Enhancing the Capability of Molecular Dynamics with Machine Learning

Felipe Vaiano Calderan[1], Rafael Besse[2], Otaviano da C. Neto[3], João Paulo A. de Mendonça[2], Tuanan C. Lourenço[2], Marcos G. Quiles[1], and Juarez L. F. Da Silva[2]

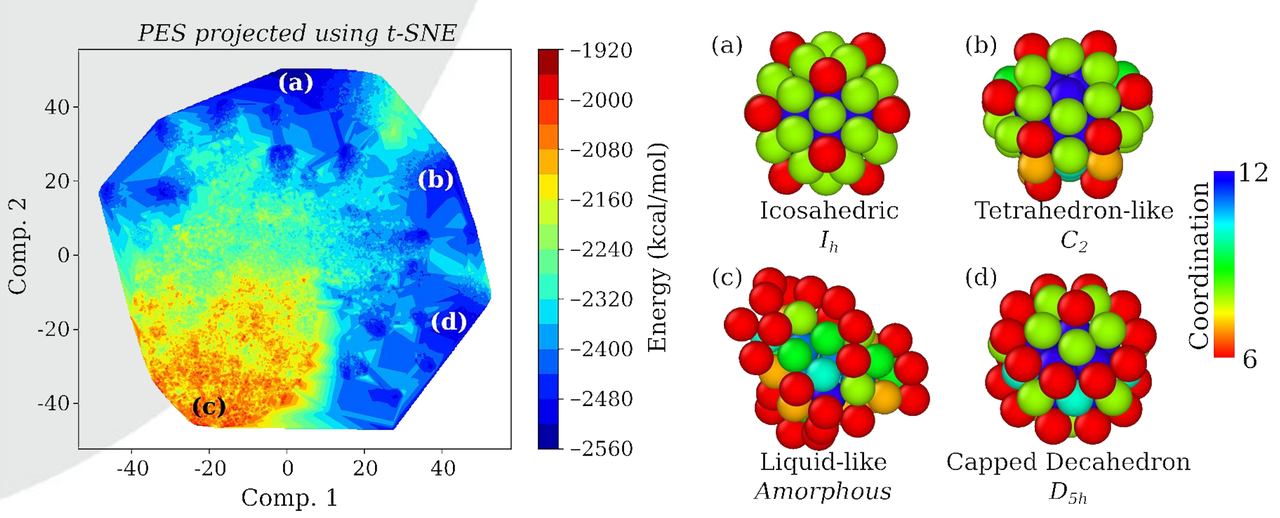
[1]Institute of Science and Technology, Federal University of São Paulo, [2]São Carlos Institute of Chemistry, University of São Paulo, [3]National Institute for Space Research

Abstract

Molecular dynamics (MD) is a powerful framework employed in computational materials science. We can go beyond traditional analyses by combining it with machine learning (ML) methods, such as clustering and dimensionality reduction. For instance, we obtained insights into nanoparticle potential energy surface (PES) and configuration space to characterize the temperature-morphology relationship and phase transitions. We also can improve the accuracy of MD simulations to near quantum accuracy avoiding the high cost of ab initio MD, creating a direct relation between atomic coordinates and energies with high-dimensional neural networks, where data to represent the PES is obtained with density functional theory. In this context, active selection of training data is fundamental to building a more significant data set aiming at better performance, and thus we explore active learning in the training of the ML force fields.

ML for Analysis of MD Data

Nanocluster processes occur under a variety of experimental conditions, therefore we provide a theoretical framework to explore the **potential energy surface** and configuration space of nanoclusters to map the most important **morphologies** presented and **phase transitions** between them.



t-SNE reveals the basins present in the MD trajectories from ReaxFF. These basins are automatically identified through **DBSCAN** clustering. As the systems heat up, the structures transition from basin to basin until they fall into the amorphous zone. Sampling from these basins reveals different morphologies.

Validating this theoretical framework for the test case of several **Cu_n nanoclusters**, we concluded that many of our observations, such as the transition temperatures, number of basins, and complexity of the interpolated PES, show that the Cu_n nanoclusters with n = 13, 55, 147 have special configuration spaces, which agrees with the magic numbers for icosahedral nanoclusters.

Machine-Learning Force Fields

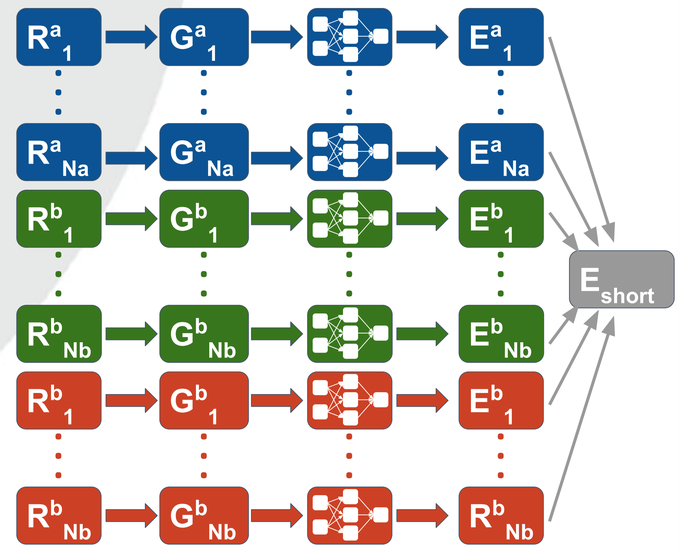
Force fields: description of interatomic interactions

$$F_{\alpha,i} = \frac{\partial E(\{\mathbf{R}_\alpha\})}{\partial R_{\alpha,i}}$$

MLFFs: numerical fitting of $E(\{\mathbf{R}_\alpha\})$ from high-quality ab initio data, aiming to achieve MD simulations with **quantum chemical accuracy and lower computational cost.**

Behler-Parrinello High Dimensional Neural Network Potentials:

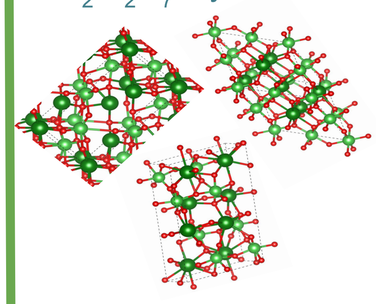
- Atomic neural networks;
- Finite atomic environments;
- Atom-centered symmetry functions for structural fingerprints.



$$E_i = f_a^2\left[w_{01}^2 + \sum_{j=1}^{3} w_{j1}^2 f_a^1\left(w_{0j}^1 + \sum_{\mu=1}^{2} w_{\mu j}^1 G_i^\mu\right)\right].$$

Mixed oxides $La_2B_2O_7$ (B = Ti, Zr, Ce etc); catalysis for clean energy and environment protection, e.g. methane dry-reforming, oxidative coupling of methane.

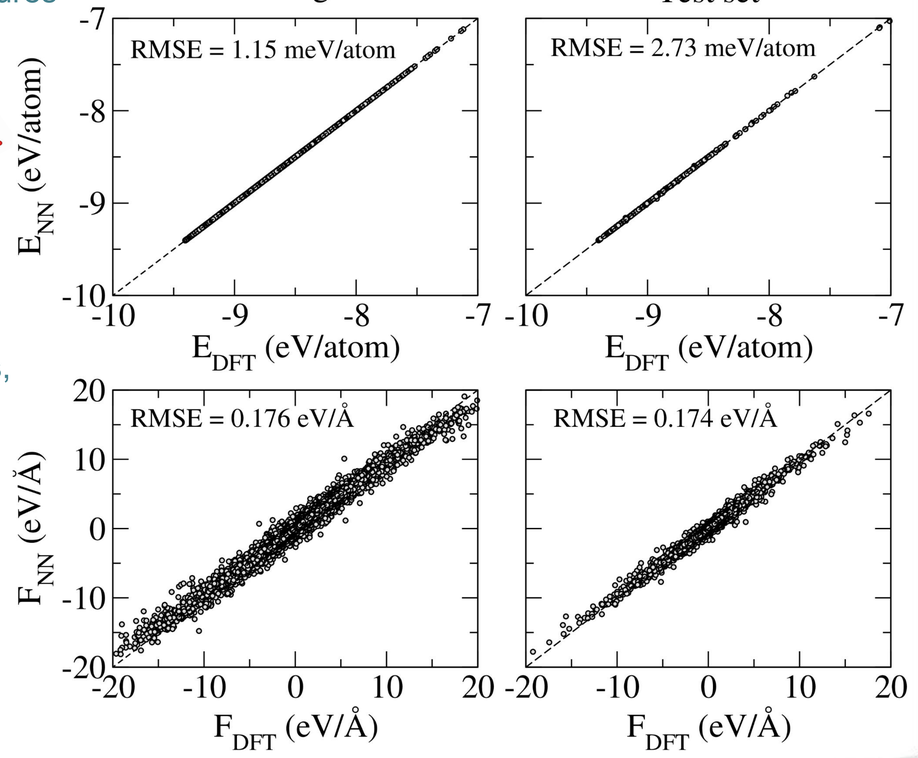
$La_2Zr_2O_7$ crystal structures



DFT → Energies, Forces

Data set

7398 structures
204248 atomic environments
614744 force components



Autoencoder as active learning query

As a way to optimise the NNP training process, we introduce new methods to estimate representativeness and informativeness using autoencoder reconstructions loss. This process can be presented as follows :

Conclusions and Perspectives

For the analysis of MD data, we report an automated procedure to explore the potential energy surface of finite-size particles, which can be applied to small and large particles. An HDNNP parametrization for solid $La_2Zr_2O_7$ has been obtained, which will be improved by optimizing the chemical space representation and validated for application in the study of thermodynamic properties. The optimization of chemical space representation will be enhanced by employing active learning via autoencoder.

Acknowledgements

CINE CENTER FOR INNOVATION ON NEW ENERGIES
FAPESP
USP
IQSC
UNIFESP
INPE