Advancing machine learning for organic material simulations with quantum accuracy

Leonardo Medrano Sandonas[1], Macarena Tolmos[2], Mirela Puleva[3], Alexandre Tkatchenko[3], and Gianaurelio Cuniberti[1]

[1]Institute for Materials Science and Max Bergmann Center of Biomaterials, TUD Dresden University of Technology, 01062 Dresden, Germany

[2]Escuela Profesional de Física, Facultad de Ciencias, Universidad Nacional de Ingeniería, Lima, Peru

[3]Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg, Luxembourg

leonardo.medrano@tu-dresden.de

Abstract

The rising demand for sustainable solutions to technological and societal challenges has driven significant research and development efforts to integrate machine learning (ML) techniques in computational physics and chemistry. As ML becomes more prevalent in interdisciplinary research, the amount of comprehensive quantum-mechanical (QM) property data generated in recent years to train robust predictive models has significantly increased. Recently, we introduced high-fidelity property data at the level of non-empirical hybrid density-functional theory (DFT) with a many-body treatment of vdW dispersion interactions (i.e., PBE0+MBD) for both small [1] and large [2] drug-like molecules in equilibrium and non-equilibrium states. These datasets have been instrumental in advancing QM-based ML interatomic potentials (e.g., SO3LR [3]) and enhancing semi-empirical methods (e.g., third-order density functional tight-binding DFTB3 [4]), enabling accurate (bio)molecular simulations. In this presentation, we will discuss our recent efforts to improve the transferability and generalizability of the ML-corrected DFTB3 method, see Figure 1. Within the DFTB method, the pairwise repulsive component has certain shortcomings, which we will address by training a many-body repulsive potential using neural networks (NNs). Indeed, we have demonstrated that equivariant NNs (e.g., SpookyNet and MACE) significantly enhance the accuracy and scalability of ML-based many-body repulsive potentials trained on energies and forces of small organic systems and molecular dimers. The developed framework, namely EN4TB, facilitates the calculation of the energetic and structural properties of large drug-like molecules and molecular dimers at a higher level of theory such as PBE0+MBD. Preliminary results are shown in Figure 1. Additionally, we have expanded this approach to investigate the structural and thermodynamic properties of potential candidates for organic electrodes in Li-battery applications [5]. For comparison, our results are compared with these obtained by ML force fields trained on full DFT reference data. Hence, EN4TB highlights the benefits of integrating ML with semi-empirical methods to achieve both high accuracy and computational efficiency, thereby paving the way for diverse applications in organic material simulations. See the EN4TB GitHub repository for examples of how to use our approach.

References

[1] J. Hoja, L Medrano Sandonas et al., Scientific Data 8, 43, (2021).
[2] L. Medrano Sandonas, D. Van Rompaey et al., Scientific Data 11, 742, (2024).
[3] A. Kabylda, J. Thorben Frank et al., ChemRxiv, (2024). DOI: 10.26434/chemrxiv-2024-bdfr0.
[4] M. Stöhr, L. Medrano Sandonas, and A. Tkatchenko, The Journal of Physical Chemistry Letters 11, 6835 (2020).
[5] C. Lu and X. Chen, EcoMat, 4, e12181, (2022).

Figures

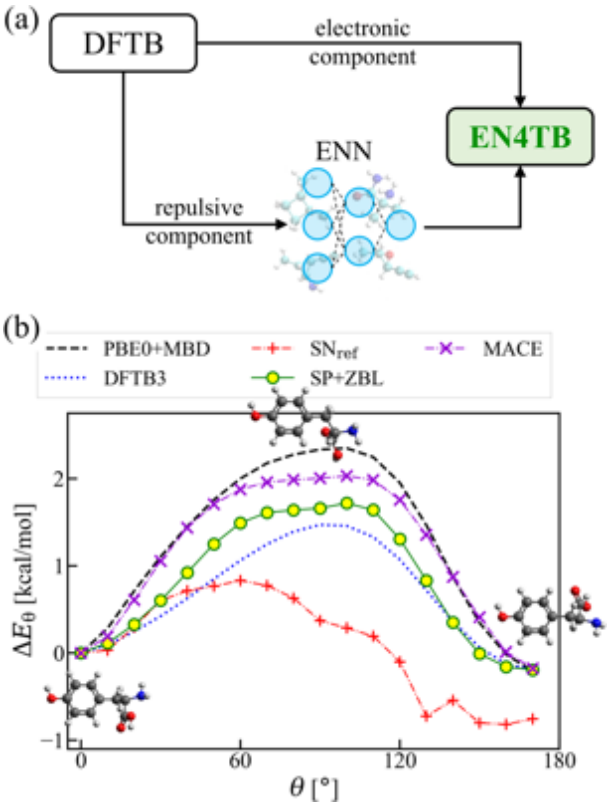

Figure 1. (a) Schematic of the EN4TB framework incorporating the ML-based many-body DFTB repulsive potentials. (b) Results of the predicted reaction path of tyrosine computed using nudged elastic band calculations with equivariant many-body repulsive potentials SP+ZBL (SpookyNet) and MACE. SN_{ref} is the potential trained in our previous work using SchNet architecture [4]. For comparison, PBE0 (reference method) and DFTB3 results are included. All calculations considered many-body dispersion correction.