How Robust are Modern Graph Neural Network Potentials in Long and Hot Molecular Dynamics Simulations?

Sina Stocker,[1, 2, a)] Johannes Gasteiger,[2, a)] Florian Becker,[2] Stephan Günnemann,[2] and Johannes T. Margraf[1, b)]
[1)] *Fritz-Haber-Institute of the Max-Planck-Society, Germany*
[2)] *Technical University of Munich, Germany*

(Dated: 21 March 2022)

Graph neural networks (GNNs) have emerged as a powerful machine learning approach for the prediction of molecular properties. In particular, recently proposed advanced GNN models promise quantum chemical accuracy at a fraction of the computational cost. While the capabilities of such advanced GNNs have been extensively demonstrated on benchmark datasets, there have been few applications in real atomistic simulations. Here, we therefore put the robustness of GNN interatomic potentials to the test, using the recently proposed GemNet architecture as an example. Models are trained on the QM7-x database of organic molecules and used to perform extensive MD simulations. We find that low test set errors are not sufficient for obtaining stable dynamics and that severe pathologies sometimes only become apparent after hundreds of ps of dynamics. Nonetheless, highly stable and transferable GemNet potentials can be obtained with sufficiently large training sets.

Atomistic simulations are an invaluable tool for gaining mechanistic and structural insight into chemical systems, including solid materials[1], interfaces[2,3], liquids[4] or even complex biological systems like the SARS-CoV-2 virus[5]. They are also becoming increasingly important in the design of new materials and drugs[6,7]. In many ways, the prototypical atomistic simulation is a Molecular Dynamics (MD) trajectory, which propagates the atomic coordinates of a system in time, starting from some initial conditions. MD simulations are extremely common, both by themselves and as part of more elaborate sampling procedures like parallel tempering or metadynamics.

In principle, highly accurate MD trajectories can be obtained from electronic structure methods like density functional theory (DFT). Unfortunately, such *ab initio* MD (AIMD) simulations require the (approximate) solution of the electronic Schrödinger equation at every time step. This makes them very expensive from a computational perspective and ultimately limits the applicability of AIMD to a few hundreds of atoms and relatively short (*i.e* ps) timescales. For many scientific questions, simulations of much larger systems, longer timescales or (usually) both are required. To this end, empirical interatomic potentials are typically used. These provide an analytical expression for high-dimensional potential energy surfaces which can be evaluated in a small fraction of the time required for a DFT calculation. This gain in efficiency invariably comes at the expense of a decrease in accuracy and/or transferability, however.

To bridge this gap between computational cost and accuracy, machine learned interatomic potentials have recently gained popularity in computational chemistry[8–11] and materials science[12–14]. In particular, a range of neural network[15–17] and kernel based potentials[18,19] have been developed and applied to a wide variety of chemical systems. While somewhat more expensive than classical force fields, these potentials are able to predict energies and forces with DFT accuracy and have thus become an important part of the toolbox of computational chemistry.

One of the most recent additions to this family of methods are potentials based on graph neural networks (GNNs), such as SchNet, DimeNet, GemNet and NequIP.[20–31] Here, much progress towards ever more accurate and expressive potentials has been made, e.g. by using equivariant formulations or embedding atom pairs and triplets. While such efforts naturally focus on established benchmark databases like QM9[32,33], MD17[34] or OC20[35], comparatively little research has been conducted to show the applicability of such advanced GNN potentials in real atomistic simulations. A notable exception to this is a recent paper of Batzner and coworkers[22], which demonstrated that potentials based on the equivariant NequIP architecture could be used in stable and accurate MD simulations, when trained on AIMD data for the respective system.

In this contribution, we aim to provide an in-depth exploration of the robustness of state-of-the-art GNN potentials based on the GemNet architecture[21] in MD simulations. To this end, we ran a total of 245 ns of dynamics (around 500 million timesteps) across a wide range of temperatures and organic molecules. By checking samples from these large ensembles with DFT reference calculations, the extrapolative capabilities of these potentials in configuration and chemical space was tested simultaneously. Furthermore, the impact of training set size on the robustness of the potentials was explored.

GNNs treat chemical systems as graphs, with nodes representing atoms and edges representing interactions between atom pairs. While traditional chemical graph representations usually equate edges with covalent bonds, GNNs assume edges between all atoms within a given cutoff. All potentials discussed in the following are based on the geometric message passing neural network (GemNet)[21], which shows excellent performances on established benchmark data sets like MD17 and OC20 as well as QM7-x (see Fig. 1). GemNet embeds both the atoms and the interatomic edges via high-dimensional

[a)] The first two authors contributed equally
[b)] Electronic mail: margarf@fhi.mpg.de

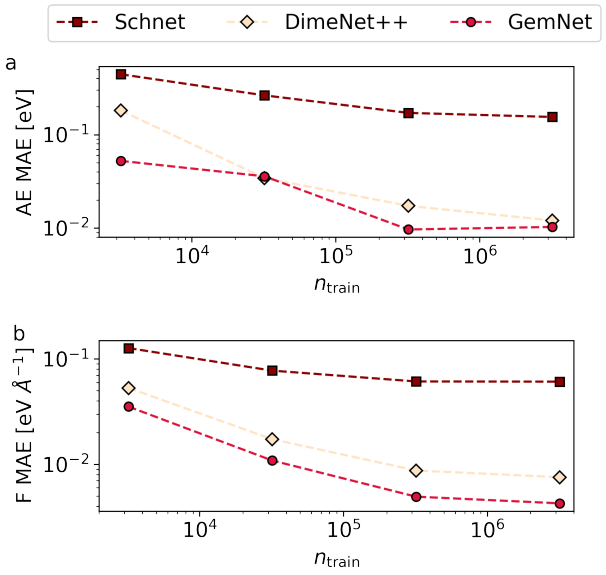

FIG. 1: **Learning curves.** (a) Mean absolute errors (MAEs) of atomization energies (AE) and (b) forces (F) against the number of training configurations. MAEs are calculated for a test set consisting of 10,100 random configurations from the QM7-x database.

vectors. Both kinds of embeddings are then updated in multiple layers using learnable weight matrices and by passing messages between the edges and atoms within a given cutoff distance. GemNet leverages the full geometric information for this: The interatomic distances, the angles between neighboring edges, and the dihedral angles defined via triplets of edges. From the learned embeddings, energy contributions for each atom and layer are obtained, which are subsequently summed up to calculate the total energy of the system. The whole model is continuously differentiable, which allows calculating the forces via $\mathbf{F}_i = -\frac{\partial}{\partial \mathbf{x}_i} E$. As for all GNNs, the use of a finite cutoff and per-atom energy contributions makes the predictions size-extensive and the computational cost scale linearly with the number of atoms.

Herein, we trained several GemNet potentials on different subsets of the recently published QM7-x database.[36] This dataset consists of around 4.2 million configurations sampled from small organic molecules consisting of up to seven non-hydrogen atoms (*i.e.* C, O, N, S, Cl), with 4-23 atoms in total. Importantly, QM7-x covers both equilibrium and non-equilibrium structures. Starting from 6,950 structural formulas, it contains around 41,500 equilibrium structures (including stereoisomers and conformers) and 100 additional non-equilibrium structures for each equilibrium geometry. The latter were generated by applying linear combinations of normal mode displacements to each configurations, thus approximately mimicking molecular dynamics within the harmonic approximation. For each configuration, total energies and forces at the hybrid DFT (PBE0)[37] level with a many-body dispersion correction (MBD)[38] are provided, computed with tightly converged numerical atom-centered basis sets and integration grids[39,40] (see Ref. [36] for full details).

GemNet potentials were trained on atomization energies (AE) and forces (F) simultaneously. Since forces are ultimately the driver of MD simulations and contain more fine-grained information than energies, forces were weighted more strongly in our fits, so that the AE only contributes 0.1% to the loss function (see SI for details). This essentially follows the philosophy of gradient domain machine learning,[34,41] which exclusively uses forces. However, we do include a small AE contribution to the loss, as energy differences across chemical space cannot be learned effectively from forces alone.[42] For training, the QM7-x dataset was randomly split into a test set of 10,100 configurations, training sets of 3.2k, 32k, 320k and 3.2Mio configurations and corresponding validation sets of 800, 8k, 80k and 800k configurations (the latter being used for hyperparameter selection, see SI). In the interest of simplification, we will denote models trained on small (3.2k and 32k) and large (320k and 3.2Mio) training sets as 'sparse' and 'exhaustive' models respectively.

In Fig. 1, the corresponding learning curves for AE and F are shown. The force curve shows a roughly linear decrease on the log-log scale between 3.2k and 320k training configurations but levels off between 320k and 3.2Mio configurations. This indicates that the more exhaustive models approach the intrinsic accuracy that is possible given the precision of the data and limitations of the models themselves (e.g. due to the cutoffs employed). Due to the lower weighting of energies in the loss the AE curve is somewhat more noisy but follows the same trend.

To put this performance into perspective, the most exhaustive GemNet model yields a force MAE of 0.0043 eV Å^{-1}, which can be compared with an MAE of 0.015 eV Å^{-1} for the recently developed SpookyNet[23] architecture (in this case trained on 4.2Mio molecules). In addition, GemNet outperforms SchNet[28] and DimNet++[25] on QM7-x for nearly all points of the learning curve (with the only exception being the AE error of the 32k model). Importantly, the energy errors are very low (0.01 eV = 0.23 kcal mol^{-1}) despite the low weighting of AEs in the loss. It is furthermore notable that even the model trained on 3.2k configurations displays quite good performance with MAEs of around 0.035 eV Å^{-1} and 0.05 eV ($\approx$ 1 kcal mol^{-1}).

To explore the robustness of the GemNet potentials within the scope of their training set, constant temperature MD simulations were performed for 20 representative molecules from QM7-x (see FIG. 2a). Here, care was taken to include all atom types in the dataset. For each molecule, 1 ns trajectories were generated with a 0.5 fs timestep at three different temperatures (300 K, 600 K and 1200 K), using all models presented in the learning curve (see SI for details on the MD simulations). The rationale for using these temperatures is that they lead to increasingly extensive exploration of phase space. Indeed, it is not uncommon to use high temperature dynamics for this purpose, *e.g.* in replica exchange MD.[43] From each trajectory, 72 configuration were uniformly sampled and the corresponding energies and forces computed with identical DFT settings to the ones used for the QM7-x set.

Figures 2b and 2c show the AE and F MAEs for these

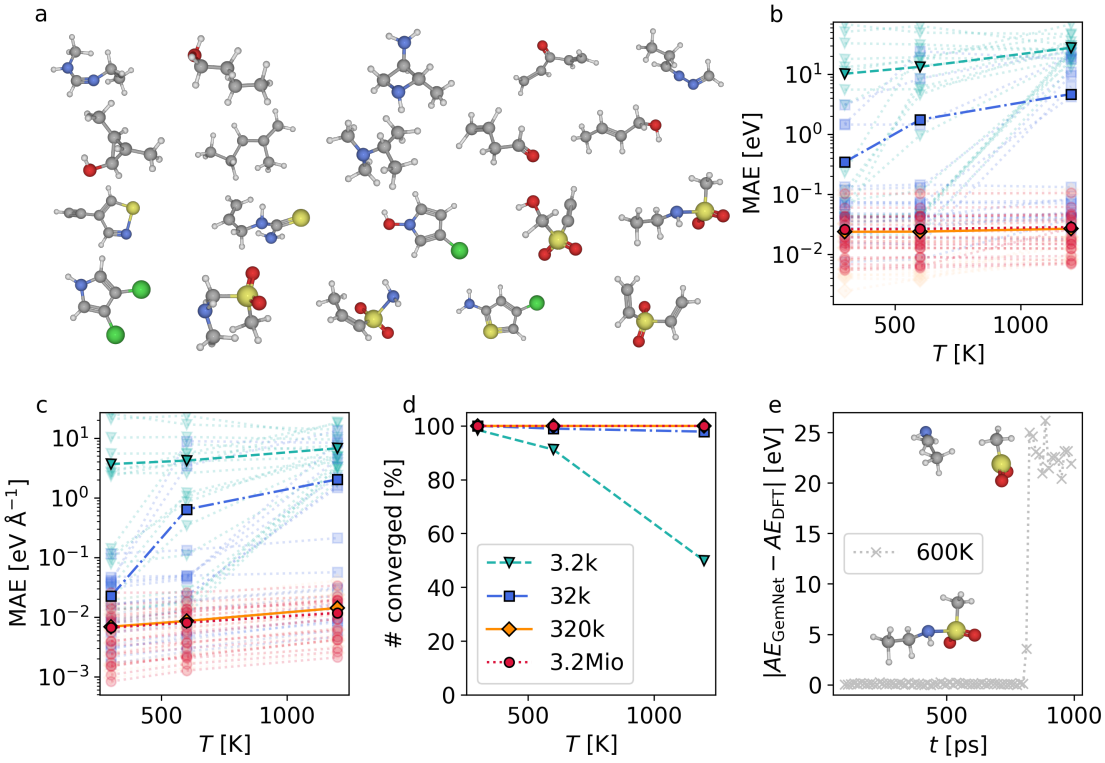

FIG. 2: **Robustness of GemNet potentials in molecular dynamics (MD).** (a) Representative molecules from QM7-x used in the MD tests. (b), (c) Mean absolute errors (MAEs) on atomization energy (AE) and force (F) predictions as a function of training set size and temperature. Opaque lines and symbols represent the average MAE over all molecules in (a). Transparent lines and symbols represent the MAE of one specific molecule. (d) Percentage of converged DFT calculations for configurations generated with different potentials and temperatures. (e) AE error as a function of simulation time for a 1 ns MD trajectory at 600 K, using a potential trained with 3.2k configurations. This sub-panel shows a drastic deterioration in energy predictions after around 700 ps, when the molecule dissociates into fragments that cannot be accurately described by the model.

samples as a function of temperature and training set size. Here, opaque symbols and lines represent MAEs averaged over 20 different trajectories corresponding to a given model and temperature. Transparent symbols and lines illustrate the MAEs for each trajectory individually, to provide some insight into the spread of MAEs for different molecules (see SI for additional illustrations of the respective error distributions). Overall, we find quite consistent trends for both AE and F predictions. Whereas the exhaustive models (320k and 3.2Mio) only display a very slight increase of the MAEs with temperature, the errors of the sparse models (3.2k and 32k) increase dramatically. This is expected, as higher temperature MD simulations more extensively explore the phase space and consequently move away from the training configurations.

Notably, the 3.2k model already displays a very large AE error of more than 10 eV at 300 K. The MD error is thus orders of magnitude larger than the test set error, even though these configurations should arguably fall within the scope of the training set. This mainly stems from the fact that the trajectories for certain molecules lead to completely unphysical configurations, for which the potential then displays extremely large errors. Such unphysical configurations also commonly lead to convergence issues in the reference DFT calculations. To quantify this, the percentage of converged DFT calculations for configurations obtained with a given potential and simulation temperature is shown in Fig. 2d. We find that all DFT calculations converge for the 320k and 3.2Mio potentials, while the sparse models generate increasingly unphysical configurations with rising temperature. This is particularly evident for the 3.2k model at 1200 K, where only about half of the DFT calculations converge.

The marked discrepancy between the test set and MD performance of the 3.2k model underscores the limitations of using test configurations that are not generated by the potential itself. For a ML model to be useful in atomistic simulations, it is not sufficient to show that it provides accurate fits for physically reasonable configurations. It is equally important that the model avoids unphysical configurations in its own simulations. Note that testing this requires sufficiently long trajectories. This is illustrated for a representative example in Fig. 2e. Here, the error of the 3.2k model is actually quite low for the first 700 ps of the simulation after which it rises sharply to more than 20 eV due to an unphysical bond dissociation event. This behaviour can be understood as a kind

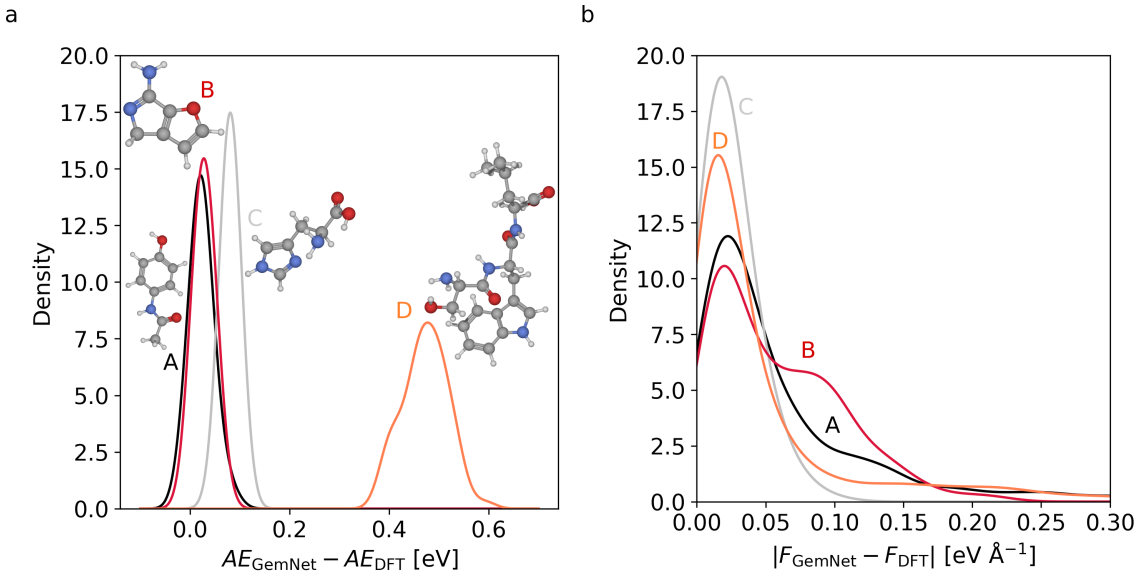

FIG. 3: **Out-of-sample validation of the GemNet potential trained on 3.2Mio configurations.** (a) Kernel density estimates of atomization energy (AE) error distributions for four out-of-sample molecules (A=Paracetamol: red, B=4H-Furo[2,3-c]pyrrol-6-amin: black, C=Histidin: grey and D=Ser-Trp-Leu-tripeptide: orange), obtained at 1200 K using the 3.2Mio potential. (b) Kernel density estimates of the corresponding force (F) error distributions.

of 'hole' in the potential energy surface of the model. This hole can be rather small, but once the simulation reaches such a configuration the trajectory is completely unreasonable. The 'robustness' of a ML potential can thus be understood as a measure of how frequent and how large such holes in the potential energy surface are. Ultimately this can only be quantified by performing MD simulations with the corresponding potential.

It should be stressed that this notion of robustness is not necessarily correlated with the test MAE, despite the fact that the robust GemNet models also display much lower MAEs. Indeed, the robustness of traditional bio-organic forcefileds with fixed topologies is very high. However, in this case robustness is gained at the expense of model flexibility. The challenge for ML potentials is that they must be robust without sacrificing flexibility. Our tests show that this is not trivial. On a more positive note, we do find that GemNet potentials with sufficiently large training sets are very robust across the phase space of the QM7-x dataset and beyond.

Another way to illustrate this is to consider the performance of the 3.2k model for a trajectory generated with the 3.2Mio potential in comparison with its own trajectory. Specifically this means that we generate two independent trajectories with the 3.2k and 3.2Mio model and evaluate MAEs of the 3.2k model for configurations drawn from each trajectory. Taking the molecule in Fig. 2e at 1200 K, the F MAE of the 3.2k potential is 6.8 eV Å^{-1} for the 3.2k trajectory but only 0.16 eV Å^{-1} when it is evaluated on the 3.2Mio trajectory. Again, the sparse model performs quite well for the physically reasonable configurations generated with the 3.2Mio model. The problem only becomes apparent when testing the sparse model on its own trajectory.

Having established the robustness of the exhaustive models within the scope of QM7-x, we now turn to the simultaneous exploration in chemical and configuration space. To this end, we consider four molecules consisting of 9-29 heavy atoms (*i.e.* which are significantly larger than the training molecules). Again, 1 ns MD trajectories were generated with the 3.2Mio potential at 1200 K. Figure 3 shows the corresponding AE and F error distributions. Strikingly, the AE errors are systematically more positive than the DFT reference energies, most prominently for the large Ser-Trp-Leu tripeptide. Here, the mean AE is shifted by 0.47 eV with respect to the reference, which is substantial when compared to an MAE of 0.0284 eV at 1200 K in Fig. 2b.

This shift can be explained by the absence of attractive long-range interactions (*e.g.* dispersion or electrostatics) in the GemNet potential. While message-passing neural networks can in principle include information from beyond their cutoff distance, the QM7-x database exclusively consists of small molecules so that long-range interactions simply cannot be learned from it. Methods to include long-range interactions are proposed in literature[23,27,44,45] and could also be applied to the GemNet architecture. Nonetheless, GemNet and DFT energies are highly correlated ($R^2 = 0.998$, see SI) and the standard deviation of the AE error distribution is only 0.045 eV so that the MD trajectory for this system should still be considered to be of high quality. While the long-range interactions are thus considerable in magnitude, they do not fluctuate very strongly.[46] This is also the case for the other molecules, which display very narrow AE error distributions. Similarly, force component errors are consistently small, with MAEs between even 0.012 eV Å^{-1} and 0.036 eV Å^{-1}.

In conclusion, we have explored the robustness of GNN potentials based on the recent GemNet architecture in

MD simulations. We find that sufficiently large training sets are key to obtaining robust GNN potentials and that a low test set error does not guarantee that stable trajectories can be generated. Interestingly, in some cases severe instabilities were only discovered after hundreds of ps of dynamics. The test set error should thus not be taken at face value as a measure for the error one can expect in 'real' applications. Demonstrating 'chemical accuracy' on a test set is by itself not enough.

With large enough training sets, the GemNet potentials used herein do display impressive performance, however. This is demonstrated by applications in high-temperature MD simulations of systems that are significantly larger than the training molecules. In this extrapolative regime, errors are mostly systematic and explainable and no instabilities were observed. Interestingly, no significant improvements in terms of accuracy or robustness were observed when training on 3.2Mio instead of 320k samples, indicating that all relevant information about the underlying potential energy surface can be learned from less than 10% of the dataset. This is significant because robust ML potentials are often associated with iterative training procedures. Due to their size and complexity (the models used herein fit 2.2 million parameters), GNN models are *a priori* not ideal for such settings. Indeed, training times of several GPU weeks are not unusual, which is clearly impractical in an iterative workflow. Well curated databases like QM7-x and powerful model architectures like GemNet circumvent this issue.

As a final point, we note that the potentials discussed herein (as well as the underlying code) are freely available at `https://www.daml.in.tum.de/gemnet`. We recommend the 3.2Mio GemNet potential as a general-purpose force field for exploring the conformational space of small to medium organic molecules. Indeed, the accuracy and the robustness of the 320k and 3.2Mio models is high enough that they can be considered as a cost effective replacement of DFT calculations for this application. It remains to be seen whether equally accurate and robust models can be obtained for larger chemical spaces, broader sections of the periodic table and chemical reactions.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through TUM International Graduate School of Science and Engineering (IGSSE) (GSC 81)

[1]V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, "Origins of structural and electronic transitions in disordered silicon," *Nature*, vol. 589, no. 7840, pp. 59–64, 2021.

[2]S. Stegmaier, R. Schierholz, I. Povstugar, J. Barthel, S. P. Rittmeyer, S. Yu, S. Wengert, S. Rostami, H. Kungl, K. Reuter, R.-A. Eichel, and C. Scheurer, "Nano-scale complexions facilitate li dendrite-free operation in latp solid-state electrolyte," *Adv. Energy Mater.*, vol. 11, no. 26, p. 2100707, 2021.

[3]J. Timmermann, F. Kraushofer, N. Resch, P. Li, Y. Wang, Z. Mao, M. Riva, Y. Lee, C. Staacke, M. Schmid, C. Scheurer, G. S. Parkinson, U. Diebold, and K. Reuter, "iro$_2$ surface complexions identified through machine learning and surface investigations," *Phys. Rev. Lett.*, vol. 125, p. 206101, 2020.

[4]B. Cheng, G. Mazzola, C. J. Pickard, and M. Ceriotti, "Evidence for supercritical behaviour of high-pressure liquid hydrogen," *Nature*, vol. 585, no. 7824, pp. 217–220, 2020.

[5]M. I. Zimmerman, J. R. Porter, M. D. Ward, S. Singh, N. Vithani, A. Meller, U. L. Mallimadugula, C. E. Kuhn, J. H. Borowsky, R. P. Wiewiora, M. F. D. Hurley, A. M. Harbison, C. A. Fogarty, J. E. Coffland, E. Fadda, V. A. Voelz, J. D. Chodera, and G. R. Bowman, "Sars-cov-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome," *Nat. Chem.*, vol. 13, no. 7, pp. 651–659, 2021.

[6]M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C.-S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S.-C. Lo, A. Ip, Z. Ulissi, and E. H. Sargent, "Accelerated discovery of co2 electrocatalysts using active machine learning," *Nature*, vol. 581, no. 7807, pp. 178–183, 2020.

[7]N. Artrith, A. Urban, and G. Ceder, "Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species," *Phys. Rev. B*, vol. 96, no. 1, pp. 014112–014112, 2017.

[8]O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, "Machine learning force fields," *Chem. Rev.*, vol. 121, no. 16, pp. 10142–10186, 2021.

[9]J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, "Combining machine learning and computational chemistry for predictive insights into chemical systems," *Chem. Rev.*, vol. 121, no. 16, pp. 9816–9872, 2021.

[10]J. Xu, X.-M. Cao, and P. Hu, "Accelerating metadynamics-based free-energy calculations with adaptive machine learning potentials," *J. Chem. Theory Comput.*, vol. 17, no. 7, pp. 4465–4476, 2021. PMID: 34100605.

[11]S. Stocker, G. Csányi, K. Reuter, and J. T. Margraf, "Machine learning in chemical reaction space," *Nat. Commun.*, vol. 11, no. 1, p. 5505, 2020.

[12]J. Kim, D. Kang, S. Kim, and H. W. Jang, "Catalyze materials science with machine learning," *ACS Mater. Lett.*, vol. 3, no. 8, pp. 1151–1171, 2021.

[13]V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, "Gaussian process regression for materials and molecules," *Chem. Rev.*, vol. 121, no. 16, pp. 10073–10141, 2021.

[14]J. Behler and G. Csányi, "Machine learning potentials for extended systems: a perspective," *Eur. Phys. J. B*, vol. 94, no. 7, p. 142, 2021.

[15]J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.*, vol. 98, p. 146401, 2007.

[16]J. S. Smith, O. Isayev, and A. E. Roitberg, "Ani-1: an extensible neural network potential with dft accuracy at force field computational cost," *Chem. Sci.*, vol. 8, pp. 3192–3203, 2017.

[17]J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning," *Nat. Commun.*, vol. 10, no. 1, pp. 2903–2903, 2019.

[18]A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Phys. Rev. Lett.*, vol. 104, p. 136403, 2010.

[19]A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," *Phys. Rev. B*, vol. 87, p. 184115, 2013.

[20]K. Schütt, O. Unke, and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," in *ICML*, 2021.

[21]J. Gasteiger, F. Becker, and S. Günnemann, "Gemnet: Universal directional graph neural networks for molecules," in *NeurIPS*, 2021.

[22]S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E(3)-

equivariant graph neural networks for data-efficient and accurate interatomic potentials," 2021.
[23]O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda, and K.-R. Müller, "Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects," 2021.
[24]J. Gasteiger, J. Groß, and S. Günnemann, "Directional message passing for molecular graphs," in *ICLR*, 2020.
[25]J. Gasteiger, S. Giri, J. T. Margraf, and S. Günnemann, "Fast and uncertainty-aware directional message passing for non-equilibrium molecules," in *ML for Molecules workshop, NeurIPS*, 2020.
[26]C. W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky, and J. P. Mailoa, "Accurate and scalable multi-element graph neural network force field and molecular dynamics with direct force architecture," 2020.
[27]O. T. Unke and M. Meuwly, "Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges," *J. Chem. Theory Comput.*, vol. 15, no. 6, pp. 3678–3693, 2019.
[28]K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "Schnet: A continuous-filter convolutional neural network for modeling quantum interactions," in *NeurIPS*, 2017.
[29]K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nat. Commun.*, vol. 8, no. 1, pp. 13890–13890, 2017.
[30]K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, "SchNetPack: A Deep Learning Toolbox For Atomistic Systems," *J. Chem. Theory Comput.*, vol. 15, no. 1, pp. 448–455, 2019.
[31]R. Zubatyuk, J. S. Smith, J. Leszczynski, and O. Isayev, "Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network," *Sci. Adv.*, vol. 5, p. eaav6490, 2019.
[32]L. Ruddigkeit, R. van Deursen, L. C. Blum, and J.-L. Reymond, "Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17," *J. Chem. Inf. Model.*, vol. 52, no. 11, pp. 2864–2875, 2012. PMID: 23088335.
[33]R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," *Sci. Data*, vol. 1, 2014.
[34]S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," *Sci. Adv.*, vol. 3, no. 5, p. e1603015, 2017.
[35]L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, and Z. Ulissi, "Correction to the open catalyst 2020 (oc20) dataset and community challenges," *ACS Catal.*, vol. 11, no. 21, pp. 13062–13065, 2021.
[36]J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio Jr., and A. Tkatchenko, "Qm7-x, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules," *Sci. Data*, vol. 8, no. 1, p. 43, 2021.
[37]J. P. Perdew, M. Ernzerhof, and K. Burke, "Rationale for mixing exact exchange with density functional approximations," *J. Chem. Phys.*, vol. 105, no. 22, pp. 9982–9985, 1996.
[38]A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, "Accurate and efficient method for many-body van der Waals interactions," *Phys. Rev. Lett.*, vol. 108, no. 23, pp. 1–5, 2012.
[39]V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, "Ab initio molecular simulations with numeric atom-centered orbitals," *Comput. Phys. Commun.*, vol. 180, no. 11, pp. 2175–2196, 2009.
[40]X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, "Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions," *New J. Phys.*, vol. 14, no. 5, pp. 053020–053020, 2012.
[41]S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields," *Nat. Commun.*, vol. 9, no. 1, pp. 3887–3887, 2018.
[42]A. S. Christensen and O. A. von Lilienfeld, "On the role of gradients for machine learning of molecular energies and forces," *Mach. Learn. Sci. Technol.*, vol. 1, p. 045018, 2020.
[43]R. Petraglia, A. Nicolaï, M. D. Wodrich, M. Ceriotti, and C. Corminboeuf, "Beyond static structures: Putting forth REMD as a tool to solve problems in computational organic chemistry," *J. Comput. Chem.*, vol. 37, no. 1, pp. 83–92, 2016.
[44]T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, "A Fourth-Generation High-Dimensional Neural Network Potential with Accurate Electrostatics Including Non-local Charge Transfer," *Nat. Commun.*, vol. 12, p. 398, 2021.
[45]C. Staacke, S. Wengert, C. Kunkel, G. Csányi, Reuter, K, and J. Margraf, "Kernel Charge Equilibration: Efficient and Accurate Prediction of Molecular Dipole Moments with a Machine-Learning Enhanced Electron Density Model," *Mach. Learn. Sci. Technol.*, vol. 3, p. 015032, 2022.
[46]C. Staacke, H. Heenen, C. Scheurer, G. Csányi, K. Reuter, and J. Margraf, "On the role of long-range electrostatics in machine-learned interatomic potentials for complex battery materials," *ACS Appl. Energy Mater.*, vol. 4, no. 11, pp. 12562–12569, 2021.

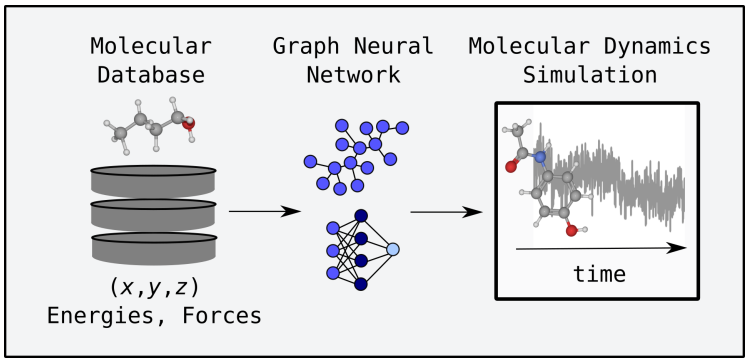

Table of Content Graphic