Physically Motivated Recursively Embedded Atom Neural Networks: Incorporating Local Completeness and Nonlocality

Yaolong Zhang, Junfan Xia, and Bin Jiang*

Hefei National Laboratory for Physical Science at the Microscale, Key Laboratory of Surface and Interface Chemistry and Energy Catalysis of Anhui Higher Education Institutes, Department of Chemical Physics, University of Science and Technology of China, Hefei, Anhui 230026, China

(Received 21 June 2021; accepted 7 September 2021; published 8 October 2021)

Recent advances in machine-learned interatomic potentials largely benefit from the atomistic representation and locally invariant many-body descriptors. It was, however, recently argued that including three-body (or even four-body) features is incomplete to distinguish specific local structures. Utilizing an embedded density descriptor made by linear combinations of neighboring atomic orbitals and realizing that each orbital coefficient physically depends on its own local environment, we propose a recursively embedded atom neural network model. We formally prove that this model can efficiently incorporate complete many-body correlations without explicitly computing high-order terms. This model not only successfully addresses challenges regarding local completeness and nonlocality in representative systems, but also provides an easy and general way to update local many-body descriptors to have a message-passing form without changing their basic structures.

DOI: 10.1103/PhysRevLett.127.156002

Over the past years, machine learning has achieved enormous success in many scientific fields, especially in the development of more accurate interatomic potentials based on *ab initio* data for chemical systems [1], including molecules and reactions [2–7], excited states [8–11], condensed phase materials [12–16], etc. Besides using different machine learning algorithms, these MLIPs mainly differ in their structural descriptors (or features) which should distinguish diverse molecular configurations and be invariant with respect to translation, rotation, and permutation of identical atoms. In small molecular and reactive systems, it is well known that a global descriptor like permutationally invariant polynomials in terms of interatomic distances [5] of a sufficiently high order, or equivalently fundamental invariants [2], well satisfy both invariance and distinguishability requirements [17]. However, the size of polynomials scales factorially with the number of permutations, preventing their applications in large systems.

On the other hand, most MLIPs for large molecules and materials rely on an atomic decomposition of total energy, namely $E = \sum_{i=1}^{N} E_i$, as first proposed by Behler and Parrinello in their high-dimensional neural network (BPNN) approach [12]. In this representation, each atomic energy is dependent on the corresponding local environment (within a certain cutoff radius) described by a set of locally invariant many-body features between the central and neighboring atoms [18–27]. Because of the high costs of evaluating higher-order terms, these features are typically truncated up to three- or four-body correlations. However, it was recently shown that some local atomic structures in a system as small as CH_4 become indistinguishable by the third-order (or even fourth-order) correlations [28]. This would introduce a distortion of the feature space and intrinsically limit the representability of the MLIP [28]. While some approaches [24–26,29] could in principle resolve this atomic structural degeneracy by systematically including higher-order terms, the computational cost would, however, increase dramatically.

An alternative way to describe an atom-centered environment is to repeatedly convolute feature vectors between every atom and its neighbors by neural networks (NNs) [3], allowing the information progressively passed among the central atom, the neighbors, the neighbors' neighbors, and so on so forth. Such so-called message-passing neural networks (MPNNs) [3,30–32] can learn an increasingly more sophisticated feature-property correlation from the training data. However, it is less clear that how this type of models incorporate many-body correlations by iteratively integrating (mostly) two-body terms [30,31] (and angular terms [33,34]) and whether they can resolve the local structural degeneracy issues discussed in Ref. [28].

In this Letter, to address this challenge, we propose a physically inspired recursive neural network model that naturally integrates the message-passing concept into a well-defined three-body descriptor. We derive that this model can formulate a complete atomic representation of the local environment without explicitly computing high-order correlations and incorporate some nonlocal interactions beyond the cutoff radius, both validated by numerical tests. Like in conventional MPNN models, however, the nonlocal charge transfer [35] and conjugated effects [1] are not yet included and will not be discussed here.

Let us start with the embedded atom neural network (EANN) model which adopts the atomistic representation of total energy and encodes the information of local environment by the symmetry-invariant embedded density descriptor [20] inspired by the embedded atom method [36]. For simplicity, an embedded density invariant (ρ_i) at the position of atom i is given by the square of the linear combination of atomic orbitals of its neighbors,

$$\rho_i = \sum_{l_x,l_y,l_z}^{l_x+l_y+l_z=L} \frac{L!}{l_x!l_y!l_z!}\left[\sum_{j\neq i}^{N_c} c_j \varphi(\hat{\mathbf{r}}_{ij}) f_c(r_{ij})\right]^2, \quad (1)$$

where $\hat{\mathbf{r}}_{ij}=\hat{\mathbf{r}}_i-\hat{\mathbf{r}}_j$, with $\hat{\mathbf{r}}_i=(x_i,y_i,z_i)$ and $\hat{\mathbf{r}}_j=(x_j,y_j,z_j)$ being the Cartesian coordinate vectors of the central atom i and a neighbor atom j, $r_{ij}=|\hat{\mathbf{r}}_{ij}|$ is the distance between them, $\varphi(\hat{\mathbf{r}}_{ij})$ is the Gaussian-type orbital centered at atom j parametrized by its center (r_s), width (α), and angular momenta ($L=l_x+l_y+l_z$),

$$\varphi(\hat{\mathbf{r}}_{ij}) = (x_i-x_j)^{l_x}(y_i-y_j)^{l_y}(z_i-z_j)^{l_z}\exp[-\alpha(r_{ij}-r_s)^2], \quad (2)$$

$f_c(r_{ij})$ is a cutoff function continuously damping the invariant to zero at the cutoff radius (r_c), and N_c is the number of atoms within r_c. Clearly, ρ_i corresponds to the embedded density contribution from a given type of atomic orbital and expresses two-body ($L=0$) and three-body ($L>0$) interactions in a uniform way. This can be seen by explicitly rewriting Eq. (1) in terms of interatomic distances and angles according to the multinomial theorem [20,37],

$$\rho_i = \sum_{j,k\neq i} c_j \exp[-\alpha(r_{ij}-r_s)^2] f_c(r_{ij}) \times c_k \exp[-\alpha(r_{ik}-r_s)^2] \times f_c(r_{ik}) r_{ij}^L r_{ik}^L (\cos\theta_{ijk})^L. \quad (3)$$

Indeed, Eq. (1) allows the evaluation of atom-centered three-body terms at a cost of atom-centered two-body ones, resulting in a linear scaling with respect to the number of neighbors. As a result, this EANN model is more efficient than many other descriptor-based MILPs [38], and accurate in predicting energies [20] and tensorial properties [39,40].

In the CH_4 example, the C-centered embedded density invariants and corresponding atomic energies are essentially identical, when two C-centered structures of CH_4 have the same list of distances and angles, as displayed in Fig. 1. This problem intrinsically exists in other three-body (or lower-order) atomic descriptors [18,19,22,23]. It can be seen from Eq. (3) as orbital coefficients are fixed after training (like NNs' parameters) so that ρ_i are determined by these distances and angles only. However, considering the linear combination of atomic orbitals in Eq. (1) as an analog of a molecular orbital, it is a matter of fact in quantum

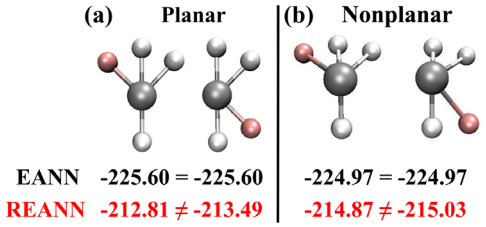

FIG. 1. Two representative pairs of CH_4 molecules that have the same set of distances and angles between central C atom (silver) and neighboring H atoms (white for identical ones and light red for the different one), for which the EANN (REANN) atomic energies (in electron volts) for C are identical (distinct).

chemistry that c_j should virtually vary with the molecular configuration. One simplest way to cast this physical concept into the descriptor is to make c_j itself a function of the jth atom's neighbor environment behaving like the atomic energy. In this scenario, orbital coefficients of the four H atoms in the two CH_4 molecules can be different since their respective H-centered environments are different. This leads to nonequivalent C-centered embedded density invariants and atomic energies for the two indistinguishable atomic structures by three-body correlations in Fig. 1. Importantly, atomic orbitals in the vicinity of atom j have been calculated for obtaining the atomic energy (E_j), thus need not be recalculated to evaluate the environment-dependent c_j.

Apparently, the orbital coefficient can be recursively embedded in this way whenever necessary and a generalized expression is

$$c_j^t = g_j^{t-1}[\boldsymbol{\rho}_j^{t-1}(\mathbf{c}_j^{t-1},\mathbf{r}_j^{t-1})], \quad (4)$$

where $\mathbf{c}_j^{t-1}$ and $\mathbf{r}_j^{t-1}$ are the collections of orbital coefficients and atomic positions in the neighborhood of the central atom j in the $(t-1)$th iteration, $\boldsymbol{\rho}_j^{t-1}$ is the corresponding embedded density feature vector, g_j^{t-1} is an atomic NN mapping $\boldsymbol{\rho}_j^{t-1}$ to c_j^t, namely the orbital coefficient of atom j as a neighbor of other atoms in the tth iteration. This procedure is schematically displayed in Fig. 2(a). One may immediately realize that this recursively EANN (REANN) model has an effective message-passing form [1], except that here the orbital coefficients, rather than the whole feature vectors, iteratively pass the environmental information between an atom and its neighbors. This is an intriguing result that links up, perhaps for the first time, the local many-body descriptors and the less physically intuitive message-passing features.

Next, we turn to discuss how higher-order correlations are incorporated in this recursion, an issue rarely discussed in previous studies on MPNNs. Supposing that the iteration undergoes T times ($T>0$), it is convenient to use a simplified version of Eq. (3),

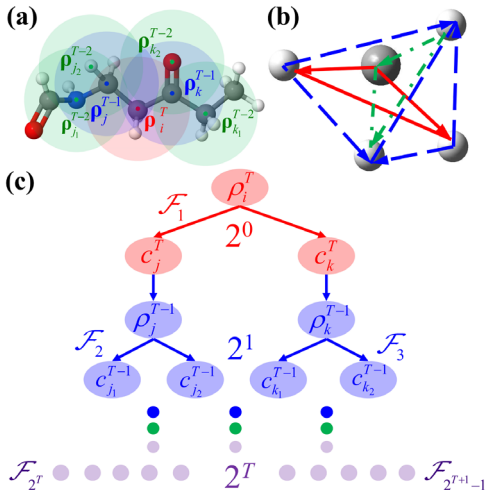

FIG. 2. (a) Schematic diagram of the REANN model showing how the density descriptor is recursively embedded. (b) An example of CH_4 showing how the C-centered complete five-body correlation is achieved by twice iteration, where the path going through all atoms corresponds to the product of F functions and arrows point from the central atom to neighbor atoms. (c) An illustration that how the number of three-body terms (F) increases in each iteration (2^T). Different colors correspond to different iteration times, namely $T = 0$ (red), $T = 1$ (blue), and $T = 2$ (green).

$$\rho_i^T = \sum_{j,k} c_j^T c_k^T F(r_{ij}, r_{ik}, r_{jk}), \tag{5}$$

where the orbital coefficients are now T dependent (c_j^T and c_k^T) and $F(r_{ij}, r_{ik}, r_{jk})$ represents a generalized three-body correlation term collecting these functions in Eq. (3). Substituting Eq. (4) into Eq. (5) and assuming no hidden layer in g_j^{T-1} (i.e., a linear function), we have

$$\rho_i^T = \sum_{j.k} F(r_{ij}, r_{ik}, r_{jk}) \sum_{n_1=1}^{N_\rho} w_{n_1} \rho_j^{T-1,n_1} \sum_{n_2=1}^{N_\rho} w_{n_2} \rho_k^{T-1,n_2}, \tag{6}$$

where w_{n_1} and w_{n_2} are linear weights of the corresponding features, N_ρ is the number of embedded density invariants. Note that using a nonlinear g_j^{T-1} here would not alter our conclusion but will complicate this equation. We then substitute Eq. (5) in the $(T-1)$th iteration back to Eq. (6) and reorder the summations,

$$\rho_i^T = \sum_{j,k} \sum_{n_1,n_2}^{N_\rho} w_{n_1} w_{n_2} \sum_{j_1,j_2} \sum_{k_1,k_2} c_{j_1}^{T-1,n_1} c_{j_2}^{T-1,n_1} c_{k_1}^{T-1,n_2} c_{k_2}^{T-1,n_2} \times F(r_{ij}, r_{ik}, r_{jk}) F(r_{jj_1}, r_{jj_2}, r_{j_1 j_2}) F(r_{kk_1}, r_{kk_2}, r_{k_1 k_2}), \tag{7}$$

As orbital coefficients are expanded, the number of three-body functions doubles in each iteration till the last environment-independent ones, as illustrated in Fig. 2(c). This will make ρ_i^T eventually the sum of products of $(2^{T+1}-1)$ three-body F functions after T iterations, which can be generalized as

$$\rho_i^T = \sum_m \eta_m \prod_{(i,j,k)}^{2^{T+1}-1} F(r_{ij}, r_{ik}, r_{jk}), \tag{8}$$

where m collects all indexes of the summation, η_m is the collection of all weights and orbital coefficients, and i, j, k span over all atomic indexes involved. According to above discussion, ρ_i^T will contain at least one highest-order correlation term involving $3(2^{T+1}-1)$ nonredundant interatomic distances in the neighborhood of atom i with sufficient neighbors, along with some lower-order terms due to repeated interatomic distances. Regarding atoms as nodes and interatomic distances as edges, the highest-order correlation term can be viewed as an analog of the Eulerian path in graph theory (a path in a finite graph passing every edge just once), except that in our case this path can pass the same edge more than once. Figure 2(b) illustrates such a path walking through all edges in CH_4 after two iterations. Examples for lower-order correlations are provided in Supplemental Material (SM) [41].

By definition, a complete many-body descriptor has to correlate all atoms in the system [46]. This implies that ρ_i^T will involve a complete correlation of an atom-centered environment, if $3(2^{T+1}-1) \geq N_c(N_c-1)/2$. The minimum number of iterations to warrant this is thus given by $T_{\min} = (\log_2\{[N_c(N_c-1)]/6+1\}) - 1$, where $()$ rounds up the value to its nearest integer. Recall that the cost of each iteration scales linearly with N and atomic orbitals need be calculated only once. This is a striking finding that the complete atomic representation can be achieved with $\sim O(\log_2 N_c)$ complexity, instead of the exponential scaling with the body order when explicitly computing high-order correlations [29]. Our approach will be increasingly more favorable as N_c increases.

Similarly, this analysis can also estimate the required number of interaction blocks (or the time of message passed) in other MPNN models, which was often empirically specified without a guidance. This number has to be greater than $N_c(N_c-1)/2$, theoretically, if only two-body features were recursively embedded (e.g., in SchNet [30]), because each iteration now introduces only one more interatomic distance toward the higher-body correlation. It is even worse that using radial functions alone actually does not warrant the local completeness, because atoms with distances greater than r_c cannot be correlated in any way. Examples are given in the SM. It is also found in other more recent MPNN models that including angular information in the feature update is beneficial [33,34], consistent with our derivation. Note that our practical implementation remains based on Eqs. (2) and (4) for numerical efficiency.

To validate our derivation numerically, we use the CH_4 dataset provided by Ceriotti and co-workers as a stringent test [28]. This dataset includes $\sim 7.7 \times 10^6$ configurations with randomly distributed atoms excluding structures with too close contacts. Because of the existence of near degenerate manifolds and many unphysical configurations with energies up to 70 eV, this dataset has been claimed to be the best touchstone of the representability and completeness of the descriptor. Since there are only five atoms in CH_4, we estimate that many-body correlations become complete at $T_{\min} = 2$. We have optimized r_s, α, and c_j together with all NN parameters, as readily implemented in PyTorch [42], yielding an end-to-end deep learning framework. To demonstrate the performance of the features themselves, we also train linear models by removing all hidden layers of NNs (for both orbital coefficients and atomic energies). Details of training are given in the SM.

Figure 3(a) compares the test root-mean-square errors (RMSEs) of various linear models as a function of the number of training configurations (n_{train}). The learning curve of $T = 0$ (including three-body correlations only) exhibits a clear saturation with respect to n_{train}, which is fully consistent with the result of Ref. [28] using three-body power spectrum features. Recursively expanding orbital coefficients steepens the learning curve and reduces the error significantly. The result with a single iteration ($T = 1$) obviously outperforms that from Ref. [28] obtained with the mix of three- and four-body ($3B + 4B$) correlations. With two iterations ($T = 2$), which are supposed to offer a complete correlation, we observe a saturated error of ~ 0.6 kcal/mol with 10^6 points. This is in good agreement with that of Nigam *et al.* [29] who used an iterative contraction algorithm to select up to five-body ($5B$) invariants (the highest-body correlation for CH_4). These results clearly indicate the local completeness of our recursively embedded density descriptor.

Incorporating the nonlinearity of NNs substantially increases the flexibilities of all models. As shown in Fig. 3(b), $3B + 4B$ correlations in Ref. [28] trained with 3×10^6 points led to an RMSE of ~ 0.5 kcal/mol. Impressively, our EANN model ($T = 0$) gives a much lower learning curve, exhibiting its superior performance despite its three-body nature. The lower error may be due to the deeper NNs used in our EANN model, but one shall note that much fewer invariants are used as the input (45) here than that (2000) in Ref. [28]. The model accuracy increases with T, although the improvement from $T = 1$ to $T = 2$ is less significant than that from $T = 0$ to $T = 1$. This is consistent with the fact that $T = 1$ already includes eight interatomic distances of CH_4 [see Fig. 2(b)] that are close to complete (10 distances in total). The learning curve more or less converges at $T = 2$, whose errors are one order of magnitude smaller than those with $3B + 4B$ features [28], and those with the contracted $5B$ features [29]. We find that other MPNN models [30,33,34] also perform better than the purely local descriptor-based model [28] (detailed in the SM). This provides more convincing evidence that iterative message passing can include more complete correlations to better represent the atomic environment.

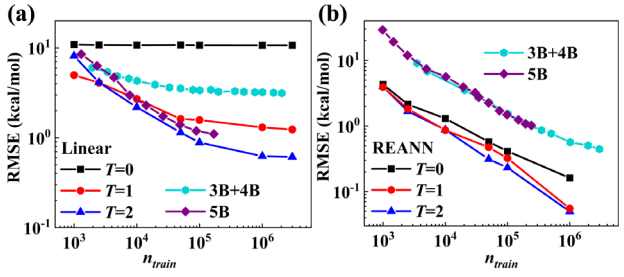

FIG. 3. (a) Comparison of the RMSEs for energies of random CH_4 configurations of linear fits in Ref. [28] (with $3B + 4B$ correlations) and [29] (with $5B$ correlation), and that with the recursively embedded density descriptor ($T = 0, 1, 2$). (b) Similar to (a) but all results are now based on nonlinear NN fits.

An additional advantageous feature of the REANN model is its effective description of some nonlocal effects. This is because the correlations between atoms inside and outside the cutoff sphere have been implicitly encoded when iteratively updating orbital coefficients, as illustrated in Fig. 2(a). We demonstrate this in bulk water, an important benchmark to demonstrate the power of MLIPs. We first use a dataset with 1593 structures of 64 water molecules computed by Cheng *et al.* [45] for developing a BPNN potential [12]. The cutoff radius of BPNN potential was set long enough ($r_c = 6.2$ Å) to describe the strong hydrogen bond interactions. With an optimal selection of symmetry functions ($3B$ features), the reported RMSEs of the BPNN potential are comparable to those of the EANN model [38] with the same r_c, as listed in Table I. Impressively, our REANN model ($T = 3$) greatly outperforms these two purely local descriptor-based counterparts, leading to a smaller error of force with merely half of the cutoff radius ($r_c = 3$ Å). Apparently, this cutoff only incorporates the interactions between a water molecule and some nearest neighbors, but the second neighboring shell is implicitly correlated by the message-passing way of updating orbital coefficients. The performance of the REANN model further improves with the increasing r_c and saturates at $r_c \approx 5.5$ Å, yielding less than half of the RMSEs of the

TABLE I. Test RMSEs of energies (meV/atom) and forces (meV/Å) for bulk water using the dataset in Ref. [45].

Model	REANN ($T = 3$)					EANN[a]	BPNN
r_c (Å)	3.0	3.5	4.5	5.5	6.2	6.2	6.2
Energy	2.8	1.5	1.1	0.9	0.8	2.1	2.3
Force	104.4	73.1	58.0	51.1	53.2	129.0	120

[a]Values taken from Ref. [38].

BPNN potential. It is worth noting that the SchNet model performs less well in this condensed phase system, due presumably to that more interatomic distances are greater than r_c and corresponding atoms cannot be correlated by two-body terms. To avoid any data bias, we simply test another dataset of water trained by Zhang *et al.* using the deep potential molecular dynamics (DPMD) method [14] with $r_c = 6.0$ Å. Our REANN model ($T = 2$) with $r_c = 4.5$ Å gives RMSEs of 0.2 meV/atom (energies) and 15.9 meV/Å (forces), again less than half of the reported values in Ref. [14]. These results suggest that the REANN model captures nonlocal interactions more efficiently than simply extending the effective cutoff radius in complex systems.

Summarizing, we make a physical adaption of the local descriptor-based EANN model to generate the REANN model and reveal its connection with other less physically intuitive MPNNs often inspired from graph neural networks in computer science. We formally derive that how the many-body correlations are introduced by iteratively passing messages (updating orbital coefficients here) and prove that this is a more efficient way to achieve a complete description of the local environment, without explicitly computing high-order features. Numerical tests demonstrate the local completeness and nonlocality of this new model, warranting its superior accuracy among existing ML models. Our strategy can be easily adapted to improve other sophisticated many-body descriptors without changing their basic structures, for example, by making atomic weights of the weighted atom-centered symmetry functions variable with its local environment [23] or adding such learnable coefficients to the DPMD descriptors [14]. We believe this will open a new window for developing more accurate and efficient ML models of more complicated physical systems.

All datasets can be found in the original publications. The reported machine-learned models are openly available from the GitHub respository [47].

This work is supported by National Key R&D Program of China (No. 2017YFA0303500), CAS Project for Young Scientists in Basic Research (YSBR-005), National Natural Science Foundation of China (No. 22073089 and No. 22033007), Anhui Initiative in Quantum Information Technologies (No. AHY090200), and The Fundamental Research Funds for Central Universities (No. WK2060000017). Calculations have been done on the Supercomputing Center of USTC. We thank Professor Michele Ceriotti and Dr. Sergey Pozdnyakov for kindly explaining their dataset for CH_4.

*bjiangch@ustc.edu.cn

[1] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Chem. Rev. **121**, 10142 (2021).
[2] K. Shao, J. Chen, Z. Zhao, and D. H. Zhang, J. Chem. Phys. **145**, 071101 (2016).
[3] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Nat. Commun. **8**, 13890 (2017).
[4] J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. **8**, 3192 (2017).
[5] C. Qu, Q. Yu, and J. M. Bowman, Annu. Rev. Phys. Chem. **69**, 151 (2018).
[6] S. Manzhos and T. Carrington, Chem. Rev. **121**, 10187 (2021).
[7] B. Jiang, J. Li, and H. Guo, J. Phys. Chem. Lett. **11**, 5120 (2020).
[8] W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral, and G. Cui, J. Phys. Chem. Lett. **9**, 6702 (2018).
[9] J. Westermayr, M. Gastegger, M. Menger, S. Mai, L. Gonzalez, and P. Marquetand, Chem. Sci. **10**, 8100 (2019).
[10] Y. Guan, D. H. Zhang, H. Guo, and D. R. Yarkony, Phys. Chem. Chem. Phys. **21**, 14205 (2019).
[11] J. Westermayr and P. Marquetand, Chem. Rev. **121**, 9873 (2021).
[12] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[13] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
[14] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Phys. Rev. Lett. **120**, 143001 (2018).
[15] S. D. Huang, C. Shang, X. J. Zhang, and Z. P. Liu, Chem. Sci. **8**, 6327 (2017).
[16] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Sci. Adv. **3**, e1701816 (2017).
[17] B. Jiang and H. Guo, J. Chem. Phys. **141**, 034109 (2014).
[18] J. Behler, J. Chem. Phys. **134**, 074106 (2011).
[19] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).
[20] Y. Zhang, C. Hu, and B. Jiang, J. Phys. Chem. Lett. **10**, 4962 (2019).
[21] M. J. Willatt, F. Musil, and M. Ceriotti, J. Chem. Phys. **150**, 154110 (2019).
[22] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, J. Chem. Phys. **148**, 241717 (2018).
[23] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, and P. Marquetand, J. Chem. Phys. **148**, 241709 (2018).
[24] A. V. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).
[25] R. Drautz, Phys. Rev. B **99**, 014104 (2019).
[26] A. E. A. Allen, G. Dusson, C. Ortner, and G. Csányi, Mach. Learn. **2**, 025017 (2021).
[27] S. D. Huang, C. Shang, P. L. Kang, and Z. P. Liu, Chem. Sci. **9**, 8644 (2018).
[28] S. N. Pozdnyakov, M. J. Willatt, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Phys. Rev. Lett. **125**, 166001 (2020).
[29] J. Nigam, S. Pozdnyakov, and M. Ceriotti, J. Chem. Phys. **153**, 121101 (2020).
[30] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).
[31] O. T. Unke and M. Meuwly, J. Chem. Theory Comput. **15**, 3678 (2019).
[32] K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, and R. J. Maurer, Nat. Commun. **10**, 5024 (2019).

[33] B. Anderson, T. S. Hy, and R. Kondor, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach *et al.* (Curran Associates, Inc., Vancouver, Canada, 2019), p. 14537.
[34] K. Johannes, G. Janek, and G. Stephan, in *Proceedings of the International Conference on Learning Representations* (2020).
[35] J. Behler, Chem. Rev. **121**, 10037 (2021).
[36] M. S. Daw and M. I. Baskes, Phys. Rev. B **29**, 6443 (1984).
[37] A. Takahashi, A. Seko, and I. Tanaka, Phys. Rev. Mater. **1**, 063801 (2017).
[38] Y. Zhang, C. Hu, and B. Jiang, Phys. Chem. Chem. Phys. **23**, 1815 (2021).
[39] Y. Zhang, R. J. Maurer, and B. Jiang, J. Phys. Chem. C **124**, 186 (2020).
[40] Y. Zhang, S. Ye, J. Zhang, C. Hu, J. Jiang, and B. Jiang, J. Phys. Chem. B **124**, 7284 (2020).
[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.127.156002 which includes Refs. [14,28–31,33,34,38,42–45], for training details, illustrations of high-order correlations, incompleteness of two-body feature-based message-passing neural networks, and comparison with other message-passing neural networks.
[42] A. Paszke *et al.*, in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, edited by H. Wallach *et al.* (Curran Associates Inc., Vancouver, Canada, 2019).
[43] I. Loshchilov and F. J. A. E.-P. Hutter, arXiv:1711.05101.
[44] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. J. A. E.-P. Salakhutdinov, arXiv:1207.0580.
[45] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, Proc. Natl. Acad. Sci. U.S.A. **116**, 1110 (2019).
[46] E. J. L. Borda and A. Samanta, arXiv:2004.14442.
[47] Y. Zhang, J. Xia, and B. Jiang, Physically-motivated-recursively-embedded-atom-neural-networks (2021), https://github.com/zhangylch/Physically-motivated-Recursively-Embedded-Atom-Neural-Networks.