Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules

Johannes Klicpera, Shankari Giri, Johannes T. Margraf, Stephan Günnemann
Technical University of Munich
{klicpera,guennemann}@in.tum.de, shankari.giri@tum.de,
johannes.margraf@ch.tum.de

Abstract

Many important tasks in chemistry revolve around molecules during reactions. This requires predictions far from the equilibrium, while most recent work in machine learning for molecules has been focused on equilibrium or near-equilibrium states. In this paper we aim to extend this scope in three ways. First, we propose the DimeNet^{++} model, which is 8x faster and $10\,\%$ more accurate than the original DimeNet on the QM9 benchmark of equilibrium molecules. Second, we validate DimeNet^{++} on highly reactive molecules by developing the challenging COLL dataset, which contains distorted configurations of small molecules during collisions. Finally, we investigate ensembling and mean-variance estimation for uncertainty quantification with the goal of accelerating the exploration of the vast space of non-equilibrium structures. Our DimeNet^{++} implementation as well as the COLL dataset are available online.[1]

1 Introduction

Modern machine learning models for molecular property prediction typically focus on molecules in equilibrium (e.g. QM9 [13]) or close to the equilibrium (e.g. MD17 [2], ANI-1 [15], QM7-X [6]). However, this precludes their application to the dynamics during chemical reactions, which involve transition states far away from the equilibrium. Making reliable predictions for these states requires models that are able to cover a much broader range of chemical and configurational space, i.e. including open-shell electronic structures, stretched bonds and distorted angles. In this work we aim at making progress on this problem from three directions.

First, we propose a model that is fast, accurate, and generalizes well both to different configurations and different molecules. This model predicts both the molecule's energy and the forces acting on each atom, since the latter are crucial for the molecule's dynamic behavior. To this end, we start from the recently proposed Directional Message Passing Neural Network (DimeNet) [7], which fulfills all of these properties except one: It is comparatively slow to compute. We perform a thorough model analysis to fix this and propose the DimeNet^{++} model, which achieves an 8x runtime improvement while also improving predictions by $10\,\%$ on average and by $20\,\%$ for energies.

Second, we develop a new dataset that contains highly reactive non-equilibrium systems. The new COLL dataset contains $140\,000$ configurations of pairs of molecules reacting at high kinetic energies. It only consists of small molecules but covers the space of reactions much better and includes a significantly wider range of energies and forces than previous benchmarks, as shown in Fig. 4.

Due to the vast number of possible non-equilibrium configurations it is crucial that we are able to detect when we move out of the region covered by the training data and react appropriately (e.g. via
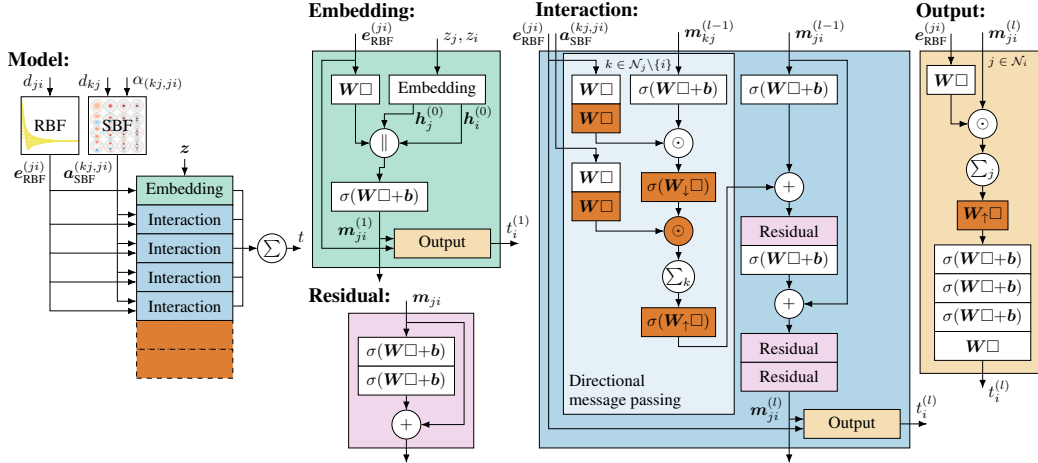
---

[1]`https://www.daml.in.tum.de/dimenet`

Figure 1: DimeNet^{++} architecture. $\square$ denotes the layer's input and $\|$ denotes concatenation. Changes to regular DimeNet are highlighted in **red**.

active learning). To achieve this we investigate ensembling [4] and mean-variance estimation [11]. We conclude that both are insufficient, due to their overhead and inability of reliably predicting the energy and force uncertainties.

2 DimeNet^{++}

DimeNet. DimeNet is a recently proposed Graph Neural Network (GNN) for molecular property prediction [7]. It improves upon regular GNNs in two ways. Normal GNNs represent each atom i separately via its embedding $\boldsymbol{h}_i$ and update these in each layer l via message passing. DimeNet instead embeds and updates the messages between atoms $\boldsymbol{m}_{ji}$, which enables it to consider directional information (via bond angles $\alpha_{(kj,ji)}$) as well as interatomic distances d_{ji}. DimeNet furthermore embeds distances and angles jointly using a spherical 2D Fourier-Bessel basis, resulting in the update

$$\boldsymbol{m}_{ji}^{(l+1)} = f_{\text{update}}(\boldsymbol{m}_{ji}^{(l)}, \sum_{k \in \mathcal{N}_j \setminus \{i\}} f_{\text{int}}(\boldsymbol{m}_{kj}^{(l)}, \boldsymbol{e}_{\text{RBF}}^{(ji)}, \boldsymbol{a}_{\text{SBF}}^{(kj,ji)})), \tag{1}$$

where f_{update} denotes the update function, f_{int} the interaction function, $\boldsymbol{e}_{\text{RBF}}^{(ji)}$ the radial basis function (RBF) representation of d_{ji} and $\boldsymbol{a}_{\text{SBF}}^{(kj,ji)}$ the spherical basis function (SBF) representation of d_{kj} and $\alpha_{(kj,ji)}$. In this work we do not touch either of those contributions and instead focus on the model architecture. The updated DimeNet^{++} architecture is illustrated in Fig. 1.

Combinatorial representation explosion. DimeNet embeds every message, i.e. every interacting *pair*, separately and thus uses many times as many embeddings as a regular GNN. This combinatorial explosion becomes even worse in the interaction block, where we need to embed every *triplet* to represent bond angles. On the QM9 dataset (with 5 Å cutoff) we found that DimeNet uses around 15x as many message embeddings as there are atoms and again around 15x as many triplet representations. Operations in the "directional message passing" block are thus 15x more expensive than elsewhere in the model, while those in the output block (which uses atom embeddings) are 15x cheaper.

Figure 2: DimeNet's original "directional message passing" block.

Fast interactions. We therefore first focus on the expensive "directional message passing" block. It is DimeNet's centerpiece, modelling the interaction between embeddings $\boldsymbol{m}_{kj}$ and basis representations $\boldsymbol{e}_{\text{RBF}}^{(ji)}$ and $\boldsymbol{a}_{\text{SBF}}^{(kj,ji)}$. As such, it requires an adequately expressive transformation. The original DimeNet accomplishes this with a bilinear layer, as shown in Fig. 2. Unfortunately, this layer is very expensive, which is exacerbated by being used in the model's most costly component.
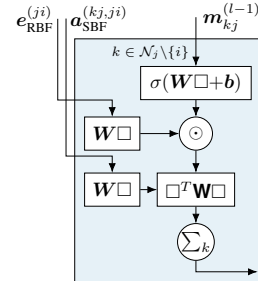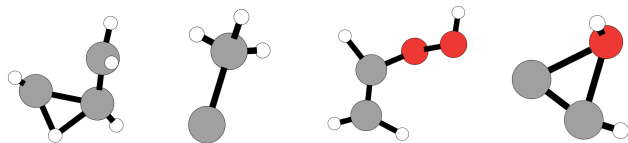
Figure 3: Example configurations from the COLL dataset. COLL covers a much broader range of configurational space, including stretched bonds and distorted angles.
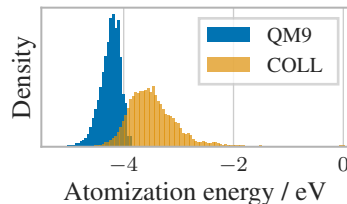

Figure 4: Distribution of atomization energy per atom. COLL covers a much wider range.

We alleviate this by replacing it with a simple Hadamard product and compensate for the loss in expressiveness by adding multilayer perceptrons (MLPs) for the basis representations. This recovers the original accuracy at a fraction of the computational cost (see Sec. 5).

Embedding hierarchy. We can directly leverage the fact that certain parts of the model use a higher number of embeddings by reducing the embedding size in these parts via down- and upprojection layers $W_\downarrow$ and $W_\uparrow$. This both accelerates the model and removes information bottlenecks, since we no longer aggregate information to a smaller number of equally sized embeddings.

Other improvements. We furthermore found that using 4 layers performs en par with the original 6 for U_0. Moreover, larger batch sizes significantly slowed down convergence, and mixed precision caused the model's precision to break down completely. Considering that DimeNet's relative error is below float16's machine precision ($5 \cdot 10^{-4}$), the latter might be expected.

3   COLL Dataset

The COLL dataset consists of configurations taken from molecular dynamics simulations of molecular collisions. To this end, collision simulations were performed with the cost-effective semiempirical GFN2-xTB method [1]. Subsequently, energies and forces for $140\,000$ random snapshots taken from these trajectories were recomputed with density functional theory (DFT). These calculations were performed with the revPBE functional and def2-TZVP basis, including D3 dispersion corrections [16].

Exemplary structures from the COLL set are shown in Fig. 3. Unlike established molecular benchmark sets (e.g. QM9), which consist of equilibrium or near-equilibrium configurations, the structures in COLL can be highly distorted. In particular, stretched bonds and angles, as well as open-shell electronic structures are prevalent. All calculations are preformed with broken spin-symmetry.

Overall, the configurations in COLL represent a challenge for electronic structure calculations, since such systems may display multiple self-consistent-field (SCF) solutions. This can pose a significant problem for ML-based models, as the corresponding reference potential energy surfaces can be discontinuous. To avoid this issue, multiple calculations from randomized initial wavefunctions were conducted, and the lowest energy solution selected. Furthermore, Fermi-smearing with an electronic temperature of $5000\,\mathrm{K}$ was applied, which is helpful both for SCF convergence and the approximate description of static correlation effects [3].

4   Uncertainty Quantification

The vast number of non-equilibrium states reachable in high-energy molecular dynamics simulations (such as reactions) means that systems will often move outside the space covered by our training set. We therefore need a reliable way of detecting a degradation in predictive performance. Most uncertainty quantification (UQ) methods are focused on providing an uncertainty estimate for the direct prediction [5, 10]. However, out-of-equilibrium dynamics require uncertainty estimates for both the energy E and the force $\boldsymbol{F} = -\frac{\partial E}{\partial \boldsymbol{x}}$. Many non-differentiable methods (e.g. combining GNNs with a random forest) are therefore not applicable. Ensembling is a notable exception but introduces a large computational overhead since we need to calculate predictions using multiple separate models.

3

Table 1: U_0 validation MAE on the QM9 dataset after each DimeNet^{++} improvement.

Model	Time per epoch / min	Val MAE U_0 / meV
DimeNet	35.4	8.27
& Hadamard product	6.6	9.45
& 2-layer MLP for RBF and SBF	7.1	8.42
& reduced to 4 layers	4.7	8.54
& triplet embedding size to 64	4.2	7.60
& output (atom) embedding size to 256	4.3	7.05

Table 2: MAE on the QM9 dataset. DimeNet^{++} performs best on average and for most properties individually, despite being 8x faster than DimeNet.

Target	Unit	SchNet	MGCN	DeepMoleNet	DimeNet	**DimeNet^{++}**
μ	D	0.0330	0.0560	**0.0253**	0.0286	0.0297
α	$a_0{}^3$	0.235	**0.0300**	0.0681	0.0469	0.0435
ϵ_{HOMO}	meV	41.0	42.1	**23.9**	27.8	24.6
ϵ_{LUMO}	meV	34.0	57.4	22.7	19.7	**19.5**
$\Delta\epsilon$	meV	63.0	64.2	33.2	34.8	**32.6**
$\left\langle R^2 \right\rangle$	$a_0{}^2$	**0.073**	0.110	0.680	0.331	0.331
ZPVE	meV	1.70	**1.12**	1.90	1.29	1.21
U_0	meV	14.0	12.9	7.70	8.02	**6.32**
U	meV	19.0	14.4	7.80	7.89	**6.28**
H	meV	14.0	14.6	7.80	8.11	**6.53**
G	meV	14.0	16.2	8.60	8.98	**7.56**
c_{v}	$\frac{\text{cal}}{\text{mol K}}$	0.0330	0.0380	0.0290	0.0249	**0.0230**
std. MAE	%	1.76	1.86	1.03	1.05	**0.98**
logMAE	-	-5.17	-5.26	-5.46	-5.57	**-5.67**

Even if the method is differentiable it might only provide a mean and standard deviation, i.e. μ_E and σ_E (e.g. mean-variance estimation (MVE)). This allows us to obtain the force prediction via

$$\mu_{\boldsymbol{F}} = \left\langle -\frac{\partial E}{\partial \boldsymbol{x}} \right\rangle = -\frac{\partial}{\partial \boldsymbol{x}} \left\langle E \right\rangle = -\frac{\partial}{\partial \boldsymbol{x}} \mu_E. \tag{2}$$

However, performing the same operation on σ_E does not yield the analogous result:

$$\frac{\partial}{\partial \boldsymbol{x}} \sigma_E^2 = \frac{\partial}{\partial \boldsymbol{x}} \left(\left\langle E^2 \right\rangle - \left\langle E \right\rangle^2 \right) = -2 \left(\left\langle -E \frac{\partial}{\partial \boldsymbol{x}} E \right\rangle - \left\langle E \right\rangle \left\langle -\frac{\partial}{\partial \boldsymbol{x}} E \right\rangle \right) = -2\text{Cov}(E, \boldsymbol{F}). \tag{3}$$

There is thus no general way of estimating $\sigma_{\boldsymbol{F}}$ for these kinds of models. Instead, we have to rely on σ_E as the uncertainty measure and hope that it correlates with the force error.

5   Experiments

DimeNet^{++} In Table 1 we evaluate each of the proposed DimeNet improvements separately on the U_0 validation set of QM9 [13]. We see that each change either reduces the runtime or improves the error. Exchanging the bilinear layer for a Hadamard product has by far the largest impact, single-handedly decreasing the runtime by a factor of 5. Interestingly, decreasing the embedding size both accelerates the model and improves the accuracy. This is either due to the additional down- and upprojection layers improving expressiveness or to the smaller embedding size improving generalization.

We evaluate the final DimeNet^{++} model on all QM9 targets and compare it to the state-of-the-art models SchNet [14], MGCN [9], and DeepMoleNet [8]. Table 2 shows that DimeNet^{++} performs better for most targets and best overall, in addition to being 8x faster than DimeNet. OrbNet [12] performs better on U_0, but has not published results for the other properties. Note that the DFT-based representations introduced by DeepMoleNet and OrbNet can also be incorporated into DimeNet^{++}.

COLL dataset. Table 3 shows that DimeNet^{++} is only 17 % slower than SchNet (reference implementation), while reducing the error by 76 % on average. As expected, the COLL dataset is

Table 3: Performance on the COLL dataset. MAE is given in eV and $\mathrm{eV}/\text{Å}$. ρ denotes the correlation coefficient and Δ the absolute error. DimeNet^{++} performs significantly better than SchNet. MVE is much faster than ensembling but unable to estimate the force error.

	Time per epoch	MAE$_E$	MAE$_{\boldsymbol{F}}$	$\rho(\Delta_E, \sigma_E)$	$\rho(\Delta_{\boldsymbol{F}}, \sigma_{\boldsymbol{F}})$	$\rho(\Delta_{\boldsymbol{F}}, \sigma_E)$
SchNet	**8.9 min**	0.198	0.172	-	-	-
DimeNet^{++}	10.4 min	0.047	0.040	-	-	-
Ensemble (DimeNet^{++})	31.2 min	0.050	**0.038**	0.42	0.85	0.64
MVE (DimeNet^{++})	13.6 min	**0.033**	0.041	0.16	-	0.05

significantly more challenging than QM9. Both SchNet and DimeNet^{++} exhibit an MAE that is around 10x higher than on QM9.

Uncertainty quantification. Ensembling and MVE both struggle with estimating the energy uncertainty, as shown for DimeNet^{++} in Table 3. The force error is very well estimated by the ensemble, but not by the energy uncertainty – especially for MVE. The energy uncertainty is thus not as good a proxy for the force error as one would expect. While the ensemble does perform decently, its computational overhead is still considerable. Reliable and fast uncertainty estimates thus remain an important direction for future work.

Acknowledgments and Disclosure of Funding

References

[1] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.*, 15 (3):1652–1671, 2019.

[2] Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017.

[3] Stefan Grimme and Andreas Hansen. A Practicable Real-Space Measure and Visualization of Static Electron-Correlation Effects. *Angew. Chemie - Int. Ed.*, 54(42):12308–12313, 2015.

[4] L.K. Hansen and P. Salamon. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):993–1001, 1990.

[5] Lior Hirschfeld, Kyle Swanson, Kevin Yang, Regina Barzilay, and Connor W. Coley. Uncertainty Quantification Using Neural Networks for Molecular Property Prediction. *CoRR*, 2005.10036, 2020.

[6] Johannes Hoja, Leonardo Medrano Sandonas, Brian G. Ernst, Alvaro Vazquez-Mayagoitia, Robert A. DiStasio Jr., and Alexandre Tkatchenko. QM7-X: A comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules. *arXiv*, 2006.15139, 2020.

[7] Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional Message Passing for Molecular Graphs. In *ICLR*, 2020.

[8] Ziteng Liu, Liqiang Lin, Qingqing Jia, Zheng Cheng, Yanyan Jiang, Yanwen Guo, and Jing Ma. Transferable Multi-level Attention Neural Network for Accurate Prediction of Quantum Chemistry Properties via Multi-task Learning. *ChemRxiv*, 12588170.v1, 2020.

[9] Chengqiang Lu, Qi Liu, Chao Wang, Zhenya Huang, Peize Lin, and Lixin He. Molecular Property Prediction: A Multilevel Quantum Interactions Modeling Perspective. In *AAAI*, 2019.

[10] Félix Musil, Michael J. Willatt, Mikhail A. Langovoy, and Michele Ceriotti. Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. *Journal of Chemical Theory and Computation*, 15(2):906–915, 2019.

[11] D.A. Nix and A.S. Weigend. Estimating the mean and variance of the target probability distribution. In *ICNN*, 1994.

[12] Zhuoran Qiao, Matthew Welborn, Animashree Anandkumar, Frederick R. Manby, and Thomas F. Miller. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of Chemical Physics*, 153(12):124111, 2020.

[13] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):1–7, 2014.

[14] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NeurIPS*, 2017.

[15] J.S. Smith, O. Isayev, and A.E. Roitberg. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science*, 8(4):3192–3203, 2017.

[16] Yingkai Zhang and Weitao Yang. Comment on "generalized gradient approximation made simple". *Phys. Rev. Lett.*, 80(4):890, 1998.