MACHINE LEARNING
Science and Technology

PURPOSE-LED PUBLISHING™ Part of the MACHINE LEARNING Series

PAPER • **OPEN ACCESS**

Machine learning surrogate models of many-body dispersion interactions in polymer melts

To cite this article: Zhaoxiang Shen *et al* 2026 *Mach. Learn.: Sci. Technol.* **7** 025040

View the article online for updates and enhancements.

You may also like

- Beyond-dipole van der Waals contributions within the many-body dispersion framework
Dario Massa, Alberto Ambrosetti and Pier Luigi Silvestrelli
- Porous PVA/Zn–Fe–Mn oxide nanocomposites: methylene blue dye adsorption studies
Buzuayehu Abebe, Ananda Murthy H C, Enyew Zerefa et al.
- Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL system
M Schenkelberger, S Shanak, M Finkler et al.

This content was downloaded from IP address 158.64.192.102 on 15/04/2026 at 13:12

MACHINE LEARNING
Science and Technology

Check for updates

PAPER

OPEN ACCESS

RECEIVED
19 March 2025

REVISED
4 March 2026

ACCEPTED FOR PUBLICATION
18 March 2026

PUBLISHED
1 April 2026

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Machine learning surrogate models of many-body dispersion interactions in polymer melts

Zhaoxiang Shen[1], **Raúl I Sosa**[2], **Jakub Lengiewicz**[1,3], **Alexandre Tkatchenko**[2] and **Stéphane P A Bordas**[1,4,*]

[1] Department of Engineering; Faculty of Science, Technology and Medicine University of Luxembourg, Esch-sur-Alzette, Luxembourg
[2] Department of Physics and Materials Science, University of Luxembourg, Luxembourg City, Luxembourg
[3] Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
[4] Department of Engineering, University of Exeter, Exeter, United Kingdom
[*] Author to whom any correspondence should be addressed.

E-mail: stephane.bordas@uni.lu and s.bordas@exeter.ac.uk

Keywords: many-body dispersion, van der Waals interaction, machine learning force field, surrogate modeling, polymer melts, deep neural network

Abstract

Accurate prediction of many-body dispersion (MBD) interactions is essential for understanding the van der Waals forces that govern the behavior of many complex molecular systems. However, the high computational cost of MBD calculations limits their direct application in large-scale simulations. In this work, we introduce a machine learning surrogate model specifically designed to predict MBD forces in polymer melts, a system that demands accurate MBD description and offers structural advantages for machine learning approaches. Our model is based on a trimmed SchNet architecture that selectively retains the most relevant atomic connections and incorporates trainable radial basis functions for geometric encoding. We validate our surrogate model on datasets from polyethylene, polypropylene, and polyvinyl chloride melts, demonstrating high predictive accuracy and robust generalization across diverse polymer systems. In addition, the model captures key physical features, such as the characteristic decay behavior of MBD interactions, providing valuable insights for optimizing cutoff strategies. Characterized by high computational efficiency, our surrogate model enables practical incorporation of MBD effects into large-scale molecular simulations.

1. Introduction

Van der Waals (vdW) dispersion interactions play a critical role in diverse physical and chemical phenomena, from layered material cohesion [1] and protein folding [2] to gecko adhesion [3]. Accurately capturing vdW dispersion forces is essential for understanding and predicting the properties of numerous engineering and biological systems. Well established pairwise (PW) models, such as Lennard–Jones (LJ) potential, Grimme's dispersion corrections [4], or the Tkatchenko–Scheffler (TS) method [5], are widely used due to their simplicity and computational efficiency. While effective for certain systems, they inherently neglect the quantum many-body nature of vdW dispersion interactions, often leading to discrepancies with experimental observations [6–8]. For this reason, the many-body dispersion (MBD) method [9, 10] becomes more preferable because it offers a more sophisticated and accurate approach by explicitly capturing collective long-range electron correlations. Numerous studies have shown that MBD surpasses PW models in reproducing experimental or nearly exact quantum-mechanical results, achieving superior accuracy in diverse scenarios, including thin-layer delamination [11], modeling of supramolecular systems [6], and polymorphs prediction for molecular crystals [12].

The major obstacle of using MBD is its computational complexity, which scales as $O(N^3)$ with the number of atoms, N, thus severely limits its application to large-scale simulations. Recent efforts have attempted to accelerate MBD through physically motivated approximations and algorithmic optimizations of the analytical formulation [13, 14], but these approaches still require further refinement to

© 2026 The Author(s). Published by IOP Publishing Ltd

achieve practical applicability in large-scale simulations. An alternative and more common approach is to develop a surrogate model that offers a fast yet accurate approximation, addressing the problem from the efficiency side while gradually improving accuracy. In molecular simulations, this is typically referred to as a force field (FF) model, which aims to approximate the high-fidelity energy and forces of molecular systems using functional forms based on simple geometric descriptors, such as interatomic distances, bond angles, and torsions [15–18].

While classical FF models are computationally efficient, they suffer from limited accuracy due to their inadequate description of polarization and many-body interactions. To overcome these drawbacks, machine learning FFs (MLFFs) [19–24] have emerged as promising solutions, offering quantum-level accuracy with competitive computational efficiency. MLFFs, which include a wide range of models from kernel-based methods to artificial neural networks (e.g. sGDML [25], SchNet [26, 27], PhysNet [28], GEMS [29], among others [30–42]), aim to learn the statistical relationship between chemical structures and the corresponding potential energy as well as atomic forces. Despite their advantages, many MLFFs are primarily trained for short-distance interactions and continue to rely on classical PW models for vdW interactions [43]. Some approaches, such as SpookyNet [44] and SO3LR [45], incorporate MBD-corrected reference data into their training while explicitly embedding parametric PW vdW functions within their architectures. These approaches, however, remain essentially parametric fittings to MBD and are limited in capturing many-body vdW effects beyond intramolecular interactions. To the best of our knowledge, there is currently no MLFF model explicitly designed as an MBD surrogate capable of capturing many-body vdW effects in large-scale systems.

The objective in this work is to develop an ML surrogate model for MBD interactions that is specifically designed for *large-scale systems*. Here, we focus on a particular class of molecular systems, the polymer melts, which can benefit from an accurate MBD description of vdW interactions, and which also requires a fast model for large-scale molecular simulations. This choice of systems follows the conclusions made in our previous work [46], where we have identified classes of systems where MBD effects are most pronounced, namely, extended systems with flexible degrees of freedom whose physical properties are predominantly governed by vdW interactions, such as polymer crystals, polymer melts, and proteins. Among these, polymer melts emerge as the ideal initial target for our surrogate modeling. Polymer melts [47, 48] are characterized by their disordered amorphous arrangement of polymer chains, where the chains exhibit significant mobility and fluid-like behavior while maintaining their high-molecular-weight structure. Accurately capturing vdW interactions is essential for predicting the mechanical and thermodynamic properties of polymer melts, as vdW forces directly influence key behaviors such as chain mobility, density, and viscosity [49, 50], while the desired molecular simulations of polymer melts often involve more than 100k atoms [51–54]. Despite extensive efforts that have been made to enable such large-scale molecular simulations [55–59], incorporating MBD into these simulations still requires a dedicated ML architecture to provide an efficient surrogate model.

In this paper, we introduce a trimmed SchNet architecture that is specifically tailored for MBD force predictions in polymer melts. It is based on the SchNet architecture [26, 27], which we found an optimal choice for our problem due to its simplicity and superior adaptability (see also a more detailed discussion in section 2.2). SchNet has demonstrated strong performance in molecular energy and force predictions, which is attributed to its radial basis function (rbf) encoding of structural information and continuous-filter convolutions. It remains scalable with respect to system size and ensures rotational invariance for energy prediction and rotational equivariance for force prediction. In our adaptation, we reduce the dimension and depth of the model while trimming the connection graph to retain only the most relevant interactions, which are then encoded using a reduced number of trainable rbfs. In addition, inspired by the repetitive structure of polymer chains, we introduce a unit-specific batching strategy that enhances training convergence. To validate the predictive capacity of the model, we apply it to three types of polymer melts: polyethylene (PE), polypropylene (PP), and polyvinyl chloride (PVC) and further assess its generalization capacity and robustness on mixed datasets and advanced MBD variants. A preliminary study is also conducted to evaluate temperature effects on the model's performance. Furthermore, we demonstrate the physical interpretability of the model through analysis of the Hessian and discuss practical implementation of the model in molecular dynamics (MD) simulations together with a preliminary stability test, showcasing its applicability to large-scale systems. To increase the impact of this work, we open-source our codes and dataset, which are available in our repository [60].

The remainder of this paper is structured as follows: section 2 introduces the formalism of MBD and outlines the requirements for MBD surrogate modeling in polymer melts. It then details the proposed trimmed SchNet architecture, describes the data generation process, and provides implementation details for training. In section 3, we evaluate the accuracy, generalization capability, and robustness of

the model through multiple examples, including three types of polymer melts, and analyze the key components of its architecture. In addition, we perform an analysis of the Hessian in this section and discuss the implementation of the model in MD simulations. Finally, section 4 summarizes the work and discusses potential directions for future research.

2. Methods

In this section, we will first introduce the intricate formalism of MBD in comparison to the classical PW model, see section 2.1. Then, in section 2.2, we will outline the requirements for an MBD surrogate, which motivate the choice of a SchNet-based architecture. The proposed trimmed SchNet model, along with details on the trimmed connections, will be elaborated in section 2.3. In section 2.4, we will specify the selected polymer melt systems and describe the corresponding data generation process. Finally, section 2.5 will present the training setup and evaluation metric.

2.1. MBD method

vdW dispersion, also know as London dispersion, is a ubiquitous attractive interaction between atoms and molecules. It originates from the temporary fluctuations in the mean-field electron density, leading to instantaneous dipoles that then interact with other dipoles induced by them. In this work, we focus exclusively on the dispersion interaction, which corresponds to the attractive part of the conventional vdW model. Traditional vdW modeling approaches rely on PW additive methods, where the dispersion energy $E^{\mathrm{disp,PW}}$ for an N-atom molecular system is calculated as a simple sum of PW atomic interactions:

$$E^{\mathrm{disp,PW}} = -\sum_{j>i}^{N} f^{\mathrm{damp}} \frac{C_{6,ij}}{r_{ij}^{6}}, \tag{1}$$

where $r_{ij} = \|\boldsymbol{r}_i - \boldsymbol{r}_j\|$ is the interatomic distance between atoms i and j, and $\boldsymbol{r}$ represents the atomic position. The material parameters $C_{6,ij}$ are determined experimentally or through high-fidelity numerical calculations. A damping function f^{damp} is employed to mitigate non-physical singularities at short distances, while the function value converges to 1 at large distances. The interatomic force of a given atom i in the system is obtained by taking the negative gradient of the energy with respect to its atomic position $\boldsymbol{r}_i$:

$$\boldsymbol{F}_{\boldsymbol{r}_i}^{\mathrm{disp,PW}} = -\nabla_{\boldsymbol{r}_i} E^{\mathrm{disp,PW}}. \tag{2}$$

However, the vdW dispersion interaction is inherently many-body in nature, as it arises from correlated electron fluctuations. The MBD method [9, 10] provides an efficient quantum-mechanical treatment of this phenomenon, offering a more accurate description compared to PW models. The MBD energy reads:

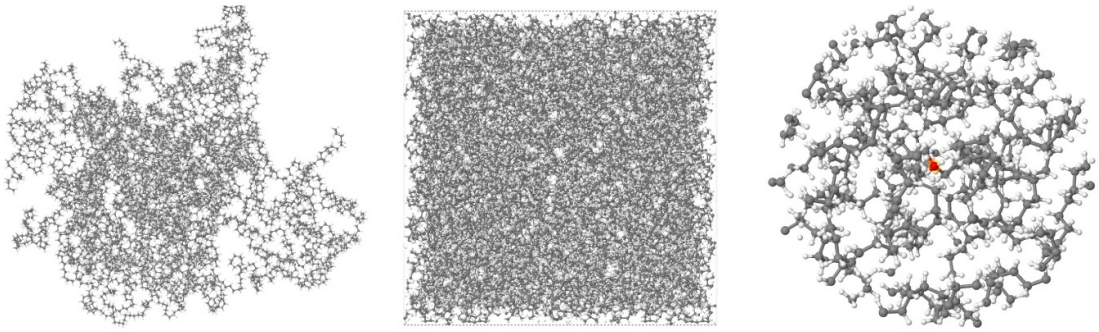
$$E^{\mathrm{disp,MBD}} = \frac{1}{2}\sum_{l=1}^{3N} \sqrt{\lambda_l} - \frac{3}{2}\sum_{i=1}^{N} \omega_i, \tag{3}$$

where λ_l is the eigenvalue of a $3N \times 3N$ matrix $\boldsymbol{C}^{\mathrm{MBD}}$ that couples each pair of atoms (see equation (A.5)), and ω_i is the atomic characteristic frequency. To enhance readability, we present the detailed methodology and formalism of MBD in appendix. Taking the negative gradient of the energy with respect to the atomic position, after linear algebra transformations, the MBD force acting on atom i can be expressed as:

$$\boldsymbol{F}_{\boldsymbol{r}_i}^{\mathrm{disp,MBD}} = -\frac{1}{4}\mathrm{Tr}\left[\boldsymbol{\Lambda}^{-1/2}\boldsymbol{S}^{\mathrm{T}}\left(\nabla_{\boldsymbol{r}_i}\boldsymbol{C}^{\mathrm{MBD}}\right)\boldsymbol{S}\right], \tag{4}$$

where $\boldsymbol{\Lambda}_{lq} = \lambda_l \delta_{lq}$ is obtained from the diagonalization of $\boldsymbol{C}^{\mathrm{MBD}}$ via a suitable SO($3N$) rotation matrix $\boldsymbol{S}$ that satisfies $\boldsymbol{\Lambda} = \boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{MBD}}\boldsymbol{S}$. Summing over the eigenmodes reflects the collective contributions from the whole system, qualitatively differing from the PW force splitting in simpler PW models.

As discussed in [46], the correlated MBD model is fundamental for capturing cooperative motions that can significantly influence the dynamics of systems such as polymer melts. However, it involves the diagonalization operation for MBD forces as expressed by equation (4), which introduces a computational complexity of $O(N^3)$, along with substantial memory demands compared to the additive PW formula. This makes direct application of MBD infeasible for large-scale polymer melt simulations, which



(a) Unfolded unit cell. (b) Folded unit cell. (c) Cluster.

Figure 1. A PE melt structure. (a) is the original unfolded configuration of the unit cell. (b) is the folded unit cell according to the periodic boundary condition. (c) shows an atomic cluster with 1000 atoms cut off from (b), in which the center atom is marked as red.

can exceed 100k atoms. To overcome this challenge, an ML surrogate model with a tailored modeling strategy is required. In the following section, we analyze this modeling challenge from a computational perspective and demonstrate the requirements for an effective MBD surrogate for polymer melts, leading to the development of the proposed model architecture, as detailed in section 2.3.

2.2. Surrogate modeling requirements for MBD in polymer melts

An effective surrogate model should be informed by both the numerical properties of the underlying physical model and the structural characteristics of the target system, in this case, MBD and polymer melts. As a reference, figure 1 presents a representative PE melt structure. The dense and periodic system consists of a finite number of polymer chains within the unit cell, with chain lengths ranging from a few hundred monomers to millions depending on the molecular weight.

Even with conventional PW formulas, such as equation (1), computational feasibility is achieved by leveraging the decay properties of vdW dispersion to impose a cutoff range of interactions for each atom. This ensures that the overall computational complexity remains $O(N \times N_{\text{cut}})$, where N_{cut} is the number of atoms within the cutoff neighborhood. To maintain efficiency, the MBD model must adopt a similar strategy. As illustrated in (a)–(c) of figure 1, MBD calculations are performed on the nearly spherical atomic clusters extracted from the full folded system, a strategy directly employed in the proposed surrogate modeling approach. Specifically, the surrogate model processes an atomic cluster as input and outputs the MBD force acting on its center atom, which is repeated for all atoms in the system. Moreover, accurately predicting MBD forces requires a significantly larger cutoff range compared to PW models, due to the inherently long-range nature of many-body correlations, as discussed in [11, 61, 62]. This presents an additional challenge, as N_{cut} scales cubically with the cutoff distance. On the positive side, the structural characteristics of polymer melts introduce notable regularity in the cutoff region. For a fixed N_{cut}, the extracted atomic clusters closely resemble spheres with similar radii. This structural uniformity benefits statistical ML surrogate modeling, as the limited number of atom types enables the model to focus primarily on encoding local geometric details rather than learning higher-level chemical features.

Taking these considerations into account, we conclude that an optimal architecture for our problem, beyond fundamental requirements such as rotational equivariance for force predictions, should exhibit the following properties:

- Lightweight design to enable efficient repetitive and parallel execution in large-scale simulations.
- Decomposability to allow per-atom force contributions, aligning with the center-focused nature of the problem.
- Strong capability to encode local geometric features while effectively capturing long-range interactions.

Several well-established MLFF models, such as SchNet, NequIP [41], MACE [42], SpookyNet [44], and SO3LR [45], satisfy these requirements and demonstrate strong general-purpose performance. However, many of these architectures are advanced derivatives of one another, designed for broad chemical spaces

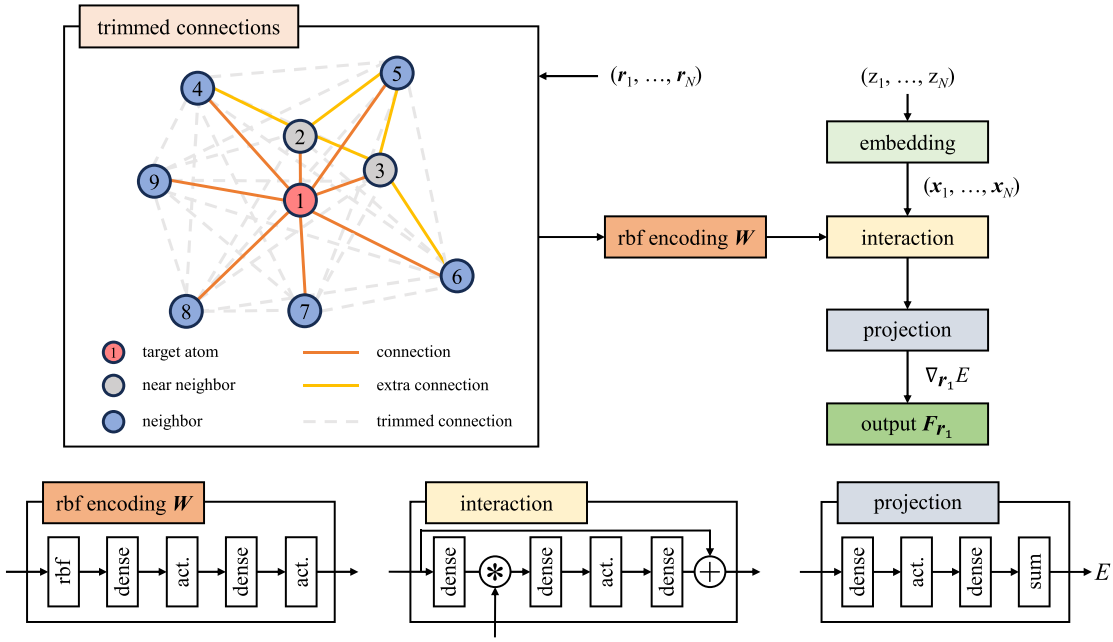


Figure 2. Illustration of the trimmed SchNet. The trimmed connections graph is depicted for a toy system that consists of $N = 9$ atoms, where only connections to the central target atom (red solid lines) and a few extra connections to its two nearest neighbors (yellow solid lines) are retained, while all other peripheral connections (gray dashed lines) are trimmed. The detailed architectures of rbf encoding, interaction, and projection are shown in the bottom.

and enriched prediction targets, often incorporating higher-order features together with multiple layers of physically motivated interaction modules (e.g. attention-based modules [63]). While powerful, these additions introduce substantial computational overhead. In our case, the task is restricted to predicting MBD forces, a non-bonded interaction with a smoother landscape than the full interatomic physics, and is formulated for a single target atom per cutoff cluster. This narrower scope does not benefit from the additional complexity of such general-purpose models. Adapting them would likely involve removing most of their specialized components, effectively converging towards a model similar to SchNet. For this reason, we adopt a SchNet-based design, starting from a minimal yet robust baseline that avoids the overhead of unused functionality while retaining the required accuracy. This deliberate trade-off of generality for scalability makes the model well matched to our task. In the following section, we present our proposed trimmed SchNet and the dedicated modifications tailored to the MBD surrogate problem.

2.3. Trimmed SchNet

We now present the architecture of the trimmed SchNet, designed for surrogate modeling of MBD forces in polymer melts. The section begins with an overview of the proposed architecture, followed by a detailed discussion of key modifications to the original SchNet, including trimmed connections (section 2.3.2) and trainable rbf encoding (section 2.3.3). Finally, section 2.3.4 presents the modified model output and the force-based loss function.

2.3.1. Architecture overview

The trimmed SchNet model processes an atomic system by taking atomic positions and atom types as inputs, encoding their structural information, and predicting the MBD force of the center atom. Given an atomic system with N atoms, the model takes the atomic positions $\boldsymbol{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ and the corresponding atom types $\boldsymbol{Z} = (z_1, \ldots, z_N)$ as inputs. For expressiveness, each atom is represented by a P-dimensional feature embedding, $\boldsymbol{x}_i^l \in \mathbb{R}^P$, where $\boldsymbol{x}_i^l$ is the feature vector at layer l. These embeddings are initialized based on the atom types $\boldsymbol{Z}$ and iteratively updated through interaction layers.

Figure 2 illustrates the architecture of the trimmed SchNet. The model follows the fundamental structure of the original SchNet [26, 27], where atomic interactions are modeled through learned feature representations. It encodes structural information based on specified atomic connections, processes interactions via continuous-filter convolutions, and projects the learned representations into a scalar energy value. The force output is then obtained as the gradient of the energy with respect to atomic positions.

While preserving SchNet's core structure, the proposed architecture introduces several modifications tailored to the MBD surrogate problem. These include trimming interaction connections, incorporating

trainable rbf encoding, using a single interaction block, and reducing the network width. These adaptations simplify the model while maintaining its functionality, enabling efficient implementations for large-scale molecular simulations.

2.3.2. Interactions by trimmed connections

Formally, the atomic feature updates of atom i in SchNet are computed by an interaction block with so called continuous-filter (cf) convolution:

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \sum_{j\neq i}^{N} \boldsymbol{W}^l\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right) * \boldsymbol{x}_j^l, \tag{5}$$

where $\boldsymbol{W}$ is the filter-generating function that encodes the atomic connections via rbfs, and $*$ denotes element-wise multiplication.

In the application to MBD force prediction for polymers, we only need to calculate for the center atom in a given cutoff range. To maintain a fixed input dimension for the neural network, the atomic cluster is cut off by a fixed number of atoms $N = N_{\text{cut}}$, while the position tuple $\boldsymbol{R}$ is ordered by the atomic distance to the center atom. In the original version, a full SchNet leverages all $\sum_{n=2}^{N_{\text{cut}}}(n-1)$ connections between atoms, which restricts the cutoff range in practice. However, our specific goal, focusing only on the center atom, implies that a reduction of required connections is potentially feasible. In this paper, we implement a trimmed SchNet that remains limited connections, as shown in connection graph in figure 2 (top left), which updates the atom feature as follows:

- Keep the full convolution (orange lines) to the atom 1, i.e. the center atom (red circle):

$$\boldsymbol{x}_1^{l+1} = \boldsymbol{x}_1^l + \sum_{j\neq 1}^{N_{\text{cut}}} \boldsymbol{W}\left(\boldsymbol{r}_1 - \boldsymbol{r}_j\right) * \boldsymbol{x}_j^l. \tag{6}$$

- Add N_{extra} connections (yellow lines) for the p nearest neighbors (gray circles) of atom 1:

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \sum_{j\neq i}^{i+N_{\text{extra}}} \boldsymbol{W}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right) * \boldsymbol{x}_j^l,\ i = 2,\ldots,p+1. \tag{7}$$

- The rest of the atoms (blue circles) only take reverse connections from the previous two updates:

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \sum_{j=1}^{p+1} \boldsymbol{W}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right) * \boldsymbol{x}_j^l,\ i = p+2,\ldots,2+N_{\text{extra}}. \tag{8}$$

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \sum_{j=i-N_{\text{extra}}}^{p+1} \boldsymbol{W}\left(\boldsymbol{r}_i - \boldsymbol{r}_j\right) * \boldsymbol{x}_j^l + \boldsymbol{W}(\boldsymbol{r}_i - \boldsymbol{r}_1) * \boldsymbol{x}_1^l,\ i = 3+N_{\text{extra}},\ldots,p+1+N_{\text{extra}}. \tag{9}$$

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \boldsymbol{W}(\boldsymbol{r}_i - \boldsymbol{r}_1) * \boldsymbol{x}_1^l,\ i = p+2+N_{\text{extra}},\ldots,N_{\text{cut}}. \tag{10}$$

The trimmed connections, i.e. those disregarded from the original SchNet, are depicted as dashed gray lines. Trimming these peripheral connections effectively reduces the computational cost of the model, while the remaining extra connections to the nearest neighbors enhance the capability of the trimmed SchNet. The results presented in section 3.2.1 highlight the importance of these additional connections for achieving highly accurate fitting. Moreover, in practice, a SchNet model can incorporate multiple interaction blocks in sequence, each with separate rbf encodings, allowing it to capture more complex and hierarchical features as the model depth increases. In this work, we opt for a single interaction block to balance accuracy with computational efficiency.

2.3.3. Rbf encoding

The filter-generating function $\boldsymbol{W}$ calculates the distance of input atoms $d_{ij} = \|r_i - r_j\|$, and then expands it by N_{rbf} Gaussian rbfs:

$$\boldsymbol{e}_k = \exp\left(-\gamma_k |d_{ij} - \mu_k|^2\right), \tag{11}$$

where γ_k is the rbf coefficient and μ_k is the center of an rbf, which forms the weighted encoding bases over the range of interest. In contrast to the original SchNet, we set both parameters as trainable to

reduce the requirement of N_{rbf}, improving the convergence of training. The mechanism is that the tunable rbf centers tend to cluster to some critical distances over the cutoff range and are weighted accordingly by γ_k, therefore, encode the atomic structure more effectively. A numerical analysis for the trainable rbfs is shown in section 3.2.2.

2.3.4. Force-based loss function

The output of this trimmed SchNet is directly and only the force of the center atom, that is, atom 1. This is derived by taking the gradient of E w.r.t. the atomic position:

$$\boldsymbol{F}_{\boldsymbol{r}_1} = -\nabla_{\boldsymbol{r}_1} E, \tag{12}$$

where E here is only an energy-like latent quantity obtained after the projection layer of the network. Importantly, this quantity is not intended to represent a physically meaningful or conserved potential energy, as the model is trained on truncated atomic clusters extracted from a larger polymer melt system, for which a well-defined global energy does not exist. Nevertheless, the rbf encoding introduced above ensures rotationally invariance of E, so by construction, the force defined by equation (2) is rotationally equivariant.

With this force-only output, our trimmed SchNet is trained with a reduced loss function compared to the one for the original SchNet. If we denote this neural network as $h : (\boldsymbol{R}, \boldsymbol{Z}) \to \boldsymbol{F}_{\boldsymbol{r}_1}$ and its trainable parameters as $\boldsymbol{\theta}$, the mean squared error loss function used in this paper can be formulated as:

$$\mathcal{L}\left(\mathcal{D}_{\text{train}}, \boldsymbol{\theta}\right) = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \left\| h\left(\left(\boldsymbol{R}_i, \boldsymbol{Z}_i\right), \boldsymbol{\theta}\right) - \hat{\boldsymbol{F}}_{\boldsymbol{r}_1 i} \right\|^2, \tag{13}$$

where $\mathcal{D}_{\text{train}} = \{((\boldsymbol{R}_1, \boldsymbol{Z}_1), \hat{\boldsymbol{F}}_{\boldsymbol{r}_1 1}), \ldots, ((\boldsymbol{R}_{N_{\text{train}}}, \boldsymbol{Z}_{N_{\text{train}}}), \hat{\boldsymbol{F}}_{\boldsymbol{r}_1 N_{\text{train}}})\}$ is a given training dataset with N_{train} data points, and $\hat{\boldsymbol{F}}_{\boldsymbol{r}_1}$ is the reference MBD force. This force loss requires the model to be at least twice differentiable for implementing gradient-based optimization methods. Therefore, SchNet uses a shifted softplus as the activation function:

$$\text{ssp}\left(x\right) = \ln\left(e^x + 1\right) - \ln 2, \tag{14}$$

which promises a smooth energy surface and improves the convergence by ensuring $\text{ssp}\left(0\right) = 0$.

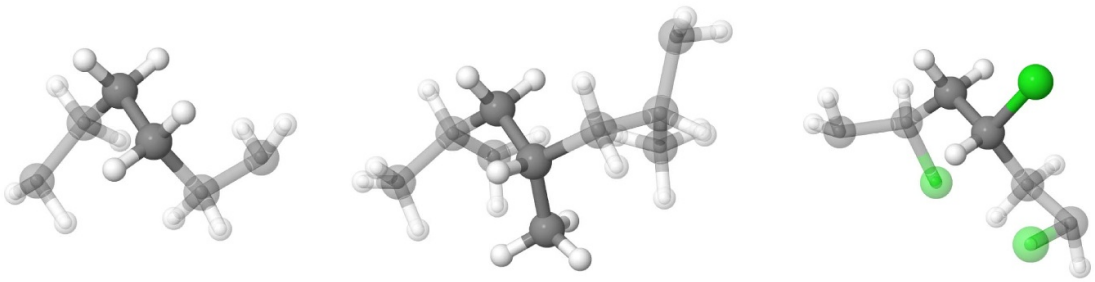
In principle, a SchNet model that solely outputs energy is simpler and requires only half the computational cost [26]. However, our goal is to develop a surrogate model for MBD forces that can be directly integrated into MD simulations. If the model predicts energies only, forces must be obtained at every MD step via an additional differentiation operation, either through a finite difference algorithm or through automatic differentiation (AD). This extra step would inevitably introduce higher computational costs and pose additional challenges for deployment in external MD packages.

2.4. Three polymer melt systems: description and data generation

To construct a dataset with diverse structural variability, three types of polymer melts, PE, PP, and PVC, are generated using the polymer builder in CHARMM-GUI [64, 65]. These polymers span from the simplest structure (PE) to more chemically complex systems: PP incorporates additional side-chain functional groups, while PVC features one more atom type, as presented in figure 3.

For each polymer, multiple configurations with varying chain lengths are considered while maintaining a similar unit cell size across cases: 24×100, 12×200, 8×300, 6×400, and 4×600, where $n \times m$ denotes n polymer chains, each with m monomers. All of these systems are equilibrated using CHARMM-GUI at 300 K to ensure uniform morphology within each polymer type[5]. In this setting, the dataset should naturally capture various levels of crystallinity across the three polymers, arising from their intrinsic molecular structures and thermal behaviors [66, 67], which further enriches the dataset. For example, PE, with a glass transition temperature (T_g) of approximately 160 K, can exhibit more crystalline regions at 300 K due to the high chain mobility above its T_g and its simple and highly symmetric structure for chain packing. In contrast, PVC, with a T_g around 350 K, remains largely amorphous at 300 K as its polymer chains are relatively rigid below T_g and its bulky chlorine side groups further hinder the regular chain packing required for crystallization. However, we acknowledge that the morphological effects, particularly crystallinity, require a more detailed investigation. Due to the limited number

[5] PVC at 300 K is below its glass transition temperature ($T_g \approx 350$ K) and thus technically in a glassy state with very limited mobility. However, we broadly retain the term 'polymer melt' throughout this study for simplicity and consistency. PVC is intentionally included to assess the model's generalization across polymers with different structural and chemical characteristics.



(a) PE. (b) PP. (c) PVC.

Figure 3. Segments of the three types of polymer used in this work, with carbon atoms in dark gray, hydrogen atoms in white, and chlorine atoms in green. The monomer of each polymer (ethylene for PE, propylene for PP, and vinyl chloride for PVC) is highlighted in non-transparent shading.

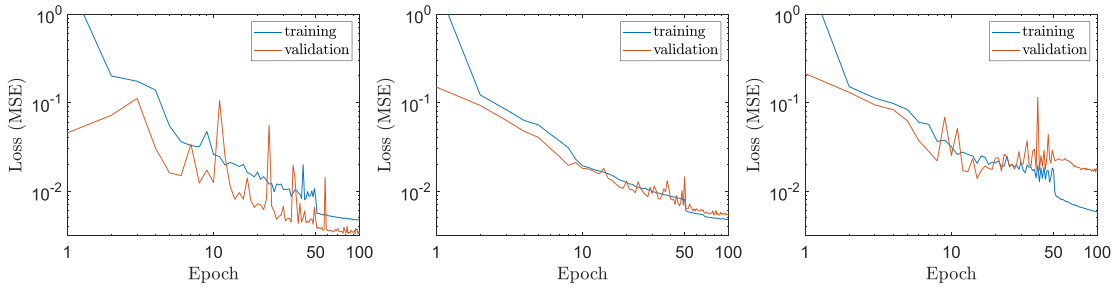
of time steps allowed by CHARMM-GUI, the polymer systems may not achieve sufficient crystallization by the online equilibrating procedure. In this paper, we present a preliminary study on this topic in 3.4, while leaving a comprehensive exploration of crystallinity effects for future work.

For training the neural network, we collect 60 k data points for PE, 72k for PP, and 60 k for PVC, with 10% of each reserved for validation. These data points are extracted equally from different configurations for each type of polymer. Each data point is a $\left((\boldsymbol{R},\boldsymbol{Z}),\hat{\boldsymbol{F}}_{r_1}\right)$ tuple, representing an atomic cluster with $N_{\text{cut}} = 1000$ atoms and the MBD force on its center atom (scaled by 10^3 for numerical stability) generated by our open-source toolkit [46]. This number of N_{cut} corresponds to a cutoff distance of approximately 14 Å, significantly larger than the typical range of 6–9 Å used in classical PW methods due to the long-range nature of MBD. While N_{cut} could be adjusted depending on the specific MBD variant and the density of the specific polymer melt, or reduced to accommodate stricter computational resource constraints, we select $N_{\text{cut}} = 1000$ as a general-purpose value for demonstration, ensuring applicability across all cases in this paper. Our cutoff strategy is consistent with a recently reported statistical analysis of MBD forces in similar PE melt systems [61]. At this cutoff, the contributions from peripheral atoms decay smoothly towards negligible magnitudes, such that the outer region effectively serves as a buffer and changes in the cutoff neighborhood during MD lead to smooth force variations rather than discontinuous jumps, analogous to standard cutoff treatments in classical PW vdW models. The selected clusters always include complete data points from the smallest repeating structural units in the polymers: methylene groups (CH_2) for PE, propylene monomers (C_3H_6) for PP, and vinyl chloride monomers (C_2H_3Cl) for PVC. This enables a special batching strategy to improve convergence of training as discussed in section 3.2.3. Regarding the edge atoms of polymer chains, although they are not directly involved in the training due to their minor representation in the structure, the model demonstrates reasonable predictive capability for them after being trained on our sufficiently extensive dataset.

2.5. Implementation details

2.5.1. Training setting

In the implementation, the embedding dimension is set to $P = 32$, and is maintained throughout the rbf encoding block and the interaction block. The projection block at the end reduces the dimension to 16 and then to 1 by two dense layers. Again, we reduce P by a factor of 2 compared to the original SchNet, which is tailored to our surrogate modeling problem. $p = 2$ nearest neighbors are included for the extra connections with $N_{\text{extra}} = 50$, and we use $N_{\text{rbf}} = 100$ for the rbf encoding. The effects of these two settings are discussed in section 3.2. For each case in the following subsection, unless otherwise specified, the model is trained for 100 epochs with a batch size of 36. The learning rate is initialized at 10^{-3} and reduces to 10^{-4} after 50 epochs. AdamW [68] is used as the optimizer with a weight decay of 0.004. All the implementations in this work are carried out using TensorFlow and executed on the HPC facilities of the University of Luxembourg [69]. The proposed training strategy is generally effective for the three polymers considered in this work, as shown in figure 4. Given the robustness of the SchNet-based architecture, our focus is on understanding the model and its adaptation to the MBD surrogate problem. Fine-tuning hyperparameters and searching for the optimal training strategy for each dataset are beyond the scope of this paper.



(a) PE. (b) PP. (c) PVC.

Figure 4. Training and validation convergence of the model on three polymer datasets, following the proposed training strategy.

2.5.2. Evaluation metrics

To evaluate the performance of the trained models, an additional test dataset consisting of 6 k data points is collected for each polymer. Due to the uneven distribution of atom types within the polymers and the generally smaller average MBD force magnitudes observed on backbone atoms, e.g. carbon atoms in this study, the neural network naturally fits better for hydrogen and chlorine atoms. This trend is evident in the results shown later. Furthermore, the dataset inherently contains near-zero force components, arising from coincidental symmetries in the molecular geometry. Considering these factors, we adopt the mean absolute relative error (MARE) as the evaluation metric (in percentage) w.r.t mean force magnitude:

$$\mathcal{E}_{\text{MARE}}\left(\mathcal{D}_{\text{test}},\boldsymbol{\theta}\right)=\frac{1}{3N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}\sum_{j=1}^{3}\frac{\left|h\left(\left(\boldsymbol{R}_i,\boldsymbol{Z}_i\right),\boldsymbol{\theta}\right)_j-\hat{F}_{\boldsymbol{r}_1 ij}\right|}{\tilde{F}_{\boldsymbol{r}_1}}\times 100\%, \tag{15}$$

where $\tilde{F}_{\boldsymbol{r}_1}=\frac{1}{3N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}\sum_{j=1}^{3}\left|\hat{F}_{\boldsymbol{r}_1 ij}\right|$ is the mean force magnitude per degree of freedom. This approach ensures an unbiased assessment of the model's predictive capability. When evaluating for one specific type of atom, $\tilde{F}_{\boldsymbol{r}_1}$ will be the mean of only that kind.

Later in this paper, we also occasionally use two other metrics to support analysis of model's performance:

$$\mathcal{E}_{\text{angle}}\left(\mathcal{D}_{\text{test},i},\boldsymbol{\theta}\right)=G\left(\boldsymbol{h}\left(\left(\boldsymbol{R}_i,\boldsymbol{Z}_i\right),\boldsymbol{\theta}\right),\hat{\boldsymbol{F}}_{\boldsymbol{r}_1 i}\right), \tag{16}$$

that measures the error of force angle (°) of an atom with $G(\boldsymbol{u},\boldsymbol{v})=\arccos\left(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right)\times\frac{180}{\pi}$, and

$$\mathcal{E}_{\text{ARE}}\left(\mathcal{D}_{\text{test},i},\boldsymbol{\theta}\right)_j=\frac{\left|h\left(\left(\boldsymbol{R}_i,\boldsymbol{Z}_i\right),\boldsymbol{\theta}\right)_j-\hat{F}_{\boldsymbol{r}_1 ij}\right|}{\tilde{F}_{\boldsymbol{r}_1}}\times 100\%, \tag{17}$$

that provides the force error per degree of freedom of an atom.

3. Results and discussions

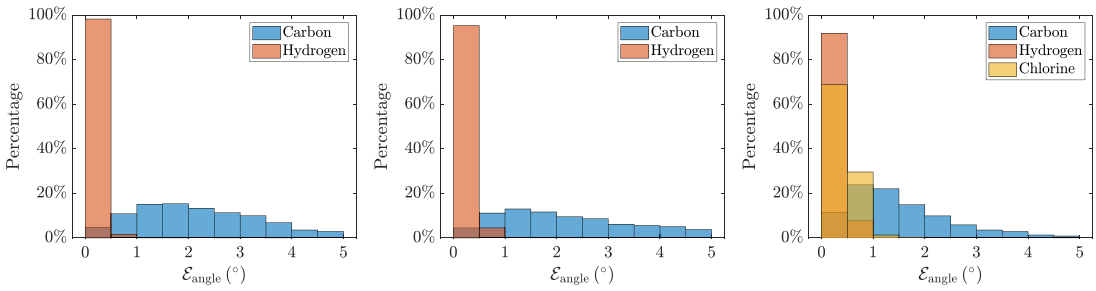
In this section, we will demonstrate the predictive capability and robustness of the proposed trimmed SchNet model, enhanced by our novel modifications. First, in section 3.1, we will present the model's performance on the three polymer melt systems, including mixed-polymer datasets. Then, in section 3.2, three key components of the model will be analyzed: trainable rbf encoding, specialized extra connections, and a dedicated unit-specific batching strategy. In section 3.3, we will further examine the performance of the model for advanced MBD variants, which will be followed, in section 3.4, by a preliminary study on temperature effects. A physically meaningful analysis of the model's Hessian will be presented in section 3.5, and finally section 3.6 will discuss the prospects of implementing our ML surrogate model in MD simulation environments.

3.1. Model performance across polymers

The ability of the model to efficiently and accurately predict forces and generalize across different polymer systems is crucial for its applicability in large-scale molecular simulations. The trimmed

Table 1. Performance ($\mathcal{E}_{\text{MARE}}$ (%)) of the model trained on single- and mixed-polymer datasets. Values in parentheses indicate standard deviations of $\mathcal{E}_{\text{MARE}}$ (%).

Train \ Test	PE			PP			PVC			
	C	H	all	C	H	all	C	H	Cl	all
PE	**5.59** (7.82)	**0.41** (0.57)	**0.71** (1.03)	59.61 (80.49)	2.04 (4.02)	5.40 (10.03)	— —	— —	— —	— —
PP	15.15 (12.26)	0.46 (1.63)	1.30 (2.41)	**6.38** (6.90)	**0.49** (0.45)	**0.83** (0.89)	— —	— —	— —	— —
PVC	43.13 (28.63)	3.53 (5.00)	5.78 (6.52)	49.81 (40.59)	3.41 (2.61)	6.12 (5.41)	**3.42** (3.40)	**0.60** (1.61)	**1.00** (0.87)	**0.98** (1.75)
PE+PP	6.54 (8.57)	0.44 (0.58)	0.79 (1.10)	8.03 (8.14)	0.56 (0.51)	1.00 (1.05)	— —	— —	— —	— —
PE+PVC	6.46 (7.68)	0.54 (0.74)	0.88 (1.15)	35.60 (30.19)	1.12 (0.88)	3.13 (3.90)	3.56 (3.40)	0.61 (0.67)	1.06 (1.50)	1.02 (1.20)
PE+PP+PVC	7.11 (8.54)	0.48 (0.69)	0.86 (1.18)	8.40 (7.30)	0.63 (0.58)	1.09 (1.03)	3.90 (3.63)	0.63 (0.75)	1.09 (1.28)	1.07 (1.12)



(a) PE. (b) PP. (c) PVC.

Figure 5. Histogram of angular error ($\mathcal{E}_{\text{angle}}$ (°)) for the model trained on each single-polymer dataset and tested on the corresponding test dataset.

SchNet model, with its lightweight architecture, enables fast inference, achieving computation times of 3 ms atom^{-1} (batch size of 1) and 0.13 ms atom^{-1} (batch size of 1000), depending on parallelization and memory constraints. This efficiency allows for extensive evaluations across different polymer systems. Furthermore, the practical execution of the trained model in MD simulations, where computational requirements are more complex and evolving, has also been proven to be efficient in section 3.6.

To assess its predictive accuracy and generalization capability, we trained the model on both single- and mixed-polymer datasets following the training setup described in section 2.5.1. The trained models were then tested on various polymer test datasets to evaluate their ability to transfer across polymer types and adapt to unseen structures. The test results are presented in table 1. For all tables in this paper, bold values denote the best performance on each test dataset.

For single-polymer datasets, the model achieves low errors on the corresponding test datasets, with higher relative errors observed for carbon atoms. This is primarily due to their smaller force magnitudes and lower representation in the polymer structure compared to other atoms. Among the three polymers, the PVC case exhibits lower errors on carbon atoms, attributed to its distinct atomic composition and slightly higher force magnitudes of carbon atoms. However, these errors remain negligible for molecular simulations, given the larger mass of carbon atoms. Notably, compared to a PW model, the MBD model not only alters the magnitude of vdW dispersion forces but also significantly twists their directions due to many-body correlations [61]. This directional alteration is crucial for accurately capturing many-body effects in the dynamical behavior of polymer melts, emphasizing the importance of force direction predictions alongside magnitudes. To assess this, we further examine the atomic force direction (angular) error $\mathcal{E}_{\text{angle}}$ (°), as defined in equation (16), for models trained on single-polymer datasets, corresponding to the bold results in table 1. As shown in figure 5, force predictions for hydrogen atoms exhibit the

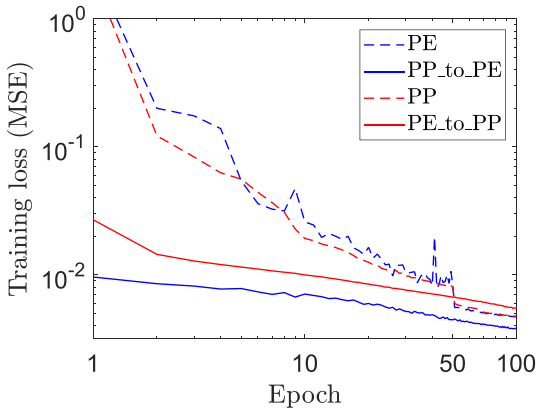


Figure 6. Training convergence of the transfer learning experiment. Solid lines represent transfer learning (learning rate 10^{-4}) between PE and PP: the PP-trained model fine-tuned on PE (blue) and the PE-trained model fine-tuned on PP (red). Dashed lines in corresponding colors indicate standalone training, which begins with a learning rate of 10^{-3} for the first 50 epochs and changes to 10^{-4} for the last 50 epochs.

highest angular accuracy across all three polymers, while chlorine atoms in PVC also maintain an angular error below 1°, in agreement with the observed $\mathcal{E}_{\text{MARE}}$. For carbon atoms, due to the aforementioned limitations, the angular error is generally higher, predominantly ranging between 1° and 5°. Moreover, the model trained on PVC performs best on carbon atoms, aligning with the trend observed in the magnitude-based evaluation. Given the strong correlation and consistency between $\mathcal{E}_{\text{MARE}}$ and $\mathcal{E}_{\text{angle}}$, we do not present angular error analysis for the remaining cases.

Regarding the generalization and transferability of our model, we first evaluate its performance on unseen polymers. In this scenario, the model trained on PP demonstrates superior generalization to PE, likely due to PP's more complex structure, which enriches its feature embeddings while retaining similarities to PE. This observation is further supported by a transfer learning experiment, as shown in figure 6, where the PP-trained model quickly adapts to the PE dataset, reaching a lower loss at 100 epochs compared to standalone training. In contrast, the PE-trained model requires more epochs to converge when transferred to PP, indicating a more challenging adaptation process. The generalization capability is further assessed by training on mixed datasets with equal ratios of different polymers, specifically: 36k PE + 36k PP, 36k PE + 36k PVC, and 24k PE + 24.3k PP + 24k PVC. For testing, we use the same test datasets as in the single-polymer cases. Although the monomer batching strategy introduced earlier is not applicable, the results in table 1 demonstrate the model's capability to generalize across multiple polymer types. Models trained on mixed datasets achieve competitive test errors across all individual polymers, highlighting their potential for applications involving diverse polymer systems.

In addition to the mean error values, table 1 includes their standard deviations in parentheses. In most cases, the standard deviations are comparable to or larger than the mean values due to the skewed and heavy-tailed nature of the error distributions, as confirmed by the histograms presented in figure 5 in this section and figure 9 in section 3.3. Such distributions reflect the intrinsic complexity of the MBD force landscape and configurational diversity in polymer melt systems. For clarity, standard deviations are omitted in subsequent tables, as they focus on model comparisons and the overall statistical trends remain consistent.

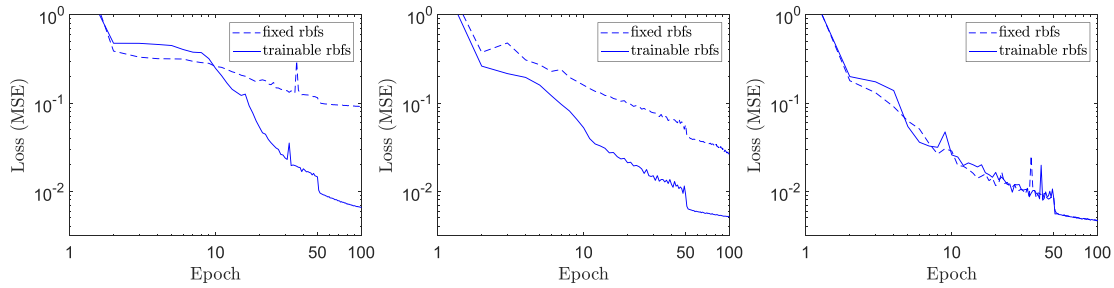
3.2. Analysis of key components in trimmed SchNet

3.2.1. Effects of extra connections

Regarding the extra connections, we use $p = 2$ neareast neighbors and $N_{\text{extra}} = 50$ for all cases. This introduces only 10% more connections to be encoded, requiring no additional parameters and having a negligible impact on computational cost. As shown in table 2, $p = 2$ provides the most consistent accuracy across all polymer types, offering notable improvements over no extra connection ($p = 0$) or only $p = 1$ nearest neighbor. While $p = 3$ may yield slight gains for certain cases, the changes are marginal under the same training strategy. The enhancement from these extra connections aligns with the many-body nature of the fitting target: by incorporating extra connections closer to the center, even if not directly linked to it, the model can more effectively capture finer details of the local atomic environment through enriched feature embeddings of nearby atoms. This is analogous to the coupling of

Table 2. Effects of extra connections on model performance ($\mathcal{E}_{\text{MARE}}$ (%)) for different polymers. Results are shown for varying p values (number of nearest neighbors) with N_{extra} fixed at 50.

p	0				1				2				3			
Test	C	H	Cl	all	C	H	Cl	all	C	H	Cl	all	C	H	Cl	all
PE	8.45	0.65	—	1.10	6.67	0.46	—	0.81	5.59	**0.41**	—	**0.71**	**5.43**	0.46	—	0.74
PP	7.48	0.64	—	1.04	10.22	0.54	—	1.10	6.38	**0.49**	—	**0.83**	**6.20**	0.56	—	0.89
PVC	4.39	0.80	1.25	1.28	3.82	**0.55**	**0.93**	0.97	3.42	0.60	0.98	1.04	**3.34**	0.57	0.99	**0.95**



(a) $N_{\text{rbf}} = 10$. (b) $N_{\text{rbf}} = 20$. (c) $N_{\text{rbf}} = 100$.

Figure 7. Training convergence of the model on PE dataset using different rbf encoding strategies. Three N_{rbf} values are considered. Solid curves represent trainable rbfs and dashed curves represent fixed rbfs.

Table 3. Effects of N_{rbf} and trainable rbfs on model performance ($\mathcal{E}_{\text{MARE}}$ (%)) for PE.

N_{rbf}	10			20			100		
Test	C	H	all	C	H	all	C	H	all
Fixed	14.92	1.98	2.71	10.15	0.89	1.40	**5.61**	**0.45**	**0.75**
Trainable	6.19	0.51	0.83	6.00	0.45	0.76	**5.59**	**0.41**	**0.71**

dipole–dipole interactions in MBD via diagonalization. In our neural network model, instead, interactions are 'coupled' through the non-linearity of the proposed shifted softplus activation function, as shown in equation (14).

3.2.2. Performance of trainable rbfs

For the rbf encoding, the original SchNet uses a single constant value for all γ_k and places μ_k at uniform intervals of 0.1 Å across the range of interest. This approach is reasonable for general molecular datasets, where structural configurations or densities vary significantly. However, for a polymer cluster, as shown in figure 1(c), evenly and finely distributed encoding bases are excessive, given the more consistent spherical geometry, the center-focused target, and the decaying behavior of vdW forces. In this work, we set $N_{\text{rbf}} = 100$, sufficient to cover all cases, and make both γ_k and μ_k trainable to optimize the encoding strategy. While trainable rbfs may not significantly improve model accuracy for large N_{rbf}, since finer fixed bases can adequately encode structure if computational resources are not constrained, they can accelerate training convergence for smaller N_{rbf}. In particular, for $N_{\text{rbf}} = 10$ and 20, the trainable rbfs exhibit a much faster loss decay and achieve lower final training loss compared to fixed rbs, as shown in figure 7. Here, the fixed rbf uses $\gamma_k = 10$ and μ_k uniformly spaced between 0 and 15 Å, which also serves as the initialization for the trainable case. This ability to maintain reasonable accuracy with very few rbfs enables the development of extremely lightweight models, as demonstrated in table 3, whereas models with fixed rbfs under the same settings experience a substantial loss in predictive power.

Moreover, with sufficient trainable rbfs, SchNet encodes atomic structures in a more physically meaningful way, enhancing the model's robustness. As shown in figure 8, rbf centers cluster naturally in the short- and mid-range distances (see black vertical lines), with only a few extending to the long range, where they are weighted by decaying γ_k values over distance (see the blue line). In the proposed Gaussian-type rbf, the coefficient γ controls the width of the function, and a small γ allows the rbf to attend a wider range of neighbors. The observed distribution pattern reveals that the model autonomously identifies regions with the most predictive information and averages long-range contributions

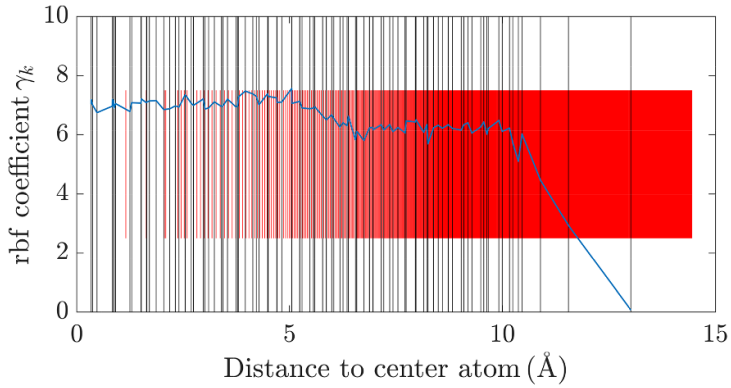


Figure 8. Visualization of the trained rbfs for PE with $N_{\text{rbf}} = 100$. The x-axis shows distance relative to the center atom, and the y-axis shows rbf coefficient γ_k. Black vertical lines indicate the learned positions of the rbf centers μ_k, red lines mark the averaged interatomic distances for each neighbor relative to the center atom, and the blue curve illustrate the corresponding γ_k values, controlling the width of each basis function.

Table 4. Effects of batching strategies on model performance ($\mathcal{E}_{\text{MARE}}$ (%)) for different polymers.

Batching	normal				unit-specific			
Test	C	H	Cl	all	C	H	Cl	all
PE	**5.53**	0.42	—	0.71	5.59	**0.41**	—	**0.71**
PP	7.2	0.52	—	0.91	**6.38**	**0.49**	—	**0.83**
PVC	3.73	0.66	1.04	1.07	**3.42**	**0.60**	**0.98**	**1.04**

without requiring fine details. Notably, at short distances, where nearest-neighbor interactions dominate, an intriguing pattern can be observed: rbf centers cluster between interatomic distances (red vertical lines) to better capture local structural features while avoiding zero gradients, as governed by equation (11).

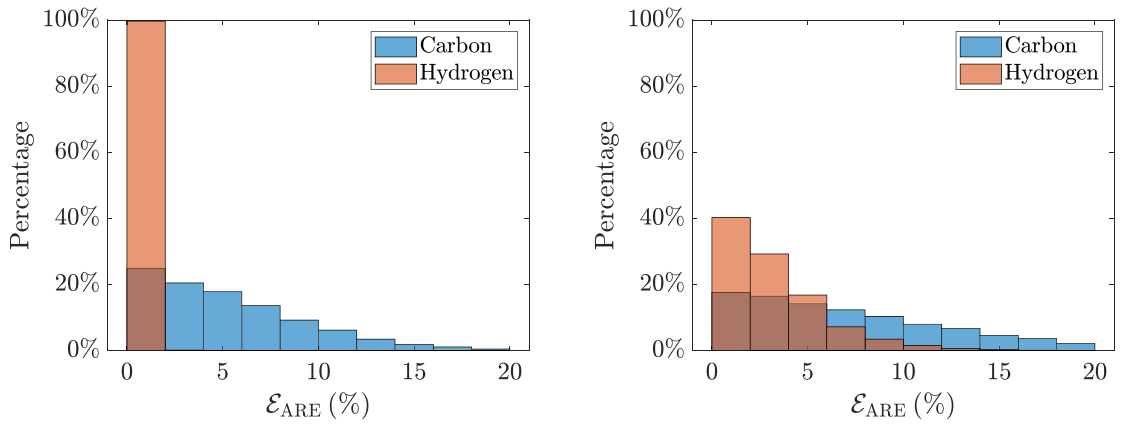
3.2.3. Training with unit-specific batching

In the extreme case where N_{extra} includes all atoms in the system, the setup effectively merges the $p+1$ trimmed connection graphs, allowing direct force predictions for all $p+1$ atoms. With a limited number of p neighbors, the concerning atoms share the same local atomic environment. As a result, using a model with a single common geometry input and a multi-force output, such as predicting the forces of a methylene group collectively, could potentially improve accuracy. This is attributed to the more complete connection graph and a more explicit 'coupling' achieved during the output phase through gradient descent optimization of the combined force predictions. However, this approach has practical limitations. It is incompatible with the neighbor-search routines in commonly used MD packages, such as JAX_MD [70, 71], where atomic geometries are not necessarily ordered in a desired manner that can align with the fixed multi-force output. Additionally, edge atoms in polymer chains require special treatment, since they do not conform to the fixed-size output structure, making it more complicated for implementation.

To simulate the multi-force output scenario while keeping a single-output model, we introduce a unit-specific batching strategy during training. This approach groups atoms from the same smallest repeating unit into a single batch. For example, in the training for PE, each batch of size 36 contains 12 complete ethylene groups (CH_2), while for PP, the same batch size includes 4 complete propylene monomers (C_3H_6). By grouping atoms this way, the optimizer can more effectively 'couple' them, as the model updates them collectively within the same optimization step. Although this technique does not upgrade the model itself to predict multiple outputs, it smoothens the optimization landscape by leveraging a physically meaningful averaging across each batch. As shown in table 4, the largest error reduction occurs for PP (9-atom unit), while the effect is marginal for PE (3-atom unit). This further demonstrates the many-body nature of the force, where coupling larger atomic groups leads to a more accurate description.

Table 5. Performance ($\mathcal{E}_{\text{MARE}}$ (%)) of the model for two MBD variants in PE.

Model	MBD (plain)			MBD@rsSCS		
Test	C	H	all	C	H	all
PE	5.59	0.41	0.71	7.61	3.25	4.49



(a) MBD (plain). (b) MBD@rsSCS.

Figure 9. Histogram of $\mathcal{E}_{\text{ARE}}$ (%) of the model for two MBD variants in PE.

3.3. Performance for the advanced MBD model

The MBD method introduced in section 2.1 is commonly referred to as the 'plain' MBD. Building on the same methodology, several MBD variants have been developed, including MBD@rsSCS [72], universal MBD (uMBD) [73], and nonlocal MBD [74]. Among these, the MBD@rsSCS version (range-separation self-consistent screening) aligns with the scope of this work that is mainly focused on organic systems. MBD@rsSCS is an advanced formulation designed to seamlessly integrate with density functional theory (DFT)-based models, providing a more accurate treatment of polarizability screening across different ranges. This is achieved by applying the range separation technique to the dipole-dipole interaction tensor $\boldsymbol{T}$, allowing a proper derivation of short-range screening and long-range correlation.

For the surrogate modeling, MBD@rsSCS represents a more complex and challenging target compared to the plain version, with an expected loss in predictive accuracy. To explore this, we collected an MBD@rsSCS dataset for PE, revealing notable significant changes in force magnitudes (generally smaller) and distributions, with the forces on carbon and hydrogen atoms becoming closer in magnitude. As shown in table 5, the SchNet model trained on MBD@rsSCS forces, with the same architecture and training strategy (scaling forces by 10^4 instead), exhibits a decreased ability to predict these forces, particularly for hydrogen atoms. The error for hydrogen increases from 0.41% to 3.25%, reflecting their reduced force values relative to carbon. The near-2:1 error ratio aligns with the atomic composition in the dataset. This disparity is further illustrated in the error histogram in figure 9, where $\mathcal{E}_{\text{ARE}}$ (%) is used as the error metric (see equation (17)). The results show that hydrogen errors for MBD@rsSCS show significantly lower concentration near zero compared to the plain MBD results, while the distribution of carbon remains similar.

The reduced performance of the model on MBD@rsSCS can be attributed to the increased intrinsic complexity of this variant. Compared to the plain MBD, MBD@rsSCS introduces short-range self-consistent screening and modifies the dipole interaction tensor at long range [72], both of which significantly alter the overall force landscape and force magnitudes. These effects likely make the learning task more challenging. To achieve a similar level of accuracy as in the plain MBD case, we would need to increase the size of the model or apply dedicated hyperparameter tuning to better capture the more complex force distribution. However, this does not indicate a limitation but rather demonstrates the expected trade-off when addressing a more complex problem. Essentially, this test demonstrates the robustness of our trimmed SchNet in handling altered force distributions, highlighting its adaptability. Furthermore, it motivates the exploration of new techniques to accommodate varying levels of target complexity in the future. Within the scope of this work, where we aim to develop a surrogate model for MBD that integrates with general FF models for molecular simulations, the plain MBD fits better than

Table 6. Performance ($\mathcal{E}_{\mathrm{MARE}}$ (%)) of the model trained on mixed-temperature dataset of PE.

Train \ Test	mixed-size			mixed-temperature		
	C	H	all	C	H	all
mixed-size	**5.59**	**0.41**	**0.71**	**5.50**	0.43	0.71
mixed-temperature	5.90	0.45	0.76	5.52	**0.42**	**0.71**

other MBD variants. Hence, the current trimmed SchNet adequately meets our requirements for this application.

3.4. Preliminary study of temperature effects

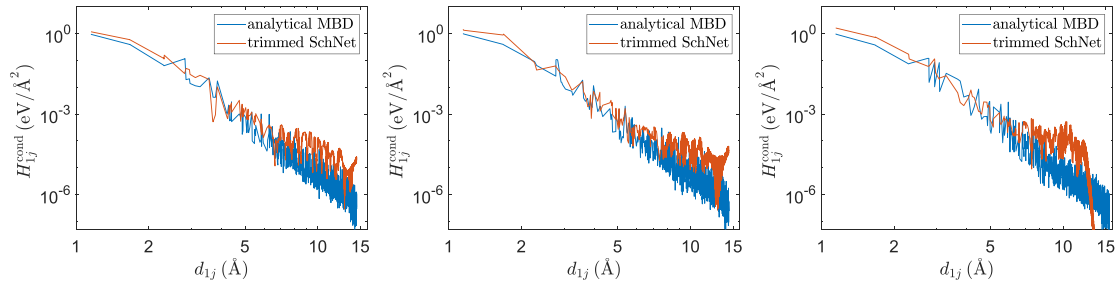
Despite the strong predictive capability of the proposed trimmed SchNet model for our polymer melt dataset, we acknowledge that the obtained melt structures may not fully represent realistic crystallization behaviors. Ideally, at the same equilibration temperature of 300 K, the three types of polymer should exhibit varying degrees of crystallinity due to their different structural and thermal properties. However, this distinction is likely underrepresented in the dataset due to the limited equilibration time steps permitted by the online CHARMM5GUI. Nevertheless, we are still able to conduct a preliminary study on the effects of temperature.

Here, we focus on one configuration of PE and construct an additional mixed-temperature dataset. We equilibrate an 8×300 PE melt at five different temperatures, ranging from 100 K to 300 K, which leads to mild volume changes, with the unit cell length spanning from 55.14 Å to 56.41 Å. From this dataset, we again collect 60k data points for training and reserve an additional 6k for testing. The model is trained on this mixed-temperature dataset using the same training strategy as in previous cases. As shown in table 6, we cross-compare this newly trained model with the model trained on the original dataset introduced in section 2.4, which is equilibrated at 300 K but includes different chain lengths (referred to as the mixed-size dataset). The two models exhibit similar performance when tested on both datasets, despite the fact that 1/5 of their data points are identical.

These results demonstrate that the proposed model generalizes well across polymer chain lengths and minor density variations. The consistent accuracy, even across a wide temperature range, further suggests that the representation of crystallinity in the mixed-temperature dataset is limited, leading to predominantly amorphous structures. This, in turn, may enhance the regularity of the cutoff cluster, thereby simplifying the surrogate modeling task. We acknowledge that a more rigorous validation would require testing on systems with pronounced crystallinity, which could be achieved through longer equilibration or dedicated crystalline sampling. Such systems will be included in future work to evaluate the model's robustness under reduced structural regularity and the presence of ordered domains.

3.5. Analysis of the Hessian

Differing from generic deep neural network models, such as fully connected neural networks, SchNet incorporates a more physically meaningful architecture by applying convolutional operations to PW atomic connections. This is analogous to the analytical MBD force formulation, where the dipole matrix $\boldsymbol{C}^{\mathrm{MBD}}$ undergoes differentiation and scaling through eigen-decomposition products. Similarly, SchNet effectively scales the convolved atomic connections through activations and computes forces via a gradient operation. Thanks to the infinite continuity of softplus activation functions and TensorFlow's AD function, it is possible to extract the second derivative of a trained SchNet model, corresponding to the Hessian $\boldsymbol{H}_{1j}$. This tensor serves as a valuable tool for analyzing atomic interactions, as it quantifies the force response on a center atom due to infinitesimal displacements of neighboring atoms within the cutoff range [75]. In figure 10, we evaluate the Frobenius norm of $\boldsymbol{H}_{1j}$, referred to as the 'condensed Hessian' $H_{1j}^{\mathrm{cond}} = \|\boldsymbol{H}_{1j}\|_{\mathrm{F}}$, and compare the condensed Hessian profiles obtained from the analytical MBD model and the trained trimmed SchNet model for a randomly chosen carbon atom in different polymer melts. This condensed form serves as a scalar representation for the effective force constants associated with MBD interactions, offering a physically intuitive and easily visualizable measure of how MBD stiffness decays with interatomic distance. For all cases, wavy behaviors are observed at long range, attributed to many-body correlation effects and varying atom types over distance. In the analytical Hessian profiles, we find a polynomial decay power below 7, which is lower than the decay power of 8 for the PW case. This further underscores the non-trivial modifications introduced by MBD to the power laws that govern the vdW dispersion interactions [62], while also justifying the need for an extended cutoff distance for MBD force calculations.



(a) PE. (b) PP. (c) PVC.

Figure 10. Condensed Hessian profiles for a random carbon atom in different polymers. In the figures, we order H_{1j}^{cond} in terms of the corresponding interatomic distance to the center atom d_{1j} for clarity.

Acknowledging the physical significance of the Hessian, our model effectively reproduces the decaying behavior observed in the analytical reference and loosely captures the intricate short-range details, further demonstrating the network's robustness. Notably, the Hessian profile for the PVC system reveals a rapid decay near the cutoff boundary, suggesting that the current N_{cut} is excessive for this polymer. This observation is reasonable given that the longer C–Cl bonds result in a larger radius for the cutoff cluster, thus reducing the number of atoms required for calculating MBD forces. This analysis provides valuable insights for optimizing the cutoff strategy, allowing us to tailor the cutoff range more precisely for different polymer systems.

3.6. Towards implementation for MD simulations

We have open-sourced the polymer melt dataset [76] and the TensorFlow code of the trimmed SchNet architecture [60] to facilitate future research on surrogate modeling of MBD interactions. One of the most direct use of our surrogate model would be its integration in MD simulations. However, such an integration can be challenging, and in this section we discuss possible approaches to do it efficiently.

Technically, as the MD simulation framework we propose using JAX_MD, with the trimmed SchNet model introduced as a custom potential implemented with Flax_NNX [77]. Both libraries provide JAX-based APIs [78], enabling a smooth and straightforward integration of the neural network into a standard MD simulation. As demonstrated in the demo provided in our repository, the user simply needs to define a force function that calls the SchNet model to calculate the MBD forces for a given atomic system $\boldsymbol{R}$. Due to VRAM limitations, inference is performed iteratively in a for-loop, processing subsets of the full system.

To estimate the computational performance of the model, we have carried out an NVT simulation of a 9k-atom system on a 32GB Nvidia V100 GPU (with an inference batch size of 3000). The model predictions achieve a speed of 0.02 ms/atom/step, leveraging JAX's high-performance computing capabilities. This is several order of magnitude faster to a single analytical MBD calculation for one atom in a polymer melt, which requires approximately 1 s.

To demonstrate that the proposed surrogate can be stably integrated into an MD workflow, we performed a preliminary MD test using a practical FF coupling strategy for polymer melts. As discussed in section 2.1, the MBD surrogate model serves as the attractive component of the vdW interaction, and must therefore be combined with additional interaction terms to enable MD simulations. In this context, the surrogate is coupled to TraPPE [79], a classical FF model for polymers, following the implementation strategy detailed in [80]. The simulation is conducted at 300 K in the NVT ensemble for a PE system with approximately 9k atoms, consisting six polymer chains. As a simple and widely used structural observable, the time evolution of the radius of gyration $R_{\mathrm{g}}(t)$ is monitored for each chain for around 150 picoseconds. The resulting trajectories shown in figure 11 exhibit smooth and bounded behavior over a relatively long simulated time window, indicating that the surrogate model behaves stably when used in a standard MD workflow. The purpose of this test is solely to demonstrate numerical stability and practical integrability of the model, rather than to validate polymer structural or thermodynamic properties.

The above mentioned practical coupling is in general not sufficient. Deploying the MBD surrogate as part of a complete and quantitatively predictive FF for polymer melts requires the inclusion of bonded interactions and a *consistent* short-range repulsive component for the vdW interaction. In principle, the surrogate can be coupled to an existing or lightly tuned LJ repulsive term within a classical FF to enable rapid, qualitative investigations of MBD effects relative to the PW model. For rigorous,

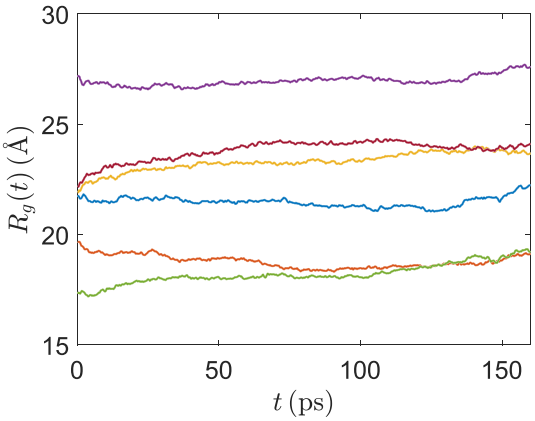


Figure 11. Time evolution of the radius of gyration $R_g(t)$ for six polymer chains in a PE melt during an NVT MD simulation using the trimmed SchNet MBD surrogate. Each curve corresponds to one polymer chain and represents an average over multiple independent random-seed simulations.

quantitative coupling, however, the repulsive parameters should be refitted using the same physical targets employed in the original FF development. For example, in TraPPE, LJ parameters are calibrated to reproduce vapor–liquid phase equilibria, and replacing the attractive component with the MBD surrogate would naturally be accompanied by re-calibrating the repulsive term to preserve these target properties.

In even longer perspective, a more physically consistent integration can be achieved by coupling the surrogate to a short-range MLFF such as SO3LR or GEMS. These general-purpose MLFFs already model bonded and short-range non-bonded interactions at high fidelity by fitting to DFT-level data, whereas explicit long-range MBD effects are not yet incorporated in these approaches. Moreover, in the context of the range-separation concept as discussed in [46], DFT-based models inherently lack a description of long-range correlation and require a vdW dispersion complement. This makes coupling to the MBD surrogate both smooth and physically consistent. Such a combination would yield a complete MLFF with accurate short-range physics and long-range MBD interactions. Developing and validating this coupling, including tests on structural, dynamic, and thermodynamic properties from MD simulations, are in progress but lie beyond the scope of the present manuscript.

4. Conclusions and future work

In this paper, we introduced a trimmed SchNet architecture as a surrogate model specifically designed for predicting MBD forces in polymer melts. Despite the widely recognized success of MBD over PW methods in accurately modeling many-body vdW effects, and the rapid growth of MLFFs, dedicated ML surrogates for MBD remain largely unexplored. Our work addresses this gap by presenting a model tailored for MBD in large-scale molecular simulations. Polymer melts were chosen as the initial application due to their strong dependence on vdW interactions, and their structural characteristic guided both our modeling strategy and architecture design.

In large-scale MD simulations, vdW forces are computed repeatedly for the center atom within a finite-size atomic cluster across the full polymer melt system, making MBD surrogate predictions follow the same pattern. Due to the near-spherical shape of the cutoff clusters, this surrogate task exhibits surprising regularity, making it suitable for ML modeling. The proposed SchNet-based architecture retains the core geometric encoding and continuous-filter convolution techniques of the original SchNet while introducing key modifications specific to our problem. The trimmed SchNet not only reduces the model size but also simplifies the connection graph by keeping only interactions between the center atom and others, with a few additional connections to the two nearest neighbors of the center. This significantly reduces computational cost while preserving essential many-body correlations. Additionally, the model employs trainable rbf encoding, reducing the number of required basis functions, and incorporates a unit-specific batching strategy that leverages the repetitive nature of polymer monomers to enhance training convergence.

The model was trained on MBD force datasets for PE, PP, and PVC melts, demonstrating high predictive accuracy across all three polymer systems. Furthermore, it generalizes well when trained on a combined dataset of multiple polymers. When applied to advanced MBD variants, such as MBD@rsSCS,

it reveals the challenges posed by more complex force distributions. In addition, a preliminary study of temperature effects shows that the model maintains accuracy against mild volume changes in the structure, further demonstrating its robustness. Moreover, the trimmed SchNet model exhibits strong physical interpretability. The analysis of the trained model's Hessian shows that it captures the characteristic decay behavior of MBD interactions, offering insights into cutoff optimization for different polymer systems. Finally, we discussed the practical implementation of the model in MD simulations, supported by a demo using the JAX_MD library, which showcases its applicability to large-scale MD simulations. To maximize the impact of this work, all the codes and datasets used in this study are available in our repository [60].

With the proposed trimmed SchNet model, one can already integrate it into MD simulations of large-scale polymer melts, provided that it is consistently coupled with an appropriate FF model for the remaining interactions in the system. While the present work demonstrates the numerical stability and practical feasibility of such integration, achieving a rigorous and quantitatively predictive coupling requires further effort. Our immediate future work will focus on this direction, with particular interest in potential emergent MBD effects that could lead to significant differences in polymer melt morphology and mechanical properties compared to the results from PW vdW models.

However, we acknowledge certain limitations of the proposed architecture. The success of the model is closely tied to the structural characteristics of the obtained polymer melts, where the resulting regularity of the cutoff clusters ensures a well-defined learning task. The impact of crystallinity on predictive performance remains untested. This also suggests that the current model may not generalize well to molecular systems with highly asymmetric and irregular geometries around each target atom, such as large protein molecules. In future work, we will first evaluate the model's performance in crystalline polymer structures and assess the influence of cutoff regularity. We also plan to extend our approach to more complex polymers and other condensed-phase systems, such as liquids. Ultimately, we aim to generalize the MBD surrogate to a broader class of molecular systems and integrate it into state-of-the-art MLFF frameworks such as SO3LR [45] and GEMS [29].

Finally, this work is part of the wider effort within our research group towards ML surrogate modeling at different scales [29, 45, 81–86] and towards effective multiscale modeling [11, 46, 87]. We hope our contribution and open-source repository will serve the community of computational engineering for future research on *ab initio* modeling of large-scale systems.

Acknowledgments

We are grateful for the support of the Luxembourg National Research Fund (C20/MS/14782078/QuaC). The calculations presented in this paper were carried out using the HPC facilities of the University of Luxembourg.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/Zhaoxiang-Shen/trimmed-SchNet [60]; https://doi.org/10.5281/zenodo.15012728 [76].

Appendix. many-body dispersion (MBD) formalism in detail

The MBD method [9, 10] is based on the adiabatic connection fluctuation dissipation theorem (ACFDT) within the random phase approximation (RPA) and models the system by quantum harmonic oscillators (QHOs) coupled through dipole-dipole interactions [10]. Unlike pairwise methods, the MBD model accounts for all orders of dipole-dipole interactions among fluctuating atoms. The ACFDT-RPA correlation energy for the MBD model is expressed as:

$$E^{\mathrm{MBD}} = \frac{1}{2\pi}\int_0^\infty \mathrm{Tr}\left[\ln\left(\mathbf{1} - \boldsymbol{A}\boldsymbol{T}\right)\right] \mathrm{d}\omega, \tag{A.1}$$

where $\boldsymbol{A}$ is a diagonal $3N \times 3N$ matrix for a system of N atoms. For isotropic QHOs, the matrix elements are defined as $A_{lm} = -\delta_{lm}\alpha_l(i\omega)$, with $\alpha_l(i\omega)$ representing the lth frequency-dependent atomic

polarizability. $\boldsymbol{T}$ is so called dipole-dipole interaction tensor:

$$T_{ij}^{ab} = \nabla_{\boldsymbol{r}_i} \otimes \nabla_{\boldsymbol{r}_j} v_{ij}^{gg}, \tag{A.2}$$

where a and b denote the Cartesian coordinates, and v_{ij}^{gg} is a modified Coulomb potential [88], incorporating overlap effects between fluctuating point dipoles:

$$v_{ij}^{gg} = \frac{\mathrm{erf}\left(r_{ij} / \left(\beta \cdot \tilde{\sigma}_{ij}\right)\right)}{r_{ij}}. \tag{A.3}$$

The parameter β in equation (A.3) is an empirical constant and $\tilde{\sigma}_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2}$ represents an effective width derived from the Gaussian widths σ_i and σ_j of atoms i and j, respectively. These widths are directly linked to the polarizabilities, α_i, as per classical electrodynamics [89] and can be computed from the dipole self-energy:

$$\sigma_i = \left(\sqrt{\frac{2}{9\pi}}\alpha_i\right)^{1/3}. \tag{A.4}$$

To improve computational efficiency, the dipole-dipole interaction is assumed to be frequency-independent using the effective polarizability $\alpha_i^{0,\mathrm{eff}}$ to calculate σ_i in equation (A.4). For a frequency-independent $\boldsymbol{T}$ tensor, the MBD energy can be obtained directly from the eigenvalues λ_p of the $3N \times 3N$ matrix $\boldsymbol{C}^{\mathrm{MBD}}$ [10, 72], which describe the coupling between each pair of atoms i and j with a 3×3 block:

$$\boldsymbol{C}_{ij}^{\mathrm{MBD}} = \omega_i^2 \delta_{ij} + \left(1 - \delta_{ij}\right)\ \omega_i \omega_j \sqrt{\alpha_i^{0,\mathrm{eff}} \alpha_j^{0,\mathrm{eff}}}\ \boldsymbol{T}_{ij}. \tag{A.5}$$

where ω_i is the characteristic frequency for atom i, defined as [5]:

$$\omega_i = \frac{4\ C_{6,i}^{\mathrm{free}}}{3\ \left(\alpha_i^{0,\mathrm{free}}\right)^2}. \tag{A.6}$$

The above 'effective' terms are related to their free-atom values with proper scaling, i.e.

$$\alpha_i^{0,\mathrm{eff}} = \alpha_i^{0,\mathrm{free}} \left(\frac{V_i^{\mathrm{eff}}}{V_i^{\mathrm{free}}}\right), \tag{A.7}$$

where the factor $V_i^{\mathrm{eff}} / V_i^{\mathrm{free}}$ represents the ratio of the effective volume occupied by an atom when interacting with its environment, as determined through the Hirshfeld partitioning scheme [90], to the reference volume of the same atom in a free (non-interacting) state. Generally, these volume ratios vary across an atomic structure and need to be retrieved from high fidelity models for accurate representation. In this work, however, we assume the ratio to be a constant value, as its variation is less than 1% in the polymer melts under consideration. The MBD energy is finally evaluated as the difference between interacting and non-interacting frequencies as presented in equation (3) and the MBD force is formulated as in equation (4), both in section 2.1.

ORCID iDs

Zhaoxiang Shen 0000-0002-5184-812X
Raúl I Sosa 0000-0003-4116-2973
Jakub Lengiewicz 0000-0003-3947-525X
Alexandre Tkatchenko 0000-0002-1012-4854
Stéphane P A Bordas 0000-0001-7622-2193

References

[1] Liu Y, Huang Y and Duan X 2019 Van der waals integration before and beyond two-dimensional materials *Nature* **567** 323–33
[2] Dill K A 1990 Dominant forces in protein folding *Biochemistry* **29** 7133–55
[3] Autumn K, Sitti M, Liang Y A, Peattie A M, Hansen W R, Sponberg S, Kenny T W, Fearing R, Israelachvili J N and Full R J 2002 Evidence for van der waals adhesion in gecko setae *Proc. Natl Acad. Sci.* **99** 12252–6
[4] Grimme S, Hansen A, Brandenburg J G and Bannwarth C 2016 Dispersion-corrected mean-field electronic structure methods *Chem. Rev.* **116** 5105–54
[5] Tkatchenko A and Scheffler M 2009 Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data *Phys. Rev. Lett.* **102** 073005
[6] Ambrosetti A, Alfé D, DiStasio R A and Tkatchenko A 2014 Hard numbers for large molecules: toward exact energetics for supra-molecular systems *J. Phys. Chem. Lett.* **5** 849–55
[7] Reilly A M and Tkatchenko A 2013 Seamless and accurate modeling of organic molecular materials *J. Phys. Chem. Lett.* **4** 1028–33
[8] Marom N, DiStasio R A, Atalla V, Levchenko S, Reilly A M, Chelikowsky J R, Leiserowitz L and Tkatchenko A 2013 Many-body dispersion interactions in molecular crystal polymorphism *Angew. Chem. Int. Edn* **52** 6629–32
[9] Tkatchenko A, DiStasio R A, Car R and Scheffler M 2012 Accurate and efficient method for many-body van der waals interactions *Phys. Rev. Lett.* **108** 236402
[10] Tkatchenko A, Ambrosetti A and DiStasio R A 2013 Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem *J. Chem. Phys.* **138** 074106
[11] Hauseux P, Nguyen T T, Ambrosetti A, Ruiz K S, Bordas S P A and Tkatchenko A 2020 From quantum to continuum mechanics in the delamination of atomically-thin layers from substrates *Nat. Commun.* **11** 1651
[12] Hoja J, Ko H-Y, Neumann M A, Car R, DiStasio R A and Tkatchenko A 2019 Reliable and practical computational description of molecular crystal polymorphs *Sci. Adv.* **5** eaau3338
[13] Poier P P, Lagardére L and Piquemal J-P 2022 O(n) stochastic evaluation of many-body van der waals energies in large complex systems *J. Chem. Theory Comput.* **18** 1633–45
[14] Muhli H, Ala-Nissila T and Caro M A 2025 Atom-wise formulation of the many-body dispersion problem for linear-scaling van der waals corrections *Phys. Rev. B* **111** 054103
[15] Cornell W D, Cieplak P, Bayly C I, Gould I R, Merz K M, Ferguson D M, Spellmeyer D C, Fox T, Caldwell J W and Kollman P A 1995 A second generation force field for the simulation of proteins, nucleic acids and organic molecules *J. Am. Chem. Soc.* **117** 5179–97
[16] Christen M *et al* 2005 The gromos software for biomolecular simulation: Gromos05 *J. Comput. Chem.* **26** 1719–51
[17] Jorgensen W L and Tirado-Rives J 1988 The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin *J. Am. Chem. Soc.* **110** 1657–66
[18] Vanommeslaeghe K *et al* 2010 Charmm general force field: a force field for drug-like molecules compatible with the charmm all-atom additive biological force fields *J. Comput. Chem.* **31** 671–90
[19] Noé F, Tkatchenko A, Müller K-R and Clementi C 2020 Machine learning for molecular simulation *Annu. Rev. Phys. Chem.* **71** 361–90
[20] Unke O T, Chmiela S, Sauceda H E, Gastegger M, Poltavsky I, Schütt K T, Tkatchenko A and Müller K-R 2021 Machine learning force fields *Chem. Rev.* **121** 10142–86
[21] Poltavsky I and Tkatchenko A 2021 Machine learning force fields: recent advances and remaining challenges *J. Phys. Chem. Lett.* **12** 6551–64
[22] Keith J A, Vassilev-Galindo V, Cheng B, Chmiela S, Gastegger M, Müller K-R and Tkatchenko A 2021 Combining machine learning and computational chemistry for predictive insights into chemical systems *Chem. Rev.* **121** 9816–72
[23] Schütt K T, Chmiela S, Lilienfeld O A V, Tkatchenko A, Tsuda K and Müller K-R 2020 Machine learning meets quantum physics *Lecture Notes in Physics* (Springer)
[24] Peivaste I *et al* 2025 Artificial intelligence in materials science and engineering: current landscape, key challenges and future trajectories *Compos. Struct.* **372** 119419
[25] Chmiela S, Sauceda H E, Müller K-R and Tkatchenko A 2018 Towards exact molecular dynamics simulations with machine-learned force fields *Nat. Commun.* **9** 3887
[26] Schütt K T, Kindermans P-J, Sauceda H E, Chmiela S, Tkatchenko A and Müller K-R 2017 Schnet: a continuous-filter convolutional neural network for modeling quantum interactions *Proc. 31st Int. Conf. on Neural Information Processing Systems* pp 992–1002 (available at: https:/dl.acm.org/doi/abs/10.5555/3294771.3294866)
[27] Schütt K T, Sauceda H E, Kindermans P-J, Tkatchenko A and Müller K-R 2018 Schnet – a deep learning architecture for molecules and materials *J. Chem. Phys.* **148** 241722
[28] Unke O T and Meuwly M 2019 Physnet: a neural network for predicting energies, forces, dipole moments and partial charges *J. Chem. Theory Comput.* **15** 3678–93
[29] Unke O T *et al* 2024 Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments *Sci. Adv.* **10** eadn4397
[30] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403
[31] De S, Bartók A P, Csányi G and Ceriotti M 2016 Comparing molecules and solids across structural and alchemical space *Phys. Chem. Chem. Phys.* **18** 13754–69
[32] Willatt M J, Musil F and Ceriotti M 2018 Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements *Phys. Chem. Chem. Phys.* **20** 29661–8
[33] Grisafi A and Ceriotti M 2019 Incorporating long-range physics in atomic-scale machine learning *J. Chem. Phys.* **151** 204105
[34] Huguenin-Dumittan K K, Loche P, Haoran N and Ceriotti M 2023 Physics-inspired equivariant descriptors of nonbonded interactions *J. Phys. Chem. Lett.* **14** 9612–8
[35] Deringer V L, Caro M A and Csányi G 2020 A general-purpose machine-learning force field for bulk and nanostructured phosphorus *Nat. Commun.* **11** 5461
[36] Christensen A S, Bratholm L A, Faber F A and von Lilienfeld O A 2020 Fchl revisited: faster and more accurate quantum machine learning *J. Chem. Phys.* **152** 044107
[37] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401

[38] Behler J 2011 Atom-centered symmetry functions for constructing high-dimensional neural network potentials *J. Chem. Phys.* **134** 074106
[39] Frank T, Unke O and Müller K-R 2022 So3krates: equivariant attention for interactions on arbitrary length-scales in molecular systems *Advances in Neural Information Processing Systems* vol **35** pp 29400–13
[40] Frank J T, Unke O T, Müller K-R and Chmiela S 2024 A euclidean transformer for fast and stable machine learned force fields *Nat. Commun.* **15** 6539
[41] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B 2022 E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13** 2453
[42] Batatia I, Kovacs D P, Simm G, Ortner C and Csányi G 2022 Mace: higher order equivariant message passing neural networks for fast and accurate force fields *Advances in Neural Information Processing Systems* vol **35** pp 11423–36
[43] Anstine D M and Isayev O 2023 Machine learning interatomic potentials and long-range physics *J. Phys. Chem. A* **127** 2417–31
[44] Unke O T, Chmiela S, Gastegger M, Schütt K T, Sauceda H E and Müller K-R 2021 Spookynet: learning force fields with electronic degrees of freedom and nonlocal effects *Nat. Commun.* **12** 7273
[45] Kabylda A, Frank J T, Dou S S, Khabibrakhmanov A, Sandonas L M, Unke O T, Chmiela S, Müller K-R and Tkatchenko A 2025 *Molecular Simulations With a Pretrained Neural Network and universal Pairwise Force Fields* (https://doi.org/10.26434/chemrxiv-2024-bdfr0-v2)
[46] Shen Z, Sosa R I, Bordas S P A, Tkatchenko A and Lengiewicz J 2024 Quantum-informed simulations for mechanics of materials: Dftb+mbd framework *Int. J. Eng. Sci.* **204** 104126
[47] Strobl G 2007 *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior* (Springer) (https://doi.org/10.1007/978-3-540-68411-4)
[48] Francis L F 2016 Chapter 3 - melt processes *Materials Processing* (Academic) pp 105–249
[49] Van Krevelen D W and Nijenhuis K T 2009 *Properties of Polymers: Their Correlation With Chemical Structure; Their Numerical Estimation and Prediction From Additive Group Contributions* (Elsevier)
[50] Bicerano J 2002 *Prediction of Polymer Properties* (CRC Press) (https://doi.org/10.1201/9780203910115)
[51] Hossain D, Tschopp M A, Ward D K, Bouvard J L, Wang P and Horstemeyer M F 2010 Molecular dynamics simulations of deformation mechanisms of amorphous polyethylene *Polymer* **51** 6071–83
[52] Baig C, Mavrantzas V G and Kröger M 2010 Flow effects on melt structure and entanglement network of linear polymers: results from a nonequilibrium molecular dynamics simulation study of a polyethylene melt in steady shear *Macromolecules* **43** 6886–902
[53] Yamamoto T 2019 Molecular dynamics simulation of stretch-induced crystallization in polyethylene: emergence of fiber structure and molecular network *Macromolecules* **52** 1695–706
[54] Romanos N, Megariotis G and Theodorou D N 2021 Molecular dynamics simulations of stretch-induced crystallization in layered polyethylene *Polym. Crystallization* **4** e10172
[55] Kotelyanskii M and Theodorou D N 2004 *Simulation Methods for Polymers* (CRC Press) (https://doi.org/10.1201/9780203021255)
[56] Theodorou D N 2007 Hierarchical modelling of polymeric materials *Chem. Eng. Sci.* **62** 5697–714
[57] Gartner T E and Jayaraman A 2019 Modeling and simulations of polymers: a roadmap *Macromolecules* **52** 755–86
[58] Schmid F 2023 Understanding and modeling polymers: the challenge of multiple scales *ACS Polym. Au* **3** 28–58
[59] Martin T B and Audus D J 2023 Emerging trends in machine learning: a polymer perspective *ACS Polym. Au* **3** 239–58
[60] Shen Z, Sosa R I, Lengiewicz J, Tkatchenko A and Bordas S P A 2025 Github repository: Ml surrogate models of many-body dispersion (available at: https://github.com/Zhaoxiang-Shen/trimmed-SchNet)
[61] Sosa R I, Galante M and Tkatchenko A 2025 Power of the many-body force: magnitudes and angles of atomic van der waals dispersion forces in extended molecular systems *Small Struct.* **6** 2500226
[62] Ambrosetti A, Ferri N, DiStasio R A and Tkatchenko A 2016 Wavelike charge density fluctuations and van der waals interactions at the nanoscale *Science* **351** 1171–6
[63] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, u. Kaiser L and Polosukhin I 2017 Attention is all you need *Advances in Neural Information Processing Systems* vol **30** pp 6000–10
[64] Jo S, Kim T, Iyer V G and Im W 2008 Charmm-gui: a web-based graphical user interface for charmm *J. Comput. Chem.* **29** 1859–65
[65] Choi Y K, Park S-J, Park S, Kim S, Kern N R, Lee J and Im W 2021 Charmm-gui polymer builder for modeling and simulation of synthetic polymers *J. Chem. Theory Comput.* **17** 2431–43
[66] Reiter G and Sommer J 2003 Polymer crystallization: obervations, concepts and interpretations *Lecture Notes in Physics* (Springer)
[67] Reiter G and Strobl G 2007 *Progress in Understanding of Polymer Crystallization* (Springer)
[68] Loshchilov I, Hutter F 2019 Decoupled weight decay regularization (arXiv:1711.05101)
[69] Varrette S, Bouvry P, Cartiaux H and Georgatos F 2014 Management of an academic hpc cluster: the ul experience *2014 Int. Conf. on High Performance Computing & Simulation (HPCS)* pp 959–67
[70] Schoenholz S S and Cubuk E D 2020 Jax m.d. a framework for differentiable physics *Advances in Neural Information Processing Systems* vol 33 pp 992–1002
[71] Schoenholz S S and Cubuk E D 2021 Jax, m.d. a framework for differentiable physics *J. Stat. Mech.* 124016
[72] Ambrosetti A, Reilly A M, DiStasio R A and Tkatchenko A 2014 Long-range correlation energy calculated from coupled atomic response functions *J. Chem. Phys.* **140** 18A508
[73] Kim M, Kim W J, Gould T, Lee E K, Lebégue S and Kim H 2020 Kim, umbd: a materials-ready dispersion correction that uniformly treats metallic, ionic and van der waals bonding *J. Am. Chem. Soc.* **142** 2346–54
[74] Hermann J and Tkatchenko A 2020 Density functional model for van der waals interactions: unifying many-body atomic approaches with nonlocal functionals *Phys. Rev. Lett.* **124** 146401
[75] Hauseux P, Ambrosetti A, Bordas S P A and Tkatchenko A 2022 Colossal enhancement of atomic force response in van der waals materials arising from many-body electronic correlations *Phys. Rev. Lett.* **128** 106101
[76] Shen Z, Sosa R I, Lengiewicz J, Tkatchenko A and Bordas S P A 2025 Machine learning surrogate modeling of many-body dispersion interactions in polymer melts *Zenodo* (https://doi.org/10.5281/zenodo.15012728)
[77] Heek J, Levskaya A, Oliver A, Ritter M, Rondepierre B, Steiner A and van Zee M 2024 Flax: a neural network library and ecosystem for JAX (available at: http://github.com/google/flax)
[78] Bradbury J, Frostig R, Hawkins P, Johnson M J, Leary C, Maclaurin D, Necula G, Paszke A, VanderPlas J, Wanderman-Milne S and Zhang Q 2018 JAX: composable transformations of Python+NumPy programs (available at: http://github.com/jax-ml/jax)
[79] Chen B and Siepmann J I 1999 Transferable potentials for phase equilibria. 3. explicit-hydrogen description of normal alkanes *J. Phys. Chem. B* **103** 5370–9

[80] Shen Z 2025 Efficient quantum-informed computational frameworks for mechanics of materials: many-body dispersion and machine learning surrogate modeling *PhD Thesis* University of Luxembourg (available at: https://hdl.handle.net/10993/66537)
[81] Deshpande S, Lengiewicz J and Bordas S P A 2022 Probabilistic deep learning for real-time large deformation simulations *Comput. Methods Appl. Mech. Eng.* **398** 115307
[82] Deshpande S, Bordas S P A and Lengiewicz J 2024 MAgNET: a graph U-Net architecture for mesh-based simulations *Eng. Appl. Artif. Intell.* **133** 108055
[83] Deshpande S, Sosa R I, Bordas S P A and Lengiewicz J 2023 Convolution, aggregation and attention based deep neural networks for accelerating simulations in mechanics *Front. Mater.* **10** 1128954
[84] Deshpande S, Rappel H, Hobbs M, Bordas S P A and Lengiewicz J 2025 Gaussian process regression + deep neural network autoencoder for probabilistic surrogate modeling in nonlinear mechanics of solids *Comput. Methods Appl. Mech. Eng.* **437** 117790
[85] Papavasileiou P, Koronaki E D, Pozzetti G, Kathrein M, Czettl C, Boudouvis A G and Bordas S P A 2023 equation-based and data-driven modeling strategies for industrial coating processes *Comput. Ind.* **149** 103938
[86] Vijayaraghavan S, Wu L, Noels L, Bordas S P A, Natarajan S and Beex L A A 2023 A data-driven reduced-order surrogate model for entire elastoplastic simulations applied to representative volume elements *Sci. Rep.* **13** 12781
[87] Chen L, Berke P Z, Massart T J, Bordas S P A and Beex L A A 2022 An adaptive multiscale quasicontinuum approach for mechanical simulations of elastoplastic periodic lattices *Mech. Res. Commun.* **126** 104019
[88] Dombroski J P, Taylor S W and Gill P M W 1996 Kwik: Coulomb energies in o(n) work *J. Phys. Chem.* **100** 6272–6
[89] Mayer A 2007 Formulation in terms of normalized propagators of a charge-dipole model enabling the calculation of the polarization properties of fullerenes and carbon nanotubes *Phys. Rev. B* **75** 045407
[90] Hirshfeld F 1977 Bonded-atom fragments for describing molecular charge densities *Theor. Chim. Acta* **44** 129–38