

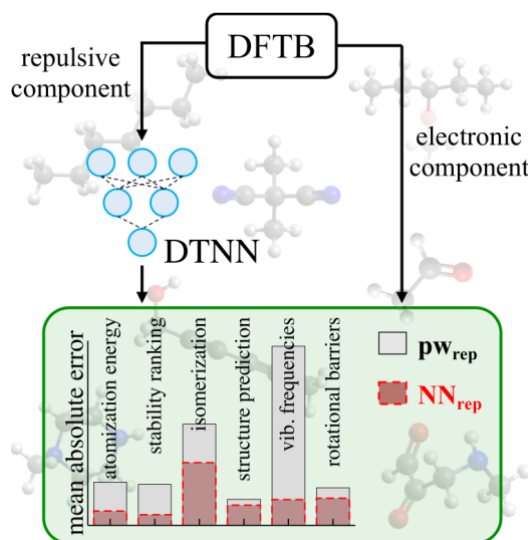
Synergy between physics and machine learning for property prediction of organic systems

Dr. Leonardo Medrano Sandonas

Institute for Materials Science and Max Bergmann Center of Biomaterials, TU Dresden, 01062 Dresden, Germany. Email: leonardo.medrano@tu-dresden.de

Abstract

Machine learning has been proven to be an extremely valuable tool for simulations with ab-initio accuracy at the computational cost between classical interatomic potentials and density-functional approximations. Similar efficiency can only be achieved by semi-empirical (SE) methods such as density-functional tight-binding (DFTB). However, shortcomings still exist in the pairwise DFTB repulsive component and the treatment of long-range (e.g., electrostatics and van der Waals) interactions in non-covalent systems. Therefore, building on our previous work (DFTB+NN_{rep}) [1], we have developed a scalable methodology that corrects the DFTB repulsive potential to a many-body potential via the use of an equivariant neural network (NN), which considers local and non-local physical interactions [2]. Moreover, a many-body dispersion treatment is applied to describe van der Waals interactions, which are crucial to investigate large/more flexible molecules and molecular dimers. Our many-body NNrep potential is trained to fit the PBE0-level data for single molecules from the QM7-X dataset [3] and the resultant model is tested rigorously. Firstly, DFTB+NN_{rep} shows an improvement in capturing intramolecular interactions as illustrated by the prediction of rotational energy profiles for organic molecules of increased size and flexibility compared to the training set. Furthermore, despite not training on non-covalent systems, our model predicts accurately the interaction energy of the molecular dimers from the s66x8 dataset as well as that of large molecular clusters extracted from the X23 molecular crystals dataset. Hence, our ML-corrected DFTB approach combines scalability and generalisability with improved accuracy. Thus, we conclude that finding an optimal synergy between SE and NN methods is key to the development of reliable models for the computation of physicochemical properties of diverse molecular systems.



Scheme our $DFTB+NN_{rep}$ approach [1] which considers the electronic information from DFTB+ code and the repulsive potential from a neural network model.

References

- [1] M. Stöhr, **L. Medrano Sandonas**, A. Tkatchenko, *J. Phys. Chem. Lett.* **11**, 6835, (2020).
- [2] O. Unke *et al.*, (2024). Accepted in Sci. Adv., arXiv:2205.08306.
- [3] **L. Medrano Sandonas**, J. Hoja, B. Ernst, Á. Vázquez-Mayagoitia, R. DiStasio, A. Tkatchenko, *Chem. Sci.*, 14, 10702-10717, (2023).