https://doi.org/10.5194/egusphere-2026-696
Preprint. Discussion started: 16 February 2026

Machine learning interatomic potentials with accurate long-range interactions for molecular dynamics collision simulations of atmospherically-relevant molecules

Ivo Neefjes[1], Jakub Kubečka[2], and Jonas Elm[1]

[1]Aarhus University, Department of Chemistry, Langelandsgade 140, Aarhus, DK 8000
[2]DTU, Department of Chemical and Biochemical Engineering, Søltofts Plads, Kongens Lyngby, DK 2800

Correspondence: Ivo Neefjes (ivo.neefjes@chem.au.dk)

Abstract. Molecular collisions and subsequent clustering events are fundamental to atmospheric cluster formation. Accurately modeling these processes requires interatomic potentials that simultaneously capture the long-range forces governing collision kinetics and the short-range quantum effects driving reactivity. In this work, we evaluate the AIMNet2 and PaiNN machine learning architectures trained on GFN1-xTB and ωB97X-3c quantum chemical data for molecular collisions involving sulfuric acid.

The models exhibit low mean absolute errors in energies and forces and accurately reproduce potentials of mean force relative to GFN1-xTB. Comparing models trained on GFN1-xTB and ωB97X-3c data reveal that while increasing the electronic structure theory level significantly alters the potential energy surface in the binding region, it has a negligible impact on the long-range shoulder and the resulting collision rate coefficients. Notably, PaiNN demonstrates superior performance in reproducing binding and repulsive regions, making it highly effective for sampling stable cluster configurations.

However, discrepancies are observed for the collision dynamics. While AIMNet2 accurately reproduces reference collision rates across all systems, PaiNN underestimates the rate for the charged sulfuric acid–bisulfate system by $\sim 50\%$. This error originates from the model's local atomic environment approximation, which neglects the strong long-range attractive forces at large intermolecular distances. Comparisons with the OPLS-AA force field demonstrate that simple fixed partial charges are sufficient to describe these interactions.

Our results highlight that while local equivariant models like PaiNN offer exceptional accuracy for thermodynamics, correctly simulating collision kinetics in systems with strong long-range interactions requires models that explicitly account for forces beyond the local environment, such as AIMNet2.

1 Introduction

Atmospheric aerosol particles influence the climate by affecting cloud formation and scattering sunlight (Haywood and Boucher, 2000), while also posing risks to human health via inhalation (Gan et al., 2013). According to the latest IPCC assessment report, interactions between aerosols and clouds remain one of the largest sources of uncertainty in current global climate models

(Chen et al., 2021). A major contributor to this uncertainty is the difficulty of accurately modeling the earliest stages of particle formation (Tröstl et al., 2016).

Most atmospheric aerosol particles form through a gas-to-particle conversion process called new particle formation (NPF), in which gas-phase molecules collide and stick together to form clusters (Kulmala et al., 2013). These clusters then grow further through condensation and coagulation (Zhang et al., 2012). The earliest stages of this formation are inherently dynamic: molecules approach each other under the influence of long-range attractive forces (e.g., van der Waals or electrostatic interactions), then rearrange and relax to accommodate one another while forming a thermodynamically stable cluster.

Cluster-forming collisions are quantified by a collision rate coefficient, which represents the frequency of collisions per unit concentration. In practical applications, such as simulations using the Atmospheric Cluster Dynamics Code (ACDC) (McGrath et al., 2012), this coefficient is typically calculated using kinetic gas theory. In this framework, colliding partners are approximated as hard spheres, and intermolecular interactions are neglected. While analytical approaches like the central field model can account for long-range forces, they require interaction parameters that are significantly more difficult to determine than standard hard-sphere radii (Neefjes et al., 2025).

Several studies have calculated canonical collision rate coefficients for atmospherically relevant systems using atomistic molecular dynamics (MD) collision trajectory simulations in the free molecular regime (Halonen et al., 2019; Neefjes et al., 2022; Yang et al., 2023; Knattrup et al., 2025; Tikkanen et al., 2025). These calculations explicitly account for long-range interactions and provide a fully atomistic description. Furthermore, the resulting trajectories offer insight into the molecular-level dynamics governing collisions and the formation of stable clusters.

Halonen et al. (2019) performed MD simulations using the classical OPLS-AA (Optimized Potentials for Liquid Simulations–All Atom) force field (Jorgensen et al., 1996), reporting an enhancement factor of 2.7 relative to kinetic gas theory for sulfuric acid dimerization. While a well-parameterized classical force field can accurately capture long-range interactions, it cannot account for short-range quantum effects, including chemical reactions such as proton transfers, which play a critical role in stabilizing atmospheric clusters.

Following the development of computationally efficient semi-empirical methods such as GFN1-xTB (Grimme et al., 2017), several studies have employed these approaches for MD collision trajectory or umbrella sampling simulations (Torrie and Valleau, 1977; Knattrup et al., 2025; Kubečka et al., 2025). However, properties derived from such methods can exhibit significant quantitative—and even qualitative—errors. Furthermore, because the computational cost of GFN1-xTB scales cubically with the number of atoms, it becomes impractical for nanosecond-scale simulations of large atmospheric cluster systems. Higher levels of theory, such as the DFT composite method ωB97X-3c (Müller et al., 2023), offer significantly better accuracy, but their computational cost makes even short MD simulations of small systems prohibitively expensive.

Recently, several machine learning (ML) architectures have been developed to construct accurate interatomic potentials for molecular systems. For instance, the polarizable atom interaction neural network (PaiNN) is an equivariant message-passing neural network capable of accelerating MD simulations while maintaining accuracy comparable to the reference training data (Schütt et al., 2021; Kubečka et al., 2024). Similarly, the second-generation atoms-in-molecules neural network (AIMNet2) has

https://doi.org/10.5194/egusphere-2026-696
Preprint. Discussion started: 16 February 2026

demonstrated high predictive accuracy for a wide range of molecular systems with remarkable efficiency, enabling simulations of systems containing up to 10^5 atoms (Anstine et al., 2025).

Machine learning interatomic potentials (MLIPs) often rely on a local atomic environment approximation, in which the model encodes the environment around each atom up to a user-defined cutoff radius. This approximation improves transferability and computational efficiency but inherently limits the model to short-range interactions. The PaiNN model addresses this through a message-passing framework, where atoms exchange information with their neighbors via message and update blocks. Through multiple iterations, the effective interaction range grows, allowing atoms to indirectly access information from beyond the immediate cutoff. However, if all atoms in one subsystem (e.g., a molecule) lie beyond the cutoff radius of another, the interaction graph becomes disconnected. Consequently, no messages are exchanged, and the model treats the subsystems as non-interacting. AIMNet2 mitigates this by supplementing message passing with explicit long-range contributions. It predicts partial charges to model analytical Coulomb interactions and adds dispersion effects via the D3(BJ) correction scheme (Grimme et al., 2010, 2011).

As illustrated by Halonen et al. (2019), long-range interactions can significantly impact properties such as the collision rate coefficient. This is particularly important in ionic systems, where interactions often extend well beyond standard atomic cutoff radii, meaning the full interaction potential is not captured (Neefjes et al., 2022). Recently, several studies have simulated the evolution of systems containing tens of particle-forming molecules to observe cluster formation dynamics (Jiang et al., 2022, 2023; Liu and Jiang, 2025). For these simulations to yield meaningful results, the MLIP must satisfy three key criteria: it must exhibit low force errors to ensure realistic trajectories; it must be computationally efficient to enable sufficiently long simulation times; and finally, it must accurately represent long-range attractive interactions to correctly model collision probabilities.

To this end, we assess the ability of the PaiNN and AIMNet2 architectures to describe collisions governed by long-range interactions. Training configurations were sampled using GFN1-xTB dynamics, with energies and forces subsequently computed at the GFN1-xTB and ωB97X-3c levels. Additionally, delta-learning was employed to upscale GFN1-xTB simulations with PaiNN corrections to the ωB97X-3c level of theory. Since sulfuric acid is a key contributor to particle formation (Sipilä et al., 2010), we studied the sulfuric acid dimer, the sulfuric acid–dimethylamine system (to investigate stabilizing proton transfers), and the sulfuric acid–bisulfate system (to examine strong ionic long-range contributions). Following hyperparameter tuning, model performance was evaluated by comparing electronic energy and force predictions on independent test sets. Furthermore, the potential of mean force (PMF) was calculated through umbrella sampling and compared to reference data obtained with GFN1-xTB. Finally, collision rate coefficients derived from MD trajectories were analyzed to examine how they evolve across different levels of theory.

The remainder of the paper is organized as follows. Section 2 details the computational framework, including the AIMNet2 and PaiNN architectures, and the methodology of the MD and umbrella sampling simulations. Section 3 presents the hyperparameter tuning and training results, followed by an analysis of the PMFs, collision probabilities, and rate coefficients predicted by each model. Finally, Section 4 summarizes our findings and outlines potential future applications.

2 Theory and methods

2.1 Collision systems

We investigated three collision systems containing the atmospherically relevant species sulfuric acid (H_2SO_4), dimethylamine ($NH(CH_3)_2$), and bisulfate (HSO_4^-) in the form of the sulfuric acid dimer (H_2SO_4–H_2SO_4), the acid–base system H_2SO_4–$NH(CH_3)_2$, and the ion–molecule system H_2SO_4–HSO_4^-. The structures of these species are shown in Fig. 1.

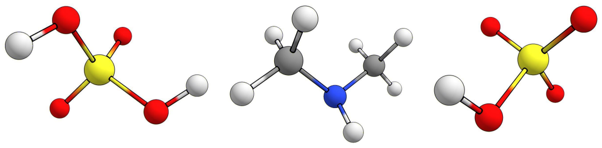

Figure 1. Ball-and-stick representations of sulfuric acid, dimethylamine, and bisulfate. This study considers collision systems of sulfuric acid paired with itself, dimethylamine, and bisulfate. Atom color code: sulfur (yellow), oxygen (red), nitrogen (blue), carbon (gray), and hydrogen (white).

2.2 Machine learning interatomic potentials

2.2.1 PaiNN

The polarizable atom interaction neural network (PaiNN) extends the message-passing formalism of the SchNet architecture by incorporating rotation-equivariant features to handle vectorial information (Schütt et al., 2021). This enables the representation of directional properties, which is essential for accurately describing forces, dipoles, and other tensorial quantities. PaiNN has been trained successfully on relatively small datasets, achieving accuracy competitive with kernel methods (Schütt et al., 2021). However, a limitation in the context of gas-phase collisions is that interactions cannot be transmitted between atoms separated by more than the cutoff distance without intermediate atoms to mediate the message passing. Increasing the cutoff can capture these long-range effects, but at the expense of higher computational cost and potentially reduced accuracy, as the model must learn to generalize over a significantly larger spatial domain.

2.2.2 AIMNet2

AIMNet2 is the second generation of the Atoms-In-Molecules Neural Network developed by Anstine et al. (2025). The model combines local atomic environments with learned "atom-in-molecule" (AIM) embeddings, which are iteratively refined through a message-passing scheme. These embeddings, available for 14 elements (H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, and I), encode both the intrinsic identity of each atom and its chemical environment, providing transferable atomic features that improve generalization across a wide range of chemical systems.

https://doi.org/10.5194/egusphere-2026-696
Preprint. Discussion started: 16 February 2026

AIMNet2 incorporates long-range physics explicitly: electrostatic interactions are modeled via learned partial charges, while dispersion is treated using the D3(BJ) correction scheme (Grimme et al., 2010, 2011). Designed for generalizability and efficiency, the architecture supports systems with varying charge states and spin multiplicities. While it may not match the data efficiency or accuracy of PaiNN for geometry-sensitive properties, its ability to explicitly account for long-range interactions and varying electronic states makes it suitable for a wider range of chemically complex systems.

2.2.3 Delta-learning

Rather than training directly on molecular properties (e.g., electronic energies and forces) at a high level of theory, one can train on the difference between the high-level target and a more computationally efficient, lower-level method. In this framework, molecular dynamics (MD) simulations are performed at the lower level of theory but are corrected to approximate the high level of theory. When the two levels of theory are correlated, this delta-learning approach can substantially reduce model errors. The main drawback is that the overall efficiency is fundamentally bounded by the cost of the lower-level baseline. In this work, we applied delta-learning using the PaiNN architecture to learn the correction between GFN1-xTB as the low-level baseline and ωB97X-3c as the high-level method. We refer to this approach as Δ-PaiNN.

2.2.4 Data generation

For each of the three collision systems, we performed collision trajectory simulations at 300 K (output frequency of 100 steps) and umbrella sampling simulations at 300 K and 500 K (output frequency of 250 steps) using the GFN1-xTB method (TBlite, version 0.2.1) (Ehlert, 2022), following the methodologies detailed in Sec. 2.3.2 and Sec. 2.3.3. All structures from a given system were pooled into a unified candidate dataset.

To construct a comprehensive training set, we employed a sampling strategy guided by the potential of mean force (PMF; see Sec. 2.3.3) along the center of mass distance between the collision partners. Candidate structures were binned by center-of-mass distance, and the number of samples selected per bin was determined by a weighted combination of the local PMF curvature and the raw sampling density. This ensures that regions where the potential energy surface (PES) changes rapidly along the collision coordinate are sampled more densely. Collision trajectory data were included to capture both non-interacting configurations and high-energy collision events, while the 500 K umbrella sampling data were incorporated to cover high-energy fluctuations and prevent model instability during thermal excursions.

Within each bin, we enforced structural diversity using a root-mean-square deviation (RMSD) filter. The RMSD threshold was dynamically relaxed near the PMF minimum to capture equilibrium fluctuations, while a stricter threshold was applied in high-energy regions to maximize configurational coverage. Finally, a global RMSD filter was applied to remove any remaining near-duplicates across the entire dataset, resulting in a final training set of 20,000 structures per system.

Gradient calculations were performed on the selected structures using GFN1-xTB (TBlite, version 0.2.1) and ωB97X-3c (ORCA, version 6.0.1) (Neese, 2012). Owing to the high computational cost, no direct MD simulations were run at the ωB97X-3c level. Instead, it was assumed that the GFN1-xTB PES sufficiently overlaps with the relevant regions of the ωB97X-3c PES.

https://doi.org/10.5194/egusphere-2026-696
Preprint. Discussion started: 16 February 2026

2.3 Molecular dynamics simulations

2.3.1 Force calculations

In MD simulations, atomistic trajectories are generated by integrating Newton's equations of motion over discrete time steps. At each step, the forces acting on the nuclei are computed, and the system is propagated classically.

In this study, we employed several methods for force evaluation. Simulations using the semi-empirical GFN1-xTB method (Grimme et al., 2017) and the trained AIMNet2 and PaiNN models were performed within the Atomic Simulation Environment (ASE) (Hjorth Larsen et al., 2017). These calculations were executed through the `tblite` (Ehlert, 2022), `aimnet2ase`, and `SchNetPack` (Schütt et al., 2024) calculators, respectively. Furthermore, we employed a Δ-learning approach (Δ-PaiNN), wherein baseline GFN1-xTB forces were corrected by a model trained on the difference between GFN1-xTB and the target ωB97X-3c theory (or GFN1-xTB itself for validation).

Additionally, classical MD simulations were performed using the OPLS-AA force field (Jorgensen et al., 1996). These simulations were carried out with the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) code (Plimpton, 1995; Thompson et al., 2022). A detailed description of the OPLS-AA parameters employed in this work can be found in Sec. S1 of the Supporting Information.

2.3.2 Collision trajectory simulations

Canonical collision rate coefficients are given by

$$\beta = 2\pi \int_0^\infty dv_0 \int_0^\infty v_0 f_{\mathrm{MB}}(v_0)\, b\, P_{\mathrm{c}}(v_0, b)\, db, \tag{1}$$

where v_0 is the initial relative velocity between the collision partners, f_{MB} represents the Maxwell–Boltzmann relative speed distribution, and $P_{\mathrm{c}}(v_0, b)$ is the collision probability. The impact parameter b is defined as the perpendicular distance between the initial velocity vectors of the two partners.

To obtain collision rate coefficients from MD simulations, we approximate Eq. (1) with a Riemann sum. The initial relative velocity v_0 is sampled from 50 to 800 $\mathrm{m\,s^{-1}}$ in steps of 50 $\mathrm{m\,s^{-1}}$. For the neutral systems (H_2SO_4–H_2SO_4 and H_2SO_4–$NH(CH_3)_2$), the impact parameter b ranges from 0 to 20 Å in steps of 1 Å. However, for the ionic H_2SO_4–HSO_4^- system, the range is extended to 60 Å to account for strong long-range forces. This velocity range covers 99% of the Maxwell–Boltzmann distribution for all systems. While a maximum impact parameter of 20 Å is sufficient for the neutral pairs (Halonen et al., 2019; Neefjes et al., 2022), the larger cutoff for the ionic system is necessary because electrostatic attraction maintains non-negligible collision probabilities at much greater distances.

For each (v_0, b) pair, we performed 100 independent trajectory simulations. The collision probability $P_{\mathrm{c}}(v_0, b)$ is estimated as the fraction of trajectories that result in a collision. A collision is defined to occur if the center-of-mass distance between the partners falls below the sum of their hard-sphere radii, derived from their liquid bulk densities, for at least one output frame. These sums are 5.5, 5.7, and 5.5 Å for the H_2SO_4–H_2SO_4, H_2SO_4–$NH(CH_3)_2$, and H_2SO_4–HSO_4^- systems, respectively.

Prior to the collision simulations, the monomers were individually equilibrated in NVT simulations using the GFN1-xTB method (TBlite, version 0.2.1). Atomic velocities were initialized from a Maxwell–Boltzmann distribution at 300 K, and the temperature was maintained by a Langevin thermostat with a friction constant of 0.1 fs^{-1}. Each equilibration run lasted 13 ns with a 1 fs timestep and an output frequency of 1,000 steps. Convergence of the total, translational, vibrational, and rotational temperatures was achieved after approximately 3 ns. The remaining 10 ns were sampled every 1,000 steps to yield 10,000 equilibrated starting configurations.

At the start of each collision trajectory, two monomers were randomly selected from their respective equilibrated structures and placed 30 Å apart along the x-axis. This value was chosen as a compromise between minimizing initial interaction forces and limiting computational cost. The collision partners were offset along the z-axis by the impact parameter b and assigned opposing velocities of $v_x = \pm v_0/2$. Each trajectory was propagated in the NVE ensemble for a duration sufficient for a non-interacting particle to traverse the initial separation plus the offset b, with an additional safety margin of 10 Å.

2.3.3 Umbrella sampling

We performed umbrella sampling simulations along the center-of-mass distance coordinate r between the collision partners (Torrie and Valleau, 1977). The reaction coordinate was discretized into 91 windows centered from 2.0 to 20.0 Å with a step size of 0.2 Å. In each window, the system was restrained by a harmonic bias potential $V_{\text{bias}} = \frac{1}{2}k_{\text{bias}}(r - r_i)^2$, where r_i is the window center and $k_{\text{bias}} = 100\,\text{kcal}\,\text{mol}^{-1}\,\text{Å}^{-2}$, consistent with the parameters used by Kubečka et al. (2025).

Starting configurations were drawn from the final 10,000 output frames of the equilibration trajectories described in Sec. 2.3.2. These structures were translated to align with the center of the target umbrella window. For windows at short distances ($r < 6.0$ Å), where direct placement might result in steric clashes, the collision partners were initially placed 6.0 Å apart. The bias potential was then gradually increased from 0 to 100 kcal,mol^{-1} Å^{-2} over the first 2,000 time steps.

To enhance sampling, ten independent simulations were performed for each window. Each simulation began with 100,000 steps of equilibration in the NVT ensemble (Langevin thermostat, friction 0.1 fs^{-1}), followed by a 500,000-step production run using the canonical sampling through velocity rescaling (CSVR) thermostat with a time constant of 25 fs (Bussi et al., 2007). The output frequency for both thermodynamic data and structural configurations was set to 250 steps.

The unbiased free energy profile was reconstructed using the umbrella integration method as implemented in the `umbrella_integration` code (Stroet and Deplazes, 2016). Because this profile represents the Helmholtz free energy $A(r)$ of finding particles at a distance r in three-dimensional space, it includes an entropic term due to the increasing volume of the spherical shell $4\pi r^2 dr$. To obtain the effective interaction potential $w(r)$ (the one-dimensional PMF), we subtracted this radial entropic contribution:

$$w(r) = A(r) - \left[-k_{\text{B}}T\ln(4\pi r^2)\right], \tag{2}$$

where k_{B} is the Boltzmann constant and T the temperature.

3 Results and discussion

3.1 Hyperparameter tuning

Training efficiency and model performance are dependent on the choice of hyperparameters. To optimize the PaiNN and AIMNet2 architectures for our systems, we performed hyperparameter tuning using the Weights and Biases (W&B) platform (Biewald, 2020). For PaiNN, we tuned the number of features, batch size, number of blocks, and radial basis size. For AIMNet2, we tuned the AIM size, number of features, batch size, batches per epoch, vector channels, radial basis size, and learning rate. Testing three values for each hyperparameter results in a total of 81 combinations for PaiNN and 729 for AIMNet2, making a systematic hyperparameter grid search computationally prohibitive. Consequently, we employed a random search.

100-epoch long tuning runs were performed on a subset of 2,000 structures labeled at the GFN1-xTB level of theory. The target was to minimize the validation loss, defined as a weighted sum of the mean squared errors (MSE) in potential energies, atomic force components, and, in the case of AIMNet2, atomic partial charges. Given the importance of accurate forces for stable molecular dynamics (MD) simulations, and the fact that force data ($3 \times N_{\mathrm{atom}}$) vastly outnumber energy data, we assigned a great weight to the force loss. For PaiNN, the energy:force_components weight ratio was set to 1:99. For AIMNet2, which also predicts partial charges, the energy:force_components:atomic_charges weights were set to 9:90:1.

We completed 70 unique tuning runs for PaiNN and 50 for AIMNet2. The results are visualized in the Supporting Information (Figs. S1 and S2). Tables 1 and 2 summarize the tested values and quantify the impact of each hyperparameter using W&B's hyperparameter importance and Pearson correlation metrics. The importance metric, derived from a random forest algorithm, quantifies the relative impact of each hyperparameter on the validation loss. A negative correlation indicates that increasing the parameter value reduces the validation loss.

For PaiNN, the batch size and number of features were identified as the most important hyperparameters, with a smaller batch size and a higher number of features correlating with improved performance. For AIMNet2, the "batches per epoch" was the dominant hyperparameter. This makes sense, as the product of the batch size and batches per epoch determines the total number of samples seen in one epoch. When this product is smaller than the training set size, increasing the batches per epoch

Table 1. Tested hyperparameter values for PaiNN, including the hyperparameter importance and Pearson correlation with respect to the validation loss obtained from Weights and Biases (Biewald, 2020).

PaiNN	Features	Batch size	Blocks	Radial basis
Values	128	2	3	64
	160	4	4	48
	192	8	5	32
Importance	0.163	0.626	0.097	0.115
Correlation	−0.402	−0.180	0.791	0.287

Table 2. Tested hyperparameter values for AIMNet2, including the hyperparameter importance and Pearson correlation with respect to the validation loss obtained from Weights and Biases (Biewald, 2020).

AIMNet2	AIM size	Features	Batch size	Batches per epoch	Vector channels	Radial basis	Learning rate
Values	512 256 128	32 16 8	16 8 4	1,000 500 100	16 12 8	20 16 12	$1 \cdot 10^{-3}$ $4 \cdot 10^{-4}$ $1 \cdot 10^{-4}$
Importance	0.039	0.087	0.022	0.586	0.087	0.142	0.038
Correlation	0.001	−0.178	−0.034	−0.621	−0.267	−0.180	−0.214

exposes the model to a wider range of data. When the product exceeds the dataset size, the model sees data multiple times per epoch (mixed into different batches), which further aids convergence.

Based on these results, we selected the following hyperparameters for the final production models. For PaiNN (and Δ-PaiNN), we used 256 features, a batch size of 2, 4 interaction blocks, and a radial basis size of 32. For AIMNet2, we selected an AIM size of 128, 16 features, a batch size of 8, 16 vector channels, a radial basis size of 20, and a learning rate of $4 \cdot 10^{-3}$. To account for the larger production dataset, the number of batches per epoch was increased to 4,000. The short-range cutoff for AIMNet2 was set to 5 Å. Since PaiNN lacks explicit long-range interactions beyond the local environment, its cutoff was extended to 10 Å. A full list of hyperparameters for each model is provided in Sec. S3 of the Supporting Information.

It is important to note that we did not necessarily identify the optimal hyperparameter combination for our systems. For instance, while our 100-epoch tuning procedure offers a reasonable indication of training behavior, some hyperparameters might converge slower but dominate during longer training. Identifying the best training settings would furthermore require a systematic search over a broader range of values, which is beyond the scope of the current study. Nevertheless, the chosen hyperparameters provide sufficiently low test errors for the collision systems studied here, as discussed in the following subsection.

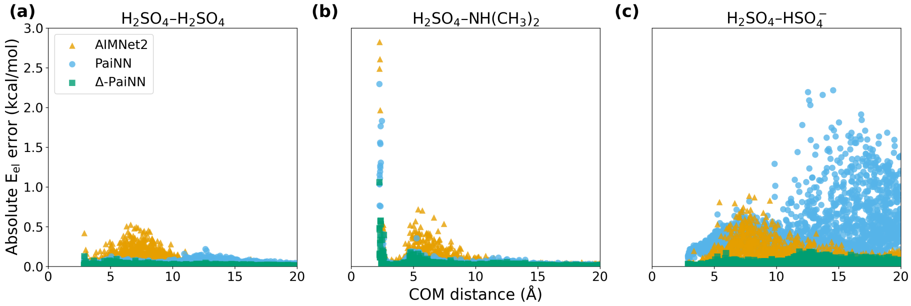

Figure 2. Electronic energy errors as a function of the center-of-mass (COM) distance across the three studied systems: H_2SO_4–H_2SO_4 (a), H_2SO_4–$NH(CH_3)_2$ (b), and H_2SO_4–HSO_4^- (c). Results are shown for AIMNet2, PaiNN, and Δ-PaiNN trained on ωB97X-3c data.

Table 3. Mean absolute errors (MAE) for electronic energies (E_{el}) and component-wise forces (F) across the three studied systems: H_2SO_4–H_2SO_4, H_2SO_4–$NH(CH_3)_2$, and H_2SO_4–HSO_4^-. Results are reported for AIMNet2, PaiNN, and Δ-PaiNN trained on GFN1-xTB and ωB97X-3c training data. Units: E_{el} in kcal mol^{-1} and F in kcal mol^{-1} Å^{-1}.

		GFN1-xTB		**ωB97X-3c**	
System	**Model**	$\mathrm{MAE}_{E_{\mathrm{el}}}$	MAE_F	$\mathrm{MAE}_{E_{\mathrm{el}}}$	MAE_F
H_2SO_4–H_2SO_4	AIMNet2	0.015	0.028	0.056	0.076
	PaiNN	0.023	0.036	0.039	0.051
	Δ-PaiNN	3.4×10^{-6}	1.5×10^{-6}	0.011	0.027
H_2SO_4–$NH(CH_3)_2$	AIMNet2	0.026	0.054	0.075	0.102
	PaiNN	0.020	0.039	0.032	0.062
	Δ-PaiNN	—	—	0.016	0.038
H_2SO_4–HSO_4^-	AIMNet2	0.020	0.036	0.109	0.092
	PaiNN	0.233	0.200	0.417	0.241
	Δ-PaiNN	—	—	0.037	0.054

3.2 Training

Using the hyperparameters determined in Sec. 3.1, we trained AIMNet2, PaiNN, and Δ-PaiNN models for each system trained on either GFN1-xTB or ωB97X-3c data (20,000 structures per system). Training durations were set to 1,000 (AIMNet2), 600 (PaiNN), and 400 (Δ-PaiNN) epochs, chosen to balance their respective convergence rates with computational cost. Model performance was evaluated on a test set of ∼2,000 structures per system from the umbrella sampling data at 300 K to ensure uniform coverage of center-of-mass distances (r) between 2.0 and 20.0 Å. We report the mean absolute errors (MAEs) for electronic energies (E_{el}) and component-wise forces (F) in Table 3. As a sanity check, we also trained a Δ-PaiNN model using GFN1-xTB as both the baseline and the target reference for the H_2SO_4–H_2SO_4 system. Since this model was trained on the difference between two identical levels of theory, it should essentially predict zero correction.

In general, the models achieve excellent accuracy. All MAEs fall below the standard chemical accuracy thresholds of 1 kcal mol^{-1} for energies and 1 kcal mol^{-1} Å^{-1} for forces. Notably, for the neutral H_2SO_4–H_2SO_4 and H_2SO_4–$NH(CH_3)_2$ systems, the errors are exceptionally low, with MAEs below 0.1 kcal mol^{-1} and 0.1 kcal mol^{-1} Å^{-1}. The Δ-PaiNN methodology proves particularly effective. It correctly predicts negligible corrections in the GFN1-xTB sanity check, and consistently achieves the lowest errors when ωB97X-3c is the target level. The highest energy error was observed for the H_2SO_4–HSO_4^- system trained with PaiNN on ωB97X-3c, as expected due to the lack of long-range interactions.

To better understand the distribution of these errors, we analyzed the energy deviations as a function of center-of-mass distance (Fig. 2) for the ωB97X-3c target level of theory. For the H_2SO_4–H_2SO_4 system, performance is consistently low across the entire coordinate. For H_2SO_4–$NH(CH_3)_2$, larger errors appear at short distances ($r < 3$ Å). However, these correspond to the highly repulsive regime of the potential energy surface (PES). In this region, steric hindrance drives the potential energy tens of kcal mol^{-1} above the minimum. Consequently, the probability of accessing these configurations during standard MD is negligible.

In contrast, the ionic H_2SO_4–HSO_4^- system illustrates the inherent limitations of applying the local atomic environment approximation to systems with strong long-range interactions. Although the global MAE for PaiNN appears reasonably low, the distance-resolved error plot (Fig. 2c) reveals significant deviations at separations $r > 10$ Å. This error results from a mismatch between the training labels, which include long-range stabilization, and the model's short-ranged (< 10 Å) representation. At these distances, the reference energies are significantly lowered by electrostatic interactions. However, due to the 10 Å cutoff, PaiNN interprets the collision partners as two non-interacting, free species. Consequently, during training, the model is forced to attribute the substantial stabilization energy of the interacting pair to the local atomic environments of the isolated monomers. In essence, the model erroneously learns that these structures, separated by more than the cutoff yet still interacting, are representative of the free molecular state, resulting in a fundamentally distorted PES.

3.3 Potentials of mean force

3.3.1 GFN1-xTB training data

The potential of mean force (PMF) along the center-of-mass distance represents the effective free energy averaged over all collision orientations accessed during the simulations, showing how the system's stability changes as the collision partners approach. The well depth and shape provide information on the binding strength, while the shoulder towards larger distances reflects the strength of long-range interactions.

Although all the models achieved low mean absolute errors and included high-energy structures (from simulations at 500 K) in the training sets, the umbrella sampling simulations occasionally explored untrained regions of the PES. This caused a breakdown in dynamics and resulted in unphysical geometries. Including these trajectories introduced large errors into the predicted PMFs. Consequently, we implemented a script to filter these simulations by monitoring the distance between any two hydrogen atoms. We discarded any simulation where the maximum distance exceeded the window center by more than 8.0 Å or the minimum distance fell below 1.0 Å. While this script does not test for every possible failure, it is a reasonable compromise between filtering out erroneous simulations and efficiently automating the process over the large amount of data generated. Approximately half of the models contained failed simulations. However, no more than three independent simulation runs were ever removed from a single window. This ensured that every window retained at least 7 simulations (500,000 steps each, saved every 250 steps), providing a dataset of at least 14,000 structures per window for constructing the PMF. A detailed list of removed simulations is provided in Sec. S4 of the Supporting Information.

Figure 3 shows the PMFs for all three systems calculated using AIMNet2 and PaiNN (trained on GFN1-xTB), compared against the GFN1-xTB reference. For the H_2SO_4–H_2SO_4 system, the Δ-PaiNN model trained on GFN1-xTB was again in-

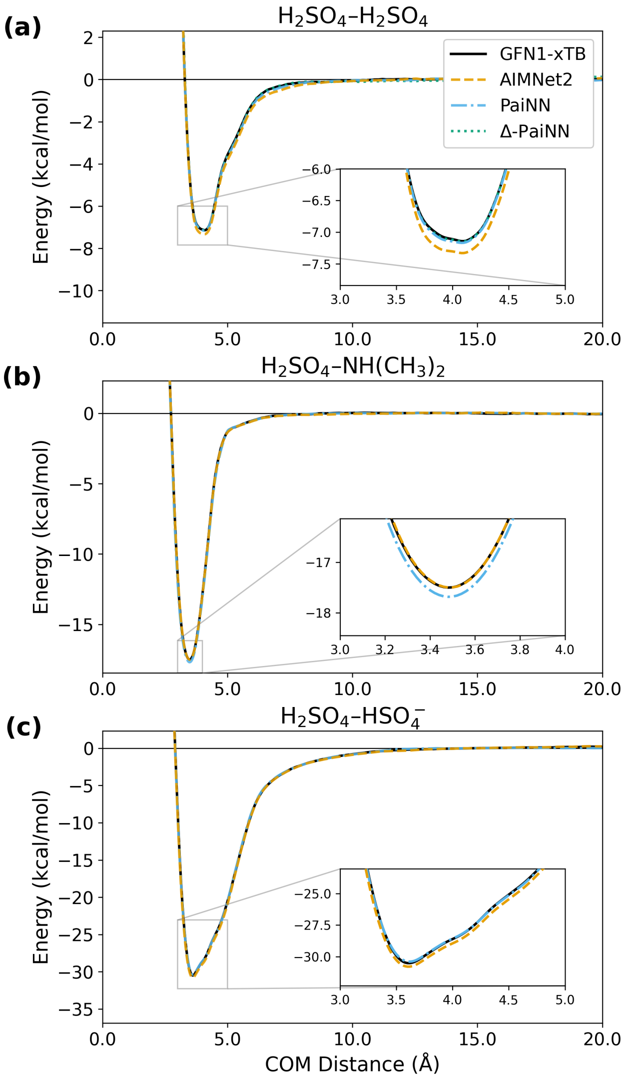

Figure 3. Potentials of mean force (PMF) as a function of center-of-mass (COM) distance for the H_2SO_4–H_2SO_4, H_2SO_4–$NH(CH_3)_2$, and H_2SO_4–HSO_4^- systems. The profiles compare the reference GFN1-xTB level of theory with predictions from AIMNet2, PaiNN, and Δ-PaiNN models trained on GFN1-xTB data.

cluded as a sanity check. While this model should theoretically reproduce the GFN1-xTB reference PMF, small deviations are nonetheless expected due to the inherent uncertainty associated with finite umbrella sampling.

Table 4 lists the root-mean-square errors (RMSEs) of the predicted PMFs relative to the reference. We report the RMSE both over the full distance range (0–20 Å) and for the shoulder region (from the PMF minimum further to the first zero-energy crossing). Theoretically, the PMF should asymptotically approach zero. However, due to small sampling deviations, the zero-energy crossing occurs at finite COM distances (< 20 Å) for these systems.

Table 4. Root-mean-square errors (RMSE) in kcal/mol for the potentials of mean force (PMFs) predicted by the machine learning models relative to the GFN1-xTB reference. The full RMSE evaluates the entire distance range (0–20 Å), while the shoulder RMSE evaluates the region from the minimum to the first zero-energy crossing, thereby excluding the repulsive region.

System	Method	full RMSE	shoulder RMSE
H_2SO_4–H_2SO_4	AIMNet2	3.3	0.15
	PaiNN	0.053	0.058
	Δ-PaiNN	3.4	0.032
H_2SO_4–$NH(CH_3)_2$	AIMNet2	0.90	0.037
	PaiNN	0.16	0.10
	Δ-PaiNN	—	—
H_2SO_4–HSO_4^-	AIMNet2	0.20	0.15
	PaiNN	0.12	0.081
	Δ-PaiNN	—	—

Comparing the full-range and shoulder RMSEs, we observe that the shoulder values are typically much lower. This indicates that the largest discrepancies occur in the repulsive region. As discussed in Sec. 3.2, the energy rises rapidly to tens of $\mathrm{kcal\,mol^{-1}}$ above the minimum as the system moves up the repulsive wall. Since it is highly unlikely that the system will visit these high-energy regions, errors here are less physically significant. Consequently, we focus our analysis on the shoulder RMSEs.

All models exhibit excellent agreement with the reference PMF, with shoulder RMSEs about an order of magnitude lower than chemical accuracy (1 $\mathrm{kcal\,mol^{-1}}$). Notably, PaiNN maintains extremely low MAEs over the full distance range, indicating it correctly models high-energy structures. It also achieves lower shoulder RMSEs than AIMNet2 for the H_2SO_4–H_2SO_4 and H_2SO_4–HSO_4^- systems. While both models correctly predict the H_2SO_4–H_2SO_4 minimum at 4.08 Å, AIMNet2 exhibits an energy error of 0.19 $\mathrm{kcal\,mol^{-1}}$ at this minimum, whereas PaiNN's error is only 0.028 $\mathrm{kcal\,mol^{-1}}$. Thus, where high accuracy is key, PaiNN appears to be the preferred model.

The performance of PaiNN on the H_2SO_4–HSO_4^- system is somewhat surprising, given that it cannot capture interactions beyond its 10 Å local atomic environment cutoff. However, Fig. 3 shows that the PMF effectively decays to zero around 13 Å. As long as any two atoms between the collision partners remain within 10 Å, the message-passing algorithm treats the system as connected. Given that the sum of the hard-sphere radii for H_2SO_4–HSO_4^- is approximately 3.89 Å (Neefjes et al., 2022), PaiNN with cutoff 10 Å can model interactions up to at least 13 Å COM distance. We note, however, that while the PMF vanishes at 13 Å, this represents an average energy. Individual trajectories with specific orientations may still exhibit longer-range interactions.

https://doi.org/10.5194/egusphere-2026-696
Preprint. Discussion started: 16 February 2026

3.3.2 ωB97X-3c training data

We subsequently trained AIMNet2, PaiNN, and Δ-PaiNN on the higher-level ωB97X-3c data to generate the PMFs presented in Fig. 4. The PMFs obtained with the GFN1-xTB method are shown for comparison. First, we observe that all three ML models are in excellent agreement with each other. While obtaining a reference PMF directly using the ωB97X-3c method is computationally prohibitive, the fact that these models, utilizing distinct algorithms, predict very similar PMFs strongly suggests that the ωB97X-3c potential energy surface is accurately reproduced.

Comparing the PMFs based on ωB97X-3c data to the GFN1-xTB reference, we observe that the shoulder regions are similar between methods, while the minima show significant differences. Most notably, ωB97X-3c predicts significantly different well depths: the minima for the H_2SO_4–H_2SO_4 and H_2SO_4–HSO_4^- systems are lower by 2.3 and 3.8 $\mathrm{kcal\,mol^{-1}}$, respectively, compared to GFN1-xTB. For the H_2SO_4–$NH(CH_3)_2$ complex, the methods differ on both the position (shifted by 0.18 Å) and the depth (difference of 1.3 $\mathrm{kcal\,mol^{-1}}$) of the minimum, indicating distinct lowest free energy geometries.

In several of our recent studies, we have integrated the PMF well to obtain binding free energy estimates (Kubečka et al., 2025; Neefjes et al., 2026). While GFN1-xTB can capture correct qualitative trends, obtaining quantitatively accurate binding energies requires higher levels of theory. Due to the computational cost of running MD with these methods, ML approaches must be employed.

3.4 Computational cost

We briefly discuss the training and evaluation of the computational costs for the three models. On an NVIDIA V100-16GB GPU, the AIMNet2, PaiNN, and Δ-PaiNN models, trained on 20,000 H_2SO_4–H_2SO_4 structures, complete approximately 30, 6, and 6 epochs per GPU hour, respectively. For evaluation, we performed umbrella sampling for the H_2SO_4–H_2SO_4 system (at the 4.0 Å window) on an Intel Xeon Gold 6248R CPU. Under these conditions, AIMNet2, PaiNN, and Δ-PaiNN achieved speeds of approximately 120,000, 130,000, and 65,000 steps per CPU hour, respectively. The performance of Δ-PaiNN is inherently limited by the computational cost of the baseline method (here GFN1-xTB), as both the baseline energy and the ML correction must be evaluated at every step. We note that these timings are highly dependent on the specific system architecture and software implementation. Specifically, our group has more extensive experience optimizing PaiNN compared to AIMNet2. Consequently, our PaiNN implementation may be more streamlined. Therefore, these timings should be considered indicative rather than a rigorous benchmark.

3.5 Collision probabilities

Collision rate coefficients are calculated via Eq. (1), where the collision probability $P_\mathrm{c}(v_0, b)$ is obtained from MD collision trajectory simulations over a relevant range of initial relative velocities v_0 and impact parameters b. Figure 5 compares the collision probabilities obtained from GFN1-xTB MD simulations with those from the AIMNet2 and PaiNN models trained on GFN1-xTB data. The AIMNet2 results are in excellent agreement with the GFN1-xTB reference, showing only a slightly lower collision probability in the tail towards higher b. Conversely, the PaiNN heat map clearly highlights the limitations of

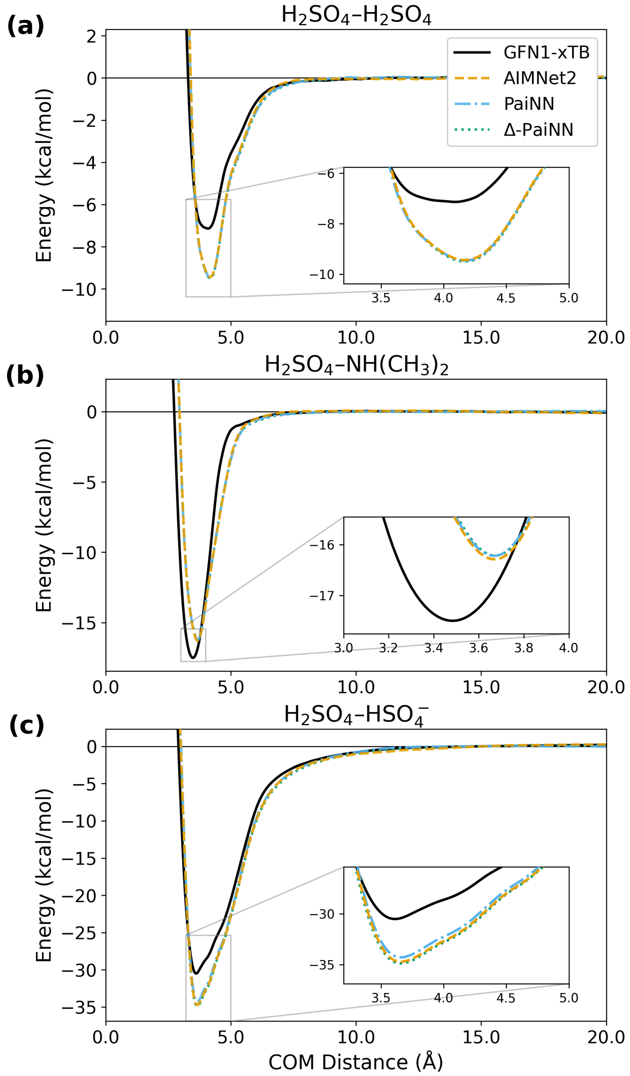

Figure 4. Potentials of mean force (PMF) as a function of center-of-mass (COM) distance for the H_2SO_4–H_2SO_4, H_2SO_4–$NH(CH_3)_2$, and H_2SO_4–HSO_4^- systems. The profiles compare the reference GFN1-xTB level of theory with predictions from AIMNet2, PaiNN, and Δ-PaiNN models trained on ωB97X-3c data.

the local environment approximation for modeling collisions. Above the cutoff plus the sum of molecular radii (∼ 13.89 Å), the message-passing algorithm no longer detects interactions between the collision partners. Consequently, zero collisions are registered past 14 Å. We also note that even below 14 Å, PaiNN appears to underestimate the collision probability compared to the reference. Therefore, for systems with strong long-range interactions, it is necessary to employ a model that accounts for interactions beyond the local atomic environment cutoff, such as AIMNet2.

The H_2SO_4–H_2SO_4 and H_2SO_4–$NH(CH_3)_2$ systems exhibit similar trends, although the discrepancies are less pronounced due to the weaker long-range interactions in these systems. These results are presented in Sec. S5 of the Supporting Information.

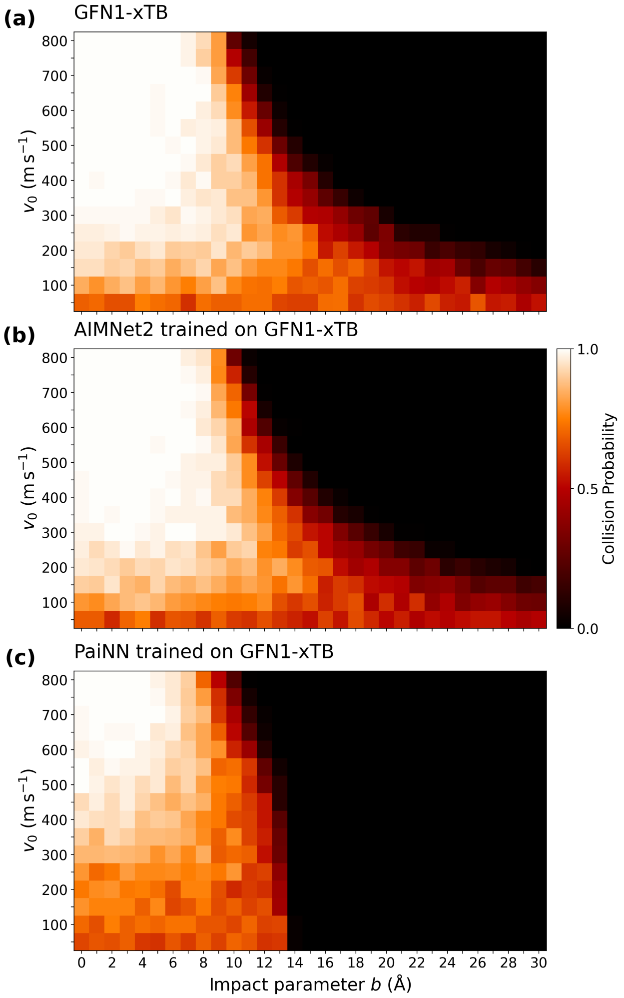

Figure 5. Comparison of collision probabilities derived from molecular dynamics simulations for the H_2SO_4–HSO_4^- system at 300 K. The heat maps show the probability distribution across impact parameter b and initial relative velocity v_0 for the reference GFN1-xTB method (a) and the AIMNet2 (b) and PaiNN (c) models trained on GFN1-xTB data.

355 3.6 Collision rate coefficients

Using the collision probabilities, we calculated the corresponding rate coefficients via Eq. (1). Table 5 presents these values, comparing the performance of the three ML models against the GFN1-xTB reference and the classical OPLS-AA force field.

Table 5. Collision rate coefficients calculated at 300 K using AIMNet2, PaiNN, and Δ-PaiNN compared to GFN1-xTB and OPLS-AA reference values. All values are in 10^{-15} $m^3\,s^{-1}$.

	Reference Methods			**ML Models**		
System	**GFN1-xTB**	**OPLS-AA**	**Training Data**	**AIMNet2**	**PaiNN**	**Δ-PaiNN**
H_2SO_4–H_2SO_4	0.771	0.796	GFN1-xTB	0.744	0.760	0.774
			ωB97X-3c	0.708	0.750	0.761
H_2SO_4–$NH(CH_3)_2$	0.721	0.741	GFN1-xTB	0.797	0.725	–
			ωB97X-3c	0.717	0.724	0.717
H_2SO_4–$HSO_4{}^-$	2.27	2.60	GFN1-xTB	2.24	1.24	–
			ωB97X-3c	3.22	1.26	2.27

For both GFN1-xTB and ωB97X-3c training data, PaiNN provides lower rate coefficients for the charged H_2SO_4–$HSO_4{}^-$ system by nearly 50% relative to the GFN1-xTB reference. As discussed in Sec. 3.5, this large error stems from the model's inability to detect collisions beyond its 10 Å cutoff, effectively neglecting the significant contribution of the long-range tail.

For the neutral systems (H_2SO_4–H_2SO_4 and H_2SO_4–$NH(CH_3)_2$), the ML models trained on GFN1-xTB data exhibit excellent agreement with the reference calculations. All three architectures reproduce the GFN1-xTB reference rate coefficients closely, with the largest deviation observed for AIMNet2 applied to the H_2SO_4–$NH(CH_3)_2$ system (∼10% discrepancy).

When targeting the higher-level ωB97X-3c theory, the predicted collision rate coefficients are generally similar to the GFN1-xTB reference values. The notable exception is the H_2SO_4–$HSO_4{}^-$ system modeled with AIMNet2, where the predicted rate is significantly higher. An analysis of the collision probabilities (Figs. S5c and S3c) shows that this difference results primarily from increased collisions at low initial relative velocities. In the GFN1-xTB reference, collision probabilities in this region are low even at small impact parameters. This behavior has previously been attributed to repulsive electrostatic interactions as the collision partners approach (Halonen et al., 2019). However, AIMNet2 trained on the higher-level ωB97X-3c data does not exhibit this repulsion. This suggests that the repulsive feature observed in GFN1-xTB could be an artifact of the semi-empirical method failing to accurately describe the long-range potential.

The classical OPLS-AA force field also produces collision rates close to the GFN1-xTB reference values. For the charged H_2SO_4–$HSO_4{}^-$ system, the OPLS-AA result lies between the GFN1-xTB reference and the AIMNet2 prediction trained on ωB97X-3c. Given that OPLS-AA relies on fixed partial charges, this agreement suggests that explicit dynamic electron density reorganization is not strictly necessary to model the collision process, provided the underlying electrostatic potential is sufficiently accurate. Simple fixed partial charges appear sufficient to model the approach, a finding consistent with observations by Knattrup et al. (2025).

Despite this accuracy in capturing collisions, classical force fields are insufficient for modeling the full nucleation process, as short-range interactions, particularly proton transfers, require an explicit quantum mechanical treatment. Proton transfers

https://doi.org/10.5194/egusphere-2026-696
Preprint. Discussion started: 16 February 2026

stabilize acid–base clusters (e.g., between sulfuric acid and amines) which play a crucial role in atmospheric particle formation. Classical force fields like OPLS-AA are fundamentally unable to model these proton transfer events. Figure 6 illustrates the geometry of the H_2SO_4–$NH(CH_3)_2$ system immediately before a collision and after cluster formation. A proton initially bound to H_2SO_4 transfers to $NH(CH_3)_2$ during the clustering process, eventually becoming separated by more than 5 Å from its original oxygen atom. Simulating this separation is impossible with classical harmonic bond potentials. In contrast, both the GFN1-xTB method and all three tested ML models successfully capture these dynamic proton transfers.

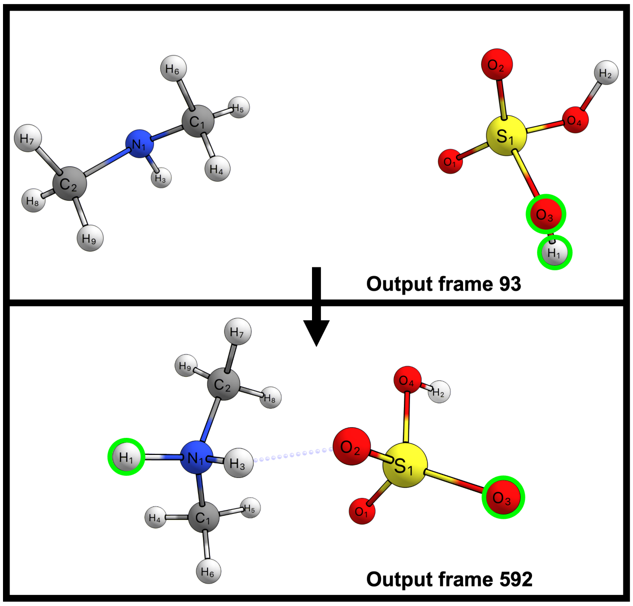

Figure 6. Stick-and-ball representations of two frames from a molecular dynamics collision trajectory simulation of the sulfuric acid–dimethylamine system. Frame 93 shows the system just before collision, with hydrogen H1 bonded to oxygen O3 of sulfuric acid (highlighted with green circles). Output frame 592 shows a post-collision configuration, where hydrogen H1 is now bonded to nitrogen N1 of dimethylamine, while oxygen O3 of sulfuric acid is more than 5 Å away. Atom color code: sulfur (yellow), oxygen (red), nitrogen (blue), carbon (gray), and hydrogen (white).

4 Conclusions

With recent advances in machine learning interatomic potentials, molecular dynamics has rapidly evolved into a field capable of directly modeling cluster formation for large systems over long timescales, all while reproducing the accuracy of high-level quantum chemistry theory. While the low concentrations of cluster-forming vapors in the atmosphere still necessitate approximations (such as artificially increased concentrations), the ability to model the inherently dynamic cluster formation process at this level of theory is a significant breakthrough. However, machine learning models are frequently treated as black boxes. It is crucial that the increased accuracy in modeling complex short-range interactions, such as cluster reconfiguration and proton transfer, does not come at the expense of the long-range interactions that govern the initial collisions.

To address this, we evaluated the AIMNet2 and PaiNN machine learning methodologies, as well as a Δ-PaiNN method (using PaiNN to learn the correlation between GFN1-xTB and high-level quantum theory), for their ability to reproduce collision dynamics for sulfuric acid with either sulfuric acid, dimethylamine, and bisufate. All models achieved low mean absolute errors on test sets and showed excellent agreement with GFN1-xTB reference potentials of mean force. PaiNN, in particular, accurately reproduced both the binding and repulsive regions of the potentials of mean force.

However, when comparing collision rate coefficients, we observe significant differences. For the charged sulfuric acid–bisulfate system, PaiNN predicted collision rate coefficients approximately 50% lower than the reference method. This error arises from the strictly short-ranged nature of the local atomic environment approximation. PaiNN only considers interactions up to a specific cutoff (here, 10 Å). In charged systems, strong long-range interactions can induce collisions from distances far beyond this cutoff. AIMNet2 avoids this issue by supplementing its local short-ranged modeling with explicit long-range Coulombic interactions (via learned partial charges) and dispersion corrections, allowing it to accurately replicate the reference rates.

Our intention is not to highlight a specific failing of PaiNN, but rather to use it as a case study. We demonstrate that low mean absolute errors on a static test set do not automatically guarantee that a model is fit for purpose in dynamic simulations. A model must always be validated against reference data for the specific physical properties of interest. We also note that while we used a generous 10 Å cutoff, many standard implementations use 5 Å. At such small cutoffs, similar discrepancies would likely appear even for neutral systems. That being said, we note that the PaiNN architecture could be adapted to also include explicit long-range interactions, similar to AIMNet2.

In conclusion, we urge researchers to validate trained machine learning models beyond simple scalar metrics like mean absolute errors and root mean square errors, as these condensed measures can mask specific physical limitations. In future work, we will employ AIMNet2 and PaiNN, applying each where it is most appropriate, to directly study the cluster formation of nucleation precursor vapors with accurate descriptions of both short- and long-range interactions.

Code and data availability. The training datasets, trained models, an AIMNet2 model definition and training configuration file, and a molecular dynamics collision trajectory submission scripts are available in the Atmospheric Cluster Database (ACDB) at:
https://github.com/elmjonas/ACDB/tree/master/Articles/

Author contributions. Conceptualization: J.E.; Methodology: I.N., J.K., J.E.; Formal analysis: I.N.; Investigation: I.N., J.K.; Resources: J.E.; Writing - original draft: I.N.; Writing - review & editing: I.N., J.K., J.E.; Visualization: I.N.; Project administration: J.E.; Funding acquisition: J.K., J.E; Supervision: J.E.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Funded by the European Union (ERC, ExploreFNP, project 101040353, and MSCA, HYDRO-CLUSTER, project 101105506). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgements. This project was funded by the European Union (ERC, ExploreFNP, project 101040353, and MSCA, HYDRO-CLUSTER, project 101105506). This work was also supported by the Danish National Research Foundation (DNRF172) through the Center of Excellence for Chemistry of Clouds. The numerical results presented in this work were obtained at the Centre for Scientific Computing, Aarhus https://phys.au.dk/forskning/faciliteter/cscaa/. We thank Haide Wu, Olexandr Isayev, and Roman Zubatiuk for their support regarding AIMNet2.

References

Anstine, D. M., Zubatyuk, R., and Isayev, O.: AIMNet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs, Chem. Sci., 16, 10 228–10 244, https://doi.org/10.1039/D4SC08572H, 2025.

Biewald, L.: Experiment Tracking with Weights and Biases, https://www.wandb.com/, software available from wandb.com, 2020.

Bussi, G., Donadio, D., and Parrinello, M.: Canonical sampling through velocity rescaling, J. Chem. Phys., 126, 014 101, https://doi.org/10.1063/1.2408420, 2007.

Chen, D., Rojas, M., Samset, B. H., Cobb, K., Niang, A. D., Edwards, P., Emori, S., Faria, S., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S., Plattner, G.-K., Edwards, P., Emori, S., Faria, S., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., and Tréguier, A.-M.: Framing, Context, and Methods, in: Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 147–286, Cambridge University Press, https://doi.org/10.1017/9781009157896.003, 2021.

Ehlert, S.: TBlite, https://github.com/tblite/tblite, version 0.2.1, 2022.

Gan, W. Q., FitzGerald, J. M., Carlsten, C., Sadatsafavi, M., and Brauer, M.: Associations of Ambient Air Pollution with Chronic Obstructive Pulmonary Disease Hospitalization and Mortality, Am. J. Respir. Crit. Care Med., 187, 721–727, https://doi.org/10.1164/rccm.201211-2004oc, 2013.

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H.: A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys., 132, 154 104, https://doi.org/10.1063/1.3382344, 2010.

Grimme, S., Ehrlich, S., and Goerigk, L.: Effect of the damping function in dispersion corrected density functional theory, J. Comput. Chem., 32, 1456–1465, https://doi.org/10.1002/jcc.21759, 2011.

Grimme, S., Bannwarth, C., and Shushkov, P.: A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86), J. Chem. Theory Comput., 13, 1989–2009, https://doi.org/10.1021/acs.jctc.7b00118, 2017.

Halonen, R., Zapadinsky, E., Kurtén, T., Vehkamäki, H., and Reischl, B.: Rate enhancement in collisions of sulfuric acid molecules due to long-range intermolecular forces, Atmos. Chem. Phys., 19, 13 355–13 366, https://doi.org/10.5194/acp-19-13355-2019, 2019.

Haywood, J. and Boucher, O.: Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: A review, Rev. Geophys., 38, 513–543, https://doi.org/10.1029/1999RG000078, 2000.

Hjorth Larsen, A., Jørgen Mortensen, J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., Friis, J., Groves, M. N., Hammer, B., Hargus, C., Hermes, E. D., Jennings, P. C., Bjerre Jensen, P., Kermode, J., Kitchin, J. R., Leonhard Kolsbjerg, E., Kubal, J., Kaasbjerg, K., Lysgaard, S., Bergmann Maronsson, J., Maxson, T., Olsen, T., Pastewka, L., Peterson, A., Rostgaard, C., Schiøtz, J., Schütt, O., Strange, M., Thygesen, K. S., Vegge, T., Vilhelmsen, L., Walter, M., Zeng, Z., and Jacobsen, K. W.: The atomic simulation environment—a Python library for working with atoms, J. Phys. Condens. Matter, 29, 273 002, https://doi.org/10.1088/1361-648X/aa680e, 2017.

Jiang, S., Liu, Y.-R., Huang, T., Feng, Y.-J., Wang, C.-Y., Wang, Z.-Q., Ge, B.-J., Liu, Q.-S., Guang, W.-R., and Huang, W.: Towards fully ab initio simulation of atmospheric aerosol nucleation, Nat. Commun., 13, 6067, https://doi.org/10.1038/s41467-022-33783-y, 2022.

Jiang, S., Liu, Y.-R., Wang, C.-Y., and Huang, T.: Benchmarking general neural network potential ANI-2x on aerosol nucleation molecular clusters, Int. J. Quantum Chem., 123, e27 087, https://doi.org/10.1002/qua.27087, 2023.

Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J.: Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc., 118, 11 225–11 236, https://doi.org/10.1021/ja9621760, 1996.

Knattrup, Y., Neefjes, I., Kubečka, J., and Elm, J.: Growth of atmospheric freshly nucleated particles: a semi-empirical molecular dynamics study, Aerosol Research, 3, 237–251, https://doi.org/10.5194/ar-3-237-2025, 2025.

Kubečka, J., Ayoubi, D., Tang, Z., Knattrup, Y., Engsvang, M., Wu, H., and Elm, J.: Accurate modeling of the potential energy surface of atmospheric molecular clusters boosted by neural networks, Environ. Sci.: Adv., 3, 1438–1451, https://doi.org/10.1039/D4VA00255E, 2024.

Kubečka, J., Knattrup, Y., Trolle, G. B., Reischl, B., Lykke-Møller, A. S., Elm, J., and Neefjes, I.: Thermodynamics of molecular binding and clustering in the atmosphere revealed through conventional and ML-enhanced umbrella sampling, ChemRxiv, https://doi.org/10.26434/chemrxiv-2025-b5xxr, 2025.

Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H. E., Nieminen, T., Petäjä, T., Sipilä, M., Schobesberger, S., Rantala, P., Franchin, A., Jokinen, T., Järvinen, E., Äijälä, M., Kangasluoma, J., Hakala, J., Aalto, P. P., Paasonen, P., Mikkilä, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., Ruuskanen, T., Mauldin, R. L., Duplissy, J., Vehkamäki, H., Bäck, J., Kortelainen, A., Riipinen, I., Kurtén, T., Johnston, M. V., Smith, J. N., Ehn, M., Mentel, T. F., Lehtinen, K. E. J., Laaksonen, A., Kerminen, V.-M., and Worsnop, D. R.: Direct Observations of Atmospheric Aerosol Nucleation, Science, 339, 943–946, https://doi.org/10.1126/science.1227385, 2013.

Liu, Y.-R. and Jiang, Y.: Predicting Composition Evolution for a Sulfuric Acid-Dimethylamine System from Monomer to Nanoparticle Using Machine Learning, J. Phys. Chem. A, 129, 222–231, https://doi.org/10.1021/acs.jpca.4c06062, 2025.

McGrath, M. J., Olenius, T., Ortega, I. K., Loukonen, V., Paasonen, P., Kurtén, T., Kulmala, M., and Vehkamäki, H.: Atmospheric Cluster Dynamics Code: a flexible method for solution of the birth-death equations, Atmos. Chem. Phys., 12, 2345–2355, https://doi.org/10.5194/acp-12-2345-2012, 2012.

Müller, M., Hansen, A., and Grimme, S.: ωB97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-ζ basis set, J. Chem. Phys., 158, 014 103, https://doi.org/10.1063/5.0133026, 2023.

Neefjes, I., Halonen, R., Vehkamäki, H., and Reischl, B.: Modeling approaches for atmospheric ion–dipole collisions: all-atom trajectory simulations and central field methods, Atmos. Chem. Phys., 22, 11 155–11 172, https://doi.org/10.5194/acp-22-11155-2022, 2022.

Neefjes, I., Reischl, B., and Yang, H.: Comparison of collision rate coefficient model predictions for different interaction strengths and temperatures, J. Aerosol Sci., 189, 106 638, https://doi.org/10.1016/j.jaerosci.2025.106638, 2025.

Neefjes, I., Knattrup, Y., Wu, H., Trolle, G. B., Elm, J., and Kubečka, J.: Thermodynamic benchmarking of hydrated atmospheric clusters in early particle formation, Aerosol Research, 4, 1–22, https://doi.org/10.5194/ar-4-1-2026, 2026.

Neese, F.: The ORCA program system, WIRES Comput. Molec. Sci., 2, 73–78, https://doi.org/10.1002/wcms.81, 2012.

Plimpton, S.: Fast Parallel Algorithms for Short-Range Molecular Dynamics, J. Comput. Phys., 117, 1–19, https://doi.org/10.1006/jcph.1995.1039, 1995.

Schütt, K. T., Unke, O. T., and Gastegger, M.: Equivariant message passing for the prediction of tensorial properties and molecular spectra, arXiv, abs/2102.03150, https://doi.org/10.48550/arXiv.2102.03150, 2021.

Schütt, K. T., Hessmann, S. S. P., Gebauer, N. W. A., Lederer, J., and Gastegger, M.: SchNetPack 2.0: A neural network toolbox for atomistic machine learning, https://github.com/atomistic-machine-learning/schnetpack, version 2.1.1, 2024.

Sipilä, M., Berndt, T., Petäjä, T., Brus, D., Vanhanen, J., Stratmann, F., Patokoski, J., III, R. L. M., Hyvärinen, A.-P., Lihavainen, H., and Kulmala, M.: The Role of Sulfuric Acid in Atmospheric Nucleation, Science, 327, 1243–1246, https://doi.org/10.1126/science.1180315, 2010.

Stroet, M. and Deplazes, E.: Umbrella Integration: Initial Version, https://doi.org/10.5281/zenodo.164996, 2016.

Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu, D. S., Brown, W. M., Crozier, P. S., in ’t Veld, P. J., Kohlmeyer, A., Moore, S. G., Nguyen, T. D., Shan, R., Stevens, M. J., Tranchida, J., Trott, C., and Plimpton, S. J.: LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Comp. Phys. Comm., 271, 108 171, https://doi.org/10.1016/j.cpc.2021.108171, 2022.

Tikkanen, V., Yang, H., Vehkamäki, H., and Reischl, B.: Gas-phase collision rate enhancement factors for acid–base clusters up to 2 nm in diameter from atomistic simulation and the interacting hard-sphere model, Atmos. Chem. Phys., 25, 17 237–17 251, https://doi.org/10.5194/acp-25-17237-2025, 2025.

Torrie, G. and Valleau, J.: Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, J. Comput. Phys., 23, 187–199, https://doi.org/10.1016/0021-9991(77)90121-8, 1977.

Tröstl, J., Chuang, W. K., Gordon, H., Heinritzi, M., Yan, C., Molteni, U., Ahlm, L., Frege, C., Bianchi, F., Wagner, R., and et al.: The Role of Low-Volatility Organic Compounds in Initial Particle Growth in the Atmosphere, Nature, 533, 527–531, https://doi.org/10.1038/nature18271, 2016.

Yang, H., Neefjes, I., Tikkanen, V., Kubečka, J., Kurtén, T., Vehkamäki, H., and Reischl, B.: Collision-sticking rates of acid–base clusters in the gas phase determined from atomistic simulation and a novel analytical interacting hard-sphere model, Atmos. Chem. Phys., 23, 5993–6009, https://doi.org/10.5194/acp-23-5993-2023, 2023.

Zhang, R., Khalizov, A., Wang, L., Hu, M., and Xu, W.: Nucleation and Growth of Nanoparticles in the Atmosphere, Chem. Rev., 112, 957–2011, https://doi.org/10.1021/cr2001756, 2012.