

Accurate Deep Learning-aided Density-free Strategy for Many-Body Dispersion-corrected Density Functional Theory

Pier Paolo Poier,^{*,†} Théo Jaffrelot Inizan,[†] Olivier Adjoua,[†] Louis Lagardère,^{†,‡}
and Jean-Philip Piquemal^{*,†,¶}

[†]*Sorbonne Université, LCT, UMR 7616 CNRS, Paris, France*

[‡]*Sorbonne Université, IP2CT, FR 2622 CNRS, Paris, France*

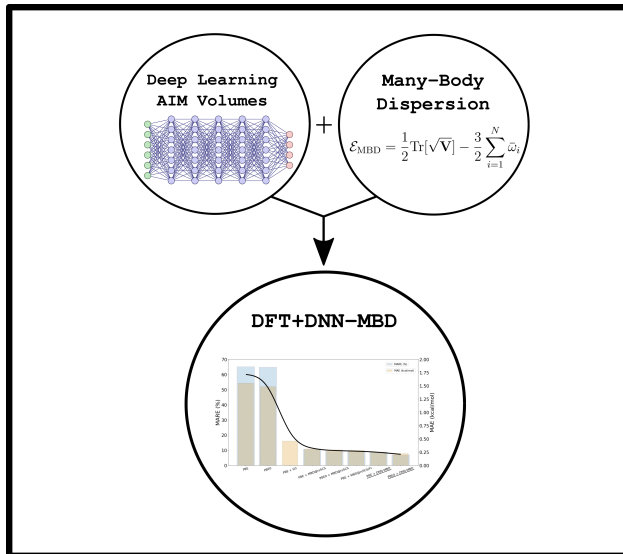
[¶]*The University of Texas at Austin, Department of Biomedical Engineering, TX, USA*

E-mail: pier.poier@sorbonne-universite.fr; jean-philip.piquemal@sorbonne-universite.fr

Abstract

Using a Deep Neuronal Network model (DNN) trained on the large ANI-1 data set of small organic molecules, we propose a transferable density-free many-body dispersion model (DNN-MBD). The DNN strategy bypasses the explicit Hirshfeld partitioning of the Kohn-Sham electron density required by MB-D models to obtain the atom-in-molecules volumes used by the Tkatchenko-Scheffler polarizability rescaling. The resulting DNN-MBD model is trained with minimal basis iterative Stockholder atomic volumes and, coupled to Density Functional Theory (DFT), exhibits comparable (if not greater) accuracy to other approaches based on different partitioning schemes. Implemented in the Tinker-HP package, the DNN-MBD model decreases the overall computational cost compared to MB-D models where the explicit density partitioning is performed. Its coupling with the recently introduced Stochastic formulation of the MB-D equations (*J. Chem. Theory Comput.*, **2022**, *18*, 3, 1633–1645) enables large

routine dispersion-corrected DFT calculations at preserved accuracy. Furthermore, the DNN electron density-free features extend MBD’s applicability beyond electronic structure theory within methodologies such as force fields and neural networks.



Since its original formulation in 1965, Kohn-Sham Density Functional Theory¹ (KS-DFT) has become the most popular family of electronic structure methods. KS-DFT represents in fact the cheapest way for introducing electronic correlation as its computational cost is similar to that of the Hartree-Fock method. KS-DFT is based on the idea of evaluating the kinetic energy from a Slater determinant thus assuming the electrons to be non-interacting. This apparently crude assumption actually leads to big improvements in describing chemical bonding compared to, for example, the use of the Thomas-Fermi kinetic energy formulation. The difference between the Slater determinant kinetic energy representation and the true one, together with the difference between the true total electronic interaction and the exchange energies represents, in KS-DFT, the key contribution to the exchange-correlation functional which remains, however, unknown.

In practice, the plethora of existing KS-DFT variants differentiate themselves in the way the exchange-correlation functional is approximated. Typically it is assumed to be a functional of the local electron density and eventually of its gradient and Laplacian. As a consequence,

only local contributions to electronic correlation are included and this explains the general inadequacy of DFT methods to describe dispersion interactions which, on the other hand, have roots in long-range electronic correlation.

To retain the pleasant computational performances of KS-DFT methods, several dispersion corrections have been proposed.² Among these, the popular and successful approach of Grimme includes dispersion via empirical pairwise C_6 terms.³⁻⁵ This is particularly appealing in virtue of its nearly zero additional computational cost.

A further approach is to replace the empirical pairwise terms with ones obtained from quantities coupled to the molecular electron density. For example, in Becke and Johnson’s model, pairwise C_6 coefficients are written in terms of atomic polarizabilities and the averaged exchange-hole dipoles corresponding to each of the two atoms in the pair.^{6,7} In the alternative approach proposed by Tkatchenko and Scheffler (TS),⁸ pairwise C_6 coefficients are instead expressed in terms of accurate free atom reference data as well as atoms-in-molecule (AIM) polarizabilities obtained from the rescaling of the corresponding AIM volumes computed via the Hirshfeld partitioning of the molecular electron density.⁹

One limitation of the above mentioned pairwise approaches is the impossibility of capturing non-additive many-body dispersion (MBD) effects, which inclusion has recently been shown important in modelling extended systems, supramolecular complexes and proteins in solutions, among others.¹⁰⁻¹³

The non-additive long-range character of dispersion interactions has been modeled via a set of coupled fluctuating dipoles^{14,15} (CFD) or alternatively by quantum Drude oscillators.¹⁶⁻¹⁹

In recent years, Tkatchenko, DiStasio Jr., Ambrosetti et al. have proposed a range-separated many-body dispersion model based on the CFD where the self-consistent screening of a set of atomic polarizabilities is performed (MBD@rsSCS).^{20,21} The MBD@rsSCS model is appealing not only for introducing non-additive many-body dispersion effects but also since it relies, *de facto*, on a single range-separation parameter which is tuned according to the choice of the exchange-correlation functional employed.

The MBD@rsSCS keeps in fact the spirit of the TS approach where AIM polarizabilities and van der Waals radii are obtained via the Hirshfeld partitioning of the density.

The Hirshfeld method leads to AIM densities which minimize the Kullback-Liebr divergence corresponding to the information loss upon molecule formation where this solid mathematical condition is used as a basis for the development of new information-theoretic partitioning methods.²²

As discussed in references^{23,24} Hirshfeld partitioning makes its resulting AIM densities as close as possible to the ones of the isolated atoms, consequently AIM's properties turn out to be as similar as possible to those of the free atoms. This is particularly evident in the magnitude of Hirshfeld atomic charges, being too small in magnitude for reproducing the molecular electrostatic potential (ESP) or in modeling AIM polarizabilities in ionic and covalent crystals where the Hirshfeld partitioning leads to unrealistically large polarizabilities of cations which can even be found larger than those of the anions.²⁵

The above mentioned shortcomings were ameliorated by the Iterative Hirshfeld (HI) scheme²⁶ where the reference atomic density employed in the partitioning is constructed as a linear combination of the two densities relative to the atomic oxidation states closest to the fractional number of electrons assigned by the partitioning at a given iteration.

The ESP computed from HI atomic charges have proven to agree remarkably well with *ab initio* computed reference.²⁷ In addition, the use of HI derived AIM polarizabilities leads to more realistic dispersion coefficients²⁵ especially in ionic systems and adsorption phenomena on surfaces of ionic solids where the HI scheme used within the TS dispersion model improves interaction energies.²⁸ HI partitioning has also been employed in the MBD@rsSCS model replacing the original Hirshfeld scheme²⁹ and its use in connection to the fractionally ionic AIM polarizabilities leads, in the just mentioned challenging systems, to reduced errors.³⁰

Despite the improvements carried by the HI partitioning, the scheme remains affected by a shortcoming arising from the density interpolation for negatively charged atoms as this procedure is, for some species, ill-defined. This arises from the fact that free anions such

as N^- and O^{2-} (or in general any doubly negative ion) are not bound and their reference electron densities, computed at a complete basis set (CBS), result in a detached electron.

The iterative Stockholder atom (ISA) scheme, on the other hand, is not affected by this problem as the partitioning does not require reference atomic densities computed from isolated atoms at different ionic states as they are rather obtained from a spherical averaging of the molecular density using nuclei as expansion points.^{31,32} The minimal basis iterative Stockholder atom (MBISA), a variant of the ISA method, have proven successful in the atomic polarizability rescaling approach employed by the TS scheme as well as in reproducing *ab initio* ESP from atomic point charges³³ and for this reasons its use in connection to the MBD@rsSCS model is particularly appealing.

AIM properties are local quantities which depend on the near chemical environment and thus carry a certain degree of transferability. In particular, the TS polarizability rescaling scheme (employed in the MBD@rsSCS) makes use of AIM volumes which are well suited to be computed via deep neural network (DNN) where the environment vector associated to an atom’s surrounding is defined within a local cutoff. The potential of deep learning in capturing local atomic properties has been proved by Isayev and co-workers whose multi-output DNN model successfully predicts AIM properties ranging from multipoles to volumes.³⁴

In this Letter we present a hybrid DNN-aided MBD@rsSCS model (DNN-MBD) where the AIM volumes ratio employed in the TS polarizability rescaling are generated by a deep neural network trained on the ANI-1 data set (approximately 4.6 million structures) containing MBISA AIM volumes.³⁵

For the common S66x8 benchmark set,³⁶ the DNN-MBD model coupled to the common PBE/PBE0 density functionals, exhibits excellent interaction energies while completely bypassing the electron density partitioning with a consequent computational cost reduction. This electron density-free DNN-MBD approach employed in connection to our recently proposed linear scaling stochastic MBD@rsSCS formulation,³⁷ allows for modelling non-additive

long-range dispersion interactions of up-to-millions atom systems at a very low computational cost without compromising the accuracy.

We note that kernel-ridge regression approaches to model AIM polarizabilities have been proposed in modelling dispersion interactions.^{38,39} This approach, however, is characterized by a $\mathcal{O}(N^2)$ and $\mathcal{O}(N^3)$ scaling of the required memory and computational cost involved in the model’s training respectively, N being the size of the data set. These non-linear scaling prevents the applicability of kernel-ridge approaches on very large and diverse data sets, necessary for the generation of general-purpose MBD models. Additionally, the poor scaling with the number of processes limits its use on large systems. Here instead we generalize the approach to model MBD interactions to a much broader class of systems thanks to the employed model’s flexibility and broad data set, without affecting the model’s accuracy and linear scalability.

We will, in the following, proceed by briefly recalling the key concepts of the standard MBD@rsSCS model before introducing the DNN-MBD hybrid model and its performances.

As a starting point in this discussion, we examine the TS polarizability rescaling in Eq.(1), where α_i and V_i represent the TS static polarizability and AIM volume respectively of the i -th atom while the zero superscript denotes free atom reference quantities.

$$\alpha_i = \left(\frac{V_i}{V_i^0} \right) \alpha_i^0 \tag{1}$$

The AIM volume V_i is obtained by solving the integral in Eq.(2) where $\rho(\mathbf{r})$ is the Kohn-Sham molecular electron density which, via the partitioning-specific weight function $w_i(\mathbf{r})$,

is decomposed into its AIM densities $\{\rho_i(\mathbf{r})\}$.

$$V_i = \int \mathbf{r}^3 \rho_i(\mathbf{r}) d^3\mathbf{r} \quad (2)$$

$$\rho_i(\mathbf{r}) = w_i(\mathbf{r})\rho(\mathbf{r})$$

Once the set of static AIM polarizabilities in Eq.(1) is obtained, a correspondent set of frequency-dependent ones is generated via Eq.(3), where this time ω_j^0 and $C_{6,j}^0$ are the free atom characteristic excitation frequency and first dispersion coefficient.

$$\alpha_j(i\nu) = \frac{\alpha_j}{1 - (i\nu/\omega_j^0)^2} \quad (3)$$

$$\omega_j^0 = \frac{4}{3} \frac{C_{6,j}^0}{(\alpha_j^0)^2}$$

These frequency-dependent polarizabilities are, in the MBD@rsSCS model, gathered as diagonal elements of the frequency-dependent superpolarizability matrix $\mathbf{A}(i\nu)$ being one of the entries in the Dyson-like equation below which solution provides the screened superpolarizability matrix $\bar{\mathbf{A}}(i\nu)$.

$$\bar{\mathbf{A}}(i\nu) = \mathbf{A}(i\nu) - \mathbf{A}(i\nu)\mathbf{T}^{\text{SR}}(i\nu)\bar{\mathbf{A}}(i\nu) \quad (4)$$

The \mathbf{T}^{SR} represents a damped dipole-dipole interaction operator applied to the Coulombic interaction of two frequency-dependent spherical Gaussian charge distributions where its explicit expression, together with the one for $\mathbf{A}(i\nu)$, can be found in reference.³⁷ We note here that the Fermi damping function employed in the definition of \mathbf{T}^{SR} makes use of AIM van der Waals radii which can also be obtained by a volume rescaling similarly to what discussed for polarizabilities.⁸

The solution of Eq.(4) for a set of frequencies, and a consequent partial contraction of the converged $\{\bar{\mathbf{A}}(i\nu)\}$, gives a set of screened frequency-dependent atomic polarizabilities $\{\bar{\alpha}_j(i\nu)\}$ which are used to approximate the Casimir-Polder integral providing screened characteristic

excitation frequencies $\{\bar{\omega}_j\}$.

$$\begin{aligned}\bar{C}_{6,j} &= \frac{3}{\pi} \int_0^\infty \bar{\alpha}_j(i\nu) \bar{\alpha}_j(i\nu) d\nu \\ \bar{\omega}_j &= \frac{4}{3} \frac{\bar{C}_{6,j}}{[\bar{\alpha}_j(0)]^2}\end{aligned}\tag{5}$$

The set of screened excitation frequencies as well as the screened static atomic polarizabilities define the MBD potential matrix shown in Eq.(6) for a general ij block. \mathbf{T}^{LR} represents the range-separated damped dipole-dipole interaction matrix which explicit expression is also found in reference.³⁷

$$\mathbf{V}_{ij} = \delta_{ij} \bar{\omega}_i^2 + (1 - \delta_{ij}) \bar{\omega}_i \bar{\omega}_j \sqrt{\bar{\alpha}_i(0) \bar{\alpha}_j(0)} \mathbf{T}_{ij}^{\text{LR}}\tag{6}$$

The trace of $\sqrt{\mathbf{V}}$ defines the interaction energy \mathcal{E}_{int} of the CFDs in the system³⁷ while its zero-point value \mathcal{E}_0 is given by the sum of all screened excitation frequencies. Finally the difference between \mathcal{E}_{int} and \mathcal{E}_0 gives the target MBD@rsSCS energy, Eq.(7), which is coupled to the KS-DFT one to include non-additive dispersion contributions.

$$\mathcal{E}_{\text{MBD}} = \mathcal{E}_{\text{int}} - \mathcal{E}_0 = \frac{1}{2} \text{Tr}[\sqrt{\mathbf{V}}] - \frac{3}{2} \sum_{i=1}^N \bar{\omega}_i\tag{7}$$

In the original MBD@rsSCS model just briefly reviewed, \mathcal{E}_{MBD} is coupled to the molecular electron density via AIM volume partitioning introduced in Eq.(2).

In this Letter instead we show that the explicit electron density partitioning can be avoided by learning AIM volumes via a DNN model without affecting the original MBD@rsSCS model's accuracy.

Bereau et al. and more recently, Mulhi et al. used ML on atomic volumes inside vdW model to capture many body effects.^{38,39} Both have developed a Gaussian approximation potential (GAP) force field on TS polarizability rescaling. While GAP has shown to outperform neural networks in predicting energies with small-sized data set, e.g few thousands of data,

its poor computational scaling $\mathcal{O}(N^3)$ prevents its use on very large training sets and thus to build a general purpose MBD model⁴⁰ Finally, these models are either restricted to pairwise interactions or do not scale linearly with respect to the number of atoms as our Stochastic reformulation of the MBD equations was introduced only recently.³⁷

Isayev et al.³⁴ recently extended their 5 million chemical conformations, the ANI-1 data set, with atomic volumes computed at the ω B97x/def2-TZVPP level with MBISA partitioning. In virtue of its size and diversity, this data set is here employed in building our DNN to be coupled to the MBD@rsSCS model. Here we restrict ourselves to structure composed of only C, H, N and O, thus reducing the actual data set size to 4.6 million conformations.

In the MBISA weight function $w_i(\mathbf{r})$, each of the reference pro-atomic densities $\rho_i^0(\mathbf{r})$ is expanded into m_i Slater functions, m_i being the number of shells of atom i placed at \mathbf{R}_i .

$$\begin{aligned}
 w_i(\mathbf{r}) &= \frac{\rho_i^0(\mathbf{r})}{\sum_{j=1}^N \rho_j^0(\mathbf{r})} \\
 \rho_i^0(\mathbf{r}) &= \sum_{\sigma=1}^{m_i} \frac{N_{i,\sigma}}{k_{i,\sigma}^3 8\pi} \exp\left(-\frac{\|\mathbf{r} - \mathbf{R}_i\|}{k_{i,\sigma}}\right)
 \end{aligned}
 \tag{8}$$

In the scheme, the population $N_{i,\sigma}$ and width $k_{i,\sigma}$ of each shell are free-variables which are optimized so that the loss of information upon molecule formation is minimized.³³

To handle such large data set, a deep neural network is the natural choice.⁴¹ In particular, we use as machine learning model a feed-forward DNN with the ANI-like symmetry functions (SFs).⁴² The ANI’s SFs are a subfamily of Behler-Parinello’s ones⁴³ which traduce an atomic local environment i into an atomic environment vector (AEV) $G_i = \{G_i^R, G_i^A\}$ where G_i^R and G_i^A represent its radial and angular contributions respectively. Although SFs development is an intensive field of research and more accurate models have been developed since (ω ACSF,⁴⁴ SOAP⁴⁵ among others), we stick to the ANI’s original SFs as they were shown to successfully predict complex local properties such as, in the case of AIMNET, multipoles and volumes.³⁴ Moreover, ANI’s SFs have the great advantage of being computationally efficient as they rely on 2-body terms thus making the overall DNN model linear scaling with the

system’s size.

The DNN part of the combined DNN-MBD model relies on Scikit-learn,⁴⁶ PyTorch⁴⁷ and TorchAni.⁴⁸ They are all included in the Tinker-HP neural network module, which implementation will be detailed in a forthcoming dedicated paper (T. Jaffrelot Inizan et al., 2022). We kept the original ANI’s SFs parameters as we did not see major differences after tuning them. We empirically tested multiple neural network architectures (further details are found in the Supporting Information (SI) Figure 3) and the best performance was obtained with 5 hidden layers. The atomic element’s neural network architectures are H 160:128:96:48:1; C 144:112:96:48:1; N 128:112:96:48:1; O 128:112:96:48:1. We observed that by adding 1 extra layer to the original ANI-1x model architecture slightly increases the performance of the model while making it more flexible. Indeed, in the original ANI-1x model, the last layer is composed of 96 neurons, and adding an extra 48 neurons layer may prevent loss of information. We used the Exponential Linear Units (ELU) activation function⁴⁹ while the model’s parameters were initialized with the so-called “He” initialization and updated with Hutter’s AdamW algorithm during the training procedure.⁵⁰ Within the AdamW algorithm, the factor was set to 0.5 and the patience to 100. The initial learning rate was set to 10^{-3} and the early stopping learning rate was set to 10^{-6} . The ANI-1 dataset was shuffled and split into training and validation set containing 80% and 20% respectively of the full dataset. The networks were trained for 6000 epochs with a batch size of 2560.

The ANI-1 data set, upon which our DNN model is trained, consists of AIM volumes computed at the ω B97x/def2-TZVPP level. The model is trained on volume ratios rather than pure AIM volumes as the narrower distribution of the former allows for a DNN’s better performance without the need for re-scaling. Indeed, the atomic volumes ratio for C, H, O, N (see Figure 1 of the SI) is between 0.1 and 1.6. Free atom volumes are computed at the same level as AIM ones. The correlation plots between the DNN model and the *ab initio* validation set reference is depicted in Figure 1. The root-mean-square-error (RMSE) and mean-absolute-error (MAE) are respectively 0.012 and 0.008 which is much less than the

smallest value of the data set showing the good accuracy of our model. The final DNN model and the dataset used for the training can be download directly via the Zenodo repository located at the following address.⁵¹

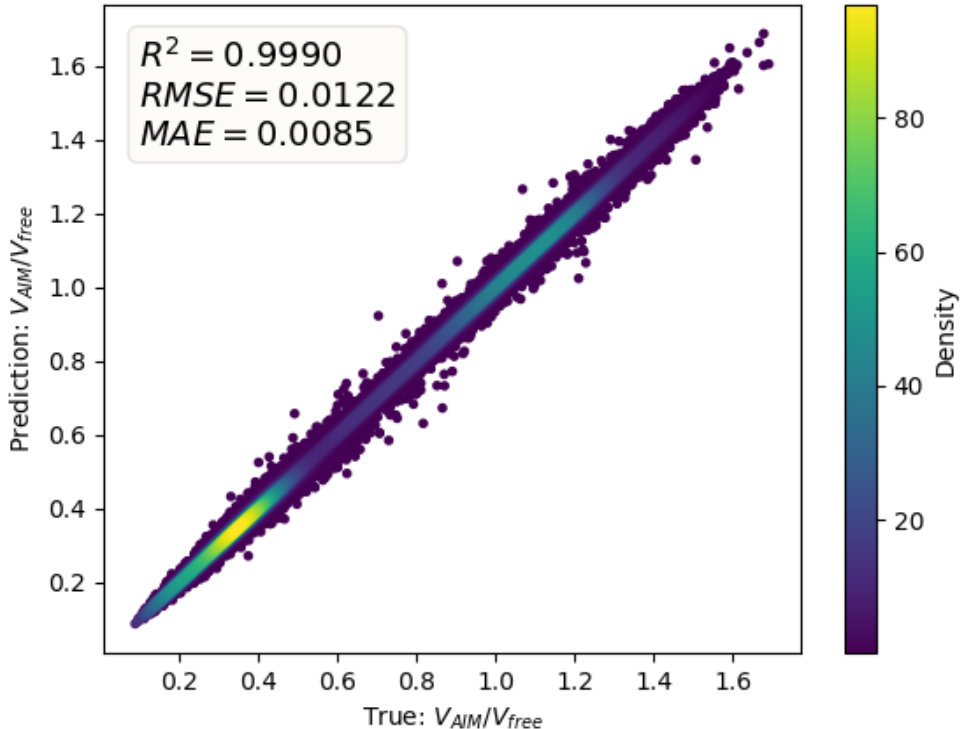


Figure 1: Atomic volume correlation plot comparing the DNN prediction to DFT reference calculations for 1/100 of the validation set. The color bar scale reflects the density of points and correlate with the atomic volumes ratio distribution (Figure 2 of the SI).

The DNN model providing AIM volumes’ ratios is embedded in the Tinker-HP package where our linear-scaling and embarrassingly parallel stochastic MBD@rsSCS is also implemented.³⁷

The outcoming DNN-MBD model is coupled to the common semi-local PBE⁵² functional as well as its hybrid PBE0 version⁵³ since this choices allow for comparisons with results ready available in literature. The optimal range-separation β parameters for both the PBE+DNN-MBD and PBE0+DNN-MBD methods are obtained by minimizing the mean absolute rel-

ative error (MARE) on the widely employed S66x8 benchmark set consisting of 66 dimers placed at 8 different intermolecular distances for a total of 528 different structures where CCSD(T) interaction energies computed at CBS are used as reference.

All DFT computations employed Jensen’s pcseg-3 basis set belonging to the family of segmented polarization-consistent⁵⁴ basis sets which, for DFT calculations, exhibits lower basis set errors than other gaussian basis sets as well as higher computational efficiency at given cardinal number as these basis sets were explicitly designed and optimized for DFT.⁵⁵

Figure 2 shows the MARE as a function of the range separation parameter for PBE+DNN-MBD and PBE0+DNN-MBD methods.

The optimal β parameters are found to be 0.75 and 0.77 for the PBE+DNN-MBD and

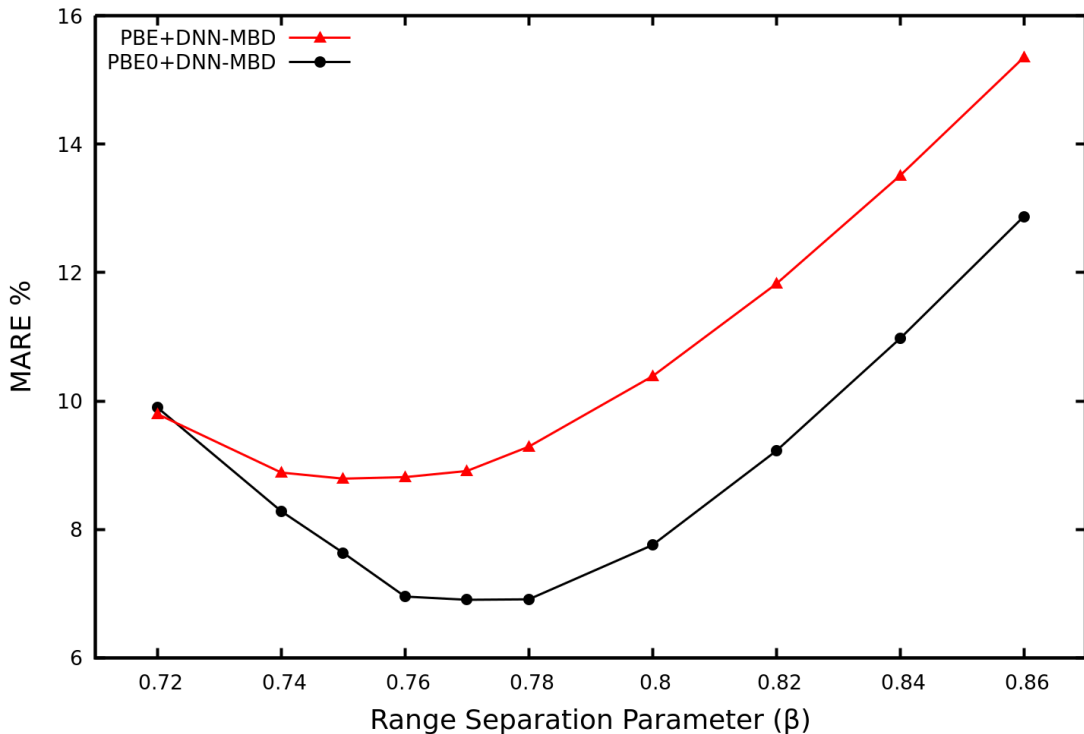


Figure 2: MARE (%) as a function of the range separation parameter for the PBE+DNN-MBD and PBE0+DNN-MBD methods.

PBE0+DNN-MBD methods respectively. These values differ from the ones optimized for the original PBE/PBE0+MBD@rsSCS models²¹ and this has to be addressed to the different partitioning scheme employed. As pointed out by Vestraelen et al.,³³ AIM densities

computed via the Hirshfeld or HI partitioning, exhibit asymmetries i.e. they are aspherical with too much density in the bonding region. This density accumulation, relatively far away from the atomic nucleus, leads to larger values of radial moments, thus leading to larger AIM volumes compared to the ones obtained via the MBISA scheme (unaffected from this asymmetry artifact) for which less screening of volume-scaled AIM quantities (smaller β) is most likely to be needed.³³

We observe, nevertheless, that both PBE0+DNN-MBD and PBE0+MBD@rsSCS methods require a larger β parameter compared to their PBE corresponding models and this is consistent with the PBE0’s improved description of short-range exchange-correlation effects due to the fraction of exact exchange included in the functional, as discussed in reference.²¹

The performance of the optimized PBE/PBE0+DNN-MBD methods is compared to different MBD models in terms of MAE and MARE for the S66x8 data set and the results are summarized in Figure 3 where actual values are reported in Table 1.

For the benchmark set here employed, the DNN-MBD model exhibits lower (although by a contained margin) errors both in its coupling to the PBE and PBE0 functionals compared to the standard MBD@rsSCS approach based on Hirshfeld AIM volumes as well the PBE+MBD@rsSCS/FI approach based on the fractionally ionic polarizabilities and HI AIM volume partitioning. For both the chosen functionals, the outcoming DNN-MBD model provides a mean absolute error in the S66x8 interaction energies which is below 0.25 kcal/mol compared to the reference CCSD(T) CBS golden standard.

To strengthen the analysis, we additionally computed the MAE and MARE for the S22 data set⁵⁷ by employing the range separation parameters previously optimized for the S66x8 set. We can, in this way, employ the S22 set as a test set to validate our conclusions, Figure 3 (bottom) and Table 2.

Compared to the S66x8 set, the MAE and MARE values of our proposed PBE/PBE0+DNN-MBD models are, for the S22 set, higher however this is not surprising as no β optimization was performed this time. Let’s note that all methods present errors that are larger in the

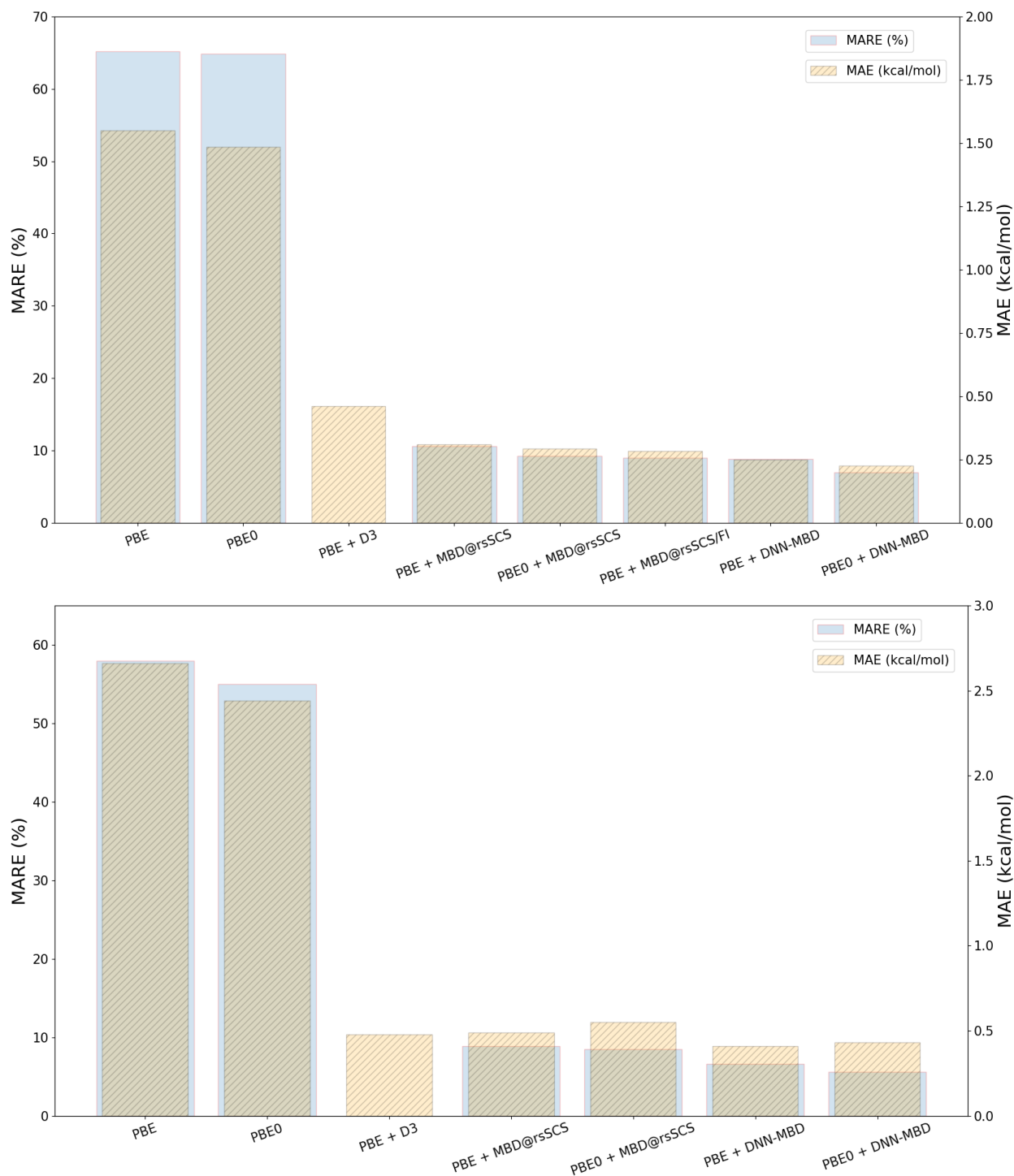


Figure 3: MARE (%) and MAE (kcal/mol) of PBE, PBE0, PBE+D3⁵⁶ and different MBD models (MBD@rsSCS,²¹ MBD@rsSCS/FI³⁰) including our DNN-MBD for the S66x8 (top) and S22 (bottom) data sets.

case of the S22 set compared to S66x8 (see Table 1 and Table 2). Indeed, there are reasons for that and we can stress that the dimers employed in the S22 set are placed at equilibrium while the S66x8 set includes out of equilibrium dimers. In our case, the DNN-MBD model trained on S66x8, appears less biased towards equilibrium structures. Overall, as one can see from Table 2, our DNN-MBD remains highly transferable and, with an error below 0.43 kcal/mol compared to the reference CCSD(T) CBS golden standard, outperforms previous S22 results obtained with others methods.

Table 1: MAE (kcal/mol) and MARE(%) relative to the S66x8 data set for our DNN-based models as well as for few other dispersion correction ones. For the MBD-based models, the method-specific range separation parameter reported in parentheses refers to the one optimised for the S66x6 set. MAE and MARE are computed taking revised CCSD(T) CBS energies.

Model	MAE[kcal/mol]	MARE%
PBE	1.55	65
PBE0	1.48	65
PBE+D3	0.44	n.a.
PBE+MBD@rsSCS ($\beta = 0.83$)	0.32	10.6
PBE0+MBD@rsSCS ($\beta = 0.85$)	0.30	9.2
PBE+MBD@rsSCS/FI ($\beta = 0.83$)	0.28	9.0
PBE+DNN-MBD ($\beta = 0.75$)	0.25	9.0
PBE0+DNN-MBD ($\beta = 0.77$)	0.23	6.9

Table 2: MAE (kcal/mol) and MARE(%) relative to the S22 data set for our DNN-based models as well as for few other dispersion correction ones. For the MBD-based models, the method-specific range separation parameter reported in parentheses refers to the one optimised for the S66x6 set. MAE and MARE are computed taking revised S22 energies where, compared to the original paper, a larger basis set was employed.⁵⁸

Model	MAE[kcal/mol]	MARE%
PBE	2.66	58
PBE0	2.44	55
PBE0+MBD@rsSCS ($\beta = 0.85$)	0.55	8.5
PBE+MBD@rsSCS ($\beta = 0.83$)	0.49	8.9
PBE+D3	0.48	n.a.
PBE0+DNN-MBD ($\beta = 0.77$)	0.43	5.6
PBE+DNN-MBD ($\beta = 0.75$)	0.41	6.6

Having been trained on a large and diverse set of AIM volumes, the outcoming DNN-MBD model inherits the strengths of the MBISA scheme discussed earlier in this Letter while completely bypassing the explicit density partitioning with a consequent decrease of the computational cost. We also note that the DNN model could be successfully trained with different AIM partitioning schemes due to the locality of the target quantities (volumes). The presented density-free DNN-SMBD model is included in the Tinker-HP package⁵⁹ and will be released with the next version of the software. There, it can benefit from the linear-scaling embarrassingly parallel performances of our stochastic formulation (SMBD) of the MBD key equations which remarkable computational performances have been recently discussed.³⁷

We believe that the present DNN-SMBD model can be beneficial in applications of dispersion-corrected DFT to large complex systems requiring an accurate yet extremely efficient inclusion of MBD effects. The DNN model, by avoiding the direct solution of the KS equations due to its electron density-free features, allows for the ready application of DNN-SMBD approach in the development of accurate *ab initio*-based force fields^{60,61} and neural networks methodologies.

Acknowledgement

This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 810367), project EMC2 (JPP). Computations have been performed at GENCI (IDRIS, Orsay, France and TGCC, Bruyères le Chatel) on grant no A0070707671.

Supporting Information Available

SI1 file contains training plots for DNN with different layers. SI2 and SI3 contain raw PBE, PBE+DNN-MBD, PBE0 and PBE0+DNN-MBD energies for the S66x8 and S22 data sets

respectively.

References

- (1) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (2) Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-corrected mean-field electronic structure methods. *Chemical Reviews* **2016**, *116*, 5105–5154.
- (3) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* **2004**, *25*, 1463–1473.
- (4) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27*, 1787–1799.
- (5) 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. *The Journal of Chemical Physics* **2010**, *132*, 154104.
- (6) Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: inclusion of higher-order corrections. *The Journal of Chemical Physics* **2006**, *124*, 174104.
- (7) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction revisited. *The Journal of Chemical Physics* **2007**, *127*, 154108.
- (8) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

- (9) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theoretica chimica acta* **1977**, *44*, 129–138.
- (10) Reilly, A. M.; Tkatchenko, A. Seamless and accurate modeling of organic molecular materials. *The Journal of Physical Chemistry Letters* **2013**, *4*, 1028–1033.
- (11) Ambrosetti, A.; Alfè, D.; DiStasio, R. A.; Tkatchenko, A. Hard numbers for large molecules: toward exact energetics for supramolecular systems. *The Journal of Physical Chemistry Letters* **2014**, *5*, 849–855.
- (12) Ambrosetti, A.; Ferri, N.; DiStasio, R. A.; Tkatchenko, A. Wavelike charge density fluctuations and van der Waals interactions at the nanoscale. *Science* **2016**, *351*, 1171–1176.
- (13) Stöhr, M.; Tkatchenko, A. Quantum mechanics of proteins in explicit water: the role of plasmon-like solute-solvent interactions. *Science Advances* **2019**, *5*, eaax0024.
- (14) Langbein, D. Microscopic calculation of macroscopic dispersion energy. *Journal of Physics and Chemistry of Solids* **1971**, *32*, 133–138.
- (15) Donchev, A. G. Many-body effects of dispersion interaction. *The Journal of Chemical Physics* **2006**, *125*, 074713.
- (16) Sommerfeld, T.; Jordan, K. D. Quantum Drude Oscillator Model for describing the interaction of excess electrons with water clusters: an application to (H₂O)₁₃-. *The Journal of Physical Chemistry A* **2005**, *109*, 11531–11538.
- (17) Jones, A. Quantum drude oscillators for accurate many-body intermolecular forces. Ph.D. thesis, University of Edinburgh, 2010.
- (18) Jones, A. P.; Crain, J.; Sokhan, V. P.; Whitfield, T. W.; Martyna, G. J. Quantum Drude oscillator model of atoms and molecules: many-body polarization and dispersion interactions for atomistic simulation. *Phys. Rev. B* **2013**, *87*, 144103.

- (19) Odbadrakh, T. T.; Jordan, K. D. Dispersion dipoles for coupled Drude oscillators. *The Journal of Chemical Physics* **2016**, *144*, 034111.
- (20) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
- (21) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *The Journal of Chemical Physics* **2014**, *140*, 18A508.
- (22) Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-theoretic approaches to atoms-in-molecules: Hirshfeld family of partitioning schemes. *The Journal of Physical Chemistry A* **2018**, *122*, 4219–4245.
- (23) Ayers, P. W. Atoms in molecules, an axiomatic approach. I. Maximum transferability. *The Journal of Chemical Physics* **2000**, *113*, 10886–10898.
- (24) Ayers, P. W.; Morrison, R. C.; Roy, R. K. Variational principles for describing chemical reactions: condensed reactivity indices. *The Journal of Chemical Physics* **2002**, *116*, 8731–8744.
- (25) Bučko, T.; Lebègue, S.; Ángyán, J. G.; Hafner, J. Extending the applicability of the Tkatchenko-Scheffler dispersion correction via iterative Hirshfeld partitioning. *The Journal of Chemical Physics* **2014**, *141*, 034114.
- (26) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *The Journal of Chemical Physics* **2007**, *126*, 144111.
- (27) Van Damme, S.; Bultinck, P.; Fias, S. Electrostatic potentials from self-consistent Hirshfeld atomic charges. *Journal of Chemical Theory and Computation* **2009**, *5*, 334–340.

- (28) Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Improved density dependent correction for the description of London dispersion forces. *Journal of Chemical Theory and Computation* **2013**, *9*, 4293–4299.
- (29) Deringer, V. L.; Csányi, G. Many-body dispersion correction effects on bulk and surface properties of rutile and anatase TiO₂. *The Journal of Physical Chemistry C* **2016**, *120*, 21552–21560.
- (30) Gould, T.; Lebègue, S.; Ángyán, J. G.; Bučko, T. A fractionally ionic approach to polarizability and van der Waals many-body dispersion calculations. *Journal of Chemical Theory and Computation* **2016**, *12*, 5920–5930.
- (31) Lillestolen, T. C.; Wheatley, R. J. Atomic charge densities generated using an iterative stockholder procedure. *The Journal of Chemical Physics* **2009**, *131*, 144101.
- (32) Misquitta, A. J.; Stone, A. J.; Fazeli, F. Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *Journal of Chemical Theory and Computation* **2014**, *10*, 5405–5418, PMID: 26583224.
- (33) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal basis iterative stockholder: atoms in molecules for force-field development. *Journal of Chemical Theory and Computation* **2016**, *12*, 3894–3912.
- (34) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science Advances* **2019**, *5*, eaav6490.
- (35) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific Data* **2020**, *7*, 134.

- (36) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced database of benchmark interaction energies relevant to biomolecular structures. *Journal of Chemical Theory and Computation* **2011**, *7*, 2427–2438.
- (37) Poier, P. P.; Lagardère, L.; Piquemal, J.-P. O(N) stochastic evaluation of many-body van der Waals energies in large complex systems. *Journal of Chemical Theory and Computation* **2022**, *18*, 1633–1645.
- (38) Bereau, T.; DiStasio, R. A.; Tkatchenko, A.; von Lilienfeld, O. A. Non-covalent interactions across organic and biological subsets of chemical space: physics-based potentials parametrized from machine learning. *The Journal of Chemical Physics* **2018**, *148*, 241706.
- (39) Muhli, H.; Chen, X.; Bartók, A. P.; Hernández-León, P.; Csányi, G.; Ala-Nissila, T.; Caro, M. A. Machine learning force fields based on local parametrization of dispersion interactions: Application to the phase diagram of C₆₀. *Phys. Rev. B* **2021**, *104*, 054106.
- (40) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. Performance and cost assessment of machine learning interatomic potentials. *The Journal of Physical Chemistry A* **2020**, *124*, 731–745.
- (41) Westermayr, J.; Chaudhuri, S.; Jeindl, A.; Hofmann, O. T.; Maurer, R. J. Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic-inorganic interfaces. 2022; <https://arxiv.org/abs/2202.13009>.
- (42) 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*, 3192–3203.
- (43) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

- (44) Gastegger, M.; Schwiedrzik, L.; Bittermann, M.; Berzsenyi, F.; Marquetand, P. wACSF—Weighted atom-centered symmetry functions as descriptors in machine learning potentials. *The Journal of Chemical Physics* **2018**, *148*, 241709.
- (45) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.
- (46) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., et al. Scikit-learn: machine learning in Python. *Journal of machine learning research* **2011**, *12*, 2825–2830.
- (47) Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. NIPS 2017 Workshop on Autodiff. 2017.
- (48) 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. *Journal of Chemical Information and Modeling* **2020**, *60*, 3408–3415.
- (49) Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (ELUs). 2015; <https://arxiv.org/abs/1511.07289>.
- (50) Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. 2017; <https://arxiv.org/abs/1711.05101>.
- (51) Poier, P. P.; Jaffrelot Inizan, T.; Adjoua, O.; Lagardère, L.; Piquemal, J.-P. ANI-1 dataset with added atomic volume ratios restricted to CHNO atoms for DNN-MBD. <https://doi.org/10.5281/zenodo.6397639>.
- (52) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

- (53) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.
- (54) Jensen, F. Polarization consistent basis sets: principles. *The Journal of Chemical Physics* **2001**, *115*, 9113–9125.
- (55) Jensen, F. Unifying general and segmented contracted basis sets. Segmented polarization consistent basis sets. *Journal of Chemical Theory and Computation* **2014**, *10*, 1074–1085.
- (56) Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking Density Functional Methods against the S66 and S66x8 Datasets for Non-Covalent Interactions. *ChemPhysChem* **2011**, *12*, 3421–3433.
- (57) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- (58) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. Basis set consistent revision of the S22 test set of noncovalent interaction energies. *The Journal of Chemical Physics* **2010**, *132*, 144104.
- (59) Lagardère, L.; Jolly, L.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J. P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956–972.
- (60) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, polarizable molecular mechanics studies of inter-, intra-molecular interactions, and ligand-macromolecule complexes. A bottom-up strategy. *Journal of Chemical Theory and Computation* **2007**, *3*, 1960–1986.

- (61) Naseem-Khan, S.; Lagardère, L.; Narth, C.; Cisneros, G. A.; Ren, P.; Gresh, N.; Piquemal, J.-P. Development of the Quantum Inspired SIBFA Many-Body Polarizable Force Field: Enabling Condensed Phase Molecular Dynamics Simulations. *J. Chem. Theory. Comput.* DOI: 10.1021/acs.jctc.2c00029 **2022**,