NeuralNEB—neural networks can find reaction paths fast

IOP Publishing Mach. Learn.: Sci. Technol. 3 (2022) 045022 https://doi.org/10.1088/2632-2153/aca23e

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
29 June 2022

REVISED
25 October 2022

ACCEPTED FOR PUBLICATION
11 November 2022

PUBLISHED
13 December 2022

PAPER

NeuralNEB—neural networks can find reaction paths fast

Mathias Schreiner*, Arghya Bhowmik, Tejs Vegge, Peter Bjørn Jørgensen and Ole Winther

Technical University of Denmark (DTU), 2800 Lyngby, Denmark
* Author to whom any correspondence should be addressed.

E-mail: matschreiner@gmail.com

Keywords: reaction barriers, computational chemistry, transition state, reaction kinetics, density function theory (DFT), nudged elastic band (NEB), neural networks (NN)

Abstract

Quantum mechanical methods like density functional theory (DFT) are used with great success alongside efficient search algorithms for studying kinetics of reactive systems. However, DFT is prohibitively expensive for large scale exploration. Machine learning (ML) models have turned out to be excellent emulators of small molecule DFT calculations and could possibly replace DFT in such tasks. For kinetics, success relies primarily on the models' capability to accurately predict the potential energy surface around transition-states and minimal energy paths. Previously this has not been possible due to scarcity of relevant data in the literature. In this paper we train equivariant graph neural network-based models on data from 10 000 elementary reactions from the recently published Transition1x dataset. We apply the models as potentials for the nudged elastic band algorithm and achieve a mean average error of 0.23 eV and root mean squared error of 0.52 eV on barrier energies on unseen reactions. We compare the results against equivalent models trained on QM9x and ANI1x. We also compare with and outperform Density Functional based Tight Binding on both accuracy and required computational resources. The implication is that ML models are now at a level where they can be applied to studying chemical reaction kinetics given a sufficient amount of data relevant to this task.

1. Introduction

Machine learning (ML) models and especially graph neural networks (GNNs) [1, 2] have turned out to be potent emulators of density functional theory (DFT) potentials for small molecules [3–7], thanks to their remarkable ability to find complex relations in high dimensional data. They have a complexity-scaling orders of magnitudes lower than classic quantum mechanics (QM) methods, but have in recent years achieved comparable accuracy [8–12]. The capability of these models is manifested by their success in tasks beyond simple prediction of molecular features such as structural optimization or studying finite-temperature dynamical properties through molecular dynamics [13, 14]. Despite their achievements, there has only been limited success in applying ML-models as potentials for transition search algorithms. The earliest work studied simple diatomic molecule dissociation and achieved acceptable accuracy with tens of thousands of data points [15]. Other works have had success by limiting their scope to studying single or few reactions but sacrificing the generality of the approach [16–18]. Attempts to study reactive systems with Gaussian processs (GPs) [19] have been successful too, but the GP is trained on the particular atomic system, sacrificing speed for generality by requiring expensive DFT calculations at inference time. Transition-states are notoriously hard to find as there is no well-defined gradient on the potential energy surface (PES) to guide traditional optimization algorithms towards them. A wealth of algorithms have been proposed to solve this problem—one is the nudged elastic band (NEB) [20] algorithm, which works by interpolating an initial path between reactant and product and iteratively updating it to minimize energy by using information about the PES. It shares a common bottleneck with other transition search algorithms—the necessity to repeatedly evaluate energy and atomic forces of molecular configurations, which is extremely costly, especially if *ab-initio* or electron DFT calculations are used [21].

Recent advances in ML have not alleviated the bottleneck as even modern neural network (NN) architectures have not proved proficient potential approximators for this type of application. The fault lies primarily with available data in the literature rather than the models' expressiveness [22]. Most quantum mechanical datasets are focused on molecular configurations in or near equilibrium [23–26]. Without configurations on and around reaction pathways in the training data, ML models cannot learn the interatomic interactions that occur during chemical reactions and cannot reliably be applied for transition-state search.

We compare ML models against Density Functional based Tight Binding (DFTB), [27] a fast approximation to DFT that is often used for fast screening of large quantities of configurations with an acceptable trade-off between accuracy and speed, and our models outperform DFTB with a factor three in accuracy and a factor two in CPU time.

In this work, we bridge generalization, speed, and accuracy for transition-state search by applying polarizable atom interaction neural network (PaiNN) models as surrogate potentials for DFT. We build on and showcase the utility of our previous paper [28] where we released Transition1x, a dataset constituted by DFT calculations for 10 million molecular configurations, all sampled around reaction pathways from 10 000 elementary, organic reactions. It is clear from the results of this paper, that for precise modeling of transition-state regions, and, consequently, transition states and barrier energies, hitherto popular benchmark datasets have had insufficient relevant data. On the other hand, training ML potentials on the Transition1x dataset allows for accurate modeling of PESs in transition-state regions, underlining that relevant and available data in the literature is as important as the efficiency of available models.

Reliable and fast analysis of reaction kinetics through ML will bring the whole field of computational chemistry a considerable step closer to the ultimate goal, a virtual laboratory, hyper-accelerating the discovery of reaction mechanisms for synthesizing drugs and materials.

2. Methods

2.1. Nudged elastic band

NEB [20] is a method for finding minimal energy path (MEP) and transition-state given product and reactant of a chemical reaction. It does so by iteratively nudging an interpolated path between the reaction endpoints in the direction of the force perpendicular to the path. Once the perpendicular force converges to zero, NEB reports the maximal-energy configuration along the path as the transition-state. The path is represented by an array of molecular configurations called images, and there is no guarantee that, at convergence, the maximal energy image corresponds to the maximal energy along the path. The maximum might lie between two images. Climbing image nudged elastic band (CINEB) [29] addresses this problem by letting the transition-state candidate (the maximal energy image) further maximize its energy by following the gradient on the PES parallel to the current path between iterations. If the current path has not converged properly, the climbing image can pull the predicted MEP off the true MEP and therefore, the path is first relaxed with regular NEB before turning on CINEB. The MEP is considered converged once the maximal perpendicular force on the path is below a threshold of 0.05 eV $Å^{-1}$. The spring constant between images on the path is set to 0.1 eV $Å^{-2}$, and ten images are used to represent the path.

2.2. Initial path generation

The endpoints of the reaction have to be minimized in their respective minima before running NEB—otherwise the energetic difference between reactant and transition-state cannot be evaluated properly. A configuration is considered relaxed if the norm of the forces acting on it is below 0.01 eV $Å^{-1}$. Once the endpoints have been minimized, the initial guess for the MEP is found by running NEB with the Image Dependent Pair Potential (IDPP) [30] on a linearly interpolated path between reactant and product. IDPP is an inexpensive potential specifically designed to generate physically realistic MEP guesses for NEB at an extremely low computational cost.

2.3. Optimizers

Reactants and products are relaxed using the Broyden–Fletcher–Goldfarb–Shannon (BFGS) [31] optimizer with $\alpha = 70$ and a maximal step size of 0.03 Å in configurational space. The MEP is found with an optimizer [32] designed to reduce the computational cost of transition-state search algorithms by applying an adaptive time step selection algorithm with $\alpha = 0.01$ and rtol $= 0.1$, and a preconditioning scheme to the PES given an estimate of its curvature.

3. Data

We train all models on ANI1x [24], QM9x [33], Transition1x [28]. All datasets are calculated with the 6-31G(d) [34] basis set and ωB97x [35] functional which has an accuracy comparable to the gold standard but expensive high-level CCSD(T) [36, 37] calculations. Given the compatibility of the datasets, it is possible to train on either dataset alone or combinations of them to leverage all of their strengths.

3.1. ANI1x

ANI1x [38] aims to provide varied data of off-equilibrium molecular configurations by perturbing equilibrium configurations with pseudo molecular dynamics. The data is collected through an active learning technique called Query by Committee; an automated data diversification process that trains an ensemble (committee) of models on a dataset and accepts or rejects new proposed data based on the disagreement of models in the committee. The assumption is that if the committee disagrees the data is sufficiently different from what has already been learned, and the proposed data should be included in the analysis. The procedure for proposing data and evaluating it with the committee is cheap compared to the calculation of data using DFT. The dataset is consecutively expanded by alternating between training committees and adding new data points based on the committee uncertainty. In total, ANI1x contains force and energy calculations for approximately 5 million configurations.

3.2. Transition1x

We have recently published Transition1x [28], a dataset providing a collection of molecular configurations on and along reaction paths for approximately 10 000 reactions. The reactions consist of up to seven heavy atoms, including C, N, and O. Transition events are rare, and it is not possible to collect sufficient data in relevant regions by simple molecular dynamics if the intention is to train NNs models to understand chemical reactions. Transition1x addresses this problem by sampling molecular configurations around reaction pathways proposed by NEB, using DFT as potential. The procedure resulted in approximately 10 million DFT calculations that were collected and saved during the process and constitute the dataset. Transition1x is available through the repository https://gitlab.com/matschreiner/Transition1x which includes data loaders and scripts for downloading the dataset and generating ASE-database files.

3.3. QM9 and QM9x

QM9 [33] is a dataset of 135 k small organic molecules with various chemical properties that has served as the benchmark for many existing ML methods for quantum chemistry. All molecules in QM9 are in equilibrium. We have recalculated QM9 with the 6-31G(d) basis set and ωB97x functional to make it compatible with Transition1x and ANI1x, and we refer to the recalculated dataset as QM9x. Molecular configurations recalculated in the new potential are not necessarily in equilibrium as the potential shifts when changing functional and basis sets. QM9x is available through the repository https://gitlab.com/matschreiner/QM9x which includes data loaders and scripts for downloading the dataset and generating ASE-database files.

3.4. Models and Training

Message Passing Neural Networks [12] are a class of GNNs [1, 2] that build their internal graph representation by running a series of message passing steps. A single message passing step consists of two distinct operations: (a) *Message Dispatching*, each node computes a message given its state (and possibly information about the edge connecting to—and the state of the receiving node) and sends it to its neighbors. (b) *State Update*, incoming messages are collected with an aggregation function, and are used to simultaneously update the internal representation of all nodes. After the message-passing phase, a readout function extracts the inner representation of the nodes and computes a final feature vector of the graph for downstream tasks. In the case of molecules, interesting properties are energy and forces where conservative force fields can be computed via the back-propagation algorithm as the negative gradient of the energy w.r.t. coordinates of the atoms.

The PaiNN model [39] was used for all experiments—it is a GNN architecture that implements rotationally equivariant representations for prediction of tensorial properties of graph structures. We refer to the literature for further details [39]. A cut-off radius of 5 Å was used to generate the initial molecular graph. All models have three message passing steps and 256 units in each hidden layer, and are trained using the ADAM [40] optimizer with learning rate 10^{-3} on training examples from QM9x, ANI1x, and Transition1x. A batchsize of 75 was used for all datasets and a maximum of 10^6 training steps was allowed—however, models training on ANI1x and Transition1x reached maximal scores on validation data after around 6×10^5

steps. In order to understand to which extent a PaiNN-model trained on Transition1x can generalize to reactions with unseen atomic compositions, building on an assessment of the substructures or elemental features, Transition1x was stratified by chemical formula such that each formula can only be found in one split. The Transition1x was split in 10, and 10 models were trained such that each split could be set aside once as testing data for the NeuralNEB algorithm and once as validation data for early stopping. ANI1x was stratified by chemical formula such that test, validation and training sets consist of chemical formulas unique to that set. QM9x was split randomly. In the case of QM9x and ANI1x, 80% of the data was used for training, 10% for testing, and 10% was used for validation and early stopping. In QM9x all configurations are unique as they are in distinct equilibria and can therefore be split randomly. No attention was paid to the molecular scaffold. For ANI1x, it is necessary to split on chemical formula to ensure that configurations across splits are significantly different. Each chemical formula contains similar configurations, since data is generated by randomly perturbing identical initial configurations.

4. Results

Table 1 shows the overall findings of the paper. Each row displays the performance of a surrogate potential, where datasets in the leftmost column indicate PaiNN models trained on the given dataset. The barrier error is the difference in barrier heights found when applying DFT as potential for NEB versus when applying the surrogate potential.

As different initializations of parameters in equivalent architectures result in variations in the trained models capabilities, five models were trained on each of QM9x and ANI1x and ten models were trained on the Transition1x dataset. QM9x and ANI1x models were used as potentials for all reactions in the Transition1x dataset, and models trained on Transition1x were used as potentials only for those reactions with atomic compositions from the test split. The best models are trained on Transition1x, with the lowest mean average error (MAE) and root mean squared error (RMSE) and the highest convergence ratio. The QM9x models have only seen data very close to equilibrium and have not learned the structure of the PES between equilibria which makes it unable to converge in most cases. In general DFT performs the best in terms of convergence rate and average iterations run, but it comes at a steep price, running almost a factor 1500 times slower than the ML potentials. DFTB is the go-to fast potential, but the models trained on Transition1x are twice as fast and three times as accurate. Figure 1, on the frontpage, displays MEPs calculated with NEB using DFT and PaiNN trained on Transition1x side by side. Each MEP is projected onto a plane in configurational space spanned by the reaction's transition-state, product, and reactant. The x and y axes are basis vectors describing the plane in units of Å, and the z-axis and color-coding show the atomization energy of configurations in the plane in eV. Not only does PaiNN trained on the Transition1x accurately calculate the barrier energy for the reaction, but it also correctly identifies the plane spanned by the configurations defining the reaction, and calculates an almost identical PES in the vicinity of the MEP. Each MEP is projected from a high dimensional space onto the plane, and therefore, only the atomization energy of equilibria and transition-states are shown correctly in the plot. At these points, the MEP intersects with the plane. The intermediate points have energies slightly shifted up the sides of the energy valley. The MEP does not necessarily lie in the plane, and since the MEP represents the energy valley, projecting it onto the plane, will increase the energy. The $\times$ symbols on the surfaces are projections of images predicted by NEB and the dashed lines connecting them are cubic spline interpolations. The importance of accurate predictions in the vicinity of the MEP is clear, as these calculations will guide the search for the transition-state. The Transition1x model predicts smooth and well-behaved PESs resembling DFT.

Figures 2 and 3 tell similar stories. Figure 2 is a histogram of barrier errors where the error is the difference between activation energy found using the surrogate potential and DFT. The Transition1x model is precise and accurate, with a sharp peak around zero, whereas DFTB and ANI1x have wider spreads with means below and above zero, respectively. The QM9x model is plotted on the histogram, but due to high errors and low convergence, only a few calculated barriers fall within an error of ± 2 eV, as shown in the figure. See appendix for an equivalent figure without truncated x-axis.

Figure 3 compares activation energies found with DFT on the x-axis with those found using various surrogate potentials on the y-axis. Each marker represents a single reaction. Predictions from the model trained on Transition1x follow the $x = y$ line with a MAE of only 0.23 eV. The QM9x model does not have a proper representation of the transition-state regions as it has not seen that type of data during training. Often, the QM9x model does not recognize nearby initial equilibria as minima on the PES, and even before optimizing the MEP, the reaction endpoints have dropped further on the PES to qualitatively different endpoints which results in the model calculating the MEP for a completely different reaction. The algorithm is not set up to detect this, and as long as the reaction converges, it is included in the analysis. Even when the QM9x model relaxes the endpoints of the reaction correctly, it either finds low energy shortcuts in the faulty

Table 1. Performance of various potentials used for Nudged Elastic Band (NEB) when compared to Density Functional Theory (DFT). ANI1x, Transition1x and QM9x indicate PaiNN models trained on the respective dataset. The Barrier column displays the Mean Average Error (MAE) and Root Mean Squared Error (RMSE) of barrier predictions, where the individual error is the difference between the barrier as predicted when using DFT as potential vs. using the surrogate potential. The convergence rate is the percentage of reactions that converged. Average CPU time is CPU time spent per reaction. Average iterations is the average number of Minimal Energy Path (MEP) updates before convergence. We have marked the lowest errors, highest convergence rate, and fastest computation in boldface.

	Barrier (eV)		NEB Convergence		
	MAE	RMSE	Rate	Avg. CPU Time	Avg. Iterations
ANI1x	0.51	1.67	69.3%	37 s	149
T1x	**0.23**	**0.52**	80.3%	33 s	135
QM9x	3.40	3.59	35.0%	**28 s**	111
DFTB	0.70	0.85	65.7%	82 s	114
DFT	—	—	**84.1%**	12 h 14 m 43 s	**100.74**

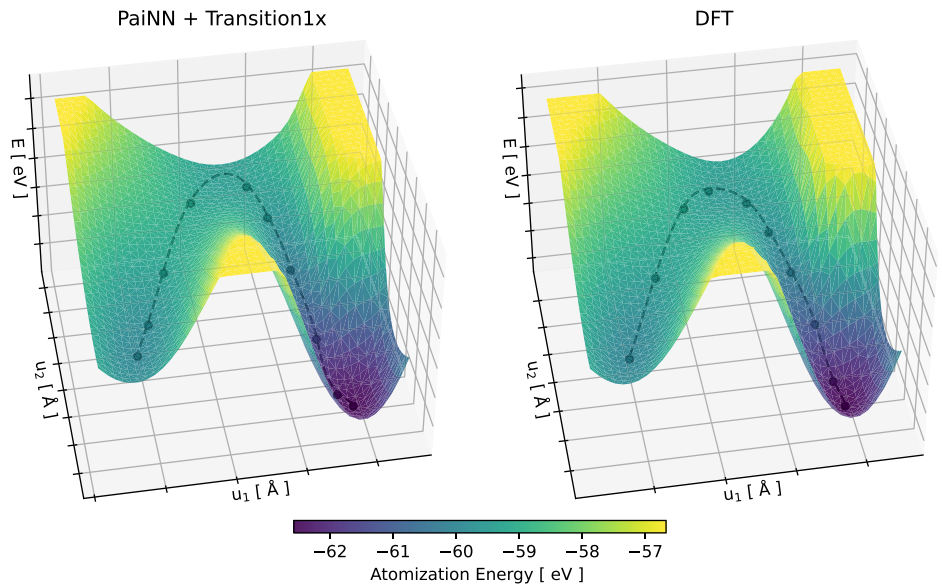

Figure 1. Minimal Energy Paths (MEPs) found with Nudged Elastic Band (NEB) applying the Graph Neural Network (GNN) architecture Polarizable Atom interaction Neural Network (PaiNN) trained on the Transition1x dataset and Density Functional Theory (DFT) as potentials. The MEPs are projected onto planes in structural space, intersecting product, reactant and transition-state of the converged MEPs. The PES has been calculated on the planes in the vicinity of the MEPs with the respective potential and is shown on the *z*-axis. The *x* and *y*-axes are basis vectors describing the plane. The reaction involves a H-transfer coupled with a C–C bond formation on C6H8. The reaction can be seen as a GIF by following this link.

potential or does not converge, and as a result the converged reactions are often only the energy difference between reactant and product. The QM9x dataset was not designed with any type of molecular dynamics or reaction kinetics in mind, and comparing it to ANI1x and Transition1x for reaction path search is perhaps inappropriate. However, given the ubiquity of QM9 in the literature, it is an important point to convey, that new datasets are required for solving higher order problems in computational chemistry. The Transition1x and ANI1x models drop in performance above 5 eV. Data becomes scarcer at higher energies and consequently, models are less accurate in high energy regions. DFTB and the ANI1x models have systematic errors in their predictions. The ANI1x models are biased towards high energies in the transition regions as they have not seen the low energy valleys connecting equilibria. The DFTB potential systematically predicts energies too low. In table 2 the systematic errors are corrected based on the training data. This leads to a lower test error for the ANI1x and DFTB, but equal test error for Transition1x underlining that Transition1x models are already very accurate.

5. Discussion

To train models that can properly step in as surrogate potentials for DFT when running NEB, it is necessary to have datasets with appropriate data in and around transition-state regions. Finding reaction barriers with ML models and NEB is a non-trivial test. ML models, and especially NNs, are known to perform poorly for out of distribution tasks [41, 42]. Table A1 illustrates this clearly with results for training and testing ML models on various datasets.

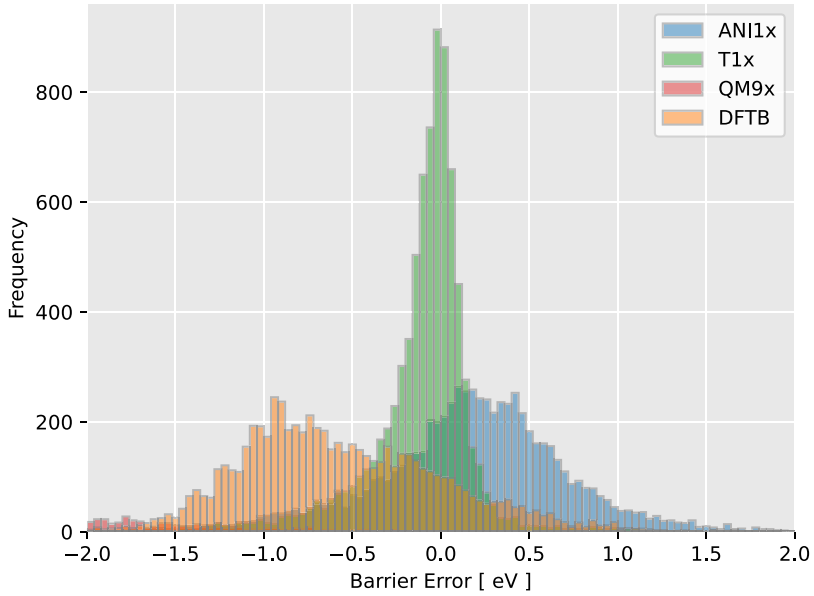

Figure 2. Histogram of barrier errors. The x-axis shows errors between reaction barriers calculated using Density Functional Theory (DFT) and surrogate potentials for Nudged Elastic Band (NEB). The x-axis has been truncated at $+/-$ 2 eV error (see appendix for the full plot). The y-axis shows the frequency of each bin. Green, red and blue display results from PaiNN models trained on Transition1x, QM9x and ANI1x, respectively. Yellow displays results from Density Functional based Tight Binding (DFTB). The QM9x model has such a low convergence frequency, and general barrier error, that the model does not show in the plot.

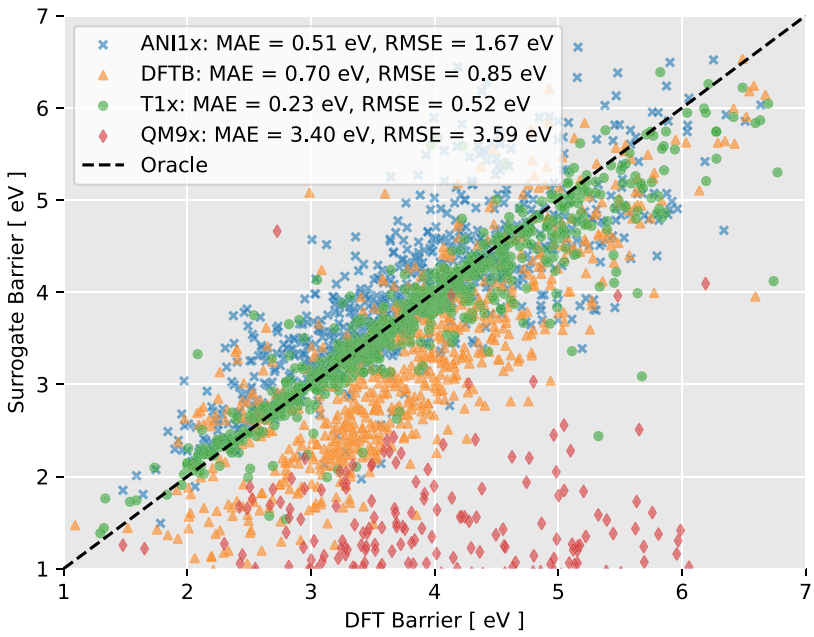

Figure 3. Comparison of reaction barriers found with Nudged Elastic Band (NEB) using Density Functional Theory (DFT) as potential on the x-axis vs. various surrogate potentials on the y-axis. Green, red and blue markers are PaiNN models trained on The Transition1x, QM9x, and ANI1x datasets respectively. Yellow is Density Functional based Tight Binding (DFTB). Points on the dashed line have been calculated perfectly. The figure displays a subsample of 500 reactions—see appendix for the full scatter plot.

Finding reaction barriers with NEB is a much more demanding test of the models' capabilities. When running NEB, the PES is swept by the path connecting endpoints, and data encountered in the process can diverge wildly from any data seen during testing and training. The model can get caught in even a small region of high error, or it can be thrown off the correct MEP and be unable to converge altogether, so the model must be accurate across the entire PES.

The reaction paths are represented by ten images in all reactions. A core strength of NNs is their ability to utilize GPUs to evaluate multiple data points at once, and in principle, NEB can be run with hundreds of

Table 2. Mean Average Error (MAE) and Root Mean Squared Error (RMSE) of barrier errors found by PaiNN trained on Transition1x and ANI1x and DFTB, after correcting for systematic error.

	Corrected Barrier (eV)		
	MAE	RMSE	Systematic Error (eV)
ANI1x	0.48	1.66	0.23
T1x	**0.23**	**0.51**	−0.10
QM9x	0.89	1.14	−3.40
DFTB	0.48	0.62	−0.58

images instead of tens at little to no additional cost when using NNs as potentials. We ran experiments with high density paths with the rest of the setup fixed but saw no improvement in neither accuracy nor convergence speed. The preconditioning scheme of the NEB optimizer relies on a sparsely populated path. But this approach could possibly produce robust results by applying other optimizers.

A clear application of this work is as a screening procedure for complex reaction networks. Cheap methods, such as permuting bond order matrices, can be used to automatically generate nodes for entire reaction networks. The individual reactions can be screened fast using the method before recalculating entire reaction networks with expensive methods. Usually this is done with DFTB [27] but running NEB with NNs is faster and more accurate.

6. Conclusion

We have trained GNN potentials on various datasets and used them as surrogate potentials for DFT when running NEB for transition-state search. A MAE of 0.23 eV and RMSE of 0.52 eV is achieved with the best model, compared against running the same set up with DFT. The models converge 80.3% of the time on unseen reactions. We show that expressive models alone are not sufficient for solving complex tasks in quantum chemistry moving forward, but just as much care has to be put into designing and generating datasets. We tested three different datasets: ANI1x, QM9x and Transition1x and only models trained on the latter could reliably solve the transition search task.

Our results show that the future development of the field of ML for quantum chemistry stands on two legs—the completeness of the available data, and the expressiveness of the available models. Transition1x deals with only four types of atoms. To apply the results of this paper to general chemistry, larger datasets with more atom types should be produced. Our results indicate that the ML approach scales: With the right amount of the right data, accuracies at a sufficient level can be achieved.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.6084/m9.figshare.19614657.v4.

Acknowledgments

The authors acknowledge support from the Novo Nordisk Foundation (SURE, NNF19OC0057822) and theEuropean Union's Horizon 2020 research and innovation program under Grant Agreement Nos. 957189 (BIG-MAP) and 957213 (BATTERY2030PLUS).

Ole Winther also receives support from Novo Nordisk Foundation through the Center for Basic Machine Learning Research in Life Science (NNF20OC0062606) and the Pioneer Centre for AI, DNRF Grant Number P1.

Code Availability

Code for training PaiNN models and running NEB is available through the repository https://gitlab.com/matschreiner/neuralneb.

Appendix A

Table A1 displays results of the models when training and testing on various datasets. In all test set-ups the models that perform best, are models that have been trained on training data from the corresponding dataset.

Table A1. Test results of PaiNN models trained on ANI1x, QM9x, Transition1x. We report RMSE and MAE on energy and forces. Force error is the Euclidian distance between the predicted and true force vector.

		Energy (eV)		Forces (eV Å^{-1})	
Trained on	Tested on	MAE	RMSE	MAE	RMSE
ANI1x		**0.02(0)**	**0.04(1)**	**0.04(0)**	**0.06(0)**
Transition1x	ANI1x	0.22(1)	0.35(2)	0.18(0)	0.42(3)
QM9x		2.32(1)	3.03(2)	1.28(1)	2.0(0)
ANI1x		0.28(2)	0.61(7)	0.16(6)	0.6(1)
Transition1x	Transition1x	**0.10(0)**	**0.15(1)**	**0.05(1)**	**0.12(0)**
QMx		1.42(1)	2.61(2)	0.23(2)	0.48(5)
ANI1x		0.12(0)	0.13(0)	0.03(0)	0.06(0)
Transition1x	QM9x	0.07(1)	0.12(0)	0.05(0)	0.08(0)
QM9x		**0.02(1)**	**0.04(2)**	**0.01(0)**	**0.01(0)**

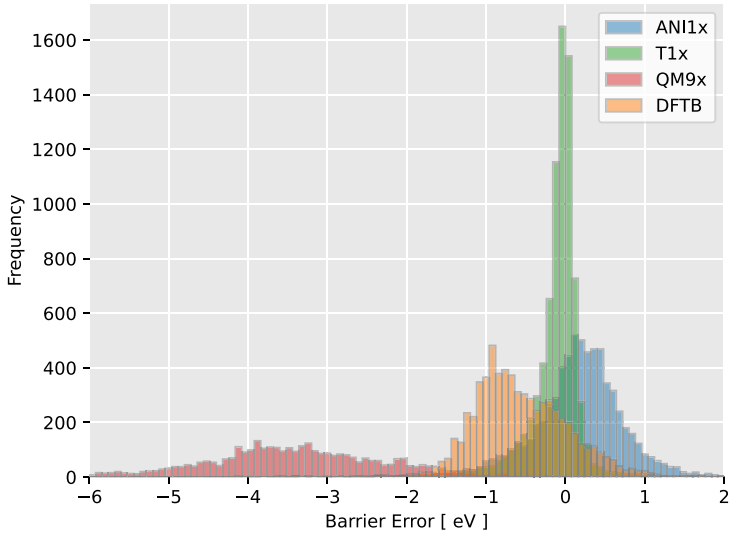

Figure B1. Histogram of barrier errors. The figure is equivalent to figure 2, but without a truncated x-axis. The x-axis shows errors between reaction barriers calculated using Density Functional Theory (DFT) and surrogate potentials for Nudged Elastic Band (NEB). The y-axis shows the frequency of each bin. Green, red and blue display results from PaiNN models trained on Transition1x, QM9x and ANI1x, respectively. Yellow displays results from Density Functional based Tight Binding (DFTB).

Appendix B. Additional figures

Figure B1 is an unbounded version of figure 2. Figure B2 is a scatter plot equivalent to figure 3, but without subsampling reactions. Figures B3–B6 are plots of MEPs and PESs comparing PaiNN trained on Transition1x with DFT for various reactions.

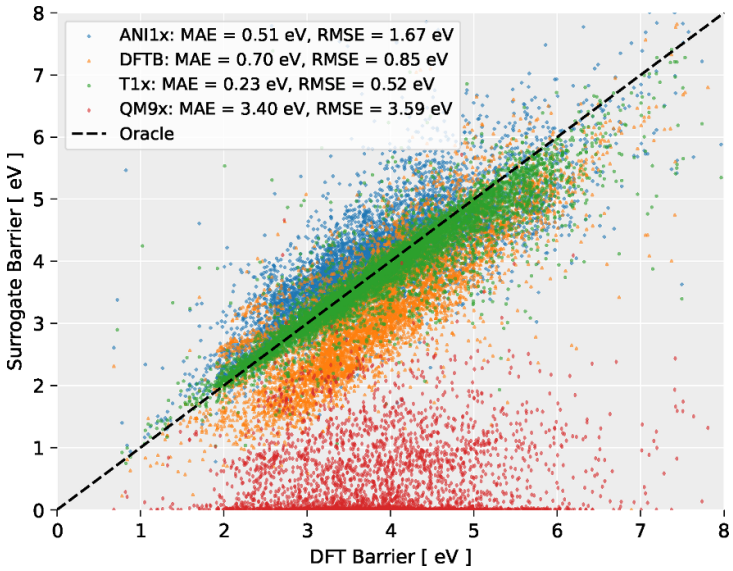

Figure B2. Comparison of reaction barriers found with Nudged Elastic Band (NEB) using Density Functional Theory (DFT) as potential on the *x*-axis vs. various surrogate potentials on the *y*-axis. Green, red and blue markers are PaiNN models trained on The Transition1x, QM9x, and ANI1x datasets respectively. Yellow is Density Functional based Tight Binding (DFTB). Points on the dashed line have been calculated perfectly.

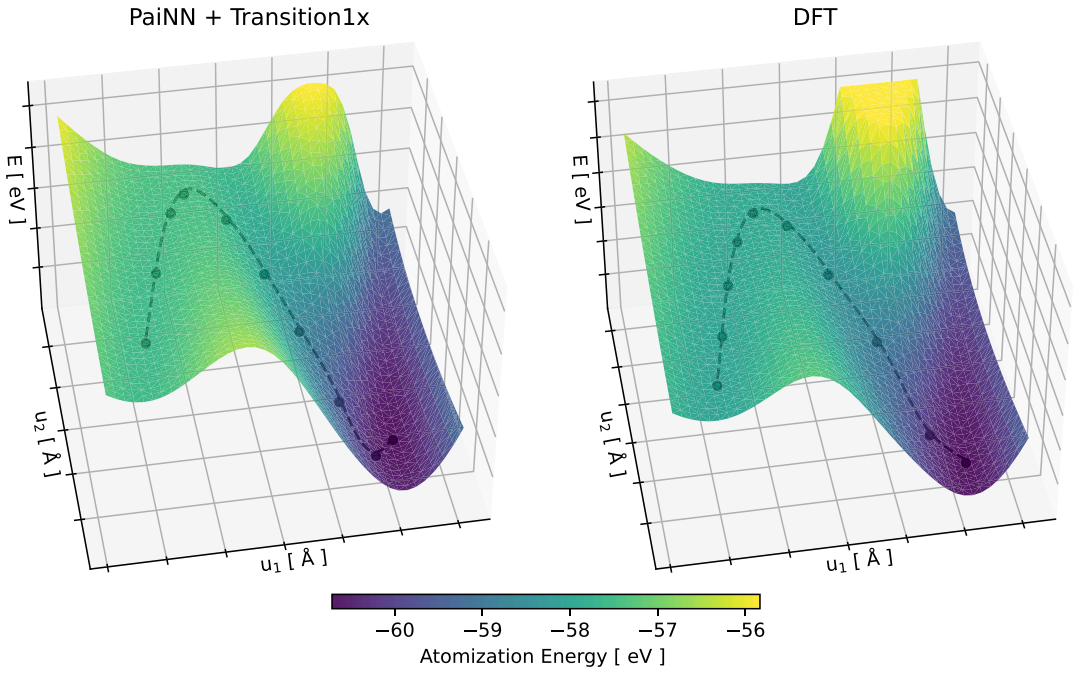

Figure B3. Reaction involving C5OH8. The reaction can be seen as a GIF by following this link.

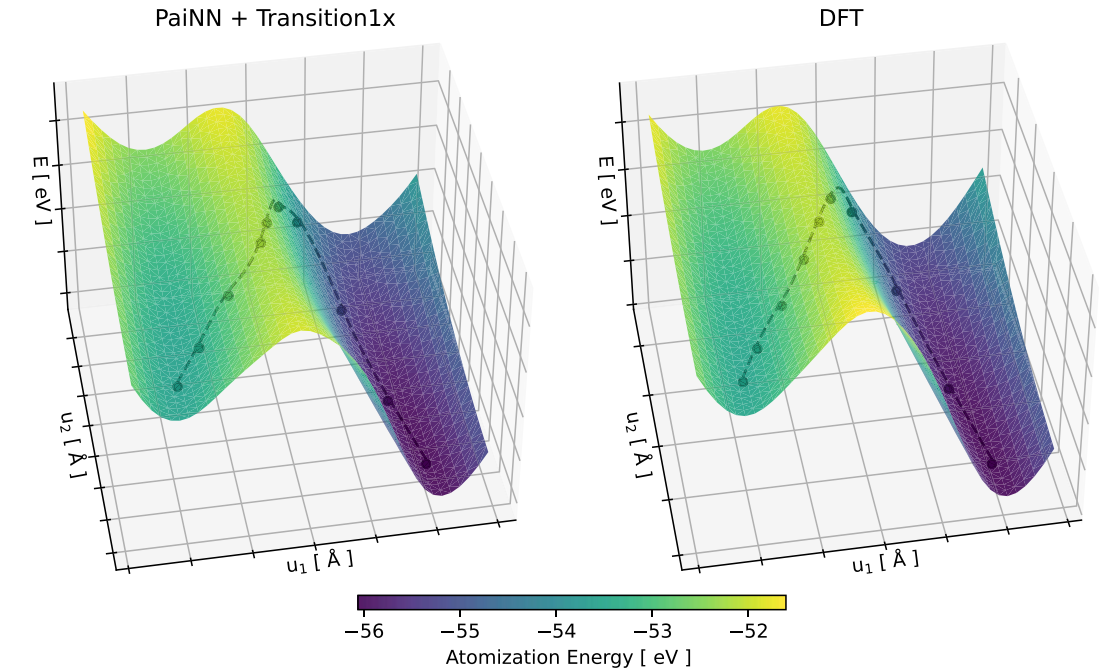

Figure B4. Reaction involving C3NCOH7. The reaction can be seen as a GIF by following this link.

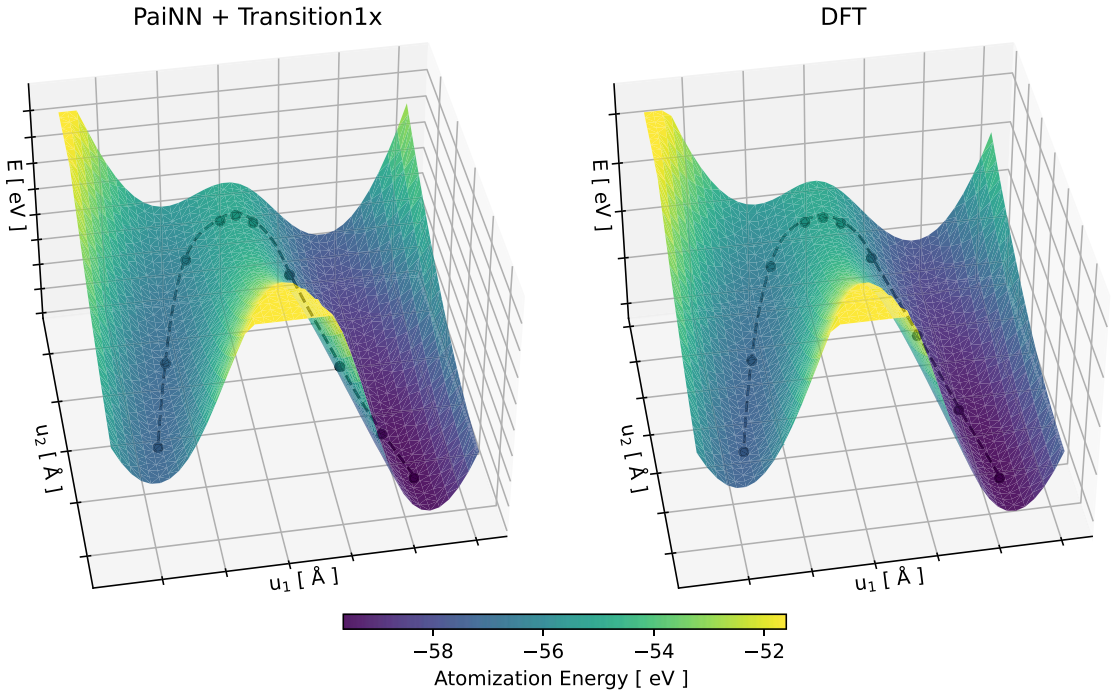

Figure B5. Reaction involving C3NCNH8. The reaction can be seen as a GIF by following this link.

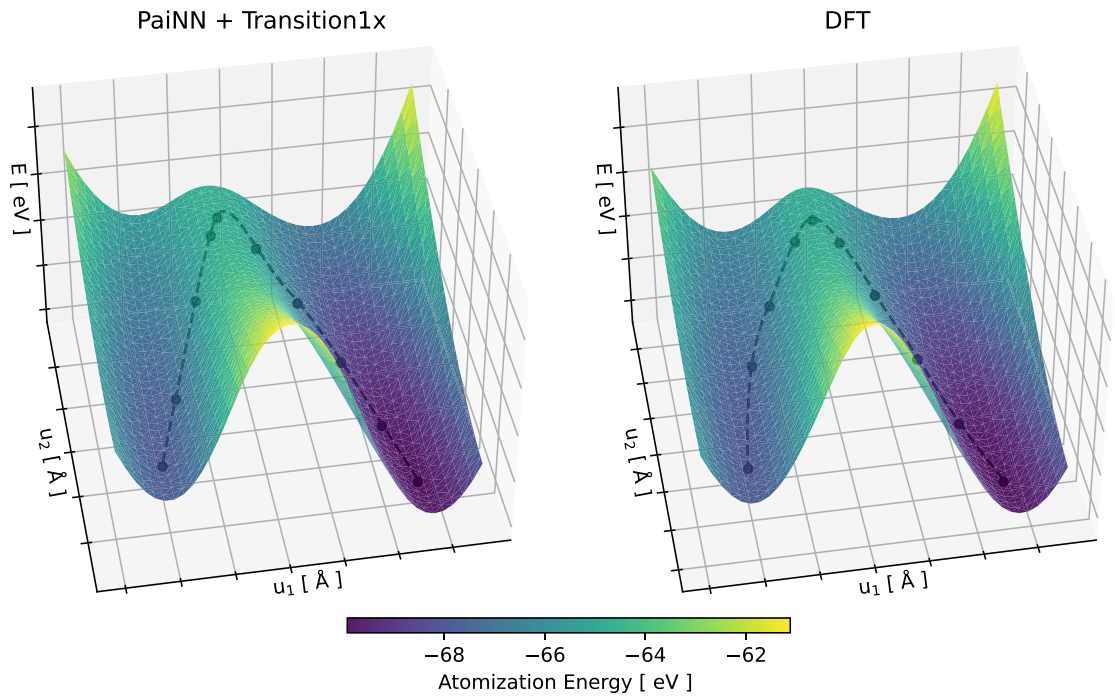

Figure B6. Reaction involving C3NC2OH9. The reaction can be seen as a GIF by following this link.

ORCID iDs

Mathias Schreiner https://orcid.org/0000-0001-7649-843X
Arghya Bhowmik https://orcid.org/0000-0003-3198-5116
Tejs Vegge https://orcid.org/0000-0002-1484-0284
Peter Bjørn Jørgensen https://orcid.org/0000-0003-4404-7276
Ole Winther https://orcid.org/0000-0002-1966-3205

References

[1] Bacciu D, Errica F, Micheli A and Podda M 2019 A gentle introduction to deep learning for graphs *Neural Netw.* **129** 203–21
[2] Zhou J, Cui G, Shengding H, Zhang Z, Yang C, Liu Z, Wang L, Changcheng Li and Sun M 2021 Graph neural networks: a review of methods and applications *AI Open* **1** 57–81
[3] Faber F A, Hutchison L, Huang B, Gilmer J, Schoenholz S S, Dahl G E, Vinyals O, Kearnes S, Riley P F and Anatole Von Lilienfeld O 2017 Prediction errors of molecular machine learning models lower than hybrid dft error *J. Chem. Theory Comput.* **13** 5255–64
[4] Westermayr J, Gastegger M, Schütt K T and Maurer R J 2021 Perspective on integrating machine learning into computational chemistry and materials science *J. Chem. Phys.* **154** 230903
[5] Campbell S I, Allan D B and Barbour A M 2020 Machine learning for the solution of the schrödinger equation *Mach. Learn.: Sci. Technol.* **1** 013002
[6] Behler Jorg and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401
[7] Westermayr J and Marquetand P 2021 Machine learning for electronically excited states of molecules *Chem. Rev.* **121** 9873–926
[8] Louis Reymond J 2015 The chemical space project *Acc. Chem. Res.* **48** 722–30
[9] Behler Jorg and Parrinello M 2007b Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401
[10] Behler Jorg 2015 Constructing high-dimensional neural network potentials: A tutorial review *Int. J. Quantum Chem.* **115** 1032–50
[11] Faber F A, Hutchison L, Huang B, Gilmer J, Schoenholz S S, Dahl G E, Vinyals O, Kearnes S, Riley P F and Anatole Von Lilienfeld O 2017 Prediction errors of molecular machine learning models lower than hybrid dft error *J. Chem. Theory Comput.* **13** 5255–64
[12] Gilmer J, Schoenholz S S, Riley P F, Vinyals O and Dahl G E 2017 Neural message passing for quantum chemistry *Int. Conf. on Machine Learning* PMLR 70 pp 1263–72
[13] Kaappa S, Larsen C and Wedel Jacobsen K 2021 Atomic structure optimization with machine-learning enabled interpolation between chemical elements *Phys. Rev. Lett.* **127** 166001
[14] Wang J, Shin S and Lee S 2020 Interatomic potential model development: finite-temperature dynamics machine learning *Adv. Theory Simul.* **3** 1900210
[15] Malshe M, Raff L M, Rockley M G, Hagan M, Agrawal P M and Komanduri R 2007 Theoretical investigation of the dissociation dynamics of vibrationally excited vinyl bromide on an ab initio potential-energy surface obtained using modified novelty sampling and feedforward neural networks. II. Numerical application of the method *J. Chem. Phys.* **127** 134105
[16] Lu X, Meng Q, Wang X, Fu B and Zhang D H 2018 Rate coefficients of the $h + h_2o_2 \rightarrow h_2 + ho_2$ reaction on an accurate fundamental invariant-neural network potential energy surface *J. Chem. Phys.* **149** 174303

[17] Young T A, Johnston-Wood T, Deringer V L and Duarte F 2021 A transferable active-learning strategy for reactive molecular force fields *Chem. Sci.* **12** 10944–55
[18] Manzhos S and Carrington T Jr 2020 Neural network potential energy surfaces for small molecules and reactions *Chem. Rev.* **121** 10187–217
[19] Pekka Koistinen O, Dagbjartsdóttir F B, Ásgeirsson Valmur, Vehtari A and Jónsson H 2017 Nudged elastic band calculations accelerated with gaussian process regression *J. Chem. Phys.* **147** 152720
[20] Sheppard D, Terrell R and Henkelman G 2008 Optimization methods for finding minimum energy paths *J. Chem. Phys.* **128** 134106
[21] Heinen S, Schwilk M, von Rudorff G F and von Lilienfeld O A 2020 Machine learning the computational cost of quantum chemistry *Mach. Learn.: Sci. Technol.* **1** 025002
[22] von Lilienfeld O A, Robert Müller K and Tkatchenko A 2020 Exploring chemical compound space with quantum-based machine learning *Nat. Rev. Chem.* **4** 347–58
[23] Smith J S, Isayev O and Roitberg A E 2017 Ani-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules *Sci. Data* **4** 170193
[24] Smith J S, Zubatyuk R, Nebgen B, Lubbers N, Barros K, Roitberg A E, Isayev O and Tretiak S 2020 The ani-1ccx and ani-1x data sets, coupled-cluster and density functional theory properties for molecules *Sci. Data* **7** 134
[25] Fink T and Reymond J-L 2007 Virtual exploration of the chemical universe up to 11 atoms of c, n, o, f: Assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes and drug discovery *J. Chem. Inf. Model.* **47** 342–53
[26] Fink T, Bruggesser H and Louis Reymond J 2005 Virtual exploration of the small-molecule chemical universe below 160 daltons *Angew. Chem., Int. Ed.* **44** 1504–8
[27] Seifert G and Ole Joswig J 2012 Density-functional tight binding—an approximate density-functional theory method *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2** 456–65
[28] Schreiner M, Bhowmik A, Vegge T, Busk J and Winther O 2022 Transition1x – a Dataset for Building Generalizable Reactive Machine Learning Potentials
[29] Henkelman G, Uberuaga B P and Jónsson H 2000 A climbing image nudged elastic band method for finding saddle points and minimum energy paths *J. Chem. Phys.* **113** 9901
[30] Smidstrup Søren, Pedersen A, Stokbro K and Jónsson H 2014 Improved initial guess for minimum energy path calculations *J. Chem. Phys.* **140** 214106
[31] Broyden C G 1970 The convergence of a class of double-rank minimization algorithms 1. General considerations *IMA J. Appl. Math.* **6** 76–90
[32] Makri S, Ortner C and Kermode J R 2019 A preconditioning scheme for minimum energy path finding methods *J. Chem. Phys.* **150** 094109
[33] Ramakrishnan R, Dral P O, Rupp M and Anatole Von Lilienfeld O 2014 Quantum chemistry structures and properties of 134 kilo molecules *Sci. Data* **1** 140022
[34] Ditchfield R, Hehre W J and Pople J A 2003 Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules *J. Chem. Phys.* **54** 724
[35] Da Chai J and Head-Gordon M 2008 Systematic optimization of long-range corrected hybrid density functionals *J. Chem. Phys.* **128** 084106
[36] Riley K E, Pitonˇcák M, Jureccka P and Hobza P 2010 Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories *Chem. Rev.* **110** 5023–63
[37] Thanthiriwatte K S, Hohenstein E G, Burns L A and David Sherrill C 2011 Assessment of the performance of dft and dft-d methods for describing distance dependence of hydrogen-bonded interactions *J. Chem. Theory Comput.* **7** 88–96
[38] Smith J S, Nebgen B, Lubbers N, Isayev O and Roitberg A E 2018 Less is more: sampling chemical space with active learning *J. Chem. Phys.* **148** 241733
[39] Schütt K T, Unke O T and Gastegger M 2021 Equivariant message passing for the prediction of tensorial properties and molecular spectra *Int. Conf. on Machine Learning* PMLR 139 pp 9377–88
[40] Kingma D P and Jimmy Lei B Adam: a method for stochastic optimization 2014 (arXiv:1412.6980)
[41] Henriksson J, Berger C, Borg M, Tornberg L, Raman Sathyamoorthy S, and Englund C 2021 Performance analysis of out-of-distribution detection on various trained neural networks (available at: www.iso.org/deliverables-all.html)
[42] Zhang L H, Goldstein M and Ranganath R 2021 Understanding failures in out-of-distribution detection with deep generative models *Int. Conf. on Machine Learning* PMLR 139 pp 12427–36