MBD-ML: Many-Body Dispersion from Machine Learning for Molecules and Materials

Evgeny Moerman, Adil Kabylda, Almaz Khabibrakhmanov, and Alexandre Tkatchenko*
Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg

(Received 2 March 2026; accepted 10 June 2026; published 28 July 2026)

van der Waals (vdW) interactions are essential for describing molecules and materials, from drug design and catalysis to battery applications. These omnipresent interactions must also be accurately included in machine-learned (ML) force fields. The many-body dispersion (MBD) method stands out as one of the most accurate and transferable approaches to capture vdW interactions, requiring only atomic C_6 coefficients and polarizabilities as input. We present MBD-ML, a pretrained message-passing neural network that predicts these atomic properties directly from structures with demonstrated transferability across molecular systems and organic condensed phases. Through seamless integration with libMBD, our method enables the immediate calculation of MBD-inclusive total energies, forces, and stress tensors. By eliminating the need for intermediate electronic-structure calculations, MBD-ML offers a practical and streamlined tool that simplifies the incorporation of state-of-the-art vdW interactions into any electronic-structure code, as well as empirical and machine-learned force fields.

DOI: 10.1103/pv86-l9h7

I. INTRODUCTION

van der Waals (vdW) dispersion interactions play a decisive role in the properties of molecular crystals, condensed matter, and biological systems [1,2]. They govern polymorphism in organic crystals [3,4], protein folding [5], and the cohesion of layered, polymeric, and porous materials [6,7]. In many such systems, dispersion interactions are not a small perturbation but a leading contribution that critically determines relative energies, equilibrium geometries, and atomistic dynamics.

Density functional theory (DFT) has become a cornerstone of atomistic simulations of molecules and solids, and increasingly serves as a reference [8–10] for the construction of machine-learning force fields (MLFFs) [11]. However, most widely used semilocal and hybrid exchange-correlation (xc) functionals, as well as local ML architectures, do not account for long-range vdW dispersion. As a consequence, explicit dispersion interactions are indispensable for achieving quantitative accuracy, unless one deals with systems where strong covalent bonds dominate the desired behavior [11,12].

Over the past two decades, a wide range of vdW methods have been proposed. Early and still widely adopted approaches include the pairwise-additive DFT-D methods [13], most notably the D3 correction by Grimme [14], based on tabulated C_6 coefficients and polarizabilities α_0. The Tkatchenko-Scheffler (TS) [15] method introduced an important conceptual advance by making these quantities environment-dependent through Hirshfeld partitioning of the electron density. Furthermore, in contrast to empirical methods, the TS method, as well as its further extensions, is a self-consistent interatomic vdW density functional [16]. Similarly, the exchange-hole dipole moment (XDM) method [17,18] derives atomic polarizabilities from the electronic density and incorporates higher-order multipole contributions to the dispersion energy through C_8 and C_{10} coefficients obtained from the exchange-hole multipole moments.

An alternative strategy for capturing long-range correlation within DFT is to use nonlocal xc functionals in place of semilocal correlation, thereby incorporating vdW effects without a separate correction step. The vdW density functional (vdW-DF) family [19,20] achieves this by constructing the correlation energy from a double spatial integral over a nonlocal kernel derived from a plasmon-pole model of the electron response, with the consistent-exchange vdW-DF-cx [21] variant reflecting the current state of the art among nonhybrid vdW-DFs [22]. The VV10 functional [23] follows a related but simplified response framework. However, vdW-DFs fundamentally require access to the electronic density, which makes them inapplicable in force-field contexts and limits their portability across simulation frameworks. Moreover, the nonlocal kernel in vdW-DF is constructed from a local approximation to the dielectric response at each point in space, neglecting screening effects and correlations beyond two-point level [1,24].

A fundamental limitation of all pairwise schemes is their inability to capture collective many-body effects in dispersion interactions [1,2]. This shortcoming is addressed by the many-body dispersion (MBD) method [25], which models dispersion interactions using a system of dipole-coupled quantum harmonic oscillators parametrized to reproduce atomic α_0 and C_6 coefficients. The MBD framework naturally accounts for nonadditive many-body effects as well as polarization anisotropy [26], and has been shown to substantially improve accuracy for molecular crystals, layered materials, and large molecular systems [27–31]. Several

*Contact author: alexandre.tkatchenko@uni.lu

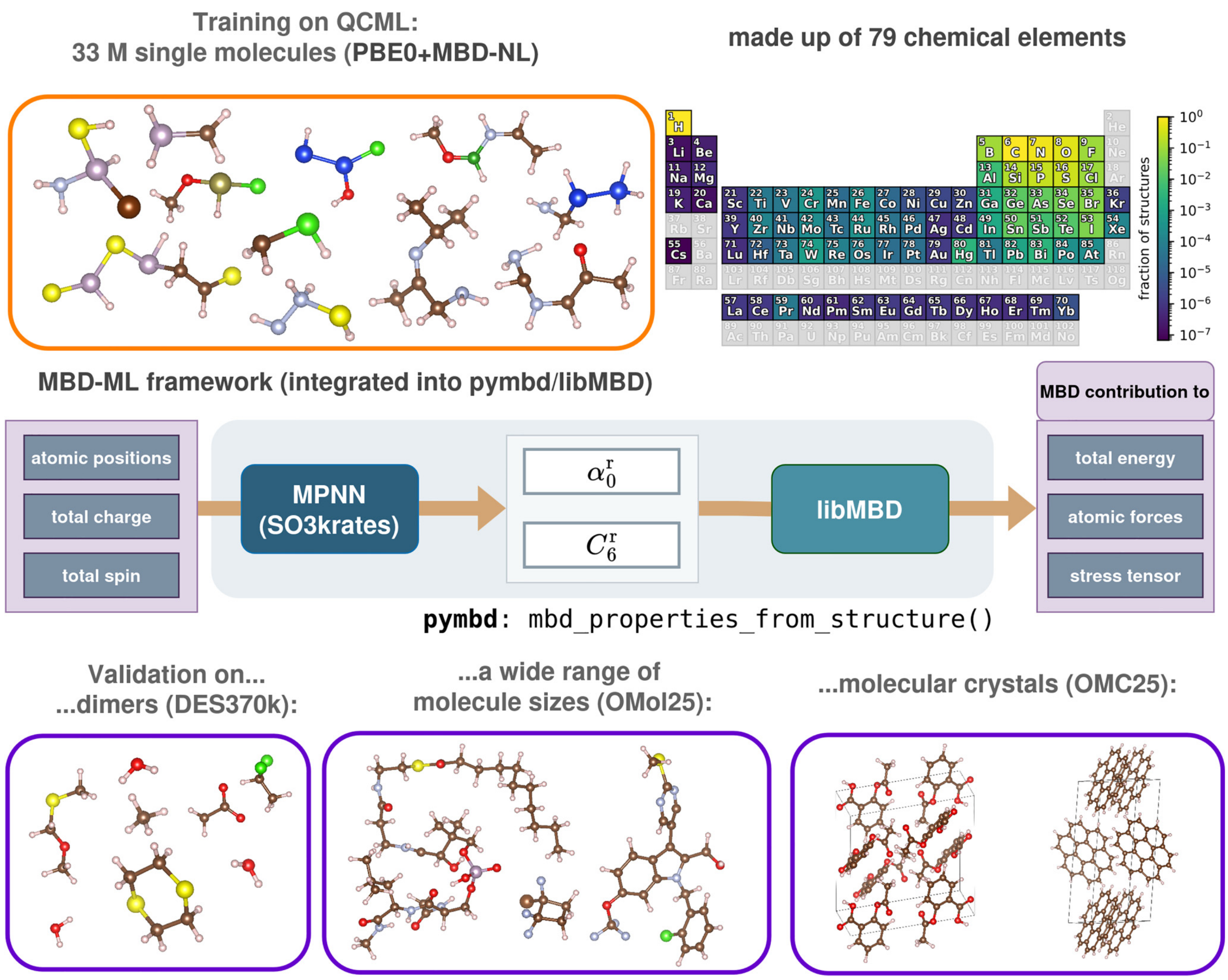

FIG. 1. Overview of the training and the validation of the MBD-ML model: (top left) MBD-ML was trained on the QCML molecular dataset spanning almost the entire periodic system (top right; reproduced from Ref. [44] under CC BY 4.0). (Middle) The model has been integrated into the libMBD python interface, pymbd, to allow direct calculation of MBD properties from the atomic structure. (Bottom) The model was validated on both molecules and molecular crystals in predicting fundamental properties including total energy and force contributions. For organic crystals, it was further used to perform geometry optimizations and recover the energy ranking of different polymorphs.

variants of MBD, differing in their parametrization and treatment of short-range screening, have been developed [27,30–33], including the state-of-the-art MBD-NL formulation [31].

Despite this rich variety of vdW methods and the demonstrated accuracy of many-body approaches, the D3 correction remains the most widely used in large-scale and high-throughput applications. This includes generation of DFT-based datasets for molecules [9] and materials [8], as well as routine applications in computational chemistry and materials science. While the limitations of D3 are well documented [30,34–40], its popularity is largely driven by its ease of use. D3 is implemented as a stand-alone library and requires only atomic positions and element types, without the need for electronic-structure information or tight coupling to DFT codes. For the MBD method, however, the reliance on electronic-structure calculations to parametrize the MBD Hamiltonian remains a major bottleneck that limits its applicability in large-scale simulations and ML architectures.

In this work, we remove this limitation for the MBD method by introducing a machine-learning-based framework, termed MBD-ML, that enables accurate many-body dispersion calculations using only atomic geometry as input. We achieve this by training a state-of-the-art message-passing neural network to predict the effective atomic polarizabilities α_0 and C_6 coefficients, using the DFT + MBD-NL method [31] as a reference. By integrating this model into the libMBD library [41] (see Fig. 1), we make it possible to compute dispersion energies, atomic forces, and stress tensors at the MBD-NL level of accuracy without performing any electronic-structure calculations. Previous machine-learned MBD models [42] employed simple deep-learning architectures to predict Hirshfeld volume ratios for the antecedent MBD@rsSCS method and were limited to small molecules containing C, H, N, and O. In contrast, the present MBD-ML model is applicable to both molecules and crystals across over 70 chemical elements and is directly integrated into the libMBD infrastructure. Moreover, α_0 and C_6 ratios from MBD-NL, which we used for training, are more robust and transferable across chemical space, facilitating treatment of ionic and metallic compounds, which are problematic for MBD@rsSCS relying on Hirshfeld volume ratios [32,43].

We demonstrate the transferability and general performance of the MBD-ML framework on subsets of five state-of-the-art datasets (see Fig. 1) covering small molecules of widely varying compositions (QCML [44]) and sizes (OMol25 [9]), biomolecular and druglike dimers (DES370k [45]), molecular crystals (OMC25 [46]), and inorganic materials (OMat24 [8]). Beyond that, we assess the robustness and practical accuracy of MBD-ML by applying the method to the prediction of structures and relative energetics of molecular-crystal polymorphs, and comparing the results directly against reference *ab initio* MBD-NL calculations. Our results demonstrate that MBD-ML retains the accuracy of MBD-NL across molecular systems and organic condensed phases while dramatically extending its usability to large-scale ML-based data generation and atomistic modeling. We further delineate the current domain of applicability of the model, identifying inorganic extended materials as the regime where generalization is not yet reliably possible and outlining directions for future work.

II. METHODS

A. The MBD framework

The MBD method [25] employs coupled quantum Drude oscillators (QDOs) [47,48] to model collective electron density fluctuations, underpinning vdW dispersion. In this framework, the electronic response of each atom is represented by a single QDO, which captures the collective fluctuations of all valence electrons. Each QDO consists of a negatively charged quasiparticle harmonically bound to a positively charged pseudonucleus. The mapping between atoms and oscillators is established via the static polarizability $\alpha_{0,i}$ and effective excitation frequency $\omega_i = 4C_{6,ii}/3\alpha_{0,i}^2$.

The interaction between oscillators is described via the long-range dipole-dipole coupling tensor $\mathbf{T}^{\mathrm{lr}}(\mathbf{R}) = f_{\mathrm{lr}}(R)\mathbf{T}(\mathbf{R})$, using the MBD@rsSCS long-range coupling

$$f_{\mathrm{lr}}(R_{ij}) = \left(1 + e^{-6[R_{ij}/\beta(R_i^{\mathrm{vdW}}+R_j^{\mathrm{vdW}})-1]}\right)^{-1}, \tag{1}$$

where R_{ij} denotes the distance between any two atoms i and j, and R_i^{vdW} and R_j^{vdW} are the atomic vdW radii, which are computed from the static polarizabilities and free-atom polarizability reference values $\alpha_{0,i}^{\mathrm{ref,free}}$:

$$R_i^{\mathrm{vdW}} = \frac{5}{2}\left(\alpha_{0,i}^{\mathrm{ref,free}}\right)^{1/7}\left(\frac{\alpha_{0,i}}{\alpha_{0,i}^{\mathrm{ref,free}}}\right)^{1/3}. \tag{2}$$

The resulting Hamiltonian for a system of N coupled QDOs is given by

$$H_{\mathrm{MBD}}(\{\mathbf{R}_i, \alpha_{0,i}, \omega_i\}) = \sum_i -\frac{1}{2}\nabla_{\boldsymbol{\xi}_i}^2 + \frac{1}{2}\omega_i^2\boldsymbol{\xi}_i^2 + \frac{1}{2}\sum_{ij}\omega_i\omega_j\sqrt{\alpha_{0,i}\alpha_{0,j}}\boldsymbol{\xi}_i\mathbf{T}_{ij}^{\mathrm{lr}}\boldsymbol{\xi}_j, \tag{3}$$

with $\boldsymbol{\xi}_i = \sqrt{m_i}\mathbf{x}_i$ denoting the mass-weighted displacements of the oscillating charge relative to the nuclei position $\mathbf{R}_i$.

Since the Hamiltonian (3) is quadratic in displacements, the many-body dispersion energy can be obtained by direct diagonalization:

$$E_{\mathrm{MBD}} = \sum_{k=1}^{3N}\frac{\tilde{\omega}_k}{2} - \sum_{i=1}^{N}\frac{3\omega_i}{2}, \tag{4}$$

where $\tilde{\omega}_k$ are the frequencies of the $3N$ collective MBD eigenmodes. The resulting MBD energy is formally equivalent to the adiabatic-connection fluctuation-dissipation correlation energy [26], while the Hamiltonian diagonalization provides a numerically efficient route to its evaluation. The computational complexity of the MBD method scales as $\mathcal{O}(N^3)$, matching that of Kohn-Sham DFT. However, we emphasize that the prefactor is substantially smaller, such that the cost of an MBD calculation constitutes only a small fraction of a single DFT self-consistency cycle, even for large systems [41].

Within this general formalism, several variants of the MBD method have been developed, which primarily differ in the determination of the atomic polarizabilities $\alpha_{0,i}$, dispersion coefficients $C_{6,ii}$, and the long-range interaction tensor $\mathbf{T}^{\mathrm{lr}}$. Among these, the nonlocal MBD approach (MBD-NL) [31] represents the most advanced and robust formulation, exhibiting uniformly high accuracy across covalent, ionic, and metallic systems. This is achieved by using the Vydrov-Van Voorhis (VV) polarizability functional [49] to obtain $\alpha_{0,i}^{\mathrm{VV}}$ and $C_{6,ii}^{\mathrm{VV}}$ coefficients that parametrize the MBD Hamiltonian.

A key feature of MBD-NL is the explicit removal of contributions from jellium-like density regions [31]. This cutoff is essential to avoid double counting of correlation energy, as most exchange-correlation functionals are exact for the homogeneous electron gas by construction. Furthermore, to compensate for the approximate nature of the VV polarizability functional, the resulting $\alpha_{0,i}^{\mathrm{VV}}$ and $C_{6,ii}^{\mathrm{VV}}$ are renormalized using high-quality free-atom reference data:

$$\alpha_{0,i} = \frac{\alpha_{0,i}^{\mathrm{VV}}}{\alpha_{0,i}^{\mathrm{VV,free}}}\alpha_{0,i}^{\mathrm{ref,free}}, \qquad C_{6,ii} = \frac{C_{6,ii}^{\mathrm{VV}}}{C_{6,ii}^{\mathrm{VV,free}}}C_{6,ii}^{\mathrm{ref,free}}, \tag{5}$$

where $\alpha_{0,i}^{\mathrm{ref,free}}$ and $C_{6,ii}^{\mathrm{ref,free}}$ denote the corresponding free-atom reference values. Further technical details of the MBD-NL formalism and its implementation can be found in the original MBD-NL publication [31].

The combination of the VV polarizability functional, the jellium cutoff, and normalization to free-atom reference data yields a robust and transferable parametrization of the MBD Hamiltonian across diverse chemical environments. In particular, the MBD-NL formulation avoids the "polarization catastrophe" observed for earlier MBD variants in transition-metal compounds [32,43] and significantly reduces overbinding error in metals and ionic solids [31], achieving balanced accuracy across the periodic table.

B. The MBD-ML framework

Building on the MBD-NL formalism described above, we now introduce the MBD-ML model, which learns to reproduce its key atomic inputs, α_0^{r} and C_6^{r}, directly from the atomic configuration. MBD-ML utilizes the SO3krates architecture

TABLE I. Performance of MBD-ML models on ratios and MBD properties for different test sets in terms of root mean square error (RMSE) and mean absolute error (MAE) values. Errors for α_0^{r} and C_6^{r} are unitless. N denotes the number of systems in each test set.

	QCML[a] ($N = 83\,226$)		DES370k[a,b] ($N = 309\,130$)		OMOL25[b] ($N = 825$)		OMC25 ($N = 200$)		OMAT24 ($N = 203$)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Ratio predictions										
α_0^{r}	0.013	0.020	0.014	0.025	0.023	0.033	0.025	0.031	0.247	0.429
C_6^{r}	0.014	0.023	0.013	0.022	0.021	0.030	0.017	0.022	0.269	0.895
MBD properties										
E_{MBD} (meV/atom)	0.158	0.228	0.117	0.165	0.198	0.226	0.713	0.841	–	–
F_{MBD} (meV/Å)	0.302	0.440	0.266	0.387	0.564	0.707	0.786	0.930	–	–
σ_{MBD} (meV/Å^3)	–	–	–	–	–	–	0.112	0.143	–	–

[a]Excluding negatively charged molecules (due to insufficient accuracy of *ab initio* treatment; see QCML discussion in Sec. III).
[b]Excluding alkali and alkaline earth metals (due to underrepresentation in the training set; see DES370k discussion in Sec. III).

[50], an equivariant message-passing neural network that represents a general and flexible framework for training state-of-the-art MLFFs. A closely related model is the recent SO3LR [12], which was trained to map atomic positions R, atomic numbers Z, and the total charge Q and spin S to atomic forces $\boldsymbol{F}_i$, partial charges q_i, and Hirshfeld ratios h_i. Here, we modified and retrained the model to predict the α_0^{r} and C_6^{r} ratios:

$$\alpha_0^{\mathrm{r}}, C_6^{\mathrm{r}} = \mathrm{SO3krates}(R, Z, Q, S), \tag{6}$$

defined as the ratios between the $C_{6,ii}^{\mathrm{VV}}$ and $\alpha_{0,i}^{\mathrm{VV}}$ atomic coefficients computed by the VV polarizability functional in a many-atom system and their free-atom equivalents $C_{6,ii}^{\mathrm{VV,free}}$ and $\alpha_{0,i}^{\mathrm{VV,free}}$,

$$\alpha_{0,i}^{\mathrm{r}} = \frac{\alpha_{0,i}^{\mathrm{VV}}}{\alpha_{0,i}^{\mathrm{VV,free}}}, \quad C_{6,ii}^{\mathrm{r}} = \frac{C_{6,ii}^{\mathrm{VV}}}{C_{6,ii}^{\mathrm{VV,free}}}, \tag{7}$$

which in the following will be simply referred to as C_6^{r} and α_0^{r} for brevity.

In contrast to polarizabilities and dispersion coefficients, broadly varying in magnitude across the periodic table, these unitless ratios are typically confined in the range $\sim$0–2 for most elements and chemical environments, which makes α_0^{r} and C_6^{r} naturally suitable targets for machine learning. We note that the MBD-ML model predicts α_0^{r} and C_6^{r} separately, while within the SO3LR model the analogous quantities are directly connected ($\alpha_{0,i}^{\mathrm{r}} = h_i$ and $C_{6,i}^{\mathrm{r}} = h_i^2$). This important distinction originates from the underlying MBD-NL method and makes the resulting MBD predictions more robust for general systems, as it was explained above. The values of $\alpha_{0,i}$ and $C_{6,ii}$, required to parametrize the MBD Hamiltonian (3), are obtained by multiplying the predicted ratios by the free-atom reference quantities according to Eq. (7). This is done internally within the libMBD interface and hence does not require any user intervention. Thus, the MBD-ML-predicted ratios are sufficient to directly compute all MBD-related quantities, including energies, atomic forces, and stress tensors, via the efficient libMBD routines [41].

Training on the atomic ratios rather than directly on the MBD properties also offers another key advantage—α_0^{r} and C_6^{r} ratios are relatively short-ranged quantities, whereas energies and forces can be very sensitive to long-range interactions. Consequently, a semilocal message-passing architecture like SO3krates is well suited for this application. In addition, the α_0^{r} and C_6^{r} ratios are not very sensitive to the underlying density functional approximation (DFA), so that application of the model in combination with other exchange-correlation functionals only requires the modification of the damping parameter β, governing the short-range behavior of the dipolar tensor $\mathbf{T}^{\mathrm{lr}}$. In the present work, β was not re-optimized for the MBD-ML model and the tabulated default values were used for the Perdew-Burke-Ernzerhof (PBE) and the PBE0 DFAs ($\beta_{\mathrm{PBE}} = 0.81$, $\beta_{\mathrm{PBE0}} = 0.83$) following Ref. [31].

The current MBD-ML implementation uses SO3krates with two message-passing layers and a cutoff radius of 4 Å, giving an effective receptive field of 8 Å, and was trained on the QCML dataset [44], comprising over 30 million molecules with up to eight nonhydrogen atoms across 79 chemical elements. QCML provides structures, energies, and forces at the PBE0 + MBD-NL level, along with MBD-NL polarizabilities $\alpha_{0,i}^{\mathrm{VV}}$ and $C_{6,ii}^{\mathrm{VV}}$ coefficients, enabling direct training of the MBD-ML model. We refer the reader to Sec. S1 of the Supplemental Material [51] for the technical details of the training procedure.

III. MODEL VALIDATION

To quantify the accuracy and identify potential limitations of the MBD-ML model in predicting MBD corrections to total energies, atomic forces, and stress tensors (for periodic systems), we first evaluate the trained model on three test sets spanning different chemical compound spaces. Beyond the hold-out validation on the QCML data, we assess performance on molecular noncovalent dimers from DES370k [45], and on molecular crystals using a subset of OMC25 [46]. Details on the selection and recalculation of the OMol25 and DES370k subsets can be found in Secs. S2 and S3 [51], respectively.

A. Molecular datasets

QCML. The MBD-ML model trained on the QCML dataset achieves strong performance for neutral and positively charged molecules (Table I and Fig. 2), with RMSEs of 0.020 and 0.023 for α_0^{r} and C_6^{r} ratios, respectively. These low errors

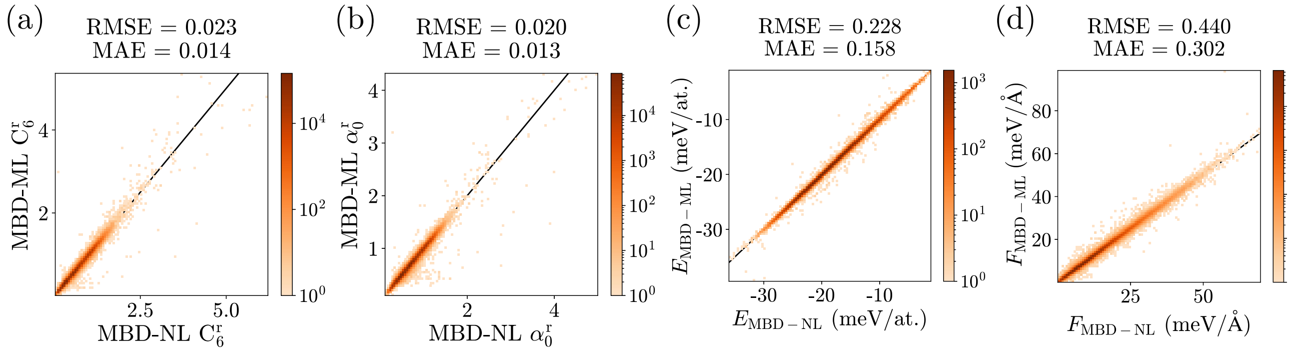

FIG. 2. Performance of PBE0 + MBD-ML in predicting the C_6 and α_0 ratios and the MBD contribution to the total energy and atomic forces of the QCML test set. MAE and RMSE values for energies and forces are given in meV/atom and meV/Å, respectively. Color indicates point density.

demonstrate that MBD-ML accurately captures the quantum mechanical effects that modulate atomic polarizabilities and dispersion coefficients in diverse chemical environments. The accurately predicted ratios yield sub-meV errors in the MBD energy (0.228 meV/atom) and force (0.440 meV/Å) contributions (Fig. 2), indicating its suitability as an accurate drop-in replacement of the *ab initio* MBD-NL method for routine electronic-structure or MLFF calculations.

Additionally, validation on the hold-out test set revealed that some negatively charged molecules are inadequately described by the employed *ab initio* treatment. The α_0^r and C_6^r ratios vary substantially with increasing confinement potential cutoff radius in FHI-AIMS (see Sec. S4.1 [51]), indicating that the underlying VV polarizability functional is highly sensitive to electronic density tails. This sensitivity arises because additional electrons in anions are often weakly bound and extremely delocalized, resulting in a far-reaching, slowly decaying electron density. Moreover, our random inspection of several pathological cases revealed that all of those anions had negative electron affinities, meaning that they are not stable thermodynamically (please consult Sec. S4 for the results [51]). For such systems, a bound solution only exists because of the finite basis set span. For these reasons, we decided to exclude negatively charged molecules from the validation, and we do not recommend using the QCML data for their vdW energies and ratios until rigorous *ab initio* benchmarks and stability-based filtering are performed. Interestingly, including anionic molecules deteriorates prediction accuracy for MBD ratios (see Sec. S4.2 [51]), with errors exceeding one order of magnitude, yet MBD energies and forces remain nearly unchanged (see Fig. S5 [51]).

Beyond these well-known issues of converging the electronic structure of molecular anions, charged systems are generally challenging for the state-of-the-art vdW methods (MBD-NL, XDM, DFT-D4) [52], in particular, charged-neutral interactions. For the MBD case, the observed interaction energy disagreements arise from the underlying DFA and can be mitigated, e.g., by combining MBD with density-corrected DFAs [53].

DES370k. To validate the MBD-ML model beyond single molecules, we tested it on molecular dimers from the DES370k dataset [45], recomputed at the PBE0 + MBD-NL level as a part of the QCell biomolecular dataset [54]. For 309 130 dimers excluding negatively charged systems and structures containing alkali or alkaline earth elements (see below), MBD-ML achieves prediction errors for MBD ratios comparable to the QCML hold-out set, with slightly improved accuracy for MBD energies and forces (RMSEs of 0.165 meV/atom and 0.387 meV/Å, respectively; see Table I). Given the substantial differences in chemical compound space between QCML (diverse small molecules) and DES370k (biomolecular and druglike dimers), the MBD-ML model proves transferable and reliable beyond its training distribution.

A more in-depth analysis of the validation results shows that accuracy substantially deteriorates for structures containing alkali (Li, Na, K) and alkaline earth elements (Be, Mg, Ca), which are severely underrepresented in the QCML training set [44]. Each element appears in only 10^{-5}% to 10^{-4}% of the 33.5 million QCML structures (approximately 3–30 structures per element), making them the rarest elements alongside the lanthanides. Consequently, we exclude these elements from our assessment. A complete overview of MBD-ML performance including these problematic cases is provided in Sec. S5 [51]. Hence, we recommend using the default version of MBD-ML with care when dealing with such underrepresented elements. However, this limitation can be overcome by fine-tuning the model on a specific dataset of interest.

B. Molecular crystals

To further test the extrapolating ability of the MBD-ML model, a random set of 200 molecular-crystal structures from the recently published OMC25 dataset was chosen and recomputed at the PBE0 + MBD-NL level of theory. To maximize the diversity of the subset, it was ensured that all 200 crystal structures featured unique molecules.

Despite being trained exclusively on small molecules, the MBD-ML predictions faithfully reproduce the PBE0 + MBD-NL reference. The ratios are predicted with an RMSE of 0.02–0.03, and the resulting MBD energies and forces are accurate to within 1 meV/atom and 1 meV/Å/atom, respectively, as illustrated in Fig. 3. While the magnitude of MBD contributions to the stress tensor σ_{MBD} ranges only between 0 and 15 meV/Å^3, the MBD-ML model predicts both diagonal and off-diagonal stress components with a very high accuracy, yielding an RMSE of 0.143 meV/Å^3 (see Fig. 3).

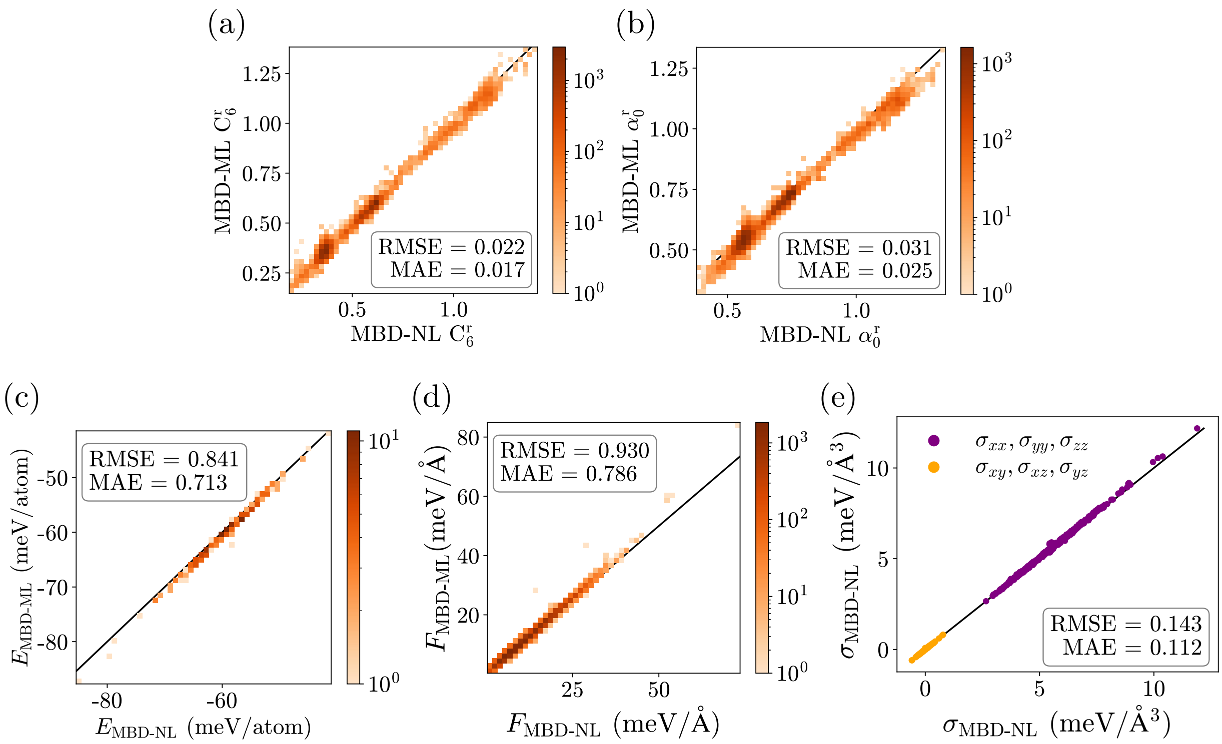

FIG. 3. Performance of PBE0 + MBD-ML in predicting the C_6 and α_0 ratios and the MBD contribution to the total energy, atomic forces, and stress tensor components in the OMC25 test set.

It is important to note that molecular-crystal unit cells often contain more than 100 atoms; consequently, even a modest per-atom error of 0.8 meV can translate into a total-energy deviation exceeding 80 meV. However, this consideration only applies to total energies, as such errors cancel out almost exactly for energy differences, resulting in correct polymorph rankings, as shown in the relevant section below. To elucidate the origin of this deviation, we performed a sensitivity analysis of the MBD energy with respect to the C_6^r and α_0^r ratios. We found that both systematic bias and random noise in these ratios influence the resulting MBD energy, with the effect of bias being approximately twice as large as that of noise. Taken together, these contributions account for the overall deviation of the MBD-ML energy. Details of this analysis are provided in Sec. S6 [51].

Overall, the MBD-ML model demonstrates robust accuracy and transferability across chemical environments. Although trained exclusively on small single-molecule structures from the QCML dataset, the predictive accuracy of the model extends to periodic molecular crystals containing multiple molecules of varying complexity.

IV. MBD-ML FORCES FOR VARYING MOLECULE SIZES

Having validated MBD-ML on fixed chemical environments, we now systematically examine whether its force accuracy degrades with increasing molecular size and benchmark it against the widely used pairwise D3 and D4 corrections. We selected 825 molecules from the OMol25 dataset [9] with sizes uniformly distributed from 3 to 350 atoms. To avoid artifacts, only uncharged, spin-non-polarized molecules were included, excluding systems containing alkali or alkaline earth elements (performance including these rare-in-data elements is reported in Sec. S5 [51]). Atomic forces were computed at the PBE0 + MBD-NL level and compared to predictions from MBD-ML, D3 [14,55], and D4 [56]. Figure 4 shows magnitude and angular deviations relative to MBD-NL forces, color coded by the magnitude of MBD-NL force vectors. An in-depth statistical analysis appears in Fig. S12 and Table S2 [51].

MBD-ML forces show excellent agreement with MBD-NL references, consistent with QCML and DES370k results. The median and mean magnitude deviations are 2%–3% with a 95th percentile of 7.7%. Angular deviations average <2° with a 95th percentile of 5.1°. While Fig. 4 shows a few MBD-ML outliers with angular deviations exceeding 150°, these occur exclusively when reference forces are nearly zero (deep purple), making such deviations practically irrelevant.

In contrast, D3 and D4 are strikingly different, with deviations in magnitude and angle being approximately ten times larger than for MBD-ML—28.2% and 17.4° for D3 and 39.5% and 19.8° for D4, respectively. This is consistent with previously reported disagreements between MBD and pairwise methods for atomic chains [57], noncovalent dimers [58], and polymers [59]. This qualitative difference also manifests itself in extreme outliers (force magnitude deviations exceeding 100%). D3 produces 3066 such outliers (2.4% of all atoms) and D4 produces 4025 (3.2%), compared to only 8 atoms (0.006%) for MBD-ML.

Interestingly, D4 shows larger deviations from MBD-NL for both force magnitudes and angles than D3. The polar representation in Fig. 4 reveals an important qualitative difference: While both methods show substantially broader deviation distributions than MBD-ML, in the case of D4

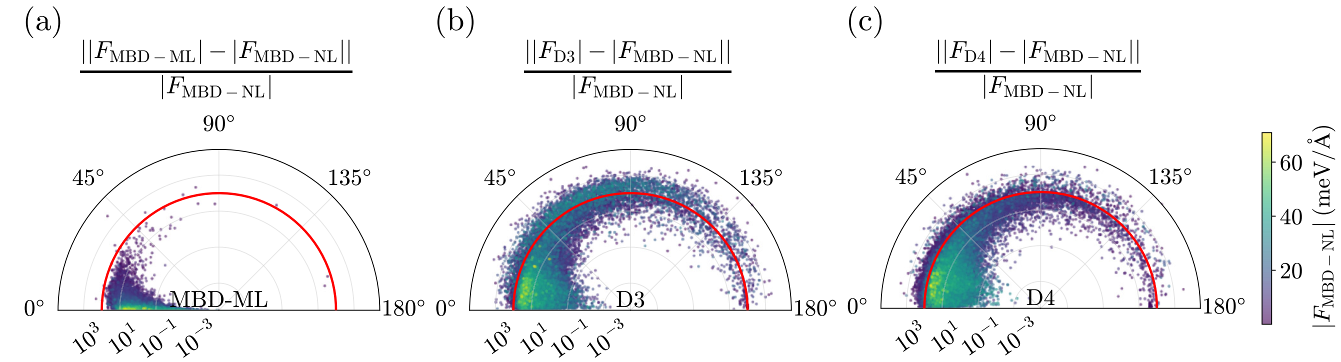

FIG. 4. Comparison of MBD-ML and DFT-D performance on predicting vdW force contributions relative to MBD-NL. Polar representation of atomic force deviations computed from the MBD-ML [panel (a)], D3 [panel (b)], and D4 methods [panel (c)]. The radial coordinate specifies the relative magnitude deviation of the atomic force vector to the MBD-NL reference in percent. The angular coordinate corresponds to the angular deviation from the reference in degrees. The point color encodes the absolute magnitude of the MBD-NL atomic force. As a guide to the eye, the radial 100% mark is highlighted in red.

large angular deviations exceeding 65° predominantly occur for very small force magnitudes (deep purple markers). In contrast, D3 forces exhibit sizable angular deviations for a higher fraction of large vdW forces (green markers) as well. This distinction is critical, as deviations in larger forces are more consequential for practical applications.

These results demonstrate that MBD-ML closely reproduces the MBD-NL reference across diverse molecular sizes, while emphasizing the importance of many-body effects beyond pairwise approximations. The substantial deviations of DFT-D methods from MBD-NL, particularly for larger forces where accuracy is most critical, underscore the importance of many-body dispersion effects that pairwise approaches fundamentally cannot capture.

V. MOLECULAR-CRYSTAL STRUCTURE AND STABILITY

Predicting the relative stability of molecular-crystal polymorphs is among the most demanding practical tests for a dispersion method, as energy differences between forms often fall below 1 kJ/mol per molecule. To assess MBD-ML in this context, we optimized geometries and computed energy rankings at the DFT + MBD-ML level of theory for a set of crystals spanning hydrogen-bonded, van der Waals, and mixed interaction types (see Table S3 and Sec. S8 [51]).

A. Geometry optimization

We first assess the quality of the optimized structures before evaluating their energy rankings. PBE + MBD-ML closely reproduces PBE + MBD-NL [60] crystal structures, with root mean square deviations (RMSDs) of atomic positions ranging from 0.001 to 0.05 Å for most structures (Fig. 5). Unit cell volumes differ by less than 1% from the *ab initio* reference, demonstrating that MBD-ML forces and stresses are sufficiently accurate to yield virtually identical equilibrium geometries. These small deviations are within typical uncertainties of experimental crystal structures and DFT geometry optimizations, confirming MBD-ML's reliability for structural predictions. In contrast, uncorrected PBE substantially overestimates unit cell volumes by over 20% due to missing attractive dispersion interactions, with RMSDs one to two orders of magnitude higher than MBD-ML (Fig. 5). This comparison underscores the necessity of correctly accounting for dispersion interactions in molecular crystals.

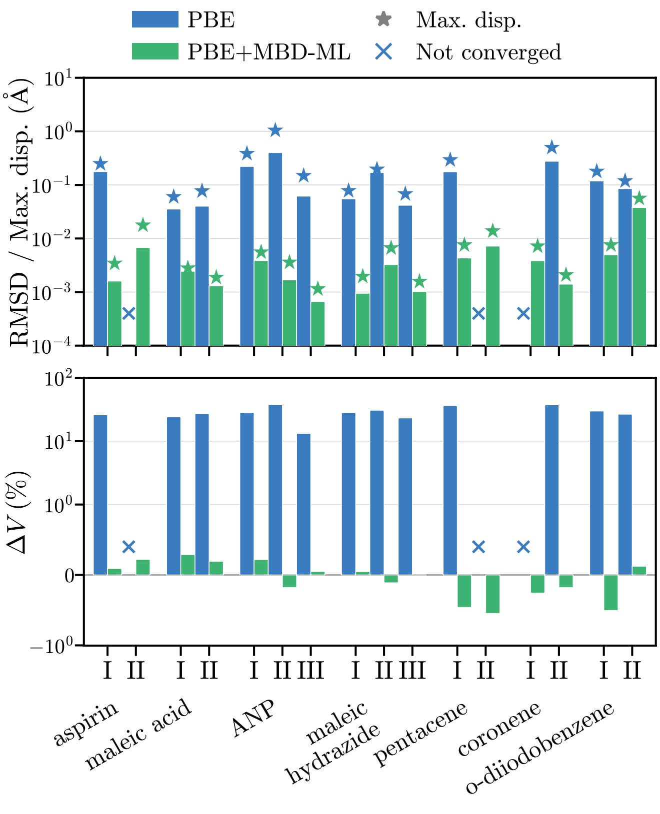

FIG. 5. Structural comparison of polymorphs relaxed with PBE and PBE + MBD-ML compared to the PBE + MBD-NL relaxed structures: (top) RMSD and maximum atomic displacement and (bottom) ratio of the unit cell volume of the relaxed structures with respect to the PBE + MBD-NL-relaxed structure. Blue crosses represent cases where the PBE geometry optimization did not converge. The RMSD and volume ratio is shown for every polymorph of each molecular crystal. ANP refers to 2-amino-5-nitropyrimidine. Numerical values are tabulated in Table S4 [51].

TABLE II. Polymorph energy differences (kJ/mol per molecule) from PBE0 + MBD-NL ($\Delta E_{\text{MBD-NL}}$) and PBE0 + MBD-ML ($\Delta E_{\text{MBD-ML}}$). Roman numerals indicate the polymorphic forms between which the energy difference is computed (e.g., I → II denotes $E_{\text{II}} - E_{\text{I}}$). Systems are grouped by dominant intermolecular interaction. ANP refers to 2-amino-5-nitropyrimidine. Incorrect rankings likely due to incomplete cancellation of total-energy errors are in bold.

Molecule		$\Delta E_{\text{MBD-NL}}$ (kJ/mol)	$\Delta E_{\text{MBD-ML}}$ (kJ/mol)
Mixed interactions			
Aspirin	I → II	−0.262	−0.291
Hydrogen bonded			
Maleic acid	I → II	0.670	0.467
ANP	I → II	0.183	**−0.292**
	I → III	0.237	0.261
Maleic hydrazide	I → II	0.974	0.870
	I → III	−0.384	**0.220**
van der Waals			
Pentacene	I → II	−0.476	−0.622
Coronene[a]	I → II	0.272	0.057
o-Diiodobenzene	I → II	0.029	0.015

[a]Structures were taken from Ref. [61].

B. Energy ranking

We evaluate MBD-ML's reliability in predicting the energetic ordering of polymorphs following the established protocol by Hoja *et al.* [4]. Energies of structures optimized at the PBE + MBD-NL and PBE + MBD-ML levels were recomputed using PBE0 + MBD-NL and PBE0 + MBD-ML, respectively, to determine lattice energy differences. Since MBD-ML is designed as a cost-effective replacement for *ab initio* MBD-NL, we use PBE0 + MBD-NL as our reference rather than experimental data, which would require free energy calculations with vibrational entropic contributions beyond the scope of this work.

For seven of nine polymorph transitions studied, MBD-ML correctly predicts the energy ranking with errors below 0.3 kJ/mol per molecule (Table II). These systems span diverse interaction types (van der Waals, hydrogen bonding, and mixed), demonstrating MBD-ML's robustness across chemical environments. Given that typical polymorph energy differences are on the order of 0–7 kJ/mol [62], the MBD-ML errors are well within the range required for reliable polymorph screening.

Two cases exhibit larger errors resulting in incorrect rankings: the I → II transition of 2-amino-5-nitropyrimidine (0.5 kJ/mol per molecule) and the I → III transition of maleic hydrazide (0.6 kJ/mol per molecule). These errors are, however, still very subtle and likely a consequence of the incomplete cancellation of total-energy errors as previously discussed for molecular crystals of the OMC25 dataset.

VI. LIMITATIONS OF THE MODEL

So far, the MBD-ML model has been shown to be an accurate and reliable dispersion method for diverse systems, including organic and inorganic molecules of widely varying sizes, molecular dimers, and organic crystals. Two limitations warrant attention, however. Molecular anions with unbound electrons can yield highly unphysical predicted α_0^{r} and C_6^{r} ratios, though this reflects the pathological electronic structure of such systems rather than a deficiency of the model itself. More substantively, the model requires further development for systems containing alkali and alkaline earth elements and for inorganic materials. While the former was discussed in the validation results above, we examine the performance of MBD-ML for inorganic materials in more detail here.

In particular, significantly larger α_0^{r} and C_6^{r} errors are observed on a random subset of 203 materials from the OMat24 dataset [8], recomputed at the HSE06 + MBD-NL level of theory. As shown in Table I, the mean absolute errors of 0.247 and 0.269 for these ratios are 10–20 times higher than those for all other chemical classes examined in this study. These errors are sufficiently large to introduce spurious negative MBD eigenvalues, preventing calculation of MBD energies and forces for most materials in the OMat24 subset.

This inadequate performance is likely due to the mismatch between the complex and diverse atomic environments in inorganic materials and those represented in the QCML training dataset. By construction, QCML molecules, derived from organic molecule databases, contain at most eight nonhydrogen atoms [44], whereas inorganic solids consist of nonhydrogen atoms in extended periodic environments, resulting in minimal overlap between the atomic environments present in the two datasets. Crystalline silicon offers a simple illustration. Each silicon atom has tetrahedral coordination with four neighbors, yet any QCML molecule is highly unlikely to contain five silicon atoms in a comparable coordination environment.

Improving MBD-ML performance in inorganic solids and systems containing alkali and alkaline earth metals requires including representative subsets of such systems in the training data, which will be pursued in future work.

VII. COMPUTATIONAL SCALING

Alongside accuracy, computational efficiency is a key requirement for any method intended for large-scale applications. Although the formal scaling of MBD with system size is cubic, its cost in practice remains a small fraction of a single DFT self-consistency cycle. Until now, however, obtaining the α_0^{r} and C_6^{r} ratios required a full DFT calculation, making this the true computational bottleneck. MBD-ML removes this bottleneck, reducing the cost of a complete MBD calculation to that of a neural network inference followed by a matrix diagonalization. To quantify the resulting performance, we characterize the scaling of MBD-ML and libMBD as a function of system size.

We computed MBD corrections to total energies and forces for water clusters $(H_2O)_n$, with $n = 3$–4321 (up to nearly 13 000 atoms) serially on a single 128-core node (AMD EPYC Rome); the results are summarized in Fig. 6. For systems up to 1000 atoms, the MBD-ML calculation time for α_0^{r} and C_6^{r} remains approximately constant at only a few seconds. The subsequent MBD energy and force calculations exhibit the theoretically predicted cubic scaling beyond 60 atoms,

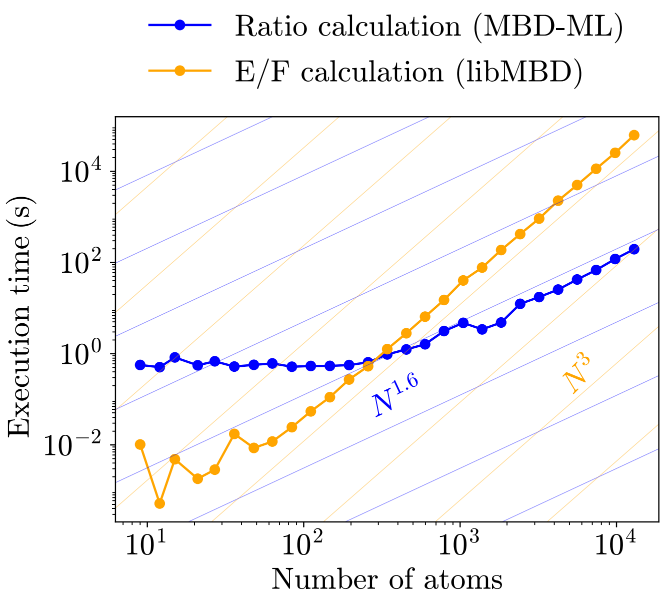

FIG. 6. Computational scaling of the α_0^r and C_6^r ratio calculation by the MBD-ML model and of the MBD property calculation via libMBD.

and are faster than the ratio calculations up to a crossover point at 250 atoms. For clusters exceeding 1000 atoms, the MBD-ML execution time grows with an approximate power of 1.6, reaching a maximum of 196 s for the $(H_2O)_{4321}$ cluster.

This analysis demonstrates that integrating MBD-ML into the libMBD infrastructure enables calculations of unprecedented scale that would otherwise be impractical with state-of-the-art DFAs. The overall cost remains cubic, dominated by libMBD's energy evaluation, which already supports systems of $\sim 10^4$ atoms [41] but ultimately caps the accessible system size. A recently proposed stochastic linear-scaling MBD solver [63] offers a route past this bottleneck. Since MBD-ML provides the required polarizability inputs independently of any electronic-structure code, it can be coupled directly with such a solver, pushing many-body dispersion calculations well beyond what is currently feasible.

VIII. SUMMARY AND CONCLUSIONS

We have developed MBD-ML, a pretrained machine-learning model that serves as an accurate and efficient replacement for *ab initio* MBD-NL calculations. By directly predicting polarizability α_0^r and C_6^r ratios from atomic coordinates, MBD-ML enables computation of MBD energies, forces, and stresses without resorting to any electronic-structure calculations. Integration with `libMBD` reduces the entire workflow to a single function call in the `pymbd` interface, facilitating incorporation of MBD-ML into a wide range of atomistic simulation workflows, including those built on the Atomic Simulation Environment (`ASE`) [64].

We have demonstrated the reliability, accuracy, and transferability of MBD-ML across diverse systems: single molecules of widely varying sizes and chemical compositions, molecular dimers, and organic crystals. Compared to *ab initio* MBD-NL, MBD-ML reproduces energies, forces, and stresses with errors below 1 meV/atom, 1 meV/Å, and 0.2 meV/Å^3, respectively. Molecular-crystal geometry optimizations further confirm this performance, with structural RMSDs of 10^{-3} to 10^{-2} Å and polymorph energy ranking errors of at most 0.6 kJ/mol.

This development addresses a practical barrier to the wider adoption of many-body dispersion methods, namely, that until now MBD-NL has required a full electronic-structure calculation. The libMBD library [41] has largely resolved the computational inefficiencies in evaluating MBD energy gradients that plagued early implementations in VASP and Quantum ESPRESSO, and MBD-ML now removes the remaining barrier by decoupling MBD-NL calculations from any underlying electronic-structure code. Beyond facilitating MBD-NL adoption in computational chemistry codes, MBD-ML also enables more accurate dispersion treatments in machine-learning force fields [12] and semiempirical methods.

Several limitations have been identified for future development. The model's applicability to inorganic solids and systems containing alkali or alkaline earth metals is currently limited by underrepresentation of these systems in the training data, and can in principle be improved by dataset expansion and fine-tuning. The treatment of molecular anions poses a more fundamental challenge, as accurately describing the electronic structure of negatively charged molecules is a well-known difficulty that, as demonstrated in this work, is further exacerbated within the MBD formalism. Generating reliable training data for such systems will require additional work to identify suitable basis set settings and density functional approximations.

ACKNOWLEDGMENTS

All calculations were performed on the Luxembourg national supercomputer MeluXina. E.M., A.K., A.Kh., and A.T. acknowledge the European Research Council under ERC-AdG grant FITMOL (101054629). A.T. acknowledges the Luxembourg National Research Fund under grant FNR-CORE MBD-in-BMD (18093472). A.K. acknowledges financial support from the Luxembourg National Research Fund (FNR AFR Ph.D. Grant No. 15720828).

A.T. conceived and supervised the project. A.K. trained the ML model. E.M. and A.Kh. integrated the MBD-ML model into the libMBD library. E.M. performed all calculations and data analysis of the main results and drafted the manuscript with input from all authors. A.Kh. analyzed the challenges involving molecular anions. E.M. and A.K. created the figures. All authors discussed the results and contributed to editing the manuscript.

DATA AVAILABILITY

The MBD-ML model [65] has been integrated into the PYTHON interface of the libMBD library, `pymbd` [66]. All *ab initio* electronic-structure calculations were performed with FHI-AIMS [67]. For the geometry optimizations, the ASE interface [64] to FHI-AIMS was employed. RMSDs were computed using the `pymatgen` [68] package. The datasets used in this work for model validation have been deposited on Zenodo [69].

[1] J. Hermann, R. A. DiStasio, Jr., and A. Tkatchenko, First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications, Chem. Rev. **117**, 4714 (2017).

[2] M. Stöhr, T. Van Voorhis, and A. Tkatchenko, Theory and practice of modeling van der Waals interactions in electronic-structure calculations, Chem. Soc. Rev. **48**, 4118 (2019).

[3] A. M. Reilly and A. Tkatchenko, Role of dispersion interactions in the polymorphism and entropic stabilization of the aspirin crystal, Phys. Rev. Lett. **113**, 055701 (2014).

[4] J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio, Jr., and A. Tkatchenko, Reliable and practical computational description of molecular crystal polymorphs, Sci. Adv. **5**, eaau3338 (2019).

[5] M. Stöhr and A. Tkatchenko, Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions, Sci. Adv. **5**, eaax0024 (2019).

[6] T. Björkman, A. Gulans, A. Krasheninnikov, and R. Nieminen, Are we van der Waals ready? J. Phys.: Condens. Matter **24**, 424218 (2012).

[7] S. A. Tawfik, T. Gould, C. Stampfl, and M. J. Ford, Evaluation of van der Waals density functionals for layered materials, Phys. Rev. Mater. **2**, 034005 (2018).

[8] L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, M. Uyttendaele, C. L. Zitnick, and Z. W. Ulissi, The open materials 2024 (OMat24) inorganic materials dataset and models, Nat. Comput. Sci. **6**, 642 (2026).

[9] D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman *et al.*, The open molecules 2025 (OMol25) dataset, evaluations, and models, arXiv:2505.08762.

[10] R. J. Maurer, C. Freysoldt, A. M. Reilly, J. G. Brandenburg, O. T. Hofmann, T. Björkman, S. Lebègue, and A. Tkatchenko, Advances in density-functional calculations for materials modeling, Annu. Rev. Mater. Res. **49**, 1 (2019).

[11] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Machine learning force fields, Chem. Rev. **121**, 10142 (2021).

[12] A. Kabylda, J. T. Frank, S. Suárez-Dou, A. Khabibrakhmanov, L. Medrano Sandonas, O. T. Unke, S. Chmiela, K.-R. Müller, and A. Tkatchenko, Molecular simulations with a pretrained neural network and universal pairwise force fields, J. Am. Chem. Soc. **147**, 33723 (2025).

[13] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, Dispersion-corrected mean-field electronic structure methods, Chem. Rev. **116**, 5105 (2016).

[14] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, 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).

[15] A. Tkatchenko and M. Scheffler, Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data, Phys. Rev. Lett. **102**, 073005 (2009).

[16] N. Ferri, R. A. DiStasio, Jr., A. Ambrosetti, R. Car, and A. Tkatchenko, Electronic properties of molecules and surfaces with a self-consistent interatomic van der Waals density functional, Phys. Rev. Lett. **114**, 176802 (2015).

[17] A. D. Becke and E. R. Johnson, Exchange-hole dipole moment and the dispersion interaction revisited, J. Chem. Phys. **127**, 154108 (2007).

[18] E. R. Johnson, The exchange-hole dipole moment dispersion model, in *Non-Covalent Interactions in Quantum Chemistry and Physics*, edited by A. Otero de la Roza and G. A. DiLabio (Elsevier, Amsterdam, 2017), Chap. 5, pp. 169–194.

[19] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Van der Waals density functional for general geometries, Phys. Rev. Lett. **92**, 246401 (2004).

[20] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Higher-accuracy van der Waals density functional, Phys. Rev. B **82**, 081101(R) (2010).

[21] K. Berland and P. Hyldgaard, Exchange functional that tests the robustness of the plasmon description of the van der Waals density functional, Phys. Rev. B **89**, 035412 (2014).

[22] V. Shukla, Y. Jiao, J.-H. Lee, E. Schröder, J. B. Neaton, and P. Hyldgaard, Accurate nonempirical range-separated hybrid van der Waals density functional for complex molecular problems, solids, and surfaces, Phys. Rev. X **12**, 041003 (2022).

[23] O. A. Vydrov and T. Van Voorhis, Nonlocal van der Waals density functional: The simpler the better, J. Chem. Phys. **133**, 244103 (2010).

[24] J. F. Dobson, Beyond pairwise additivity in London dispersion interactions, Int. J. Quantum Chem. **114**, 1157 (2014).

[25] A. Tkatchenko, R. A. DiStasio, Jr., R. Car, and M. Scheffler, Accurate and efficient method for many-body van der Waals interactions, Phys. Rev. Lett. **108**, 236402 (2012).

[26] A. Tkatchenko, A. Ambrosetti, and R. A. DiStasio, Jr., Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem, J. Chem. Phys. **138**, 074106 (2013).

[27] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, Jr., and A. Tkatchenko, Long-range correlation energy calculated from coupled atomic response functions, J. Chem. Phys. **140**, 18A508 (2014).

[28] R. A. DiStasio, Jr., V. V. Gobre, and A. Tkatchenko, Many-body van der Waals interactions in molecules and condensed matter, J. Phys.: Condens. Matter **26**, 213202 (2014).

[29] J. Hermann, D. Alfe, and A. Tkatchenko, Nanoscale π–π stacked molecules are bound by collective charge fluctuations, Nat. Commun. **8**, 14052 (2017).

[30] M. Kim, W. J. Kim, T. Gould, E. K. Lee, S. Lebègue, and H. Kim, uMBD: A materials-ready dispersion correction that uniformly treats metallic, ionic, and van der Waals bonding, J. Am. Chem. Soc. **142**, 2346 (2020).

[31] J. Hermann and A. Tkatchenko, Density functional model for van der Waals interactions: Unifying many-body atomic approaches with nonlocal functionals, Phys. Rev. Lett. **124**, 146401 (2020).

[32] T. Gould, S. Lebegue, J. G. Angyan, and T. Bučko, A fractionally ionic approach to polarizability and van der Waals many-body dispersion calculations, J. Chem. Theory Comput. **12**, 5920 (2016).

[33] A. Khabibrakhmanov, M. Gori, C. Müller, and A. Tkatchenko, Noncovalent interactions in density functional theory: All the charge density we do not see, J. Am. Chem. Soc. **147**, 40763 (2025).

[34] P. P. Paul, V. S. Thakur, and S. Banerjee, DFT-D3 for solid-state materials: A data-driven perspective on accuracy, Adv. Theory Simul. **9**, e01072 (2026).

[35] R. Sedlak and J. Rezac, Empirical D3 dispersion as a replacement for *ab initio* dispersion terms in density functional

theory-based symmetry-adapted perturbation theory, J. Chem. Theory Comput. **13**, 1638 (2017).

[36] J. Proppe, S. Gugler, and M. Reiher, Gaussian process-based refinement of dispersion corrections, J. Chem. Theory Comput. **15**, 6046 (2019).

[37] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, A generally applicable atomic-charge dependent London dispersion correction, J. Chem. Phys. **150**, 154122 (2019).

[38] J. Witte, N. Mardirossian, J. B. Neaton, and M. Head-Gordon, Assessing DFT-D3 damping functions across widely used density functionals: Can we do better? J. Chem. Theory Comput. **13**, 2043 (2017).

[39] N. V. Tkachenko and M. Head-Gordon, Smoother semiclassical dispersion for density functional theory via D3s: Understanding and addressing unphysical minima in the D3 dispersion correction model, J. Chem. Theory Comput. **20**, 9741 (2024).

[40] L. Briccolani-Bandini, B. Gnnanapareddy, F. Labat, E. Brémond, J. C. Sancho-García, A. Otero-de-la Roza, G. Scalmani, M. J. Frisch, G. Cardini, G. A. DiLabio, *et al.*, Weak noncovalent interactions in nonequilibrium structures: How good are the dispersion corrections? J. Phys. Chem. Lett. **17**, 703 (2026).

[41] J. Hermann, M. Stöhr, S. Góger, S. Chaudhuri, B. Aradi, R. J. Maurer, and A. Tkatchenko, libMBD: A general-purpose package for scalable quantum many-body dispersion calculations, J. Chem. Phys. **159**, 174802 (2023).

[42] P. P. Poier, T. Jaffrelot Inizan, O. Adjoua, L. Lagardere, and J.-P. Piquemal, Accurate deep learning-aided density-free strategy for many-body dispersion-corrected density functional theory, J. Phys. Chem. Lett. **13**, 4381 (2022).

[43] K. R. Bryenton and E. R. Johnson, Many-body dispersion in model systems and the sensitivity of self-consistent screening, J. Chem. Phys. **158**, 204110 (2023).

[44] S. Ganscha, O. T. Unke, D. Ahlin, H. Maennel, S. Kashubin, and K.-R. Müller, The QCML dataset, quantum chemistry reference data from 33.5M DFT and 14.7B semi-empirical calculations, Sci. Data **12**, 406 (2025).

[45] A. G. Donchev, A. G. Taube, E. Decolvenaere, C. Hargus, R. T. McGibbon, K.-H. Law, B. A. Gregersen, J.-L. Li, K. Palmo, K. Siva, *et al.*, Quantum chemical benchmark databases of gold-standard dimer interaction energies, Sci. Data **8**, 55 (2021).

[46] V. Gharakhanyan, L. Barroso-Luque, Y. Yang, M. Shuaibi, K. Michel, D. S. Levine, M. Dzamba, X. Fu, M. Gao, X. Liu, *et al.*, Open Molecular Crystals 2025 (OMC25) dataset and models, Sci. Data **13**, 354 (2026).

[47] A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield, and G. J. Martyna, Quantum Drude oscillator model of atoms and molecules: Many-body polarization and dispersion interactions for atomistic simulation, Phys. Rev. B **87**, 144103 (2013).

[48] A. Khabibrakhmanov, D. V. Fedorov, A. Ambrosetti, J. Crain, K. L. Hunt, E. R. Johnson, K. D. Jordan, S. Góger, M. Gori, M. R. Karimpour, *et al.*, Accurate noncovalent interactions in atomistic systems via quantum Drude oscillators, J. Chem. Phys. **163**, 151001 (2025).

[49] O. A. Vydrov and T. Van Voorhis, Dispersion interactions from a local polarizability model, Phys. Rev. A **81**, 062708 (2010).

[50] J. T. Frank, O. T. Unke, K.-R. Müller, and S. Chmiela, A Euclidean transformer for fast and stable machine learned force fields, Nat. Commun. **15**, 6539 (2024).

[51] See Supplemental Material at http://link.aps.org/supplemental/10.1103/pv86-l9h7 for additional details on the MBD-ML training, selection of validation datasets, computational settings of electronic-structure calculations, a more detailed analysis of anions in the QCML dataset and of the performance of the MBD-ML model when anions or underrepresented species are taken into account, a sensitivity analysis of the model, and a more detailed statistical analysis on the force deviation distribution of MBD-ML, D3 and D4, which includes Refs. [70–74].

[52] C. J. Nickerson, K. R. Bryenton, A. J. Price, and E. R. Johnson, Comparison of density-functional theory dispersion corrections for the DES15K database, J. Phys. Chem. A **127**, 8712 (2023).

[53] H. Zhao, B. Lőrincz, T. Henkes, D. Berta, P. Nagy, A. Tkatchenko, and S. Vuckovic, Accurate density functional theory for noncovalent interactions in charged systems, Sci. Adv. **12**, eadz8521 (2026).

[54] A. Kabylda, S. Suárez-Dou, N. Davoine, F. N. Brünig, and A. Tkatchenko, QCell: Comprehensive quantum-mechanical dataset spanning diverse biomolecular fragments, AI Sci. **2**, 025003 (2026).

[55] S. Grimme, S. Ehrlich, and L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, J. Comput. Chem. **32**, 1456 (2011).

[56] E. Caldeweyher, C. Bannwarth, and S. Grimme, Extension of the D3 dispersion coefficient model, J. Chem. Phys. **147**, 034112 (2017).

[57] P. Hauseux, A. Ambrosetti, S. P. Bordas, and A. Tkatchenko, Colossal enhancement of atomic force response in van der Waals materials arising from many-body electronic correlations, Phys. Rev. Lett. **128**, 106101 (2022).

[58] M. Puleva, L. Medrano Sandonas, B. D. Lőrincz, J. Charry, D. M. Rogers, P. R. Nagy, and A. Tkatchenko, Extending quantum-mechanical benchmark accuracy to biological ligand-pocket interactions, Nat. Commun. **16**, 8583 (2025).

[59] R. I. Sosa, M. Galante, and A. Tkatchenko, Power of the many-body force: Magnitudes and angles of atomic van der Waals dispersion forces in extended molecular systems, Small Struct. **6**, 2500226 (2025).

[60] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[61] J. Potticary, L. R. Terry, C. Bell, A. N. Papanikolopoulos, P. C. Christianen, H. Engelkamp, A. M. Collins, C. Fontanesi, G. Kociok-Köhn, S. Crampin, *et al.*, An unforeseen polymorph of coronene by the application of magnetic fields during crystal growth, Nat. Commun. **7**, 11555 (2016).

[62] J. Nyman and G. M. Day, Static and lattice vibrational energy differences between polymorphs, CrystEngComm **17**, 5154 (2015).

[63] P. P. Poier, L. Lagardère, and J.-P. Piquemal, $O(N)$ stochastic evaluation of many-body van der Waals energies in large complex systems, J. Chem. Theory Comput. **18**, 1633 (2022).

[64] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, The atomic simulation environment—A Python library for working with atoms, J. Phys.: Condens. Matter **29**, 273002 (2017).

[65] E. Moerman, The MBD-ML project repository, version 1.0.0, GitHub, 2026, https://github.com/em819/so3lr/tree/mbd-ml.

[66] E. Moerman, The libMBD repository with MBD-ML integration, version 1.0.0, GitHub, 2026, https://github.com/em819/libmbd.

[67] J. W. Abbott, C. M. Acosta, A. Akkoush, A. Ambrosetti, V. Atalla, A. Bagrets, J. Behler, D. Berger, H. Bertschi, B. Bieniek, J. Björk, *et al.*, Roadmap on advancements of the FHI-aims software package, arXiv:2505.00125.

[68] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source Python library for materials analysis, Comput. Mater. Sci. **68**, 314 (2013).

[69] E. Moerman, A. Kabylda, A. Khabibrakhmanov, and A. Tkatchenko, Validation datasets for MBD-ML: Many-body dispersion from machine learning for molecules and materials, version 1.0.1, Zenodo, 2026, https://doi.org/10.5281/zenodo.18835427.

[70] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, *Ab initio* molecular simulations with numeric atom-centered orbitals, Comput. Phys. Commun. **180**, 2175 (2009).

[71] M. A. Rohrdanz, K. M. Martins, and J. M. Herbert, A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states, J. Chem. Phys. **130**, 054112 (2009).

[72] E. Epifanovsky, A. T. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, *et al.*, Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package, J. Chem. Phys. **155**, 084801 (2021).

[73] R. A. Kendall, T. H. Dunning, and R. J. Harrison, Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions, J. Chem. Phys. **96**, 6796 (1992).

[74] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Density-functional theory for fractional particle number: Derivative discontinuities of the energy, Phys. Rev. Lett. **49**, 1691 (1982).