

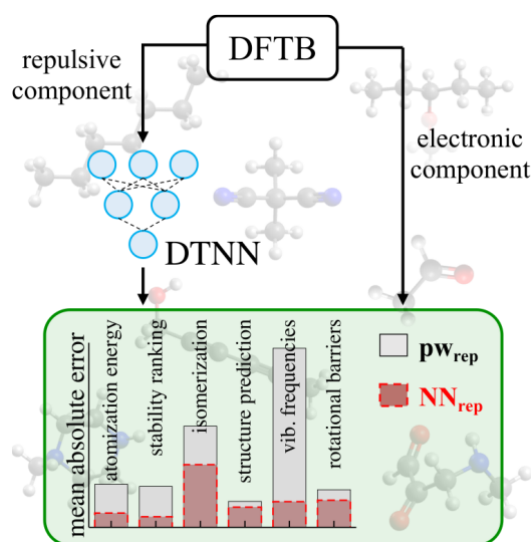
# Synergy between physics and machine learning for property prediction of organic molecules and materials

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<sub>rep</sub>) [1], we have developed a scalable methodology that corrects the DFTB overall performance via the development of many-body  $\Delta_3$  interatomic potentials using physically-inspired state-of-the-art equivariant neural networks (ENN) such as SpookyNet and MACE [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  $\Delta_3$  potentials are trained to fit the PBE0-level data for single molecules from the QM7-X dataset [3] and molecular dimers from the S66x8 dataset. Firstly, ML-corrected DFTB models show an improvement in capturing intramolecular interactions as illustrated by the prediction of rotational energy profiles and the conformational landscape for organic molecules of increased size and flexibility compared to the training set. Furthermore, depending on the physical model used in the ENN, the resulting ML model can accurately predict the dissociation curves of molecular dimers as well as the interaction energy of large molecular clusters extracted from the X23 molecular crystals dataset. Hence, our new ML-corrected DFTB models combine scalability and generalisability with improved accuracy, enabling the efficient investigation of the physicochemical properties of large covalent and non-covalent molecular complexes.



Scheme our ML- corrected DFTB approach [1] which considers the electronic information from DFTB+ code and the many-body 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.*, *Sic. Adv.* 10, eadn4397, (2024).
- [3] **L. Medrano Sandonas**, J. Hoja, B. Ernst, Á. Vázquez-Mayagoitia, R. DiStasio, A. Tkatchenko, *Chem. Sci.*, 14, 10702-10717, (2023).