JCTC Journal of Chemical Theory and Computation

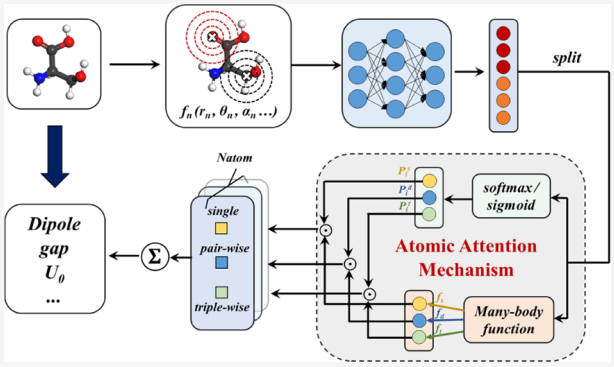
pubs.acs.org/JCTC Article

Many-Body Function Corrected Neural Network with Atomic Attention (MBNN-att) for Molecular Property Prediction

Zheng-Xin Yang, Xin-Tian Xie, Pei-Lin Kang, Zhen-Xiong Wang, Cheng Shang,* and Zhi-Pan Liu*

Cite This: https://doi.org/10.1021/acs.jctc.4c00660 Read Online

ABSTRACT: Recent years have seen a surge of machine learning (ML) in chemistry for predicting chemical properties, but a low-cost, general-purpose, and high-performance model, desirable to be accessible on central processing unit (CPU) devices, remains not available. For this purpose, here we introduce an atomic attention mechanism into many-body function corrected neural network (MBNN), namely, MBNN-att ML model, to predict both the extensive and intensive properties of molecules and materials. The MBNN-att uses explicit function descriptors as the inputs for the atom-based feed-forward neural network (NN). The output of the NN is designed to be a vector to implement the multihead self-attention mechanism. This vector is split into two parts: the atomic attention weight part and the many-body-function part. The final property is obtained by summing the products of each atomic attention weight and the corresponding many-body function. We show that MBNN-att performs well on all QM9 properties, i.e., errors on all properties, below chemical accuracy, and, in particular, achieves the top performance for the energy-related extensive properties. By systematically comparing with other explicit-function-type descriptor ML models and the graph representation ML models, we demonstrate that the many-body-function framework and atomic attention mechanism are key ingredients for the high performance and the good transferability of MBNN-att in molecular property prediction.



1. INTRODUCTION

Machine learning (ML) provides a new avenue to predict the complex properties of chemicals. In particular, the ML model to predict the total energy, an extensive property, has demonstrated its great power in accelerating atomic simulations without solving the expensive Schrodinger equation.[1−4] These ML methods with suitable extensions have been tested for other molecular and material properties using the data set from either quantum mechanics calculations or experiments.[5,6] In general, all ML methods need structural descriptors for discriminating the chemical environment as well as a specific ML model, e.g., neural network (NN), with trainable parameters for learning the target properties. The difference in the structural descriptor and the ML model led to many variants of ML methods for potential energy surface (PES) calculations, implemented in different codes, such as HDNN,[7] ANI,[8] LASP,[9,10] PIPs,[11] GAP,[12] DeepMD,[13] EANN,[14] HIP-NN,[15] ACE,[16] etc. These methods can differ greatly in the parameter size, e.g., from thousands to millions, as well as in the hardware platforms of execution, either central processing units (CPU) or graphic processing units (GPU). With such a high flexibility in choosing ML models, there is a general concern about the performance of an ML model toward a specific property. A low-cost and CPU-friendly ML framework with good performance for all types of properties is thus desirable.

The structure descriptors can be broadly classified into two types, namely, the explicit function and graph representation, as shown in Figure 1. The descriptors in the explicit function learn from classical force fields by constructing many-body functions based on internal coordinates (bonds, angles, and torsions), which satisfy the numerical invariance to the translation, rotation, and permutation of structure.[17] These explicit functions are often physically meaningful and provide a fast route to capture the interactions between atoms. Behler and Parrinello[7] first proposed a set of general-purpose numerical functions, namely, Atom-Centered Symmetry Functions (ACSFs) to describe the atomic environment, which is utilized to construct the atomic ML model, e.g., feed-forward neural network (NN) for predicting the atomic energy. The explicit functions by themselves have no learnable

Received: May 20, 2024
Revised: July 8, 2024
Accepted: July 9, 2024

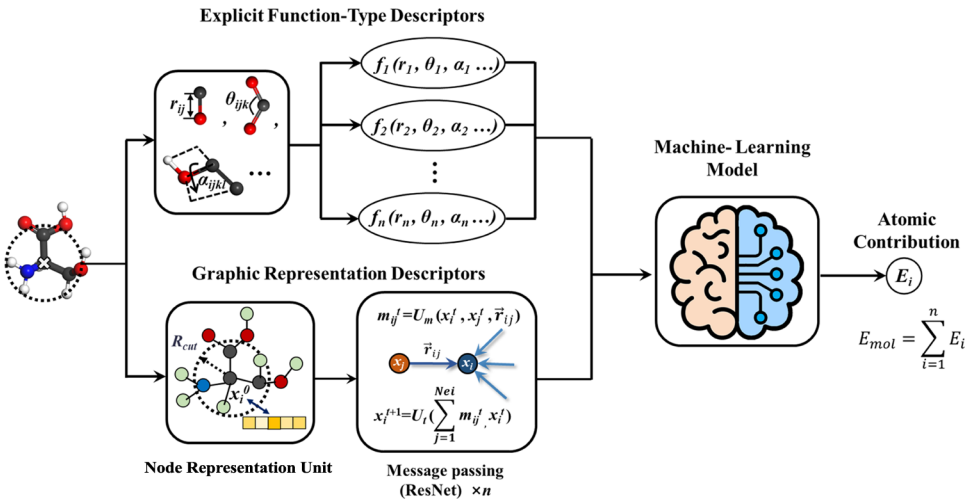


Figure 1. Illustration of the structure descriptors utilized in ML models for molecule property prediction, which can be broadly distinguished as explicit function-type descriptors and graph representation descriptors.

parameter but have a few preset empirical hyperparameters, e.g., the width in Gaussian functions in ACSF, which could be important to the performance of the ML model. All learnable parameters are in the atom-based ML model, usually one model for one element, leading to a relatively small parameter space. The atom-based ML model requiring only atom-wise operation is linear scaling (O(N)) and suited for CPU hardware devices. The representative explicit-function descriptors include ACSF,[7] SOAP,[18] PTSD,[9] MBTR,[19] etc.

On the other hand, the graph representation descriptors rely on ML models to extract structural features, often message-passing neural networks (MPNNs),[20] a type of graph network. In MPNNs, as shown in Figure 1, the feature of atom i (x_i) is initialized using an embedding vector dependent on atom type Z. During the message passing phase, a message m_{ij} between two atoms (i and j) is generated according to a message function U_m based on the atomic representations x_i and x_j, as well as the pairwise feature, e.g., that is derived from the spatial information $\vec{r}_{ij}$ between two atoms. The updated feature x_i^{t+1} at step $t + 1$ is obtained from the messages of all neighbors of atom i at step t, using the update function U_t. The message-passing process usually iterates for n time steps (e.g., $n \geq 6$) with the ResNet architecture[21] to stack the messaging passing blocks.[22,23] Both the message functions U_m and the update functions U_t can have learnable parameters, where U_m operates on paired (or more) atoms and U_t operates on atoms. While the message-passing mechanism can be extended to incorporate learnable atomic parameters in improving the explicit-function descriptors (e.g., AIMNet[24] and REANN[25,26] models), recent years have seen great efforts to incorporate more three-dimensional (3D) spatial information into the message-passing layers, leading to very complex U_m. For instance, DimeNet[27] extends U_m operation to the three-body interaction level by incorporating directional message-passing and spherical Bessel basis functions. SphereNet,[28] one step further, developed the spherical message passing and incorporated the four-body (torsion) 3D structural information. Not surprisingly, the time complexity for updating atomic feature vectors in each layer increases quickly to O(g^2N) and even O(g^3N) (g represents the number of neighboring atoms within the cutoff of the central atom and N represents the number of atoms in this system).[29−31] The high computational cost of modern MPNN determines the high-performance GPU platform as the only hardware in practice.

Here we incorporate a multihead attention mechanism into our previously developed many-body function corrected neural network (MBNN) potential, namely, MBNN-att, to realize a low-cost, general-purpose property predictor for molecules and materials. The multihead attention mechanism is designed to provide an attention weight to each of the many-body functions, extending the original MBNN framework to predict intensive properties and improving its ability to predict extensive properties. The MBNN-att can be efficiently implemented in CPU hardware and thus is compatible with most electronic structure calculation codes. Using the QM9 data set as the example, we demonstrate that MBNN-att outperforms previous explicit-function-descriptors-based methods and provides better or comparable performance compared to graph-network-based methods. In particular, the predictive ability for energy properties ranks high among all ML models.

2. METHODS

2.1. Explicit Function Structural Descriptors. Before introducing MBNN-att, we briefly overview the explicit-function structure descriptors and MBNN framework utilized in MBNN-att. To predict molecular properties, it is essential to design sensitive structural descriptors as inputs of ML models that are capable of distinguishing the subtle structural differences between structures. Unlike the classic force field that utilizes many-body functions to describe the physical terms directly, the structure descriptors for ML models are not necessarily physically meaningful as long as the symmetry invariance is maintained including translation, rotation, and permutation of atoms. This offers a large flexibility to construct explicit function-type structure descriptors, which could now be generally classified into two types, global and local descriptors.

The global descriptors encode the geometry information on the entire structure into a one-dimensional vector. As a representative global descriptor, the Many-Body Tensor Representation (MBTR)[19] decomposes structures into distributions of structural patterns of different sizes and concatenates the distributional information regarding the monomers, distances, and angles of different elements as

https://doi.org/10.1021/acs.jctc.4c00660
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

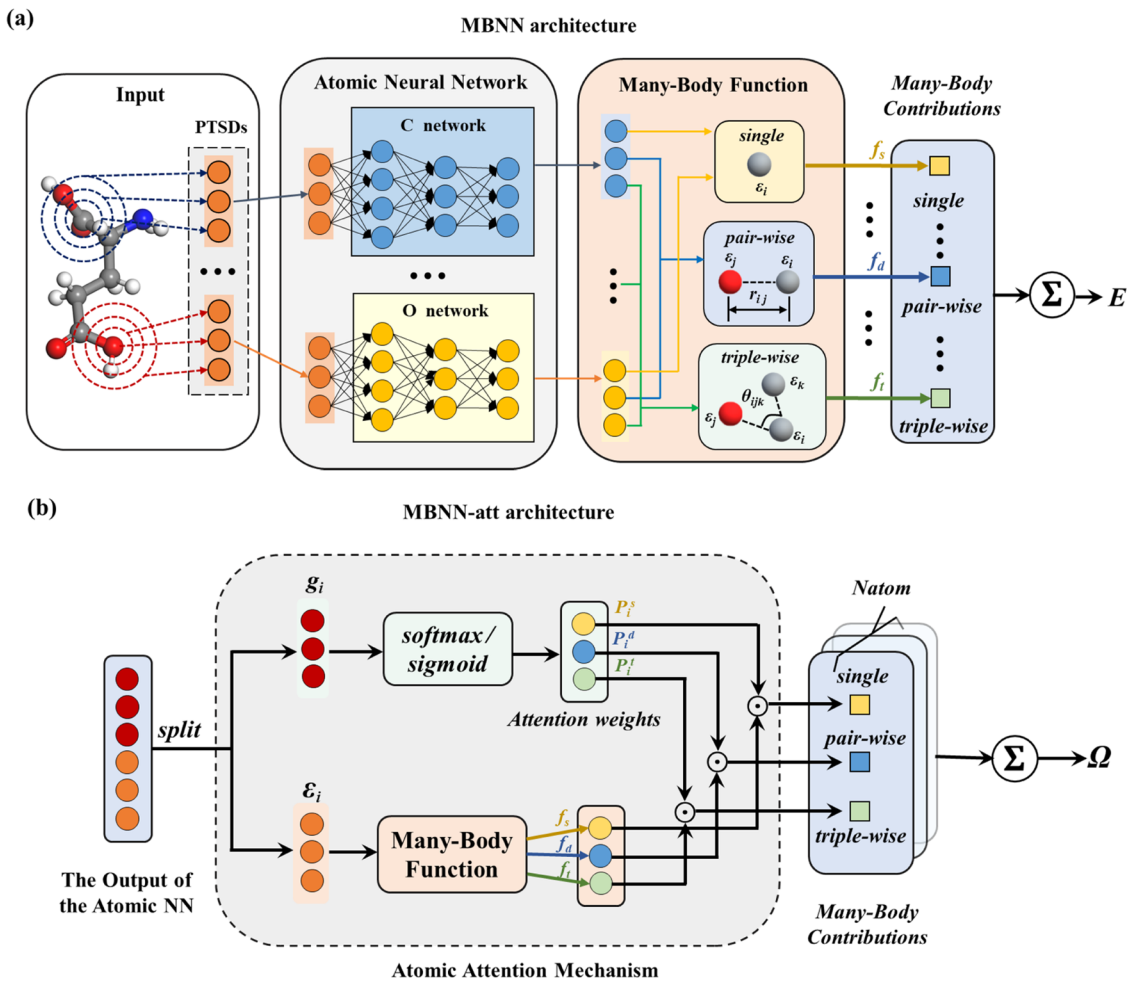


Figure 2. Illustration of the MBNN (a) and MBNN-att (b) architectures for property prediction. In MBNN (**a**), the total energy (E) is considered as the sum of many-body contributions using one-body (f_s), two-body (f_d), and three-body (f_t) functions, and so on, with the coefficients ε_i from the output of atomic NN. In MBNN-att (b), the atomic attention mechanism is introduced by splitting the output of the atomic NN, half as the many-body function coefficients ε_i, and the other half as the weights g_i for calculating the attention weight of each atom.

structural descriptors. This type of global descriptor is mainly utilized for describing entire molecular structures and can be extended to periodic systems with a fixed atom number.

On the other hand, the local descriptors describe the chemical environment of atoms and thus scale naturally with the number of atoms. This type of descriptor is more popular and can be applied generally to both aperiodic molecules and extended solids systems. Popular local descriptors include Atom-Centered Symmetry Function (ACSF),[7] Smooth Overlap of Atomic Position (SOAP),[18] and Power-Type Structure Descriptor (PTSD).[9] ACSF and PTSD are described in the following with more details, since PTSD is the structure descriptor utilized for constructing MBNN-att. More information on other descriptors can be found in the Supporting Information (SI) Part 1.

ACSF utilizes Gaussian-type functions to describe the radial part and the trigonometric function (cosine) to account for the angular part of the chemical environment around atoms. The formula for the radial function G_2 (two-body) and the angular function G_4 (three-body) are exampled in eqs 1−3, where the interatomic distance r_{ij} represents the distance between atoms i and j; θ_{ijk} is the angle between atoms i, j, and k (i being the central atom of the angle); r_s and η are the center and width of the Gaussian function respectively; and f_c is a cutoff function, ensuring that the interaction between atoms beyond the cutoff distance r_c equals zero.

$$G_i^2 = \sum_{j\neq i} \exp\left(-\eta(r_{ij} - r_s)^2\right)\cdot f_c(r_{ij}) \tag{1}$$

$$G_i^4 = 2^{1-\xi} \sum_{j,k\neq i} (1 + \lambda \cos\theta_{ijk})^{\xi}\cdot \exp\left(-\eta(r_{ij}^2 + r_{ik}^2 + r_{jk}^2)\right) \cdot f_c(r_{ij})\cdot f_c(r_{ik})\cdot f_c(r_{jk}) \tag{2}$$

$$f_c(r_{ij}) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{\pi r_{ij}}{r_c}\right)\right], & \text{for } r_{ij} \leq r_c \\ 0 \quad \text{for } r_{ij} > r_c \end{cases} \tag{3}$$

It has been observed that the performance of Gaussian functions is sensitive to the hyperparameters (center and width), and for fitting complex global PES data set, it is practically difficult to identify Gaussian parameters to cover a wide range of geometry changes.[32,33] To alleviate this problem, we designed PTSDs by combining radial power functions, spherical harmonics, and cosine functions to encode 3D structural information; particularly, the introduction of spherical harmonics greatly improves the angular description

https://doi.org/10.1021/acs.jctc.4c00660
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

of the coordination environment. The formulas for the two-body descriptors S_1 and S_2 are shown as examples in eqs 4−7, where the power function with the power n (e.g., 2, 4, 8, 16) can yield either a smooth or a sharp monotonically increasing function, and in combination with the cutoff f_c function and the spherical harmonics, can conveniently capture the radial (eq 6) and angular (eq 7) geometry changes.

$$f_c(r_{ij}) = \begin{cases} 0.5 \times \tanh^3\left[1 - \frac{r_{ij}}{r_c}\right], \text{ for } r_{ij} \le r_c \\ 0 \quad \text{for } r_{ij} > r_c \end{cases} \tag{4}$$

$$R^n(r_{ij}) = r_{ij}^n \cdot f_c(r_{ij}) \tag{5}$$

$$S_i^1 = \sum_{j \neq i} R^n(r_{ij}) \tag{6}$$

$$S_i^2 = \left[\sum_{m=-L}^{L} \left| \sum_{j \neq i} R^n(r_{ij}) Y_{Lm}(\vec{r_{ij}}) \right|^2 \right]^{1/2} \tag{7}$$

2.2. MBNN Architecture. The atom-centered structural descriptors can then be utilized as the input to construct the atomic NN that outputs the contribution term of the atom, resulting in the so-called HDNN architecture.[7] The HDNN architecture is naturally suited for extensive properties, such as the total energy, which can be expressed as the sum of individual atomic energies. The HDNN architecture has, however, limitations in describing long-range interactions and complex local interactions encountered in chemical reactions.[34−36] The former is due to the limited cutoff radius of the structure descriptors, typically within 7 Å, which is not sufficient to capture long-range interactions, for example, in the defect−defect interaction in low-doping materials; and the latter is related to the insufficient extrapolation ability of the ML model.

To solve the problems of HDNN, we recently proposed the MBNN[37] architecture, in which the atomic NN outputs a vector $\{\varepsilon_i^m, m = 1,2,\cdots\}$, instead of a single value in HDNN. The MBNN architecture is shown in Figure 2a. The output vector acts as the atomic coefficient in constructing different many-body functions to compute the energy (E) of the structure, as shown in eq 8, where f_s, f_d, and f_t stand for single-body, two-body, and three-body functions. The MBNN architecture offers an effective approach to directly capture the desired many-body interactions.

$$\begin{aligned} E = &\sum_{s=1}^{ns} \sum_{i=1}^{na} f_s(\varepsilon_i^s) + \sum_{d=1}^{nd} \sum_{i,j=1}^{npair} f_d(\varepsilon_i^d, \varepsilon_j^d, r_{ij}) \\ &+ \sum_{sp=1}^{nd} \sum_{i,j=1}^{npair} f_{sp}(\varepsilon_i^{sp}, \varepsilon_j^{sp}, r_{ij}, \vec{r_{ij}}) \\ &+ \sum_{t=1}^{nt} \sum_{i,j,k=1}^{ntriple} f_t(\varepsilon_i^t, \varepsilon_j^t, \varepsilon_k^t, r_{ij}, r_{ik}, r_{jk}, \theta_{ijk}) + \cdots \end{aligned} \tag{8}$$

Equations 9−12 list the representative many-body functions f in LASP program,[38,39] where the subscript s, d, sp, and t stand for the single-body, two-body, spherical, and three-body functions, respectively, and λ, m_d, m_{sp}, m_t, and l_t are the adjustable prefactor or power parameters in these functions.

$$f_s = \lambda \varepsilon_i^s \tag{9}$$

$$f_d = \lambda \frac{\varepsilon_i^d \varepsilon_j^d}{r_{ij}^{m_d}} f_c(r_{ij}) \tag{10}$$

$$f_{sp} = 0.5\lambda \times \sum_m (\varepsilon_i^{sp})^2 \sum_{j=1}^{na} [\varepsilon_j^{sp} r_{ij}^{m_{sp}} \cdot Y_l^m(\vec{r_{ij}}) f_c(r_{ij})]^2 \tag{11}$$

$$f_t = \lambda \frac{\varepsilon_i^t(\varepsilon_j^t + \varepsilon_k^t)(2 - \cos(\theta_{ijk}))^{l_t}}{r_{ij}^{m_t} \cdot r_{ik}^{m_t}} f_c(r_{ij}) f_c(r_{ik}) \tag{12}$$

2.3. MBNN with Atomic Attention (MBNN-att). We now describe the MBNN-att developed in this work, which is illustrated in Figure 2b. The MBNN in eq 8 provides a simple framework for predicting the extensive properties, but is not suitable for intensive properties. Inspired by the self-attention mechanism utilized in Transformer,[40] we here add the atomic attention mechanism into MBNN architecture. As shown in Figure 2b, compared to MBNN, MBNN-att outputs two sets of quantities, one set for the coefficients ε_i^m utilized in many-body functions and another set for the atomic attention g_i^m that has the same dimension as the many-body function coefficients. The atomic attention is described in eqs 13 and 14, which takes different forms for the intensive and extensive property.

For an intensive property that is invariant to the number of molecules, we employ a softmax-function-based self-attention mechanism (Figure 2b) to yield the weighting of atom P_i^m (eq 13). The input to the softmax function, g_i^m (the m-th output of the NN of atom i) is a subvector extracted from the output vector of the atomic NN, which follows the idea of Multi-Layer Perceptron Attention (MLPA).[41] The total weight is normalized to one for the entire molecule (with N atoms).

$$P_i^m = \text{softmax}(g_i^m) = \frac{\exp(g_i^m)}{\sum_N \exp(g_i^m)} \tag{13}$$

For the extensive property, the atomic attention is simply obtained by a sigmoid function (eq 14), which yields a value in a range of [0, 1].

$$P_i^m = \text{sigmoid}(g_i^m) \tag{14}$$

By assigning the atomic attention weight P_i^m to each many-body function, we can then obtain the property (Ω) by modifying eqs 8−15 by using the product of the atomic attention weights and many-body functions. Equation 15 achieves the atom-based multihead attention by assigning varied atom-based attention weights to different many-body functions.

https://doi.org/10.1021/acs.jctc.4c00660
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\Omega = \sum_{s=1}^{ns}\sum_{i=1}^{na} P_i^s f_s(\varepsilon_i^s) + \sum_{d=1}^{nd}\sum_{i,j=1}^{npair} (P_i^d + P_j^d) f_d(\varepsilon_i^d, \varepsilon_j^d, r_{ij}) + \sum_{sp=1}^{nd}\sum_{i,j=1}^{npair} (P_i^{sp} + P_j^{sp}) f_{sp}(\varepsilon_i^{sp}, \varepsilon_j^{sp}, r_{ij}, \vec{r}_{ij}) + \sum_{t=1}^{nt}\sum_{i,j,k=1}^{ntriple} (P_i^t + P_j^t + P_k^t) f_t(\varepsilon_i^t, \varepsilon_j^t, \varepsilon_k^t, r_{ij}, r_{ik}, r_{jk}, \theta_{ijk}) + \cdots \quad (15)$$

We note that the attention mechanism has been adopted in several NN architectures to predict molecular properties, such as Graph Attention Networks (GAT)[42] and Equiformer,[43] where the attention coefficients operate generally at the level of atomic pairs in updating the atomic features using the MPNN architecture. This kind of attention is computationally intensive and is currently only implemented on GPU device. By contrast, our attention mechanism is atom-based, compatible with the feed-forward NN, and can be realized readily on CPU hardware, as implemented in the LASP program.[38,39]

The MBNN-att architecture has several advantages in terms of computational efficiency. First, MBNN-att utilizes PTSDs as the descriptors, which have no learnable parameters and thus remain unchanged throughout the training process. This allows the PTSD values and their derivatives to be saved in the memory in advance. Second, the many-body functions and attention weights utilize the outputs from atomic NN, enabling atom-wise parallelization in both training and validation. We emphasize that the atom-wise operation is particularly beneficial for periodic solid systems, where the number of atoms in the cutoff sphere can be quite large. As shown in Figure 3, we plot a typical computational cost for MBNN calculations with the number of atoms ranging from 96 to 103,488 on a 96-core CPU machine (Intel(R) Xeon(R) Platinum 8168 CPU). The testing system is a periodic bulk of Ru metal (the atomic structure of the 96-atom bulk is detailed in SI Part 5). Figure 3 shows that MBNN achieves linear scaling, and the additional cost due to the many-body function calculations is less than 9.65% compared to the calculations without many-body functions. The low cost of calculating many-body functions is due to the limited number of many-body functions and their simple functional forms.

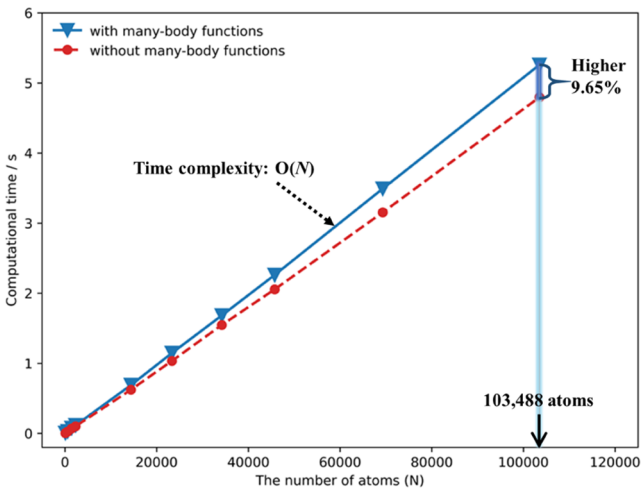


Figure 3. Linear scaling of MBNN calculations with and without the many-body functions by plotting the computational time versus the system size in atom numbers (N). The additional computational cost is less than 10% in the presence of many-body functions.

3. RESULTS AND DISCUSSION

3.1. MBNN-att Performance on QM9 Data Set. We then evaluate the performance of MBNN-att in predicting molecular properties on the QM9 data set. The QM9 data set[44] comprises 133,885 organic small molecules in their equilibrium states, consisting of elements C, H, O, N, and F, with molecular sizes ranging from 3 to 29 atoms each. For each molecule, 12 different properties, including energetic, electronic, and thermodynamic properties, were computed using density functional theory (DFT) calculations at the B3LYP/6-31G (2df, p) level. In this work, the QM9 data set was split into a training set with 121,320 samples and a test data set with 12,565 samples.

Our MBNN-att model takes 801 PTSD functions as inputs and consists of four fully connected layers, each with 240 neurons. The tanh activation function is applied to the first three layers, while the final layer utilizes a linear activation function. The final output layer varies, depending on the property being predicted. Among the 12 different properties, U_0, U, G, H, and zero-point vibrational energy (ZPVE) belong to extensive properties and the polarizability (α), heat capacity (C_v), ε_{HOMO}, and ε_{LUMO} are intensive properties, which are treated correspondingly using eqs 13−15 (highest occupied molecular orbital (HOMO); lowest unoccupied molecular orbital (LUMO)). The remaining $\Delta\varepsilon$ (energy gap) is calculated from predicted $\varepsilon_{LUMO} - \varepsilon_{HOMO}$. We use MBNN-att to predict atomic charges for properties like dipole moment (μ) and electronic spatial extent ($\langle R^2 \rangle$), and then use mathematical equations to connect atomic charges to the properties (refer to SI eqs S12−S14 for details). The limited-memory Broyden−Fletcher−Goldfarb−Shanno (L-BFGS) optimizer[45] was utilized for the training of ML models.

Table 1 compares the performance of MBNN-att with those of other ML models using different explicit function-type descriptors: MBTR, ACSF, SOAP, and PTSD. All of these ML models utilize the feed-forward NN to learn the properties (see Figure 1). The PTSD model in Table 1 would be the same as that of MBNN-att if the many-body functions and the atomic attention are removed. For the global descriptor MBTR, we utilize 3475 MBTRs as NN inputs. For the local descriptor, we use 3275 ACSFs, 3255 SOAPs, and 801 PTSDs as descriptors, respectively. The calculation of MBTR, ACSF, and SOAP is implemented using the DScribe library.[46] The NN architecture and parameter size utilized for MBTR, SOAP, and PTSD are the same as those for MBNN-att. More information about the hyperparameters of the MBTR, SOAP, and PTSD descriptors and the architecture of the feed-forward NN can be found in SI Part 1 and ref 47.[47]

We found that in general, MBNN-att achieves the best performance for all properties compared to other feed-forward NN models. For instance, the MAE of ε_{HOMO} predicted by the MBNN-att is 35.1 meV, which is substantially lower than the 54.2 meV from the PTSD model, where the same PTSD structural descriptors are utilized as the input for the feed-forward NN. Similarly, for the energy property U_0, MBNN-att achieves an MAE of 5.62 meV, almost half of the value obtained with the PTSD model (10.91 meV). This indicates

https://doi.org/10.1021/acs.jctc.4c00660
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Results on All QM9 Targets and Comparison to Neural Network Models Using Explicit Function-Type Descriptors[a]

		global descriptor-NN	local descriptor-NN			
property	unit	MBTR	ACSF	SOAP	PTSD	MBNN-att
μ	D	0.293	0.065	0.182	0.094	**0.021**
α	a_0^3	0.203	0.171	0.209	0.093	**0.066**
ε_{HOMO}	meV	79.5	68.2	83.7	54.2	**35.1**
ε_{LUMO}	meV	71.8	72.5	75.8	51.5	**29.5**
$\Delta\varepsilon$	meV	107.4	103.3	114.8	76.9	**47.7**
$\langle R^2 \rangle$	a_0^2	2.9	3.5	5.8	1.32	**0.093**
ZPVE	meV	4.20	2.00	2.74	1.79	**1.16**
U_0	meV	49.09	22.83	25.38	10.91	**5.33**
U	meV	48.90	22.85	25.05	11.44	**5.34**
H	meV	48.91	23.51	24.86	11.58	**5.32**
G	meV	49.68	23.02	24.84	11.72	**5.32**
C_v	cal/(mol·K)	0.070	0.070	0.063	0.033	**0.024**
number of descriptors		3475	3275	3255	801	801

[a]Scores are reported as mean absolute error (MAE), and the best outcomes are highlighted in bold.

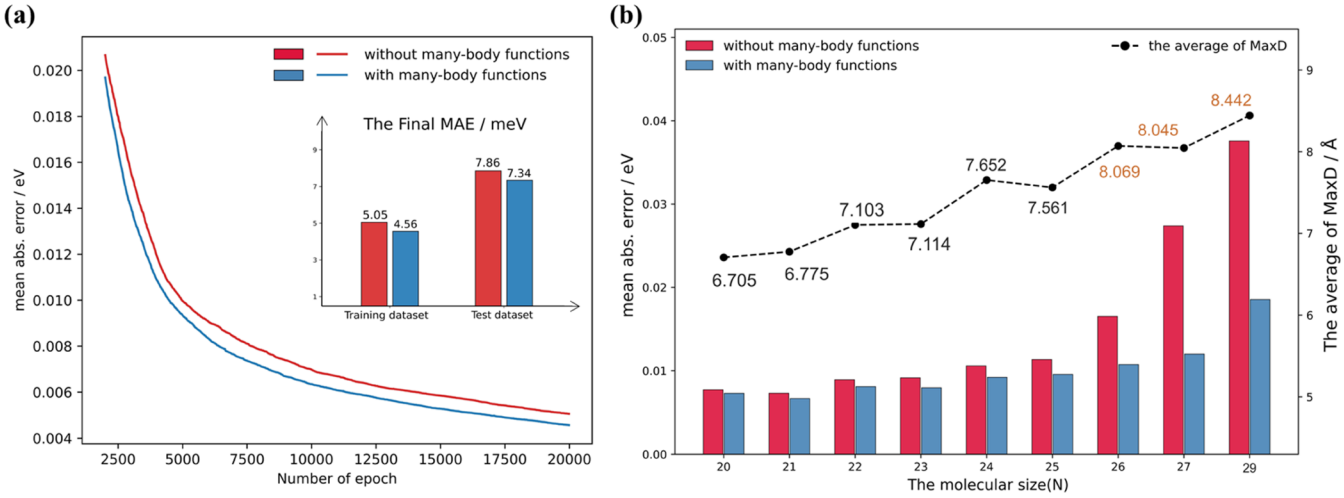


Figure 4. Ablation study for many-body functions of MBNN-att in predicting the extensive property G. (a) Comparison of learning curves with and without many-body functions on the training and test data sets. (b) Transferability test, where the models with and without many-body functions are trained with $N < 20$ and compared for predicting the property G for large molecules. The average of the maximum distance (Å) between atoms (MaxD) in the larger molecules is also shown (8.0 Å is beyond the cutoff radius of PTSD).

that the introduction of many-body functions and the multihead self-attention mechanism substantially improves the ability to predict molecular properties.

For the other ML models, SOAP and PTSD-based NN models can achieve the chemical accuracy for the energy properties (below 43 meV according to ref 48). The local descriptor representation (ACSF, SOAP, and PTSD) outperforms the global descriptor (MBTR) in predicting energy properties. Specifically, the MAE values for U_0 obtained using ACSF, SOAP, and PTSD are 46.87, 25.38, and 10.91 meV, respectively, all lower than the MAE of 49.09 meV achieved by MBTR. The better performance of local representations has been noted in the previous study utilizing kernel ridge regression as the ML model,[49] where the SOAP descriptor, with an MAE of 39.0 meV, is better than the MBTR descriptor, which has an MAE of 58.1 meV. The slightly better results achieved in this work for SOAP and MBTR could be due to NN as the ML model. For the intensive molecular properties such as ε_{HOMO}, ε_{LUMO}, and $\Delta\varepsilon$, all previous models, including the PTSD ML model, did not manage to reach the chemical accuracy (lower than 43 meV according to ref 48). This indicates the importance of the atomic attention mechanism in predicting the intensive properties—the intensive properties cannot be described simply using the direct sum of atomic contributions.

3.2. Ablation Studies for Many-Body Functions and Attention Mechanism. Ablation studies are further carried out to explore the impact of many-body functions and attention mechanisms on molecular property prediction. Two representative properties, i.e., the extensive quantity, Gibbs free energy (G), and the intensive quantity, ε_{HOMO}, are selected for analysis.

For the extensive G property, we compare the learning curves of models with and without many-body functions. The parameters of the many-body functions used in the MBNN-att model are listed in SI Table 2, which include three two-body functions with cutoff radii 12.0, 8.0, and 2.2 Å and two three-body functions with cutoff radii 2.5 and 3.0 Å. Figure 4a shows that the many-body functions speed up the training procedure and improve the accuracy in both the training and the test sets. The MBNN-att achieves the MAE of 4.56 meV on the training data set and 7.34 meV on the test data set after 20,000 training

https://doi.org/10.1021/acs.jctc.4c00660
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

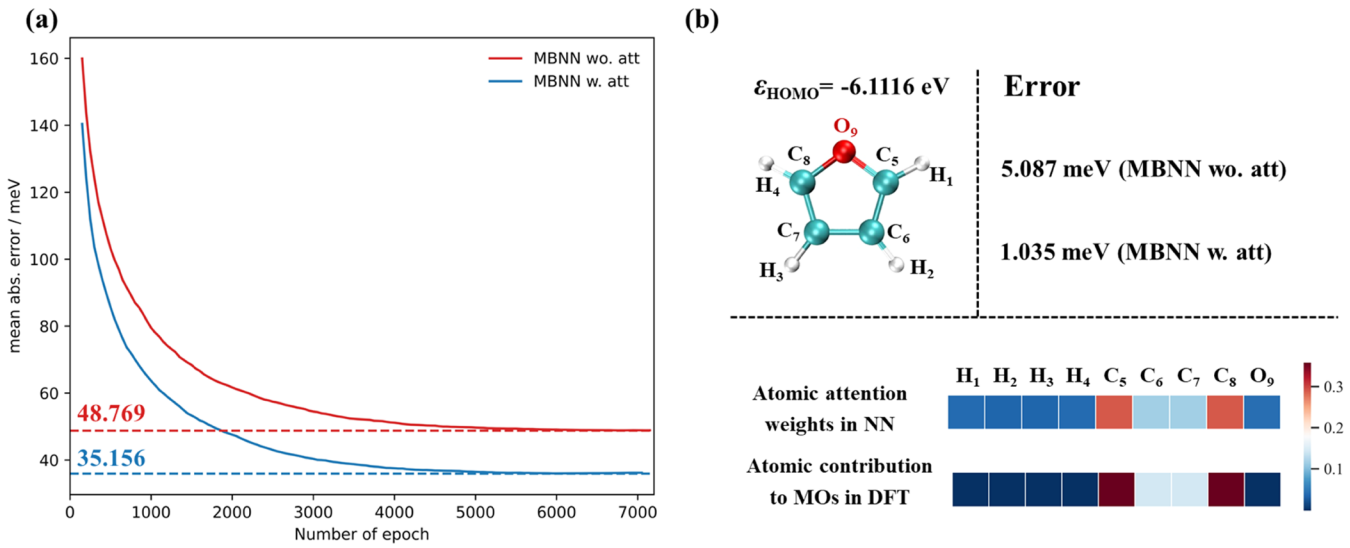


Figure 5. Ablation study for attention mechanism of MBNN-att in predicting the intensive property ε_{HOMO}. (a) Comparison of the learning curve in predicting ε_{HOMO} on test data sets for the MBNN model with and without the attention mechanism. (b) Prediction of ε_{HOMO} property for the furan molecule using MBNN with and without the attention mechanism (top), and the head map of the atomic attention weights from MBNN-att and that of the atomic contribution of HOMO from DFT. The color indicates the relative magnitude (bottom).

epochs, outperforming the model without many-body functions (5.05 and 8.06 meV). This demonstrates that the incorporation of many-body functions can better capture the intricate interactions between atoms and improve the predictive ability.

We then investigated the transferability of the models with and without many-body functions, as shown in Figure 4b. The molecules in the QM9 data set range in size from 3 to 29. We trained models on a data set containing all molecules with the number of atoms $N \leq 19$ (approximately 94,000 structures) and tested it on molecules with more than 19 atoms. It shows that the addition of many-body functions systematically improves the model's prediction ability for larger molecules, especially for those with many more atoms compared to the training set. Specifically, MBNN-att achieves MAE values of 10.73, 11.98, and 18.53 meV for predicting the property G of molecules with 26, 27, and 29 atoms, respectively. They are consistently better than the ML model without many-body functions (16.51, 27.38, and 37.56 meV). The improvement of MBNN-att in predicting large-size molecules could be understood as follows. For large molecules more than 25 atoms, the molecule size (the maximum distance between atoms) generally exceeds 8 Å, which is already beyond the cutoff radius of PTSD functions in the input layer (7.5 Å). This indicates the atom at one end could not "see" the atom at the other end if using the standard feed-forward NN model. However, by incorporating many-body functions with cutoffs of 8.0 and 12.0 Å, we increase effectively the range of atomic interactions and the long-range interactions within molecules are thus included, which boosts the model transferability in large molecules.

For the intensive ε_{HOMO} property, we compare the performance of the MBNN models with (wat) and without attention (wo. att). Figure 5a shows the learning curves for the test data set of the two ML models (SI Figure S2 shows the curve for the training set). It is obvious that the attention mechanism greatly improves the MAE on the test data set. Without the attention, the MAE of ε_{HOMO} on the train data set and test data set are 12.22 and 48.77 meV, respectively. These values are systematically higher than those (11.82 and 35.16 meV) with the attention.

Interestingly, we found that the attention weight could be related to the physical quantity in the molecular property. According to molecular orbital theory, molecular orbital (MO) wave functions are constructed by linear combinations of atomic orbitals, and the contribution of atomic orbitals to MOs is reflected by the square of the atomic orbital coefficients. The atomic contributions to MOs are obtained by summing the contributions from individual atomic orbitals to MOs. Taking furan molecule as an example, as shown in Figure 5b, the HOMO in furan is predominantly composed of p_z orbitals of four carbon atoms and thus the atomic orbital contribution for hydrogen and oxygen atoms are close to 0, while those for carbon atoms are 0.35, 0.14, 0.14, and 0.35 according to the Natural Atomic Orbitals (NAO) analysis.[50] By averaging the attention weights P_i^m assigned to each atom i in every many-body function, we can obtain the atomic attention weight for each atom and compare it with the orbital coefficient. Figure 5b shows the atomic attention weight maps, which compare reasonably with the map of the atomic orbital contribution to ε_{HOMO}. The attention weight predicted from MBNN-att is 0.038 for each of the four hydrogen atoms and 0.039 for the oxygen atom, both being much smaller than those for the carbon atoms, being 0.27, 0.13, 0.13, and 0.27 (C_5, C_6, C_7, and C_8). This suggests that the attention weights assigned by MBNN-att can capture the importance of atoms ε_{HOMO} by mimicking the atomic orbital coefficients, although these orbital coefficients are not explicitly learned in ML. With reasonable attention weights, it is no wonder that MBNN with attention predicts ε_{HOMO} with an error of only 1.035 meV, while that without attention yields a 5.087 meV error. The similarity between attention weight with orbital coefficients is not unique to furan molecule but is widely present in QM9 molecules as shown in SI Figure S3. It suggests that the atomic attention correctly identifies the underlying physics on the importance of the individual atom to the specific property, which explains nicely why MBNN-att model has a good performance in predicting molecular properties.

https://doi.org/10.1021/acs.jctc.4c00660
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Comparisons between MBNN-att (Implemented on CPU Device) and Other Graph Representation Models (Implemented on GPU Device) for Predicting QM9 Data Set Properties[a]

property	unit	MBNN-att	SchNet	DimeNet++	SphereNet	PaiNN	ET	equiformer
μ	D	0.021	0.033	0.030	0.025	0.012	**0.011**	**0.011**
α	a_0^3	0.066	0.235	**0.044**	0.049	0.045	0.059	0.046
ε_{HOMO}	meV	35.1	41	24.6	22.8	27.6	20.3	**15.0**
ε_{LUMO}	meV	29.5	34	19.5	18.9	20.4	17.5	**14.0**
$\Delta\varepsilon$	meV	47.7	63	32.6	31.1	45.7	36.1	**30.0**
$\langle R^2 \rangle$	a_0^2	0.093	0.073	0.331	0.268	**0.066**	0.033	0.251
ZPVE	meV	1.16	1.7	1.21	**1.12**	1.28	1.84	1.26
U_0	meV	**5.33**	14	6.32	6.26	5.85	6.15	6.59
U	meV	**5.34**	19	6.28	6.36	5.83	6.38	6.74
H	meV	**5.32**	14	6.53	6.33	5.98	6.16	6.63
G	meV	**5.32**	14	7.56	7.78	7.35	7.62	7.63
C_v	cal/(mol·K)	0.024	0.033	0.023	**0.021**	0.024	0.026	0.023

[a]The best outcomes are highlighted in bold.

3.3. Performance Comparison between MBNN-att and Graph Representation ML Models. As mentioned in Section 1, graph representation descriptors using MPNN offer another powerful solution to capture complex molecular features. Unlike explicit function-type descriptors, graph representation descriptors initialize atomic representations using embedding layers and update them iteratively through hidden layers in MPNN. The message-passing process involves the operation on atom pairs (as node-to-node in the graph), containing learnable parameters to determine the contribution of the pair. Because the pairwise operation is computationally intensive, graph representation ML models have to be performed on GPU devices.

Before we compare our MBNN-att performance with those of graph representation ML models, we briefly overview a few representative graph representation ML models for molecular property prediction, which differ in the model complexity in incorporating the 3D structure information into the message passing (more information can be found in SI Part 5). SchNet proposed by Schütt et al.[51] is the first graph representation ML model that only utilizes the pairwise distance (two-body term) in the message function U_m and residual connections in the update functions U_t for updating atomic representations (also see Figure 1 for U_m and U_t). The U_m in SchNet operates on distance (two-body term) between atoms with learnable parameters. Based on SchNet architecture, DimeNet++[27] integrates three-body information into the model by encoding angle information using Bessel and spherical harmonic basis functions, and the message function U_m in DimeNet++ operates on angles (three-body term) and distance (two-body term) between atoms with learnable parameters. One step further, SphereNet[28] incorporates torsion angle information (four-body term) into the message function U_m using a spherical message passing framework, which enhances the 3D representation of molecules. Recently, Gastegger et al.[31] proposed the Polarizable atom Interaction Neural Network (PaiNN) that divides message passing into scalar information and vector information paths. In the vector information path, equivariant representations based on spherical functions are first introduced to improve the description of the many-body interactions. Later, the Equivariant transformer (ET)[52] and Equiformer[43] modified the equivariant network architecture by adding the attention mechanism in the message function U_m, i.e., the pairwise attention, which implements a set of learnable parameters to control the passing of vector and scalar messages based on the equivariant representation.

Table 2 summarizes the performance of the above-mentioned mainstream graph representation ML models, including SchNet,[51] DimeNet++,[27] SphereNet,[28] PaiNN,[31] ET,[52] and Equiformer[43] from the literature, which are compared with our MBNN-att results. From the table, three general trends for the graph representation ML models can be summarized, which can help to build further understanding of the performance of MBNN-att. First, the incorporation of more complex 3D information into the message function U_m, i.e., from SchNet to DimeNet++ and SphereNet, results in better performance in predicting molecular properties, as also found in the literature.[53−55] Taking the property U_0 as an example, the MAE for SchNet, which relies only on the two-body distance to represent a structure, is the highest, being 14 meV. With three-body information, DimeNet++ reduces the MAE to 6.32 meV, and SphereNet further reaches 6.26 meV after adding the four-body information. Second, the addition of the equivariant frameworks, including PAINN, ET, and Equiformer, incorporate effectively the many-body information in message passing, showing the performance being close to, if not better than, those without equivariant representations but with four-body information (SphereNet).[31,56] Third, attention appears to be important for intensive properties including ε_{HOMO}, ε_{LUMO}, and $\Delta\varepsilon$.[43,57] The lowest MAE occurs at ET (20.3 meV) and Equiformer (15.0 meV), and both have the attention mechanism.

Compared with graph representation ML models, our MBNN-att achieves the best performance in predicting four energy properties: U_0, U, H, and G. This is encouraging since the atomic simulation using ML potential is one of the most applications of ML in chemistry and can be conveniently implemented on commonly available CPU hardware using the MBNN-att framework. For the other properties, MBNN-att consistently outperforms SchNet and is comparable with other advanced GPU models. We note that MBNN-att performs relatively poorly in the intensive properties, ε_{HOMO}, ε_{LUMO}, and $\Delta\varepsilon$. This suggests that the intensive properties, ε_{HOMO}, ε_{LUMO}, and $\Delta\varepsilon$, can be better described by the pairwise-based message passing block U_m, instead of the atom-based architecture in MBNN-att. This finding is also consistent with the exceptional performance of ET and Equiformer in predicting the intensive properties, where the attention mechanism is further added in the messaging passing block U_m.

https://doi.org/10.1021/acs.jctc.4c00660
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Finally, we also compared the transferability of MBNN-att with the graph representation of ML models. Similar to Section 3.2 (Figure 4), we trained the models on molecular data sets with fewer than 20 atoms utilized to predict U_0 and C_v of molecules with more than 19 atoms. Figure 6 shows the U_0 and

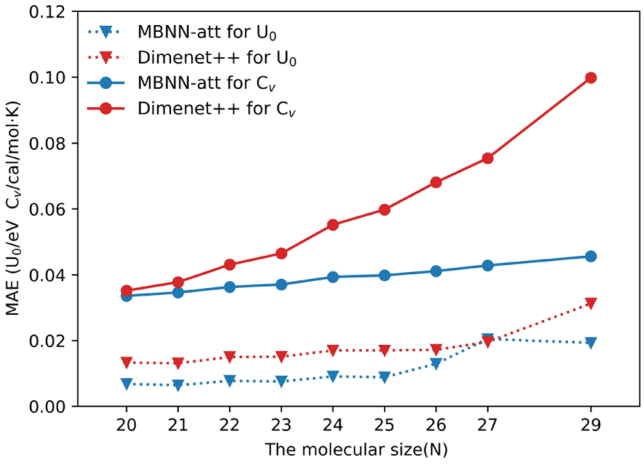


Figure 6. Transferability study of MBNN-att and Dimenet++ on two molecular properties of the QM9 data set, where the models are trained with $N < 20$ and compared for predicting larger molecules' properties.

C_v performance for MBNN-att and DimeNet++ (other MPNN models such as PAINN and SphereNet are shown in Figure S4). We found that MBNN-att exhibits smaller errors for the U_0 and C_v properties of molecules than DimeNet++ does. For the intensive C_v property, as the number of atoms increases from 20 to 29, the MAE predicted by DimeNet++ increases sharply from 0.035 to 0.099, while the MAE predicted by MBNN-att is quite steady, from 0.033 to 0.045. For the extensive U_0 property, both models perform quite well, although the MAE of MBNN-att is still consistently smaller than that of DimeNet++.

Our results (Figures 4, 6, and SI Figure S5) demonstrate that MBNN-att performs well in predicting the properties of large-sized molecules that are not present in the training data set. This could be mainly attributed to the many-body functions to directly capture the long-range contributions of molecular properties, as discussed in Section 3.2. It should be emphasized that the MPNN employed in graph networks generally has a small cutoff range (5 Å) and thus the long-range interaction can only be learned after stacking multiple message-passing layers with a considerably large number of fitting parameters. This suggests the long-range interaction in graph representation models is learned mainly through the numerical ML framework, which does not extrapolate well for the data sets not included in the training set.

4. CONCLUSIONS

In summary, this work develops the MBNN-att ML model by adding a multihead self-attention atomic attention mechanism in MBNN architecture, which achieves low-cost and general-purpose molecular property prediction. In MBNN-att, the molecular property is obtained by the sum of the product between each many-body function and the associated atomic attention weight. MBNN-att is implemented in LASP code and exhibits linear scaling efficiency on a CPU device.

MBNN-att shows the best performance on the QM9 data set compared to the representative explicit-function descriptor ML models. The ablation studies indicate that both the many-body functions and the attention mechanism are important to molecular property prediction. Many-body functions can capture the intricate atom−atom interactions, including the long-range interactions, leading to better extrapolation ability for structures outside the training set. Atomic attention is found to be important for intensive properties, where atomic attention can rank the importance of the individual atom to the molecular property.

By comparing with mainstream graph representation ML models, we show that MBNN-att remains at the top in predicting the energy properties, which demonstrates the great promise of MBNN-att for accurate atomic simulations in complex PES problems. Graph representation ML models with attention mechanisms show the best performance in the intensive properties, demonstrating the key importance of pairwise operation and the attention mechanism. Our results show that atom-based MBNN-att performs well for extensive properties, and is also good enough for intensive properties.

■ ASSOCIATED CONTENT

Data Availability Statement

The data that support the findings of this study are publicly available. The benchmark data set QM9 was obtained from the Web site http://quantum-machine.org/datasets/. The GitHub repository https://github.com/yzx1999/MBNN-att-for-Predicting-Molecular-Properties/tree/main/mbnn/ contains the used train and test data sets and the data sets in LASP training data format.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c00660.

Computational formula for prediction of properties related to the spatial distribution of electrons; ablation studies for many-body functions and attention mechanism; coefficient of determination (R^2) of MBNN-att on QM9 test data set; transferability comparison of MBNN with the GPU model; and XYZ coordinate of the 96-atom Ru bulk (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Cheng Shang − *Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China; Shanghai Qi Zhi Institution, Shanghai 200030, China;* orcid.org/0000-0001-7486-1514; Email: cshang@fudan.edu.cn

Zhi-Pan Liu − *Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China; Key Laboratory of Synthetic and Self-Assembly Chemistry for Organic Functional Molecules, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China; Shanghai Qi Zhi Institution, Shanghai 200030, China;* orcid.org/0000-0002-2906-5217; Email: zpliu@fudan.edu.cn

Authors

Zheng-Xin Yang − *Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China*

Xin-Tian Xie − *Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China;* orcid.org/0009-0009-5166-2825

Pei-Lin Kang − *Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China;* orcid.org/0000-0003-2147-2472

Zhen-Xiong Wang − *Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c00660

Notes

The authors declare no competing financial interest.

The code that was used to generate the results in this study is publicly available at https://github.com/yzx1999/MBNN-att-for-Predicting-Molecular-Properties. The code was run with Python with version 3.9.16, Pytorch with version 2.0.0, Pytorch Geometric with version 2.3.1, DScribe with version 2.0.1, and LASP (available at https://github.com/yzx1999/MBNN-att-for-Predicting-Molecular-Properties).

■ ACKNOWLEDGMENTS

This work was supported by the National Science Foundation of China (12188101, 22033003, 22122301, 91945301, 91745201), the Fundamental Research Funds for the Central Universities (20720220011), the National Key Research and Development Program of China (2018YFA0208600), and the Tencent Foundation for XPLORER PRIZE. The computation of this research was performed using the CFFF platform of Fudan University.

■ REFERENCES

(1) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121* (16), 10142−10186.

(2) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547−555.

(3) Fedik, N.; Zubatyuk, R.; Kulichenko, M.; Lubbers, N.; Smith, J. S.; Nebgen, B.; Messerly, R.; Li, Y. W.; Boldyrev, A. I.; Barros, K.; et al. Extending Machine Learning beyond Interatomic Potentials for Predicting Molecular Properties. *Nat. Rev. Chem.* **2022**, *6* (9), 653−672.

(4) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31* (46), No. 1902765.

(5) Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; Wang, H. Geometry-Enhanced Molecular Representation Learning for Property Prediction. *Nat. Mach. Intell.* **2022**, *4* (2), 127−134.

(6) Röcken, S.; Zavadlav, J. Accurate Machine Learning Force Fields via Experimental and Simulation Data Fusion. *npj Comput. Mater.* **2024**, *10* (1), No. 69, DOI: 10.1038/s41524-024-01251-4.

(7) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), No. 146401.

(8) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192−3203.

(9) Huang, S.-D.; Shang, C.; Kang, P.-L.; Liu, Z.-P. Atomic Structure of Boron Resolved Using Machine Learning and Global Sampling. *Chem. Sci.* **2018**, *9* (46), 8644−8655.

(10) Huang, S.-D.; Shang, C.; Zhang, X.-J.; Liu, Z.-P. Material Discovery by Combining Stochastic Surface Walking Global Optimization with a Neural Network. *Chem. Sci.* **2017**, *8* (9), 6327−6337.

(11) Xie, Z.; Bowman, J. M. Permutationally Invariant Polynomial Basis for Molecular Energy Surface Fitting via Monomial Symmetrization. *J. Chem. Theory Comput.* **2010**, *6* (1), 26−34.

(12) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104* (13), No. 136403.

(13) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), No. 143001.

(14) Zhang, Y.; Hu, C.; Jiang, B. Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation. *J. Phys. Chem. Lett.* **2019**, *10* (17), 4962−4967.

(15) von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring Chemical Compound Space with Quantum-Based Machine Learning. *Nat. Rev. Chem.* **2020**, *4* (7), 347−358.

(16) Drautz, R. Atomic Cluster Expansion for Accurate and Transferable Interatomic Potentials. *Phys. Rev. B* **2019**, *99* (1), No. 014104.

(17) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. Systematic AB Initio Gradient Calculation of Molecular Geometries, Force Constants, and Dipole Moment Derivatives. *J. Am. Chem. Soc.* **1979**, *101*, 2550−2560.

(18) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87* (18), No. 184115.

(19) Huo, H.; Rupp, M. Unified Representation of Molecules and Crystals for Machine Learning. *Mach. Learn.: Sci. Technol.* **2022**, *3* (4), No. 045017.

(20) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. In *Neural Message Passing for Quantum Chemistry*, International conference on machine learning; PMLR, 2017; pp 1263−1272.

(21) He, K.; Zhang, X.; Ren, S.; Sun, J. In *Deep Residual Learning for Image Recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; IEEE, 2016; pp 770−778.

(22) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. *Nat. Commun.* **2023**, *14* (1), No. 579.

(23) Atz, K.; Grisoni, F.; Schneider, G. Geometric Deep Learning on Molecular Representations. *Nat. Mach. Intell.* **2021**, *3* (12), 1023−1032.

(24) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.* **2019**, *5* (8), No. eaav6490.

(25) Zhang, Y.; Xia, J.; Jiang, B. Physically Motivated Recursively Embedded Atom Neural Networks: Incorporating Local Completeness and Nonlocality. *Phys. Rev. Lett.* **2021**, *127* (15), No. 156002.

(26) Zhang, Y.; Xia, J.; Jiang, B. REANN: A PyTorch-Based End-to-End Multi-Functional Deep Neural Network Package for Molecular,

Reactive, and Periodic Systems. *J. Chem. Phys.* **2022**, *156* (11), No. 114801, DOI: 10.1063/5.0080766.
(27) Gasteiger, J.; Groß, J.; Günnemann, S. In *Directional Message Passing for Molecular Graphs*, International Conference on Learning Representations; ICLR, 2020.
(28) Liu, Y.; Wang, L.; Liu, M.; Zhang, X.; Oztekin, B.; Ji, S. In *Spherical Message Passing for 3D Molecular Graphs*, International Conference on Learning Representations; ICLR, 2021.
(29) Wang, Y.; Wang, T.; Li, S.; He, X.; Li, M.; Wang, Z.; Zheng, N.; Shao, B.; Liu, T.-Y. Enhancing Geometric Representations for Molecules with Equivariant Vector-Scalar Interactive Message Passing. *Nat. Commun.* **2024**, *15* (1), No. 313.
(30) Kovács, D. P.; Batatia, I.; Arany, E. S.; Csanyi, G. Evaluation of the MACE Force Field Architecture: From Medicinal Chemistry to Materials Science. *J. Chem. Phys.* **2023**, *159* (4), No. 044118, DOI: 10.1063/5.0155322.
(31) Schütt, K. T.; Unke, O. T.; Gastegger, M. In *Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra*, International Conference on Machine Learning; PMLR, 2021.
(32) Bircher, M. P.; Singraber, A.; Dellago, C. Improved Description of Atomic Environments Using Low-Cost Polynomial Functions with Compact Support. *Mach. Learn.: Sci. Technol.* **2021**, *2* (3), No. 035026.
(33) Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. Comparison of Permutationally Invariant Polynomials, Neural Networks, and Gaussian Approximation Potentials in Representing Water Interactions through Many-Body Expansions. *J. Chem. Phys.* **2018**, *148* (24), No. 241725.
(34) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3* (12), No. e1701816, DOI: 10.1126/sciadv.1701816.
(35) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A Fourth-Generation High-Dimensional Neural Network Potential with Accurate Electrostatics Including Non-Local Charge Transfer. *Nat. Commun.* **2021**, *12* (1), No. 398.
(36) Shaidu, Y.; Pellegrini, F.; Küçükbenli, E.; Lot, R.; de Gironcoli, S. Incorporating Long-Range Electrostatics in Neural Network Potentials via Variational Charge Equilibration from Shortsighted Ingredients. *npj Comput. Mater.* **2024**, *10* (1), No. 47, DOI: 10.1038/s41524-024-01225-6.
(37) Kang, P.-L.; Yang, Z.-X.; Shang, C.; Liu, Z.-P. Global Neural Network Potential with Explicit Many-Body Functions for Improved Descriptions of Complex Potential Energy Surface. *J. Chem. Theory Comput.* **2023**, *19*, 7972−7981, DOI: 10.1021/acs.jctc.3c00873.
(38) Kang, P.-l.; Shang, C.; Liu, Z. Recent Implementations in LASP 3.0: Global Neural Network Potential with Multiple Elements and Better Long-Range Description. *Chin. J. Chem. Phys.* **2021**, *34* (5), 583−590.
(39) Huang, S.; Shang, C.; Kang, P.; Zhang, X.; Liu, Z. LASP: Fast Global Potential Energy Surface Exploration. *WIREs Comput. Mol. Sci.* **2019**, *9* (6), No. e1415, DOI: 10.1002/wcms.1415.
(40) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds.; Curran Associates, Inc., 2017; Vol. *30*.
(41) Bahdanau, D.; Cho, K. H.; Bengio, Y. In *Neural Machine Translation by Jointly Learning to Align and Translate*, 3rd International Conference on Learning Representations; ICLR, 2015.
(42) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. In *Graph Attention Networks*, International Conference on Learning Representations; ICLR, 2018.
(43) Liao, Y.-L.; Smidt, T. In *Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs*, Eleventh International Conference on Learning Representations; ICLR, 2023.
(44) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1* (1), No. 140022.
(45) Liu, D. C.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Program.* **1989**, *45* (1−3), 503−528.
(46) Laakso, J.; Himanen, L.; Homm, H.; Morooka, E. V.; Jäger, M. O.; Todorović, M.; Rinke, P. Updates to the DScribe Library: New Descriptors and Derivatives. *J. Chem. Phys.* **2023**, *158* (23), No. 234802, DOI: 10.1063/5.0151031.
(47) Profitt, T. A.; Pearson, J. K. A Shared-Weight Neural Network Architecture for Predicting Molecular Properties. *Phys. Chem. Chem. Phys.* **2019**, *21* (47), 26175−26183.
(48) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; von Lilienfeld, O. A. Machine Learning Prediction Errors Better than DFT Accuracy. 2017, arXiv:1702.05532. arXiv.org e-Printarchive. https://arxiv.org/abs/1702.05532.
(49) Langer, M. F.; Goeßmann, A.; Rupp, M. Representations of Molecules and Materials for Interpolation of Quantum-Mechanical Simulations via Machine Learning. *npj Comput. Mater.* **2022**, *8* (1), No. 41, DOI: 10.1038/s41524-022-00721-x.
(50) Weinhold, F. Natural Bond Orbital Analysis: A Critical Overview of Relationships to Alternative Bonding Perspectives. *J. Comput. Chem.* **2012**, *33* (30), 2363−2379.
(51) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet − A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148* (24), No. 241722.
(52) Thölke, P.; De Fabritiis, G. In *Equivariant Transformers for Neural Network Based Molecular Potentials*, International Conference on Learning Representations; ICLR, 2021.
(53) Gasteiger, J.; Becker, F.; Günnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. In *Advances in Neural Information Processing Systems*; Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P. S.; Vaughan, J. W., Eds.; Curran Associates, Inc., 2021; Vol. *34*, pp 6790−6802.
(54) Joshi, C. K.; Bodnar, C.; Mathis, S. V.; Cohen, T.; Lio, P. In *On the Expressive Power of Geometric Graph Neural Networks*, International Conference on Machine Learning; PMLR, 2023; pp 15330−15355.
(55) Wang, L.; Liu, Y.; Lin, Y.; Liu, H.; Ji, S. In *ComENet: Towards Complete and Efficient Message Passing for 3D Molecular Graphs*, Advances in Neural Information Processing Systems; NeurIPS, 2022; pp 650−664.
(56) Passaro, S.; Zitnick, C. L. In *Reducing so (3) Convolutions to so (2) for Efficient Equivariant Gnns*, International Conference on Machine Learning; PMLR, 2023; pp 27420−27438.
(57) Pelaez, R. P.; Simeon, G.; Galvelis, R.; Mirarchi, A.; Eastman, P.; Doerr, S.; Thölke, P.; Markland, T. E.; De Fabritiis, G. TorchMD-Net 2.0: Fast Neural Network Potentials for Molecular Simulations. *J. Chem. Theory Comput.* **2024**, *20*, 4076−4087, DOI: 10.1021/acs.jctc.4c00253.