) [10,11] descriptor which unifies, extends, and generalizes the aforementioned representations, scales as S^ν for terms with correlation order ν (i.e., a body order of $\nu + 1$). This poor scaling severely restricts the use of these representations in many applications. For example, in the case of machine learned interatomic potentials for systems with many (e.g. more than five) different chemical elements, the large size of the models results in memory limitations being reached during parameter estimation as well as significantly reducing evaluation speed.

Multiple strategies to tackle this scaling problem have been proposed including element weighting [12,13] or embedding the elements into a fixed small dimensional space [14,15], directly reducing the element-sensitive correlation order [16], low-rank tensor-train approximations for lattice models [17] and data-driven approaches for selecting the most relevant subset or combination of the original features for a given dataset [18–20]. A rather different class of machine learning methods is message passing neural networks (MPNNs) [21,22]. Instead of constructing full tensor products, these models also embed chemical element information in a fixed size latent space using a learnable transformation $\mathbb{R}^S \rightarrow \mathbb{R}^K$ where K is the

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

dimension of the latent space, and thus avoid the poor scaling with the number of chemical elements. Recently these methods have achieved very high accuracy [23–25], strongly suggesting that the true complexity of the relevant chemical element space does not grow as S^ν .

In this Letter we introduce a general approach for significantly reducing the scaling of density-based representations like SOAP and ACE. We show that by exploiting the tensor structures of the descriptors and applying low-rank approximations we can derive new tensor-reduced descriptors which are systematically convergeable to the original full descriptor limit. We also verify this with numerical experiments on real data. We also show that there is a natural generalization to compress not only the chemical element information but also the radial degrees of freedom, yielding an even more compact representation. When fitting interatomic potentials for organic molecules and high-entropy alloys, we achieve a tenfold reduction in the number of features required when using linear (ACE) and nonlinear kernel models [Gaussian approximation potential (GAP) using SOAP]. We also fit a linear model to a dataset with 37 chemical elements which would be infeasible without the tensor-reduced features.

All many-body density-based descriptors can be understood in terms of ACE [10]. In ACE, the first step in describing the local neighborhood $\mathcal{N}(i) = \{j: r_{ij} < r_{\text{cut}}\}$ around atom i is forming the one-particle basis $\phi_{znlm}(\mathbf{r}_{ij}, Z_j)$ as a product of radial basis functions R_n , spherical harmonics Y_l^m , and an additional element index shown in Eq. (1), where \mathbf{r}_{ij} and Z_j denote the relative position and atomic number of neighbor j . Permutation invariance is introduced by summing over neighbor atoms in Eq. (2) after which $(\nu + 1)$ -body features are formed in Eq. (3) by taking tensor products of the atomic basis $A_{i,znlm}$ with itself ν times. Finally, Eq. (4) shows how the product basis $A_{i,znlm}$ is rotationally symmetrized using the generalized Clebsch-Gordon coefficients $C_m^{l\eta}$, where η enumerates all possible symmetric couplings for a given \mathbf{l} resulting in invariant features [10,11,18].

$$\phi_{znlm}(\mathbf{r}_{ij}, Z_j) = R_n(r_{ij})Y_l^m(\hat{\mathbf{r}}_{ij})\delta_{zZ_j}, \quad (1)$$

$$A_{i,znlm} = \sum_{j \in \mathcal{N}(i)} \phi_{znlm}(\mathbf{r}_{ij}, Z_j), \quad (2)$$

$$A_{i,znlm} = \prod_{t=1}^{\nu} A_{i,z,n,l,m_t} \quad (3)$$

$$B_{i,znl\eta} = \sum_m C_m^{l\eta} A_{i,znlm}. \quad (4)$$

A linear ACE model can then be fit to an invariant atomic property φ_i as

$$\varphi_i = \sum_{znl\eta} c_{znl\eta} B_{i,znl\eta} \quad (5)$$

where $c_{znl\eta}$ are the model parameters and for practical reasons the expansion is truncated using ν_{max} , l_{max} , and $n_{\text{max}} := N$, the number of radial basis functions. Note that as $B_{i,znl\eta}$ is invariant under $(z_a, n_a, l_a) \leftrightarrow (z_b, n_b, l_b)$ symmetrically equivalent terms are usually omitted from Eq. (5); again see Refs. [10,11,18] for the details. For some \mathbf{l} there is no invariant component of $A_{i,znlm}$ so that $C_m^{l\eta} = 0$ and the corresponding trivial $B_{i,znl\eta}$ are removed from Eq. (5).

Crucially, the tensor product in Eq. (3) causes the number of features (and therefore the number of model parameters) to grow rapidly as $\mathcal{O}(N^\nu S^\nu)$. Previous work [10,14,19] has reduced this to $\mathcal{O}(K^\nu)$ by first embedding the chemical and radial information into K channels [Eq. (6)] and then taking a full tensor product across the $\bar{A}_{i,klm}$ [Eq. (7)]:

$$\bar{A}_{i,klm} = \sum_{zn} W_{zn}^k A_{i,znlm}, \quad k = 1 \dots K \quad (6)$$

$$\bar{A}_{i,klm} = \prod_{t=1}^{\nu} \bar{A}_{i,k,l,m_t}. \quad (7)$$

This approach is also used in moment tensor potentials [15,26] and in Gaussian moment descriptors [27]. The embedding can be identified in Eq. (3) of Ref. [26], where μ indexes the embedded channels and ν is similar to l in ACE. Then taking tensor products across the embedded channels corresponds to forming products of the moments $M_{\mu\nu}$. In general, the embedding weights are optimized either before or during fitting [10,14,26] with the latter causing the models to be nonlinear.

We propose a principled approach to further reduce the size of the basis to $\mathcal{O}(K)$ which can be understood from two different angles. First, we identify the model parameters $\mathbf{c}_\eta \equiv c_{znl\eta}$ in Eq. (5) as a symmetric tensor, invariant under $(z_a, n_a, l_a) \leftrightarrow (z_b, n_b, l_b)$, which can be expanded as a sum of products of rank-1 tensors [28] as

$$\mathbf{c}_\eta = \sum_{k=1}^K \lambda_{k\eta} \underbrace{\mathbf{w}_k \otimes \mathbf{w}_k \dots \otimes \mathbf{w}_k}_{\nu \text{ times}}, \quad (8)$$

or in component form

$$c_{znl\eta} = \sum_k \lambda_{k\eta} \prod_{t=1}^{\nu} W_{z,n,l_t}^k, \quad (9)$$

where W_{z,n,l_t}^k are the components of \mathbf{w}_k . This expansion is exact for finite K , as \mathbf{c} is finite due to basis truncation and is equivalent to eigenvalue decomposition of a symmetric

TABLE I. The density projection \vec{A}_i is viewed as a vector with a composite index (z, n, l, m) whereas the embedded density projections $\vec{A}_i = (\mathbf{W}\vec{A}_i)$, etc. are indexed by (k, l, m) . The most general tensor-reduced many-body density projection (“Tensor sketch”) and its special case (“Tensor decomposition”) are shown together with their scaling with the number of radial basis functions N , number of chemical elements S , and number of embedding channels K . The symbol \otimes_k^{lm} means full tensor product across l and m but elementwise product across k whereas \otimes indicates a full tensor product across all indices.

Name	Product basis	Index notation	Basis size
ACE [10]	$A_i = \vec{A}_i \otimes \vec{A}_i \dots \otimes \vec{A}_i$	$A_{i,znlm} = \prod_{t=1}^{\nu} A_{i,z_n,l_t,m_t}$	$\mathcal{O}(NS)^{\nu}$
Embedding [10,14,15,19,31]	$\vec{A}_i = (\mathbf{W}\vec{A}_i) \otimes (\mathbf{W}\vec{A}_i) \dots \otimes (\mathbf{W}\vec{A}_i)$	$\vec{A}_{i,klm} = \prod_{t=1}^{\nu} \vec{A}_{i,k,l_t,m_t}$	$\mathcal{O}(K^{\nu})$
Tensor decomposition [25]	$\vec{A}_i = (\mathbf{W}\vec{A}_i) \otimes_k^{lm} (\mathbf{W}\vec{A}_i) \dots \otimes_k^{lm} (\mathbf{W}\vec{A}_i)$	$\vec{A}_{i,klm} = \prod_{t=1}^{\nu} \vec{A}_{i,kl_t,m_t}$	$\mathcal{O}(K)$
Tensor sketch	$\hat{A}_i = (\mathbf{W}^1\vec{A}_i) \otimes_k^{lm} (\mathbf{W}^2\vec{A}_i) \dots \otimes_k^{lm} (\mathbf{W}^{\nu}\vec{A}_i)$	$\hat{A}_{i,klm} = \prod_{t=1}^{\nu} \hat{A}_{i_t,kl_t,m_t}$ $\hat{A}_{i_t,kl_t,m_t} = \sum_{z_n} W_{z_n}^{tk} A_{i_t,z_n,l_t,m_t}$	$\mathcal{O}(K)$

matrix when $\nu = 2$. Note that we choose to use the same weights W_{z_n,l_t}^k for all ν and η , which significantly reduces the number of weights that need be specified. In practice, we can choose to expand over the zn or z indices only (see the Supplemental Material [29] for details) and then substitute the expansion into Eq. (5) as

$$\varphi_i \approx \sum_{kl\eta} \lambda_{kl\eta} \left[\sum_m C_m^{l\eta} \sum_{zn} \prod_{t=1}^{\nu} W_{z_n,l_t}^k A_{i,z_n,l_t,m_t} \right] \quad (10)$$

$$= \sum_{kl\eta} \lambda_{kl\eta} \left[\sum_m C_m^{l\eta} \prod_{t=1}^{\nu} \vec{A}_{i,kl_t,m_t} \right] \quad (11)$$

$$= \sum_{kl\eta} \lambda_{kl\eta} \tilde{\mathbf{B}}_{i,kl\eta}, \quad (12)$$

where $\tilde{\mathbf{B}}_{i,kl\eta}$ are the new tensor-reduced features and the approximation arises because in practice we truncate the tensor decomposition early. The key novelty is that only elementwise products are taken across the k index of the embedded channels \vec{A}_{i,kl_t,m_t} when forming the many-body basis, rather than a full tensor product, i.e., k does not have a t subscript in Eq. (10) (see Table I for the full definitions). For completeness, we note that applying this tensor reduction to the elements only and using $K = 2$ is equivalent to the element-weighting strategies used in Refs. [12,13,30].

There are multiple natural strategies for specifying the embedding weights $W_{z_n}^{kl}$, including approximating a pre-computed $c_{znl\eta}$ or treating the weights as model parameters to be estimated during the training process, as is done in MACE [25]. Here we investigate using random weights as a simpler alternative. This ensures that Eq. (12) remains a linear model and allows the $\tilde{\mathbf{B}}_{i,kl\eta}$ to be used directly in other tasks such as data visualization.

We now show that the resulting tensor-reduced features can also be understood from the perspective of directly compressing the original $\mathbf{B}_{i,znl\eta}$ features. Random Projection (RP) [32,33] is an established technique where general high dimensional feature vectors $\{\vec{x}_1, \dots, \vec{x}_N\} \subset \mathbb{R}^d$ are compressed as $\tilde{x}_i = \mathbf{W}\vec{x}_i \in \mathbb{R}^K$, with the entries of the matrix \mathbf{W} being normally distributed. This approach is simple, offers a tuneable level of compression, and is underpinned by the Johnson-Lindenstrauss Lemma [34] which bounds the fractional error made in approximating $\vec{x}_i^T \vec{x}_j$ by $\tilde{x}_i^T \tilde{x}_j$. RP can also be used to reduce the cost of linear models, with a closely related approach recently used in Ref. [35]. In compressed least-square regression (CLSR) [36–38], features are replaced by their projections, thus reducing the number of model parameters. Loosely speaking, the approximation errors incurred in CLSR (and RP in general) are expected to decay as $1/\sqrt{K}$, and we refer to Refs. [36,38,39] for more details. The drawback of RP is that it requires the full feature vector to be constructed so that applying RP to ACE would not avoid the unfavorable $\mathcal{O}(N^{\nu}S^{\nu})$ scaling. We propose using tensor sketching [40–43], as used in Refs. [44,45] to approximate the neural tangent kernel, instead of RP. For vectors with tensor structure $\vec{x} = \vec{y} \otimes \vec{z}$ where $\vec{x} \in \mathbb{R}^{d_1 d_2}$, $\vec{y} \in \mathbb{R}^{d_1}$, and $\vec{z} \in \mathbb{R}^{d_2}$, the random projection $\mathbf{W}\vec{x}$ can be efficiently computed directly from \vec{y} and \vec{z} as

$$\mathbf{W}\vec{x} = (\mathbf{W}'\vec{y}) \odot (\mathbf{W}''\vec{z}), \quad (13)$$

where \odot denotes the elementwise (Hadamard) product. Similarly, the ACE product basis can be tensor sketched, across the zn indices, as

$$\hat{A}_i = (\mathbf{W}^1\vec{A}_i) \otimes_k^{lm} (\mathbf{W}^2\vec{A}_i) \dots \otimes_k^{lm} (\mathbf{W}^{\nu}\vec{A}_i), \quad (14)$$

where \otimes_k^{lm} denotes taking the tensor product over the upper indices lm and the elementwise product over the lower

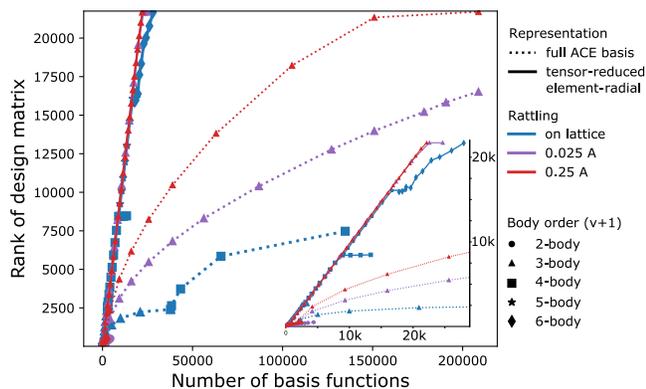


FIG. 1. The row rank of the design matrix as a function of basis set on a dataset of all symmetry inequivalent fcc lattices of five chemical elements and unit cell sizes of up to six atoms. The inset enlarges the $x = y$ region. The colors denote the size of the rattling around the crystal sites of an ideal fcc lattice, the markers correspond to increasing body order of the basis, and the line style compares the fully coupled and tensor-reduced ACE versions.

index k and \mathbf{W}^1 , \mathbf{W}^2 , etc. are independent identically distributed random matrices; see the Supplemental Material [29] for details. The $\hat{A}_{i,klm}$ can then be symmetrized as in Eq. (4), yielding

$$\hat{B}_{i,kl\eta} = \sum_m c_m^{\eta} \prod_{t=1}^{\nu} \hat{A}_{it,kl,m_t}, \quad (15)$$

where \hat{A}_{it,kl,m_t} is defined more precisely in Table I. Finally, we note that because the embedded channels are independent, the error in approximating inner products using the average across K channels is expected to decrease as $1/\sqrt{K}$, just as with standard RP. Based on this, we conjecture that similar bounds derived for the errors made in CLSR may also apply here. A summary comparing standard ACE, element-embedding, and tensor-reduced ACE is given in Table I, where it is clear that the features derived using tensor decomposition are equivalent to the tensor-sketched features with the choice of using equal weights in each factor.

We now turn to numerical results and first demonstrate that the tensor-reduced features are able to efficiently and completely describe a many-element training set. We consider a dataset comprised of all symmetry inequivalent fcc structures made up of five elements with up to six atoms per unit cell [46]. A set of features is complete on this dataset if the design matrix for a linear model fit to total energies has full (numerical) row rank, where each row corresponds to a different training configuration.

Figure 1 shows the numerical rank of the design matrix as a function of the basis set size. At a given correlation order the standard ACE basis set is grown by increasing the polynomial degree, and the tensor-reduced basis set is enlarged by increasing K , the number of independent

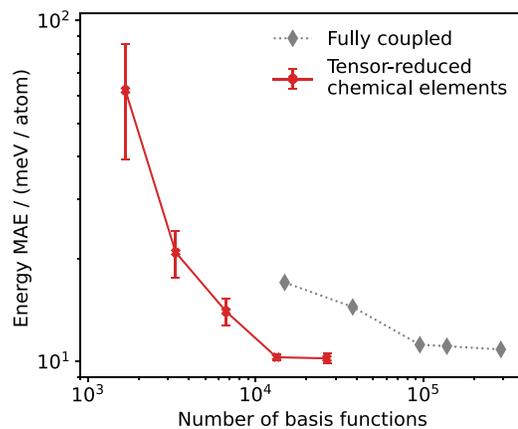


FIG. 2. Convergence of the energy mean absolute error (MAE) of organic molecules on the independent test set with respect to the number of basis functions is shown for linear models using standard and tensor-reduced (tensor decomposition) ACE features. Error bars show the standard error in the mean, computed across five fits using independently chosen random weights.

channels. In both cases once the rank stops increasing at the given correlation order we increment ν . The colors in Fig. 1 correspond to three different geometrical variations: blue contains on-lattice configurations only whilst in magenta and red the atomic positions have been perturbed by a random Gaussian displacement with mean 0 and standard deviation of 0.025 and 0.25 Å, respectively. The dotted lines corresponds to the standard ACE basis, whereas the solid lines corresponds to the tensor-reduced version from Eq. (12). Although the standard ACE basis can always achieve full row rank since it is a complete linear basis, it does this very inefficiently. In contrast, the row rank using the tensor-reduced basis grows almost linearly. Thus the tensor-reduced basis, having removed unnecessary redundancies, still retains the expressive power of the full basis.

Next, we fit a linear ACE [47] model on a training set of 400 different organic molecules of size 19–168 atoms, randomly selected from the QMugs dataset [48] that are made up of ten different chemical elements (H, C, N, O, F, P, S, Cl, Br, I). The conformers were created by running GFN2-xTB [49] 800 K NVT molecular dynamics for 1 ps starting from a published minimum energy structure. The test set is composed of 1000 *different* molecules sampled the same way. This is a small-data regime task ($\sim 60\,000$ scalar targets) that is particularly challenging due to the chemical and conformational diversity. Figure 2 shows the convergence of the energy error with the number of basis functions for the fully coupled and the tensor-reduced ACE models, both using $\nu_{\max} = 3$ (4-body) basis functions. By increasing the number of uncoupled channels K we can converge the accuracy to the previous level, whilst reducing the size of the model by a factor of 10.

The tensor-reduction techniques proposed in this Letter can be directly applied to other density-based descriptors

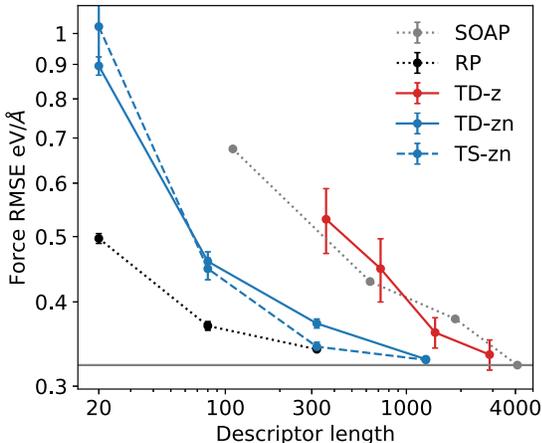


FIG. 3. Convergence of force root-mean-square error (RMSE) on the high-entropy alloy test set as a function of descriptor length for various GAP models. Error bars show the standard error in the mean, computed across fit fits using independently chosen random weights. The tensor reduction is denoted using TD for tensor-decomposition (12) and TS for tensor-sketching (15), and the reduction is performed over the indices after the hyphen.

and used with alternative models. To demonstrate this we used the tensor-reduced SOAP power spectrum to fit GAP [8] models to the quinary high-entropy alloy dataset ($\sim 400\,000$ scalar targets) from Ref. [50]. The descriptors were computed using turbosoap [51], and fitting was performed using the GAP code [52]. To provide a baseline for comparison, reference models were fit using the full power spectrum evaluated with a varying number of basis functions; $n_{\max} = 2l_{\max}$. All compressed descriptors were constructed using the largest values of $n_{\max} = 8$, $l_{\max} = 4$; see the Supplemental Material [29] for details. The force errors achieved on the independent test set are shown as a function of descriptor length in Fig. 3. For this dataset, mixing only the chemical elements did not help, which is likely due to the minimal savings made as a result of the low correlation order ν . However, using the fully element-radial tensor-reduced descriptor allowed for similar accuracy to be achieved using approximately ten times fewer features. Furthermore, the errors achieved with all compressed descriptors are shown to converge to the error achieved with the full SOAP vector they were derived from. Finally, in the high accuracy regime, descriptor length > 300 , models using the tensor-sketch element-radial reduction matched the accuracy achieved using full RP.

The information content of these SOAP based descriptors was also measured using the information imbalance [53]. It was found that the fully element-radial tensor-reduced descriptors offered comparable data efficiency to a random projection and that taking an elementwise product across embedded channels outperformed taking a full tensor product; see the Supplemental Material [29] for tests on various datasets [54–56].

TABLE II. Mean absolute errors on the HME21 dataset. Energy (E , meV) and force (F , eV/Å) errors of models. The labels, (1) or (2), of the MACE models indicate the number of message passing layers. MACE errors are an average across three seeds.

	MACE (1)	MACE (1)	MACE (1)	MACE (2)	TeaNet [59]	NequIP [23]
U_{zn}	opt	rand	rand	opt		
$V_n^{kl_i}$	opt	opt	rand	opt		
E	53.8	54.7	254.8	16.5	19.6	47.8
F	0.194	0.193	0.451	140.2	0.174	0.199

Finally, we assess the effect of optimizing the embedding weights using the MACE architecture [25]. A typical MACE model is a two-layer message passing network which utilizes tensor-reduced ACE features to efficiently represent equivariant body-ordered messages. Normally, the embedding weights are optimized using backpropagation along with all other model parameters, but it is possible to fix these weights to random values. As such, whilst a multilayer MACE is a multi-center polynomial model, a single-layer MACE with frozen embedding weights is equivalent to a linear model using tensor-reduced ACE features [24]. Furthermore, within MACE the embedding weights used in the tensor decomposition (see Table I) are approximated as $W_{zn}^{kl_i} = U_{zn} V_n^{kl_i}$ so that the element weights U_{zn} and the radial weights $V_n^{kl_i}$ can be optimized or frozen independently.

We use the HME21 dataset [57,58] ($\sim 1700\,000$ scalar targets) for this test due to its exceptional diversity, both chemically with 37 elements, and structurally with configurations including isolated molecules, bulk crystals, surfaces, clusters, and disordered materials. We fit several models (see Supplemental Material [29] for details), and the energy and force errors on the independent test set are summarized in Table II. Strikingly, using random element-embedding weights leads to almost no degradation in model accuracy compared with using optimized element-embedding weights. In contrast, randomizing both the element and radial embedding weights leads to significantly larger prediction errors. A two-layer model with optimized weights achieves state of the art accuracy on HME21, showing the power of the message passing architecture and further highlighting the effectiveness of the tensor-reduced features.

In conclusion, we introduced a tensor-reduced form of the ACE basis for modeling symmetric functions of local atomic neighbor environments that eliminates the $\mathcal{O}(N^\nu S^\nu)$ scaling of the basis set size with the number of chemical elements and radial basis functions. Intuitively, the construction can be thought of as mixing the element and radial channels and then only coupling these channels to themselves when constructing the higher order many-body basis. We derived this new embedded basis from a symmetric tensor decomposition and explored its connection to tensor sketching. We showed that this reduced basis

is also systematic, and that in practice it can enable a tenfold reduction in basis set size for diverse datasets with many elements, including organic molecules and high entropy alloys. When used in a two-layer message passing network, MACE [25], the tensor decomposition yields state of the art performance.

D. P. K. acknowledges support from AstraZeneca and the EPSRC. G. C. acknowledges discussion with Boris Kozinsky. J. P. D. and G. C. acknowledge support from the NOMAD Centre of Excellence, funded by the European Commission under Grant Agreement No. 951786. We used computational resources of the UK HPC service ARCHER2 via the UKCP consortium and funded by EPSRC Grant No. EP/P022596/1. C. O. acknowledges support of the Natural Sciences and Engineering Research Council [Discovery Grant No. IDGR019381] and the New Frontiers in Research Fund [Exploration Grant No. GR022937].

*These authors contributed equally to this work.

- [1] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, *Nature (London)* **589**, 59 (2021).
- [2] T. A. Young, T. Johnston-Wood, H. Zhang, and F. Duarte, *Phys. Chem. Chem. Phys.* **24**, 20820 (2022).
- [3] V. Kapil, C. Schran, A. Zen, J. Chen, C. J. Pickard, and A. Michaelides, *Nature (London)* **609**, 512 (2022).
- [4] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, *Chem. Rev.* **121**, 9759 (2021).
- [5] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car *et al.*, End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems, in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Zhang, 2018), pp. 4436–4446.
- [6] J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- [7] A. P. Bartók, R. Kondor, and G. Csányi, *Phys. Rev. B* **87**, 184115 (2013).
- [8] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.* **104**, 136403 (2010).
- [9] A. Thompson, L. Swiler, C. Trott, S. Foiles, and G. Tucker, *J. Comput. Phys.* **285**, 316 (2015).
- [10] R. Drautz, *Phys. Rev. B* **99**, 014104 (2019).
- [11] G. Dusson, M. Bachmayr, G. Csányi, R. Drautz, S. Etter, C. van der Oord, and C. Ortner, *J. Comput. Phys.* **454**, 110946 (2022).
- [12] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, and P. Marquetand, *J. Chem. Phys.* **148**, 241709 (2018).
- [13] N. Artrith, A. Urban, and G. Ceder, *Phys. Rev. B* **96**, 014112 (2017).
- [14] M. J. Willatt, F. Musil, and M. Ceriotti, *Phys. Chem. Chem. Phys.* **20**, 29661 (2018).
- [15] K. Gubaev, E. V. Podryabinkin, G. L. Hart, and A. V. Shapeev, *Comput. Mater. Sci.* **156**, 148 (2019).
- [16] J. P. Darby, J. R. Kermode, and G. Csányi, *npj Comput. Mater.* **8**, 166 (2022).
- [17] T. Kostiuhenko, F. Körmann, J. Neugebauer, and A. Shapeev, *npj Comput. Mater.* **5**, 55 (2019).
- [18] J. Nigam, S. Pozdnyakov, and M. Ceriotti, *J. Chem. Phys.* **153**, 121101 (2020).
- [19] A. Gosciniski, F. Musil, S. Pozdnyakov, J. Nigam, and M. Ceriotti, *J. Chem. Phys.* **155**, 104106 (2021).
- [20] C. Zeni, K. Rossi, A. Glielmo, and S. De Gironcoli, *J. Chem. Phys.* **154**, 224112 (2021).
- [21] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh (PMLR, Gilmer, 2017), pp. 1263–1272.
- [22] K. Schütt, P.-J. Kindermans, H. E. Saucedo Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, Schnet: A continuous-filter convolutional neural network for modeling quantum interactions, in *Advances in Neural Information Processing Systems*, edited by Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Long Beach, 2017).
- [23] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, *Nat. Commun.* **13**, 2453 (2022).
- [24] I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. Simm, R. Drautz, C. Ortner, B. Kozinsky, and G. Csányi, [arXiv:2205.06643](https://arxiv.org/abs/2205.06643).
- [25] I. Batatia, D. P. Kovács, G. N. Simm, C. Ortner, and G. Csányi, [arXiv:2206.07697](https://arxiv.org/abs/2206.07697).
- [26] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, *Mach. Learn.* **2**, 025002 (2020).
- [27] V. Zaverkin and J. Kästner, *J. Chem. Theory Comput.* **16**, 5410 (2020).
- [28] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, *SIAM J. Matrix Anal. Appl.* **30**, 1254 (2008).
- [29] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.131.028001> for derivations of the tensor-reduced basis and model parameters.
- [30] M. Uhrin, *Phys. Rev. B* **104**, 144110 (2021).
- [31] A. Bochkarev, Y. Lysogorskiy, S. Menon, M. Qamar, M. Mrovec, and R. Drautz, *Phys. Rev. Mater.* **6**, 013804 (2022).
- [32] S. Dasgupta, [arXiv:1301.3849](https://arxiv.org/abs/1301.3849).
- [33] E. Bingham and H. Mannila, Random projection in dimensionality reduction: Applications to image and text data, in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, San Francisco, 2001), pp. 245–250.
- [34] W. B. Johnson and J. Lindenstrauss, *Contemp. Math.* **26**, 189 (1984).
- [35] N. J. Browning, F. A. Faber, and O. A. von Lilienfeld, *J. Chem. Phys.* **157**, 214801 (2022).
- [36] A. Kabán, A new look at compressed ordinary least squares, in *IEEE 13th International Conference on Data Mining Workshops*, edited by H. Xiong, G. Karypis, B. M. Thuraisingham, D. J. Cook, and X. Wu (IEEE, Dallas, 2013), pp. 482–488.
- [37] S. E. Ahmed, *Big and Complex Data Analysis: Methodologies and Applications* (Springer, New York, 2017).

- [38] O. Maillard and R. Munos, Compressed least-squares regression, in *Advances in Neural Information Processing Systems*, edited by Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Curran Associates, Inc., Maillard, 2009).
- [39] D. Hsu, S.M. Kakade, and T. Zhang, Random design analysis of ridge regression, in *Proceedings of the 25th Annual Conference on Learning Theory, Proceedings of Machine Learning Research*, edited by S. Mannor, N. Srebro, and R.C. Williamson (PMLR, Hsu, Edinburgh, Scotland, 2012), pp. 9.1–9.24.
- [40] Y. Liu, J. Liu, Z. Long, and C. Zhu, *Tensor Computation for Data Analysis* (Springer, New York, 2022).
- [41] T.D. Ahle, M. Kapralov, J.B. Knudsen, R. Pagh, A. Velingker, D.P. Woodruff, and A. Zandieh, Oblivious sketching of high-degree polynomial kernels, in *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)* (SIAM, Ahle, 2020), pp. 141–160.
- [42] N. Pham and R. Pagh, Fast and scalable polynomial kernels via explicit feature maps, in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, Pham, 2013), pp. 239–247.
- [43] D. P. Woodruff *et al.*, *Found. Trends Theor. Comput. Sci.* **10**, 1 (2014).
- [44] D. Holzmüller, V. Zaverkin, J. Kästner, and I. Steinwart, [arXiv:2203.09410](https://arxiv.org/abs/2203.09410).
- [45] V. Zaverkin, D. Holzmüller, I. Steinwart, and J. Kästner, *Digital Discovery* **1**, 605 (2022).
- [46] G. L. W. Hart and R. W. Forcade, *Phys. Rev. B* **77**, 224115 (2008).
- [47] D. P. Kovács, C. v. d. Oord, J. Kucera, A. E. A. Allen, D. J. Cole, C. Ortner, and G. Csányi, *J. Chem. Theory Comput.* **17**, 7696 (2021).
- [48] C. Isert, K. Atz, J. Jiménez-Luna, and G. Schneider, *Sci. Data* **9**, 1 (2022).
- [49] C. Bannwarth, S. Ehlert, and S. Grimme, *J. Chem. Theory Comput.* **15**, 1652 (2019).
- [50] J. Byggmästar, K. Nordlund, and F. Djurabekova, *Phys. Rev. B* **104**, 104101 (2021).
- [51] M. A. Caro, *Phys. Rev. B* **100**, 024112 (2019).
- [52] <https://github.com/libAtoms/GAP>.
- [53] A. Glielmo, C. Zeni, B. Cheng, G. Csányi, and A. Laio, *PNAS Nexus* **1**, pgac039 (2022).
- [54] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, *Chem. Rev.* **121**, 10073 (2021).
- [55] A. S. Christensen and O. A. Von Lilienfeld, *Mach. Learn.* **1**, 045018 (2020).
- [56] V. L. Deringer, M. A. Caro, and G. Csányi, *Nat. Commun.* **11**, 5461 (2020).
- [57] S. Takamoto, [10.6084/m9.figshare.19658538](https://doi.org/10.6084/m9.figshare.19658538), High-temperature multi-element 2021 (HME21) dataset.
- [58] S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano *et al.*, *Nat. Commun.* **13**, 1 (2022).
- [59] S. Takamoto, S. Izumi, and J. Li, *Comput. Mater. Sci.* **207**, 111280 (2022).