Supplementary Materials for A Neural Network Potential with Rigorous Treatment of Long-Range Dispersion

Nguyen Thien Phuc Tu[2], Nazanin Rezajooei[1], Erin R. Johnson[3] and Christopher Rowley[1,2*]

[1*]Department of Chemistry, Carleton University, 1125 Colonel By Dr, Ottawa, K1S 5B6, Ontario, Canada.
[2*]Interdisciplinary Program in Scientific Computing, Memorial University of Newfoundland, 230 Elizabeth Ave, St. John's, A1C 5S7, Newfoundland and Labrador, Canada.
[3]Department of Chemistry, Dalhousie University, 6274 Coburg Rd, Halifax, B3H 4R2, Nova Scotia, Canada.

*Corresponding author(s). E-mail(s): Christopher.Rowley@carleton.ca;
Contributing authors: phuctu@cmail.carleton.ca; nrezajooei@mun.ca; erin.johnson@dal.ca;

1 Computational Methods

In the ANIPBE-MLXDM method, the total potential energy of the system ($\mathcal{V}$) with atomic coordinates $\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_N$ is the sum of the ANIPBE0 neural network potential term ($\mathcal{V}_{\mathrm{NNP}}$) and the MLXDM London dispersion term ($\mathcal{V}_{\mathrm{disp}}$). The NNP is intended to describe short-range interatomic interactions (e.g., Pauli repulsion, chemical bonds...), while the MLXDM term is intended to describe London dispersion interactions.

The total dispersion energy of a chemical system can be approximated as a sum of the dispersion interactions between individual atomic pairs,

$$\mathcal{V}_{\mathrm{disp}} = \sum_{i} \sum_{j=i+1} \mathcal{V}_{\mathrm{disp},ij}\left(R_{ij}\right) \tag{1}$$

The dispersion energy of the interaction between neutral atoms i and j ($\mathcal{V}_{\mathrm{disp},ij}$) can be approximated as a sum of 6^{th}, 8^{th}, and 10^{th} order terms [1]:

$$\mathcal{V}_{\mathrm{disp},ij}(R_{ij}) = -\frac{C_{6,ij}}{R_{ij}^6} - \frac{C_{8,ij}}{R_{ij}^8} - \frac{C_{10,ij}}{R_{ij}^{10}} \tag{2}$$

where $C_{6,ij}$, $C_{8,ij}$, and $C_{10,ij}$ are the 6^{th}, 8^{th}, and 10^{th} order dispersion coefficients for the atomic pair i and j. R_{ij} is the distance between atomic pair i and j.

When used with conventional DFT functionals for the calculation of the other components of the energy (or, in this case, an NNP approximating DFT), these dispersion terms must be damped at short-range to avoid "double-counting" short-range dispersion interactions that are captured by the exchange-correlation functional and to avoid a singularity as $R_{ij} \to 0$. In the XDM model, this is achieved through an order-dependent damping function, $f_n(R_{ij})$, that depends on the sum of van der Waals radii, $R_{\mathrm{vdW},ij}$ of the interacting atoms. The damping function is defined as

$$f_n(R_{ij}) = \frac{R_{ij}^n}{R_{ij}^n + R_{\mathrm{vdW},ij}^n} \tag{3}$$

With these damping functions included, Eqn. 2 becomes

$$\mathcal{V}_{\mathrm{disp},ij}(R_{ij}) = -\frac{C_{6,ij}}{R_{ij}^6} f_6(R_{ij}) - \frac{C_{8,ij}}{R_{ij}^8} f_8(R_{ij}) - \frac{C_{10,ij}}{R_{ij}^{10}} f_{10}(R_{ij}). \tag{4}$$

With the XDM model, the coefficients are related to atomic polarizabilities and atomic electric moments. The C_6 dispersion coefficient for the interaction between atoms i and j is

$$C_{6,ij} = \alpha_i \alpha_j \frac{\langle M_1^2 \rangle_i \langle M_1^2 \rangle_j}{\alpha_i \langle M_1^2 \rangle_j + \alpha_j \langle M_1^2 \rangle_i} \tag{5}$$

where $\langle M_1^2 \rangle_i$ is the expectation value of the square of the exchange-hole dipole moment for atom i, and α_i is the the atom-in-molecule polarizability,

$$\alpha_i = \frac{V_i}{V_{\mathrm{free},i}} \alpha_{\mathrm{free},i} \tag{6}$$

which is obtained using the proportionality between polarizability and atomic volume. Here $V_{\mathrm{free},i}$, and $\alpha_{\mathrm{free},i}$ are the *in vacuo* atomic volumes and polarizabilities. Similar relations provide the C_8 and C_{10} dispersion coefficients:

$$C_{8,ij} = \frac{3}{2} \alpha_i \alpha_j \frac{\langle M_1^2 \rangle_i \langle M_2^2 \rangle_j + \langle M_2^2 \rangle_i \langle M_1^2 \rangle_j}{\alpha_j \langle M_1^2 \rangle_i + \alpha_i \langle M_1^2 \rangle_j} \tag{7}$$

$$C_{10,ij} = \alpha_i \alpha_j \frac{2\langle M_1^2\rangle_i \langle M_3^2\rangle_j + 2\langle M_3^2\rangle_i \langle M_1^2\rangle_j + \frac{21}{5}\langle M_2^2\rangle_i \langle M_2^2\rangle_j}{\alpha_j \langle M_1^2\rangle_i + \alpha_i \langle M_1^2\rangle_j} \tag{8}$$

where $\langle M_2^2\rangle_i$ and $\langle M_3^2\rangle_i$ are expectation values involving the square of the exchange-hole quadrupole and octupole moments, respectively. The XDM dispersion coefficients differ from free-atom values depending on the local chemical environment because of their dependency on the electron density and its derivatives via the exchange hole [2, 3].

Finally, the van der Waals radii that appear in the damping function are

$$R_{\mathrm{vdW},ij} = a_1 R_{\mathrm{critical},ij} + a_2 \tag{9}$$

and depend on the critical radius, $R_{\mathrm{critical},ij}$, at which successive terms in the dispersion energy expansion become equal. This is determined by averaging ratios of the dispersion coefficients as

$$R_{\mathrm{critical},ij} = \frac{1}{3}\left[\left(\frac{C_{8,ij}}{C_{6,ij}}\right)^{1/2} + \left(\frac{C_{10,ij}}{C_{6,ij}}\right)^{1/4} + \left(\frac{C_{10,ij}}{C_{8,ij}}\right)^{1/2}\right]. \tag{10}$$

The values of the two constants, a_1 and a_2, are determined by DFT calculations by minimizing the root-mean-square percent error for a set of 49 molecular dimers relative to accurate reference data from wavefunction theory.[4]. These damping parameters are fit for use with a given exchange-correlation functional and basis set; their values are then kept fixed for all subsequent calculations. MLXDM uses the same damping coefficients as PBE0/aug-cc-pVTZ: $a_1 = 0.4186$ and $a_2 = 2.6791$ Å. [4]

1.1 Generation of Training Data

The dataset used to build the ANIPBE0 NNP and MLXDM NNs consisted of a total of 2.12 M configurations of chemical complexes and their corresponding DFT-calculated electronic energies and dispersion coefficients. This dataset is available for download from the FigShare repository [5]. The DFT calculations were performed using Gaussian 16 [6] with the PBE0 exchange-correlation functional [7] and the aug-cc-pVTZ basis set [8]. These calculations provide the electronic energy and XDM coefficients that are used to train the ANIPBE0 NNP and MLXDM correction, respectively. The route card of the Gaussian calculations was:

```
#PBE1PBE/aug-cc-pvtz int(grid=ultrafine) scf=tight force
```

The XDM dispersion coefficients were calculated from the wavefunction data from the PBE0/aug-cc-pVTZ calculations using the postg code [9].

The first set of 1.83 M chemical structures chosen to comprise the dataset was a subset of the ANI-1x dataset [10]. However, the NNP trained using that

Table 1 Number of systems, N, included in the training set for the various NNPs. Also shown are the resulting error statistics for the DES370K benchmark of intermolecular binding energies, in kcal/mol. MAE: mean absolute error; RMSE: root-mean-square error; R^2: coefficient of determination.

	N	MAE	RMSE	R^2
ANIPBE0a (ANI-1X data only)	1.83 M	1.20	4.66	0.59
ANIPBE0b (intermolecular data included)	2.10 M	0.86	4.08	0.68
ANIPBE0 (active learning data included)	2.12 M	0.67	1.06	0.97

dataset (ANIPBE0a) performed poorly when applied to the DES370K benchmark [11] of intermolecular interaction energies calculated using PBE0/aug-cc-pVTZ; the MAE was 1.2 kcal/mol for the binding energies the R^2 was 0.58 (Table 1). We surmised that the poor performance of ANIPBE0a was due to the lack of intermolecular complexes in the dataset. Thus, additional data for intermolecular complexes were generated by performing calculations on pairs of molecules with a separation corresponding to a random distance in the interval 2 Å to 5 Å and placed at random angles with respect to each other. An additional 0.27 M structures were added to the dataset through this process. The inclusion of these data improved the MAE for the DES370K benchmark to 0.86 kcal/mol and the R^2 of the NNP to 0.68 (ANIPBE0b, Table 1).

To increase the accuracy further, an active learning process was applied. Eight NNPs were trained for each iteration of the training set. Randomly-oriented intermolecular complexes were generated and the energies of these complexes were calculated with each NNP. If the standard deviation of this ensemble of NNPs was greater than 1 kcal/mol per atom, the DFT energy of the structure was calculated and added to the dataset. This method was based on the active learning scheme used to generate the ANI-1x dataset.[10] This process was repeated for 46 iterations, adding an additional 20,000 structures to the dataset. This process improved the R^2 of the NNP to 0.97 and decreased the MAE to 0.67 kcal/mol for the DES370K benchmark. This model was used for the NNP component (ANIPBE0) in all calculations going forward.

1.2 Construction of the Neural Network Potential

ANI-type NNPs, such as ANIPBE0, define the total potential energy of the system (E_T) as the sum of atomic energies (E_i),[1]

$$E_T = \sum_{i}^{N_{\text{atom}}} E_i \tag{11}$$

The NNP uses a query by committee approach [12, 13], where the potential energy is the average of eight energies calculated using different trained neural

[1]Here, we follow the ANI convention where the total energy calculated by the NNP is denoted as E_T. This term is equivalent to $\mathcal{V}_{\text{NNP}}$.

networks.

$$\langle E_T \rangle = \frac{1}{N_{\text{ensemble}}} \sum_{i}^{N_{\text{ensemble}}} E_{T,i} \tag{12}$$

1.3 Input Features

The same input features and network architecture are used for both the ANIPBE0 NNP and the MLXDM dispersion coefficient NNs. The Atomic Environment Vector (AEV) of an atom contains the input features used by the NNs to calculate the atomic energy or dispersion coefficients. These AEVs are composed of modified Behler–Parrinello symmetry functions (G). The AEV contains both radial (G^R, Eqn. 13) and angular symmetry functions ($G^{A_{\text{mod}}}$, Eqn. 14) corresponding to the chemical environment of an atom. The parameters of these functions are collected in Table 2. A full discussion of the construction of AEVs is available in Ref. 14. The AEVs used in this work are identical to those used in Ref. 15.

$$G_m^R = \frac{1}{4} \sum_{i \neq j}^{N} \exp\left(-\eta \left(R_{ij} - R_{R,s}\right)^2\right) f_c\left(R_{ij}\right) \tag{13}$$

$$G_m^{A_{\text{mod}}} = 2^{1-\zeta} \sum_{j,k \neq i} \left(1 + \cos\left(\theta_{ijk} - \theta_s\right)\right)^{\zeta} \times \exp\left[-\eta \left(\frac{R_{ij} + R_{ik}}{2} - R_{\theta,s}\right)^2\right] f_c\left(R_{ij}\right) f_c\left(R_{ik}\right) \tag{14}$$

$$f_c\left(R_{ij}\right) = \begin{cases} \frac{1}{2}\left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right] & \text{if } R_{ij} \leq R_c \\ 0 & \text{if } R_{ij} > R_c \end{cases} \tag{15}$$

1.4 Standardization of Values

Extraneous values were removed from the training data by calculating the means and standard deviation of $\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and V of each element. Values deviating from the mean by more than 1.5 times the standard deviation of the distribution (σ) were removed. These remaining values were standardized by the z-score method:

$$\langle M_\ell^2 \rangle' = \frac{\langle \bar{M}_\ell^2 \rangle - \langle M_\ell^2 \rangle}{\sigma_{\langle M_\ell^2 \rangle}}. \tag{16}$$

The cleaned and standardized distributions of these properties for each element are presented in Figure 4. The averages and standard deviations of for each coefficient and element are collected in Table 3.

Table 2 Parameters for the AEVs used as input features for the ANIPBE0 and MLXDM NNs.

Parameter	Value	Parameter	Value	Parameter	Value
ζ	32	θ_s	0.19634954	$R_{R,s}$	0.9
R_{cr}	5.2		0.58904862		1.16875
R_{ca}	3.5		0.98174770		1.43750
η_R	1.6		1.3744468		1.70625
η_θ	8		1.7671459		1.97500
R_θ	0.9		2.1598449		2.24375
	1.55		2.5525440		2.51250
	2.2		2.9452431		2.78125
	2.85				3.05000
					3.31875
					3.58750
					3.85625
					4.12500
					4.39375
					4.66250
	.				4.93125

Table 3 Average values and standard deviations, σ, for z-score standardization of selected quantities (in atomic units) for each element. These values were calculated for the ANIPBE0 dataset.

Coefficient	Element	Average	σ
$\langle M_1^2 \rangle$	H	1.539830	0.107733
	C	4.299883	0.341336
	N	4.643987	0.525829
	O	4.846803	0.405748
$\langle M_2^2 \rangle$	H	12.464916	1.067613
	C	54.584561	4.251465
	N	45.167788	5.537361
	O	37.321560	3.784493
$\langle M_3^2 \rangle$	H	206.44866	23.97470
	C	981.82437	100.44080
	N	594.34827	74.91799
	O	380.99753	43.10249
V	H	5.990650	0.3407511
	C	31.521044	1.0016140
	N	26.222855	1.0468942
	O	21.819846	0.8000782

The ANIPBE0-MLXDM NNP was implemented and trained using TorchANI.[15] This package provides an interface between chemical data and the PyTorch [16] machine learning library. The gradients of both the ANIPBE0 NNP and the MLXDM dispersion correction are generated by PyTorch's auto-differentiation. As a consequence, the dispersion coefficients change dynamically over the course of a molecular dynamics (MD) simulation. Further, the sum of kinetic and potential energies is conserved in microcanonical MD simulations without iterative self-convergence steps.

1.5 NN Architecture

The ANIPBE0 and MLXDM NNs both had the same network architecture as the ANI-1x NN [13], and each NN had 3 hidden layers. The number of nodes in each layer is tabulated in Table 4. Continuously Differentiable Exponential Linear Units (CELU) [17] activation functions were used in all nodes. The activation function weights were initialized with random values from a Gaussian distribution, while the biases were initialized to zero.

Table 4 The number of nodes in each level of the ANIPBE0 and MLXDM deep neural networks.

Element	Network Architecture
H	384:160:128:96:1
C	384:144:112:96:1
N	384:128:112:96:1
O	384:128:112:96:1

For the ANIPBE0 NNP, one NN was defined for each of the elements H, C, N, and O. There were 8 NNs in the ensemble, so a total of 32 NNs were trained. For the MLXDM NNs, one NN was defined for each combination of the elements H, C, N, and O with the coefficients $\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and V giving a total of 16 NNs. As the metrics for the accuracy of the MLXDM NNs were generally very good, a query by committee approach was not necessary. A flowchart of the ANIPBE0-MLXDM code is presented in Figure 1.

1.6 Pseudo-code for Computing Dispersion Energy

The pseudo-code below demonstrates the computation of dispersion energy from the atomic types and positions of the system. Detailed mathematical expression for each type are listed above. The gradient is automatically computed using **autograd** of PyTorch.

1.7 NN Training

The dataset was divided randomly into training and validation sets in a 4:1 ratio. The training of the NNP follows the work of Gao *et al.*,[15] where the loss function is a sum of the square deviation of the NNP-predicted energy and norm of the forces for every molecule (or intermolecular complex) in the validation set.

$$\delta_{\mathrm{NNP}} = \sum_{i}^{N_{\mathrm{molecules}}} \frac{(E_{i,\mathrm{NNP}} - E_{i,\mathrm{DFT}})^2}{\sqrt{N_{\mathrm{atoms}}}} + 0.1 \times \sum_{i}^{N_{\mathrm{molecules}}} \frac{(\vec{F}_{i,\mathrm{NNP}} - \vec{F}_{i,\mathrm{DFT}}) \cdot (\vec{F}_{i,\mathrm{NNP}} - \vec{F}_{i,\mathrm{DFT}})}{\sqrt{N_{\mathrm{molecules}}}} \quad (17)$$

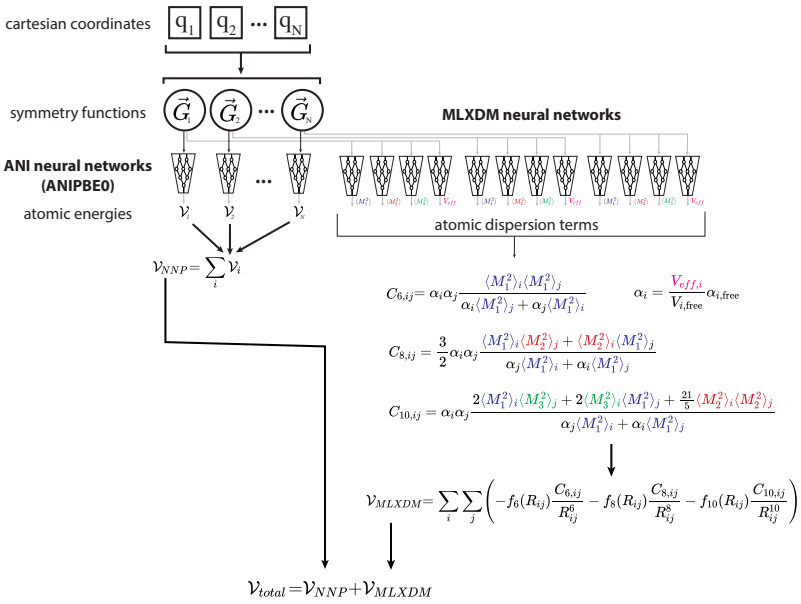

Fig. 1 Flowchart of the calculation of the total potential energy of a chemical system ($\mathcal{V}_{total}$) for a set of atomic coordinates (q_1, q_2, ..., q_N). These coordinates are used to calculate an atomic environment vector (AEV) for each atom ($\vec{G}$). These AEVs are input into NNs to calculate the atomic energy (ANIPBE0, left). The same AEVs are input into a set of three NNs to calculate the standardized $\langle M_1^2\rangle$, $\langle M_2^2\rangle$, $\langle M_3^2\rangle$, and V (MLXDM NNs, right). Both the ANIPBE0 and MLXDM NNs are element-specific (i.e., separately-trained NNs are used for atoms of the element H, C, N, and O).

The loss function for training the NN for the n^{th} order dispersion coefficient of a given element is a sum of the NN predicted coefficient ($\langle M_{n,j}^2\rangle_{\mathrm{NN}}$) of the DFT-calculated XDM coefficient for each atom of that element in the validation set.

$$\delta_{\mathrm{disp},n}=\sum_i^{N_{\mathrm{atoms}}}\frac{(\langle M_{n,i}^2\rangle_{\mathrm{NN}}-\langle M_{n,i}^2\rangle_{\mathrm{XDM}})^2}{N_{\mathrm{atoms}}} \tag{18}$$

The NNs of both the ANIPBE0 and the MLXDM correction were trained using the Adam optimizer with weight control [18, 19]. The initial learning rate coefficient was 0.001. The best model was typically found within roughly 150 epochs. An example of the change in the loss function during one of the optimizations is presented in Figure 2.

Algorithm 1 Dispersion energy computation

$\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$,
$aev \leftarrow \text{AEVComputer}(atomtype, positions)$
$R \leftarrow \text{pair\_wise\_distance}(positions)$
$\langle M_1^2 \rangle \leftarrow \text{M1\_neural\_network}(aev)$
$\langle M_2^2 \rangle \leftarrow \text{M2\_neural\_network}(aev)$
$\langle M_3^2 \rangle \leftarrow \text{M3\_neural\_network}(aev)$
$\langle V \rangle \leftarrow \text{V\_neural\_network}(aev)$
$\alpha \leftarrow \alpha_{\text{free}} \frac{V}{V_{\text{free}}}$
$C_6 \leftarrow \text{XDM\_formula}(\alpha, \langle M_1^2 \rangle)$
$C_8 \leftarrow \text{XDM\_formula}(\alpha, \langle M_1^2 \rangle, \langle M_2^2 \rangle)$
$C_{10} \leftarrow \text{XDM\_formula}(\alpha, \langle M_1^2 \rangle, \langle M_2^2 \rangle, \langle M_3^2 \rangle)$
$R_{\text{vdW}} \leftarrow \text{van\_der\_Waals}(C_6, C_8, C_{10})$
$\mathcal{V}_6 \leftarrow -\frac{C_6}{R^6} \times \text{damp\_function}(R, R_{\text{vdW}}, 6)$
$\mathcal{V}_8 \leftarrow -\frac{C_8}{R^8} \times \text{damp\_function}(R, R_{\text{vdW}}, 8)$
$\mathcal{V}_{10} \leftarrow -\frac{C_{10}}{R^{10}} \times \text{damp\_function}(d, R_{\text{vdW}}, 10)$
$\mathcal{V} \leftarrow \mathcal{V}_6 + \mathcal{V}_8 + \mathcal{V}_{10}$

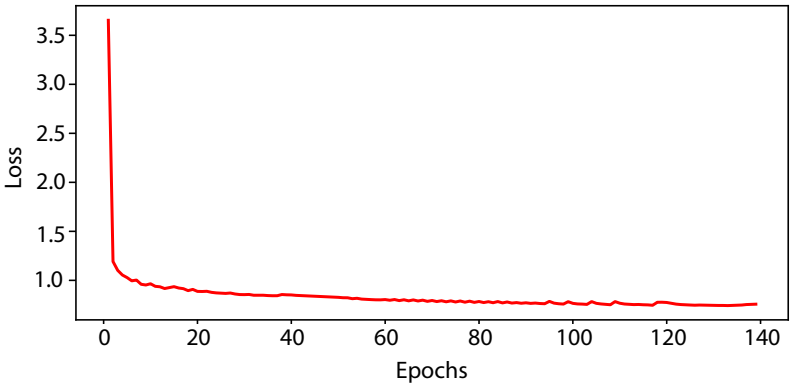

Fig. 2 Example of the training progress of the C_6 dispersion coefficient NN. The plot shows the decline of the loss function of the validation set over the epochs of the optimization.

2 Simulations

2.1 Methods

A switching function was applied to the dispersion interaction term in periodic systems so that these interactions were scaled to zero continuously over the interval $R_{\text{switch}} < R < R_{\text{dispcut}}$. In the simulations presented in this paper, this interval was 9.24 Å to 14 Å. The dispersion interaction term with the switching function included was

$$\mathcal{V}_{\mathrm{disp},ij}(R_{ij}) = -S\left(R_{ij}\right)\left(\frac{C_{6,ij}}{R_{ij}^{6}} f_6(R_{ij}) + \frac{C_{8,ij}}{R_{ij}^{8}} f_8(R_{ij}) + \frac{C_{10,ij}}{R_{ij}^{10}} f_{10}(R_{ij})\right) \tag{19}$$

$$S\left(R_{ij}\right) = \begin{cases} 1 & \text{if } R_{ij} \leq R_{\mathrm{switch}} \\ \frac{\left(R_{\mathrm{dispcut}}^2 - R_{ij}^2\right)^2 \left(R_{\mathrm{dispcut}}^2 + 2R_{ij}^2 - 3R_{\mathrm{switch}}^2\right)}{\left(R_{\mathrm{dispcut}}^2 - R_{\mathrm{switch}}^2\right)^3} & \text{if } R_{\mathrm{switch}} \leq R_{ij} \leq R_{\mathrm{dispcut}} \\ 0 & \text{if } R_{ij} > R_{\mathrm{dispcut}} \end{cases} \tag{20}$$

2.2 Technical Details

Computations for the use cases of ANIPBE0-MLXDM were performed using the Atomic Simulation Environment (ASE) [20] on a 12-core Intel(R) Core(TM) i7-8700 CPU with a Titan Xp GPU.

2.2.1 Molecular Dynamics Simulation of Liquid Toluene

A 19.3 Å × 19.3 Å × 19.3 Å simulation cell containing 36 toluene molecules was constructed. A 1 ns molecular dynamics simulation was performed using the CGenFF force field[21] to relax the structure. Beginning from this equilibrated structure, a 1 ns molecular dynamics simulation with a time step of 1 fs was performed using ANIPBE0-MLXDM. The neutron scattering structure factors were calculated using LiquidLib [22] using a set of five independent NVT trajectories.

2.2.2 Gas Adsorption in a Covalent Organic Framework

A 1 × 1 × 2 supercell was constructed from the crystallographic structure of COF-320 [23] (CCDB: 1002037) to form a 27.93 Å × 31.31 Å × 15.78 Å cell. The geometry of the host, in the absence of any gas, was optimized with a convergence criterion that the maximum norm of the atomic forces was 0.001 eV/Å. Grand Canonical Monte Carlo simulations were performed with a gas pressure of 80 bar and a temperature of 298.16 K. To place the adsorbed methane molecules, 100,000 insertion and deletion attempts were performed for ANIPBE0 and ANIPBE0-MLXDM, with 100 translation/rotation attempts in between each insertion/deletion attempt. The acceptance criteria for each action were [24]

$$P_{\mathrm{acc}}^{\mathrm{trans/rot}} = \exp\left(-\frac{\Delta\mathcal{V}}{k_B T}\right) \tag{21}$$

$$P_{\mathrm{acc}}^{\mathrm{insert}} = \frac{pV}{(N+1)k_B T} \exp\left(-\frac{\Delta\mathcal{V}}{k_B T}\right) \tag{22}$$

$$P_{\mathrm{acc}}^{\mathrm{delete}} = \frac{Nk_B T}{pV} \exp\left(-\frac{\Delta\mathcal{V}}{k_B T}\right) \tag{23}$$

where p is the pressure on the system, T is the temperature, and $\Delta\mathcal{V}$ is the change in the potential energy.

2.2.3 Timing Comparisons

To estimate the scaling of the computational cost of these simulations, we performed molecular dynamics simulations for clusters of 1–7 pentane molecules. Because the disk and memory operations to load the NNP and chemical coordinates are a significant share of the computational cost, the time required for a single energy evaluation was estimated by performing 1000 molecular dynamics steps using ANIPBE0 or ANIPBE0-MLXDM and then dividing by 1000. For the largest system examined, a cluster of 7 pentane molecules, ANIPBE0-MLXDM is 4.5 M times faster than the PBE0/aug-cc-pVTZ DFT calculation using TURBOMOLE. These benchmarks were performed using CPUs only and the largest system contained only 136 atoms. Using GPU computing and larger systems would result in an even larger advantage for ANIPBE0-MLXDM.

Table 5 CPU time in seconds for one time step of ANIPBE0 and ANIPBE0-MLXDM molecular dynamics simulation on small pentane clusters, compared to the computational cost of one energy/gradient calculation using TURBOMOLE 7.0 with the PBE0 exchange-correlation functional and the aug-cc-pVTZ basis set.

	ANIPBE0	ANIPBE0-MLXDM	PBE0/aug-cc-pVTZ
1	0.016	0.029	1782
2	0.021	0.037	8976
3	0.025	0.048	56547
4	0.029	0.056	110072
5	0.034	0.066	139996
6	0.039	0.082	270448
7	0.044	0.089	405344

The scaling of the computing time of the ANIPBE0 and ANIPBE0-MLXDM NNPs are presented in Figure 3 and Table 5. For smaller systems, ANIPBE0 is roughly four times faster than ANIPBE0-MLXDM, although the dispersion correction becomes progressively more costly in larger systems.

Although the formal scaling of ANIPBE-MLXDM is $\mathcal{O}(N^2)$ due to the calculation of the pairwise dispersion terms, this is a relatively minor component of the computational cost up to thousands of atoms. The linear-scaling calculation of the symmetry functions and evaluating the gradients of the four NNs and their gradients, are the most time-demanding steps. In very large systems, the cost of evaluating the dispersion energy could be mitigated by cutoffs and neighbor lists. The use of double-headed NNs to compute the XDM terms from a single NN or fixed coefficients for some of the less significant XDM terms could further mitigate the cost of the MLXDM calculations.

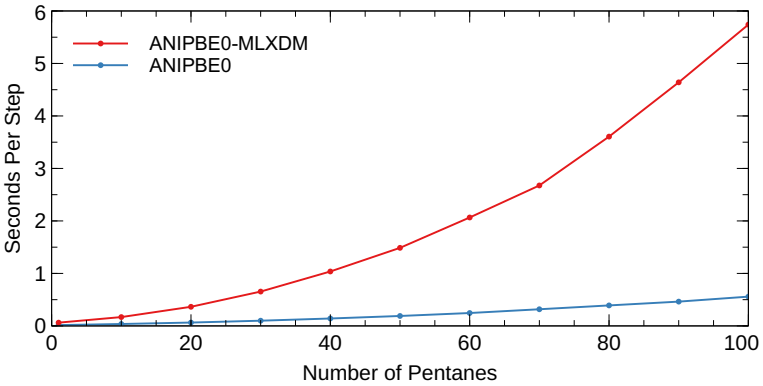

Fig. 3 CPU time in seconds needed for one time step of MD for clusters of pentane molecules.

3 Supplementary Tables and Figures

Table 6 Relative standard deviation of the terms in the MLXDM model relative to the XDM reference data for the DES370K dataset

	$\langle M_1^2 \rangle$	$\langle M_2^2 \rangle$	$\langle M_3^2 \rangle$	V
H	5.97%	11.4%	17.1%	6.47%
C	10.5%	10.3%	12.6%	3.66%
N	17.7%	18.2%	15.0%	5.46%
O	9.73%	11.3%	12.7%	4.39%

Table 7 The relative deviation in dispersion component of the interaction energies for the MLXDM and XDM-CC models (computed as $(X_{\text{MLXDM}} - X_{\text{XDM-CC}})/X_{\text{MLXDM}}$ where X is coefficients or energies), tested with 9730 randomly-selected complexes from DES370K dataset. The error in the predicted dispersion energy is smaller than the error on the individual coefficients because the error in each atomic XDM-CC term tends to cancel when the pairwise sum is performed over a number of atomic pairs.

	C_6	C_8	C_{10}
Individual coefficients	4.44%	5.66%	8.63%
Energy components	3.05%	4.12%	5.83%
Total energy		3.24%	

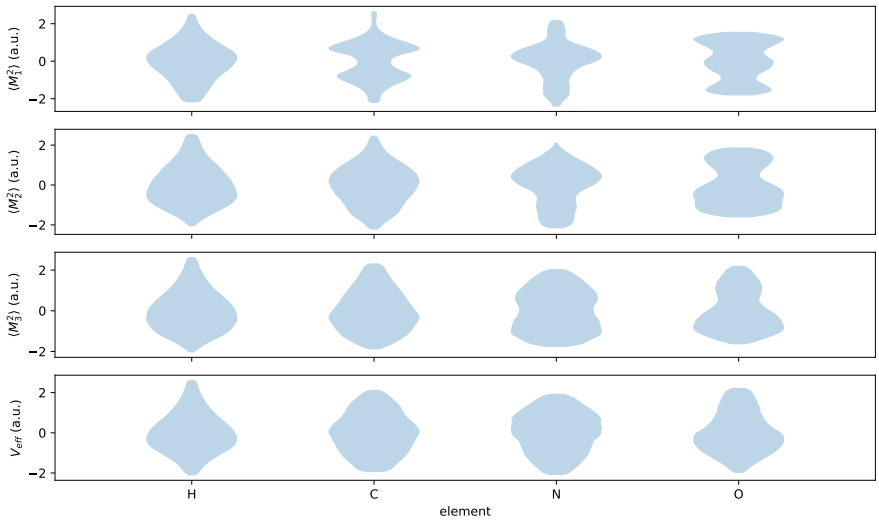

Fig. 4 Z-score Standardized distributions of the $\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$ and V terms for the training data.

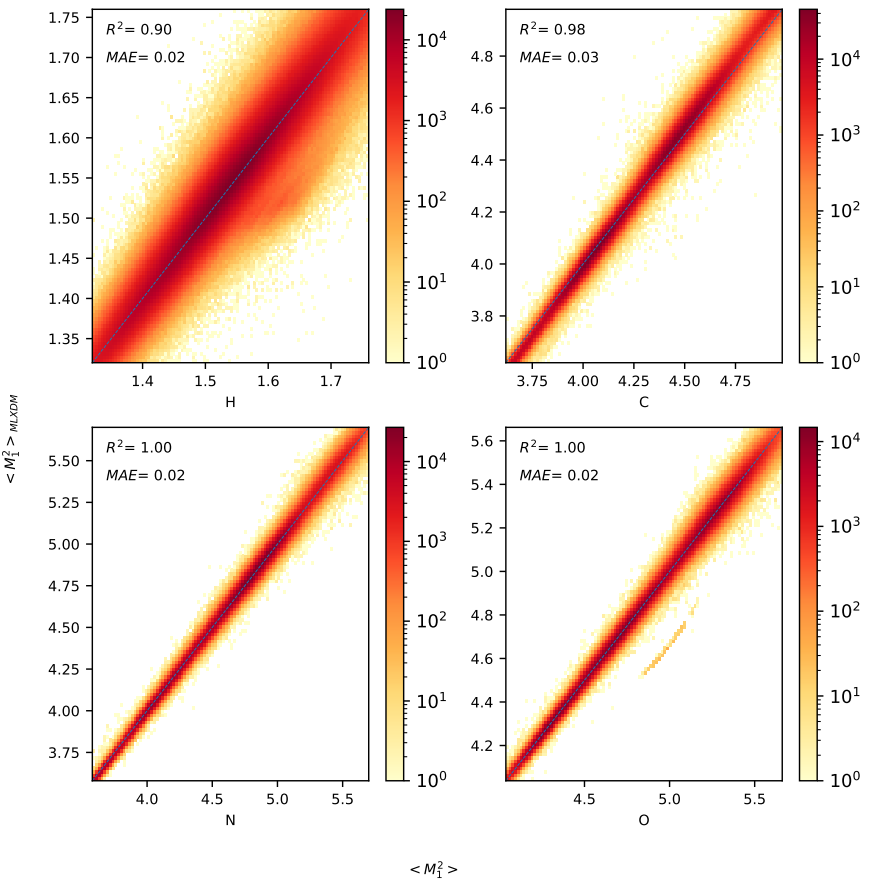

Fig. 5 Prediction of the standardized XDM atomic $\langle M_1^2 \rangle$ (PBE0/aug-cc-pVTZ) coefficients using MLXDM for elements H, C, N, and O. The outliers in the O plot are due to CO's abnormal triple bond.

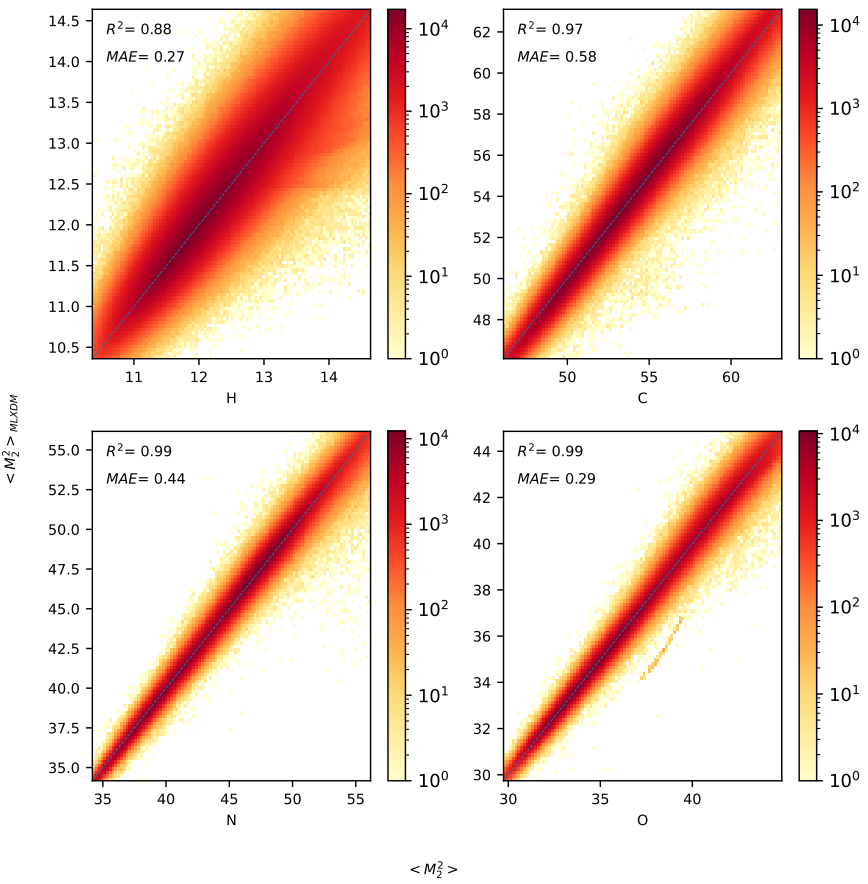

Fig. 6 Prediction of the standardized XDM atomic $\langle M_2^2 \rangle$ (PBE0/aug-cc-pVTZ) coefficients using MLXDM for elements H, C, N, and O. The outliers in the O plot are due to CO's abnormal triple bond.

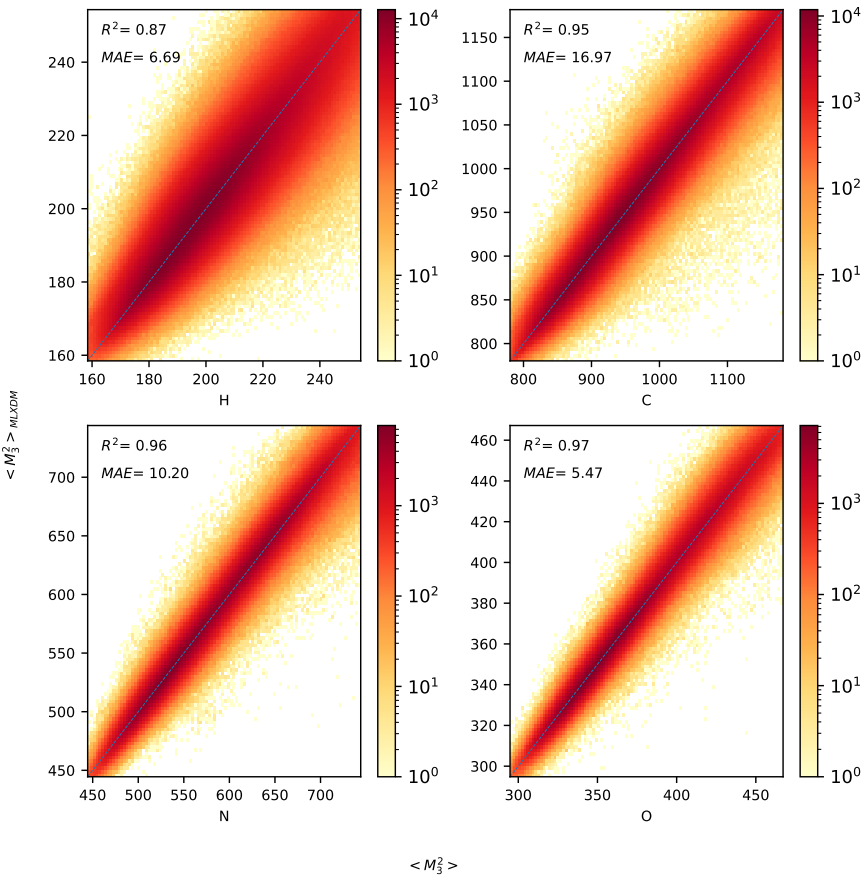

Fig. 7 Prediction of the standardized XDM atomic $\langle M_3^2 \rangle$ (PBE0/aug-cc-pVTZ) coefficients using MLXDM for elements H, C, N, and O. The outliers in the O plot are due to CO's abnormal triple bond.

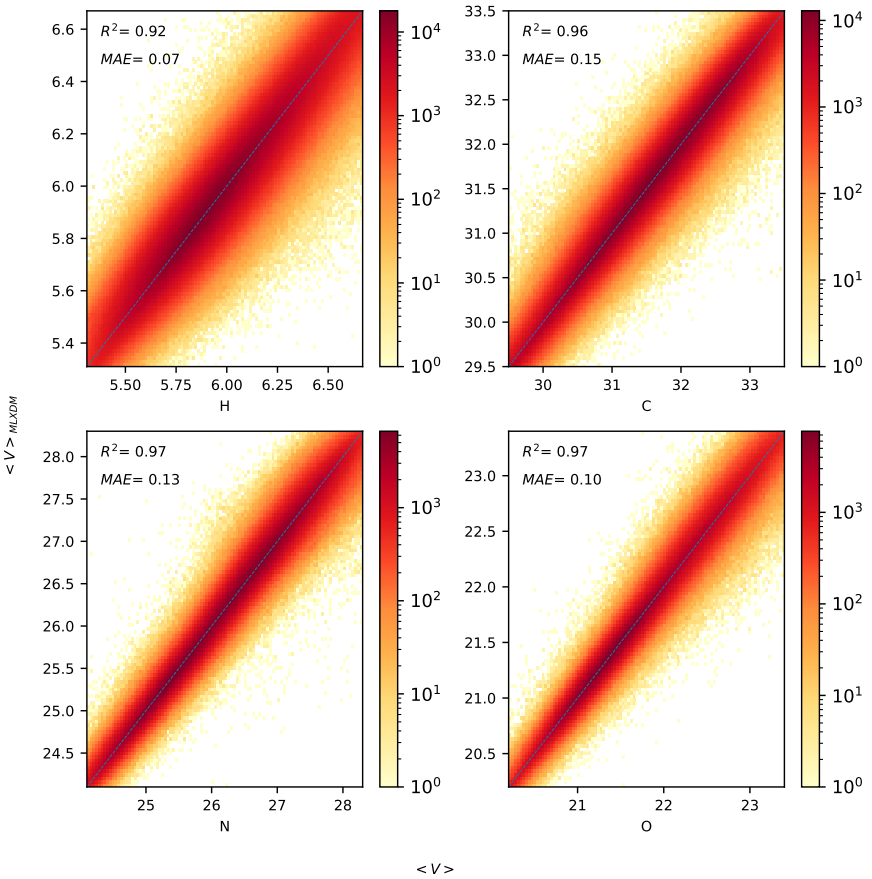

Fig. 8 Prediction of the standardized XDM atomic volumes (V) (PBE0/aug-cc-pVTZ) using MLXDM for elements H, C, N, and O.

3.1 Error Plots

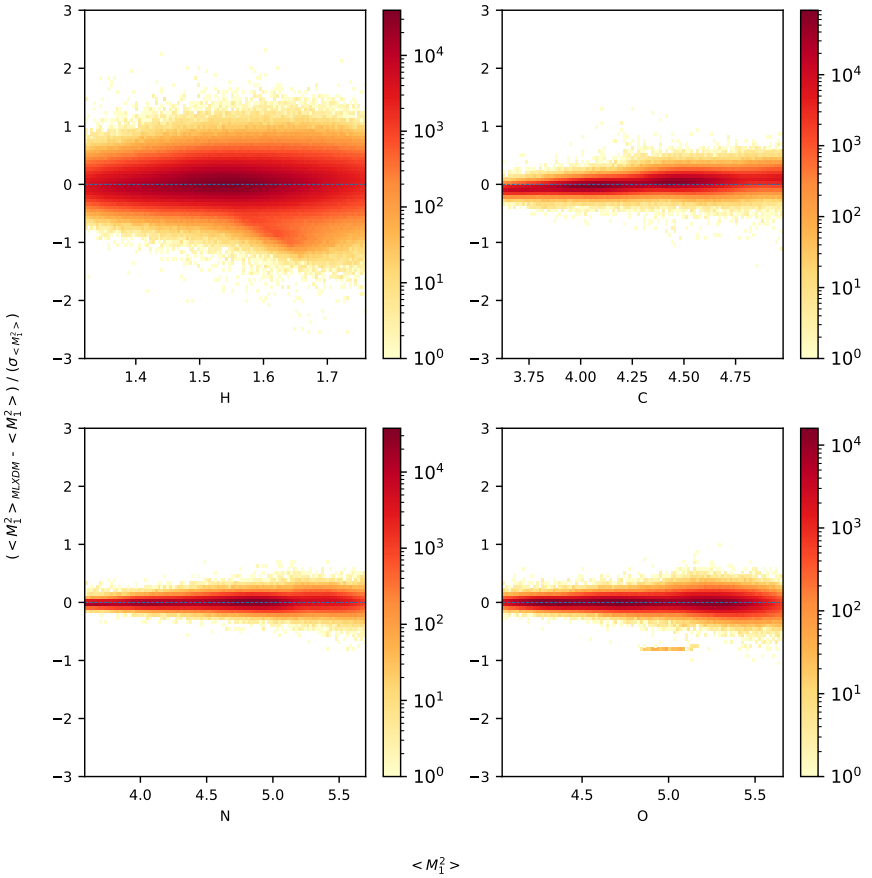

Fig. 9 The deviations of the MLXDM $\langle M_1^2 \rangle$ coefficients relative to the XDM values for the elements C, N, O, and H for the validation data set.

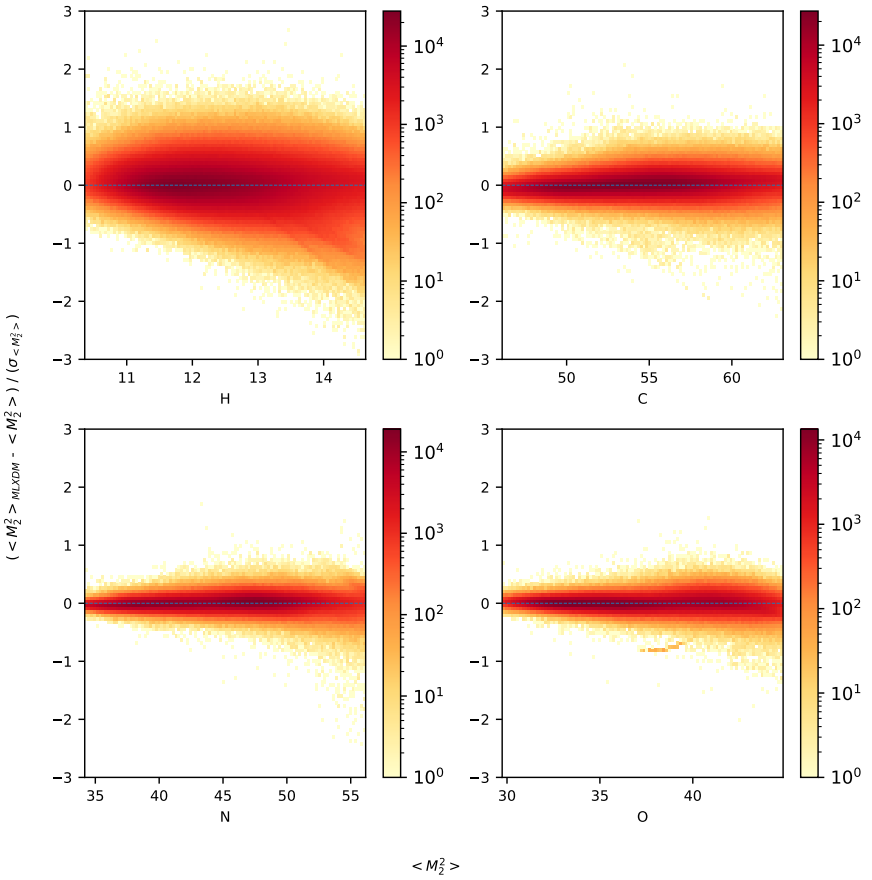

Fig. 10 The deviations of the MLXDM $\langle M_2^2 \rangle$ coefficients relative to the XDM values for the elements C, N, O, and H for the validation data set.

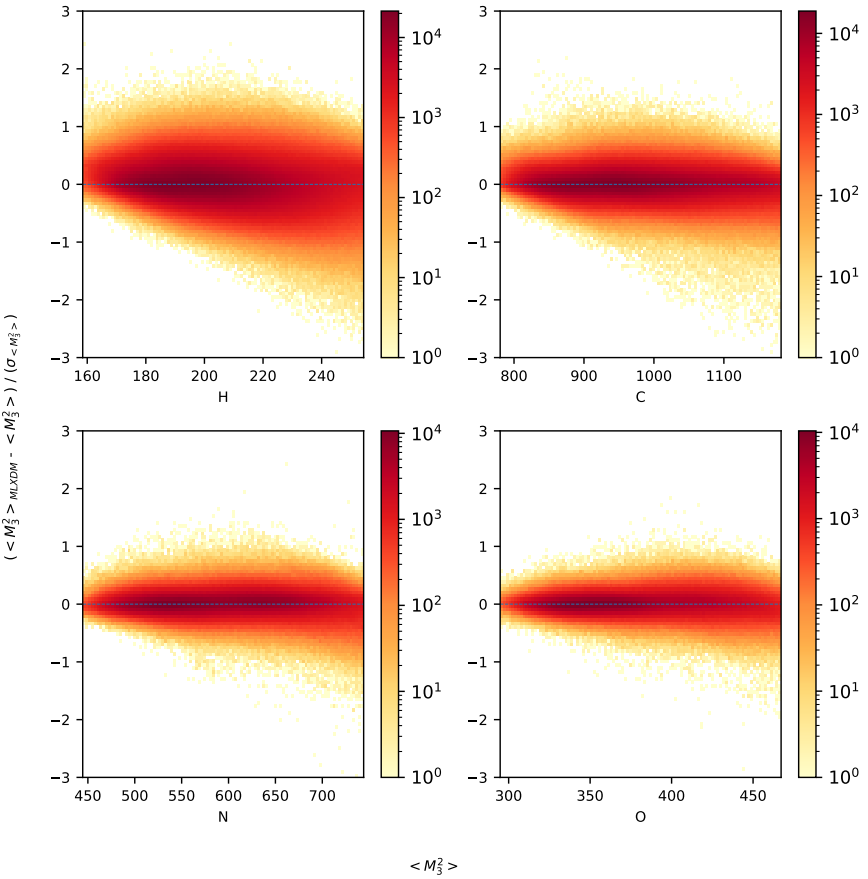

Fig. 11 The deviations of the MLXDM $\langle M_3^2 \rangle$ coefficients relative to the XDM values for the elements C, N, O, and H for the validation data set.

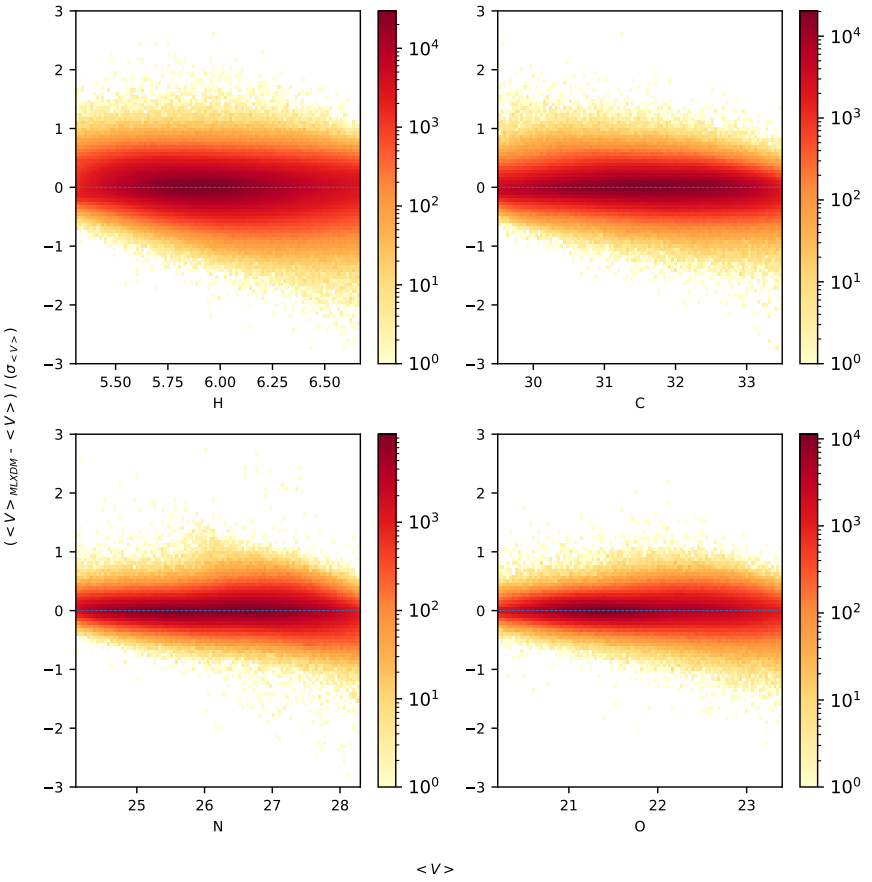

Fig. 12 The deviations of the MLXDM V coefficients relative to the XDM values for the elements C, N, O, and H for the validation data set.

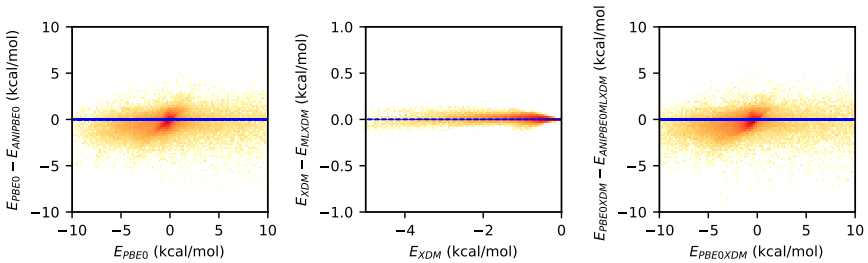

Fig. 13 The deviations of the ANIPBE0 energies from the PBE0 energies (left), the MLXDM energies from the XDM energies (center), and combined ANIPBE0-MLXDM energies from the PBE0-MLXDM energies the validation data set.

4 Baseline Comparisons

4.1 Current State-of-the-Art NNPs

The ANI-2x NNP is one of the most accurate ANI-type NNPs [25], although it is primarily designed to predict the stability of molecular structures. To assess the performance of this method, we have calculated its performance on predicting the intermolecular interactions between neutral molecules in the DES370K dataset. By including the MLXDM dispersion correction, the ANIPBE0-MLXDM NNP is more accurate than ANI-2x for this test set of intermolecular interactions (ANI-2x: $R^2 = 0.79$ MAE=1.28 kcal/mol, ANIPBE0-MLXDM: $R^2 = 0.97$, MAE = 0.69 kcal/mol).

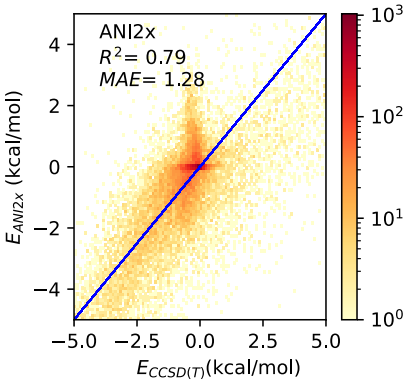

Fig. 14 Prediction of the CCSD(T)/CBS interaction energies by the ANI-2x NNP for the neutral complexes in the DES370K dataset.

The ANI-2x NNP does not include a long range correction for dispersion. Given its excellent performance for a range of intermolecular interactions and reaction energies, it would be desirable to simply add a dispersion correction to this method. However, the mean signed error of ANI-2x for these intermolecular interactions is -0.84 kcal mol^{-1}, indicating that the ANI-2x NNP tends to predict intermolecular interactions to be more attractive than they actually are. The interactions accounted for by MLXDM are universally attractive, so adding a dispersion correction to these energies would make this overestimation larger. The modest performance of ANI-2x for intermolecular interactions likely stems from their training to ωB97X/6-31G* energies. Prediction of short range (i.e., < 5 Å) intermolecular interactions using this method likely benefits from a cancellation of error between the neglect of dispersion interactions and the basis set incompleteness error resulting from the relatively small 6-31G* basis set. The ANIPBE0 NNP has an advantage here because it was trained to

QM data calculated using the larger aug-cc-pVTZ basis set and more extensive intermolecular training data.

4.2 Grimme D3 Corrections

We have also evaluated whether the Grimme D3 dispersion correction[26] with the Becke–Johnson (BJ) damping function[27] could be used with the ANIPBE0 NNP rather than using the MLXDM or XDM-CC dispersion corrections. Grimme D3-BJ model applies a pairwise damped dispersion term including C_6 and C_8 terms. On the DES370K test set, the Grimme D3-BJ correction brings ANIPBE0 interaction energies in closer agreement with the CCSD(T) energies when this correction is included. The performance of this method is comparable to MLXDM and XDM-CC, but the metrics are incrementally better for MLXDM (MAE: 0.67 kcal/mol for MLXDM vs 0.73 kcal/mol for D3-BJ). This generally suggests that a D3 dispersion correction could be effectively used with a suitable NNP for complexes in this test set, although the NNP would have accurately account for all the other components of the intermolecular interaction. Compounds with a broader range of

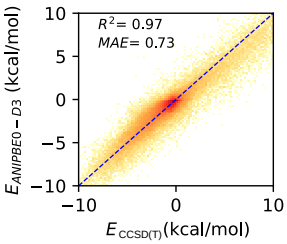

Fig. 15 Prediction of the CCSD(T)/aug-cc-pTVZ interaction energies for the neutral complexes in the DES370K dataset using ANIPBE0 with the Grimme D3-BJ dispersion correction.

5 Alternative Approach Directly Predicting C_6, C_8, and C_{10}

We have also developed an alternative approach where NN's are used to calculate atomic C_6, C_8, and C_{10} dispersion coefficients. The coefficients for the interaction between a pair of atoms can then be approximated by the geometric mean of the coefficients of the two atoms [28, 29],

$$C_{n,ij} = \sqrt{C_{n,ii}C_{n,jj}}. \tag{24}$$

$$C_{6,ii} = \frac{1}{2}\alpha_i\langle M_1^2\rangle_i \tag{25}$$

$$C_{8,ii} = \frac{3}{2}\alpha_i \langle M_2^2 \rangle_i, \tag{26}$$

$$\mathcal{C}_{10,ii} = \alpha_i \alpha_i \frac{2\langle M_1^2 \rangle_i \langle M_3^2 \rangle_i + 2\langle M_3^2 \rangle_i \langle M_1^2 \rangle_i + \frac{21}{5}\langle M_2^2 \rangle_i \langle M_2^2 \rangle_i}{\alpha_i \langle M_1^2 \rangle_i + \alpha_i \langle M_1^2 \rangle_i} \tag{27}$$

This second approach has the advantage of only requiring the evaluation of three NNs to determine the coefficients for each atoms (C_6, C_8, and C_{10}), while first approach requires evaluation of four NNs ($\langle M_1^2 \rangle$, $\langle M_2^2 \rangle$, $\langle M_3^2 \rangle$, and V). Geometric combination rules are also more amenable for simulation methods like lattice summation approaches for calculating the energy of periodic systems. We find this approach has comparable accuracy to the MLXDM.

References

[1] Stone, A.J., Stone, A.: The Theory of Intermolecular Forces. Oxford University Press, Oxford, UK (2013)

[2] Mohebifar, M., Johnson, E.R., Rowley, C.N.: Evaluating force-field London dispersion coefficients using the exchange-hole dipole moment model. J. Chem. Theory Comput. **13**(12), 6146–6157 (2017). https://doi.org/10.1021/acs.jctc.7b00522

[3] Johnson, E.R.: Dependence of dispersion coefficients on atomic environment. J. Chem. Phys. **135**(23), 234–109 (2011) https://doi.org/10.1063/1.3670015. https://doi.org/10.1063/1.3670015

[4] Otero-de-la-Roza, A., Johnson, E.R.: Non-covalent interactions and thermochemistry using XDM-corrected hybrid and range-separated hybrid density functionals. J. Chem. Phys. **138**(20), 204109 (2013). https://doi.org/10.1063/1.4807330

[5] Rowley, C., Rezajooei, N., Phúc, T.N.T.: ANIPBE0-MLXDM dataset (2022). https://doi.org/10.6084/m9.figshare.19790524.v1

[6] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A.V., Bloino, J., Janesko, B.G., Gomperts, R., Mennucci, B., Hratchian, H.P., Ortiz, J.V., Izmaylov, A.F., Sonnenberg, J.L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V.G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J.A. Jr., Peralta, J.E., Ogliaro, F., Bearpark, M.J., Heyd, J.J., Brothers, E.N., Kudin, K.N., Staroverov, V.N., Keith, T.A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A.P., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Millam, J.M., Klene, M., Adamo, C., Cammi, R., Ochterski, J.W., Martin, R.L., Morokuma, K., Farkas, O., Foresman, J.B., Fox, D.J.: Gaussian 16 Revision C.01. Gaussian Inc. Wallingford CT (2016)

[7] Adamo, C., Barone, V.: Toward reliable density functional methods without adjustable parameters: The PBE0 model. J. Chem. Phys. **110**(13), 6158–6170 (1999). https://doi.org/10.1063/1.478522

[8] Kendall, R.A., Dunning, T.H., Harrison, R.J.: Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. J. Chem. Phys. **96**(9), 6796–6806 (1992). https://doi.org/10.1063/1.462569

[9] The postg program is available from http://schooner.chem.dal.ca

[10] Smith, J.S., Isayev, O., Roitberg, A.E.: ANI-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules. Sci. Data **4**, 2052–4463 (2017). https://doi.org/10.1038/sdata.2017.193

[11] Donchev, A.G., Taube, A.G., Decolvenaere, E., Hargus, C., McGibbon, R.T., Law, K.-H., Gregersen, B.A., Li, J.-L., Palmo, K., Siva, K., Bergdorf, M., Klepeis, J.L., Shaw, D.E.: Quantum chemical benchmark databases of gold-standard dimer interaction energies. Sci. Data **8**, 55 (2021). https://doi.org/10.1038/s41597-021-00833-x

[12] Seung, H., Opper, M., Sompolinsky, H.: Query by committee. Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, pp. 287–294. ACM, New York, USA (1992). https://doi.org/10.1145/130385.130417

[13] Smith, J.S., Nebgen, B., Lubbers, N., Isayev, O., Roitberg, A.E.: Less is more: Sampling chemical space with active learning. J. Chem. Phys. **148**(24), 241733 (2018)

[14] 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. **8**, 3192–3203 (2017). https://doi.org/10.1039/C6SC05720A

[15] Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J.S., Roitberg, A.E.: TorchANI: A free and open source pytorch-based deep learning implementation of the ani neural network potentials. J. Chem. Inf. Model. **60**(7), 3408–3415 (2020). https://doi.org/10.1021/acs.jcim.0c00451

[16] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: PyTorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d' Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035 (2019). http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[17] Barron, J.T.: Continuously differentiable exponential linear units. arXiv preprint arXiv:1704.07483 (2017)

[18] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015). http://arxiv.org/abs/1412.6980

[19] Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization. arXiv (2017). https://doi.org/10.48550/ARXIV.1711.05101. https://arxiv.org/abs/1711.05101

[20] Larsen, A.H., Mortensen, J.J., Blomqvist, J., Castelli, I.E., Christensen, R., Dułak, M., Friis, J., Groves, M.N., Hammer, B., Hargus, C., Hermes, E.D., Jennings, P.C., Jensen, P.B., Kermode, J., Kitchin, J.R., Kolsbjerg, E.L., Kubal, J., Kaasbjerg, K., Lysgaard, S., Maronsson, J.B., Maxson, T., Olsen, T., Pastewka, L., Peterson, A., Rostgaard, C., Schiøtz, J., Schütt, O., Strange, M., Thygesen, K.S., Vegge, T., Vilhelmsen, L., Walter, M., Zeng, Z., Jacobsen, K.W.: The atomic simulation environment—a python library for working with atoms. J. Condens. Matter Phys. **29**(27), 273002 (2017). https://doi.org/10.1088/1361-648x/aa680e

[21] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., Mackerell Jr., A.D.: CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. **31**(4), 671–690 (2010). https://doi.org/10.1002/jcc.21367

[22] Walter, N.P., Jaiswal, A., Cai, Z., Zhang, Y.: LiquidLib: A comprehensive toolbox for analyzing classical and ab initio molecular dynamics simulations of liquids and liquid-like matter with applications to neutron scattering experiments. Comput. Phys. Commun. **228**, 209–218 (2018). https://doi.org/10.1016/j.cpc.2018.03.005

[23] Zhang, Y.-B., Su, J., Furukawa, H., Yun, Y., Gándara, F., Duong, A., Zou, X., Yaghi, O.M.: Single-crystal structure of a covalent organic framework. J. Am. Chem. Soc. **135**(44), 16336–16339 (2013). https://doi.org/10.1021/ja409033p

[24] Dubbeldam, D., Torres-Knoop, A., Walton, K.S.: On the inner workings of Monte Carlo codes. Mol. Simul. **39**(14-15), 1253–1292 (2013)

[25] Devereux, C., Smith, J.S., Davis, K.K., Barros, K., Zubatyuk, R., Isayev, O., Roitberg, A.E.: Extending the applicability of the ani deep learning molecular potential to sulfur and halogens. Journal of Chemical Theory and Computation **16**(7), 4192–4202 (2020) https://doi.org/10.1021/acs.jctc.0c00121. https://doi.org/10.1021/acs.jctc.0c00121

[26] Grimme, S., Antony, J., Ehrlich, S., Krieg, H.: A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. J. Chem. Phys. **132**, 154104 (2010)

[27] Grimme, S., Ehrlich, S., Goerigk, L.: Effect of the damping function in

dispersion corrected density functional theory. J. Comput. Chem. **32**, 1456–1465 (2011)

[28] McDaniel, J.G., Schmidt, J.R.: Physically-motivated force fields from symmetry-adapted perturbation theory. J. Phys. Chem. A **117**(10), 2053–2066 (2013). https://doi.org/10.1021/jp3108182

[29] Grimme, S.: Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J. Comput. Chem. **27**(15), 1787–1799 (2006) https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20495. https://doi.org/10.1002/jcc.20495