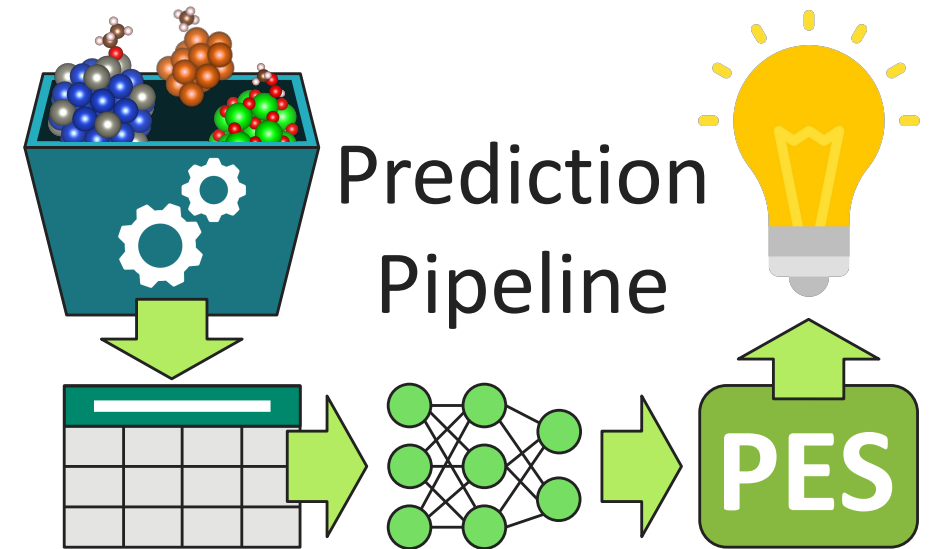
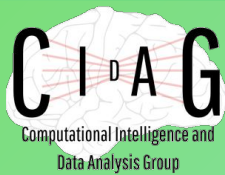


Machine Learning for Potential Energy Surface Prediction

Felipe V. Calderan
Karla F. Andriani
Priscilla Felício-Sousa
Gabriel A. Pinheiro
Juarez L. F. Da Silva
Marcos G. Quiles



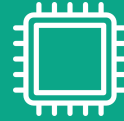
Acknowledgements:



This research tackles 3 problems in Computational Materials Science:



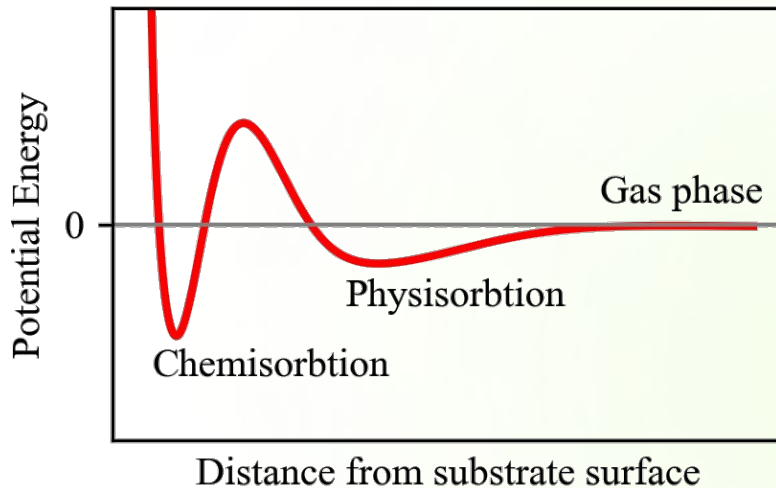
PES screening, a property that must fall within specific ranges for ideal adsorptions



Reduce number of DFT calculations by using ML to preselect likely candidates

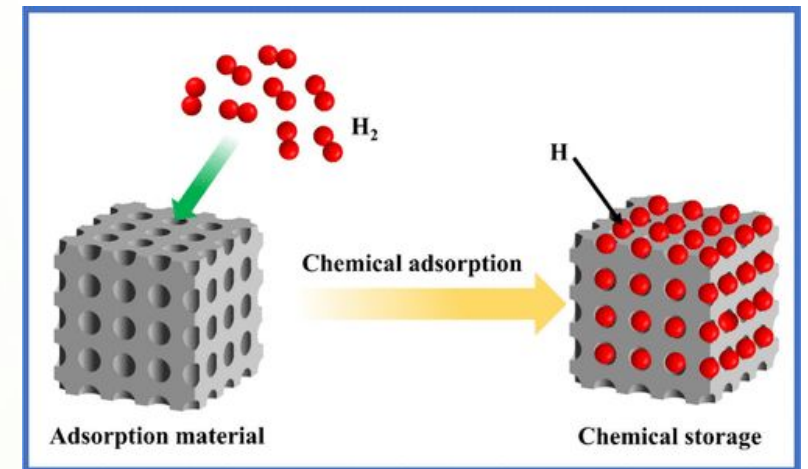


Reuse of previous datasets, not tailored for ML, to train effective predictors



Adsorption applications:

- Water & air purification
- Energy storage
- Environmental recovery
- Chromatography
- General catalysis

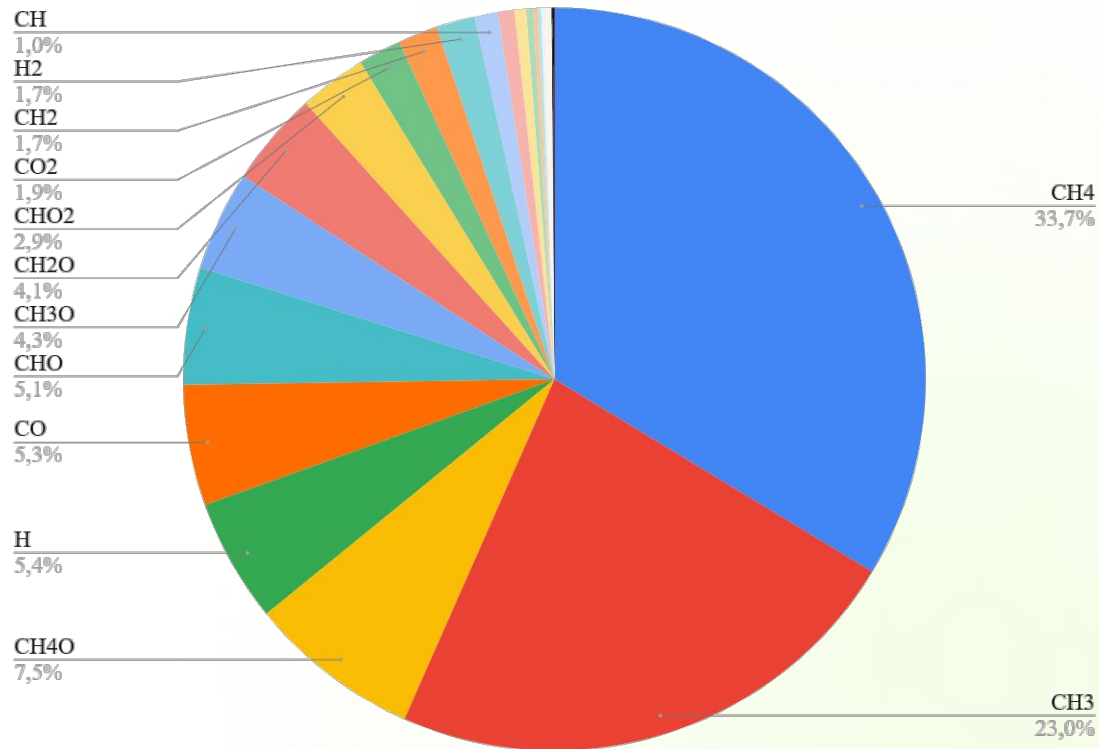


Fang, Weijie, et al. "Review of hydrogen storage technologies and the crucial role of environmentally friendly carriers." *Energy & Fuels* 38.15 (2024): 13539-13564.

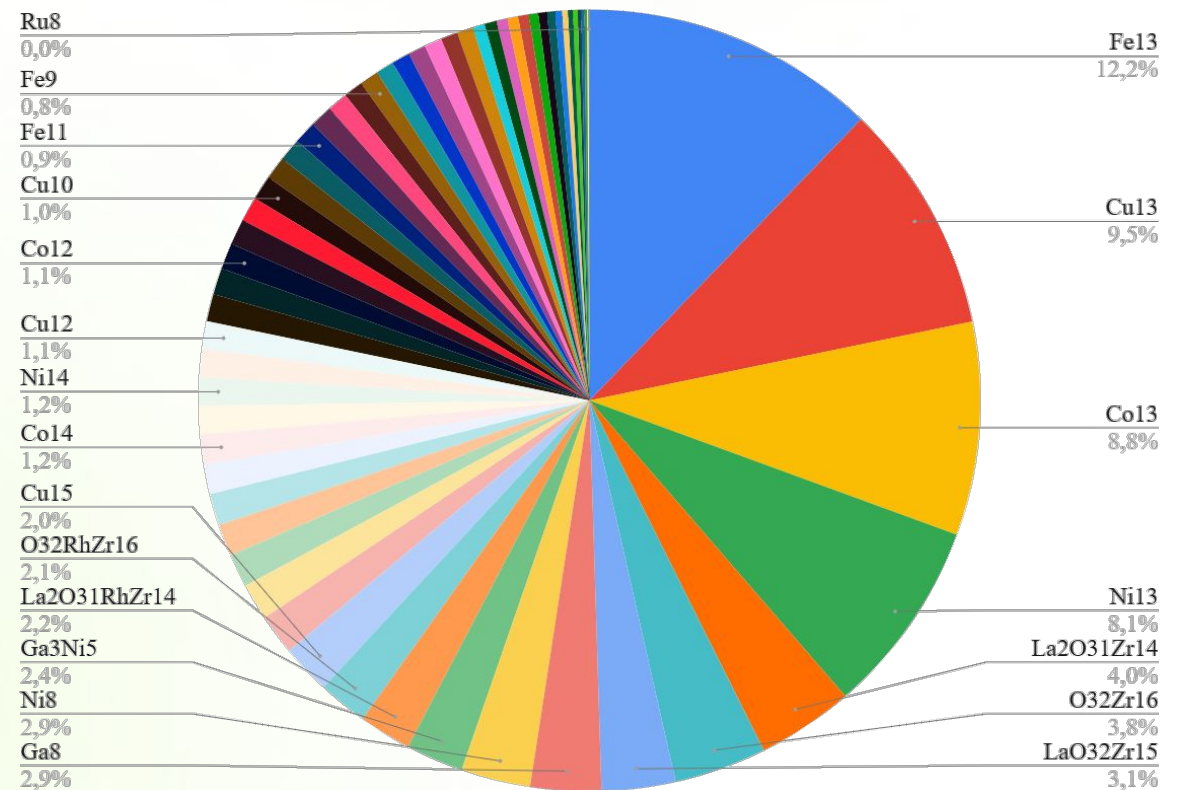
Summary of Quantum-Chemistry QTNano Dataset

Generated from prior QTNano projects, the dataset reflects real-world scenarios (not tailored for ML), making it unbalanced and diverse, with **430,000+** systems.

Adsorbates

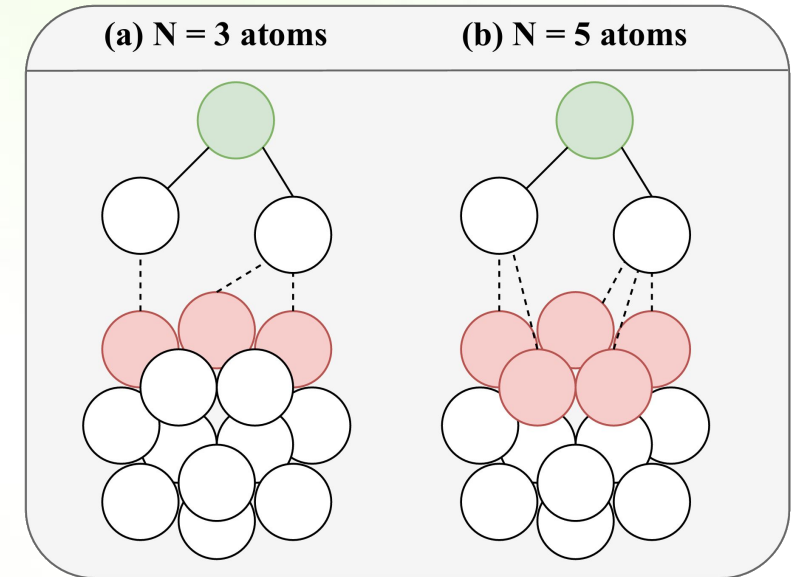


Adsorbents



Efficient chemical descriptor for adsorption systems

- XYZ lacks invariance in respect to position and rotation, making them unsuitable as features.
- Global descriptors like SOAP solve this, but are **costly** and may fail to represent **local regions of interest**.
- Cut-SOAP (our development) greatly **reduces feature count** by focusing on a **hypothetical adsorption site**.



From raw data to final feature count:



Model's predictions for newly generated systems

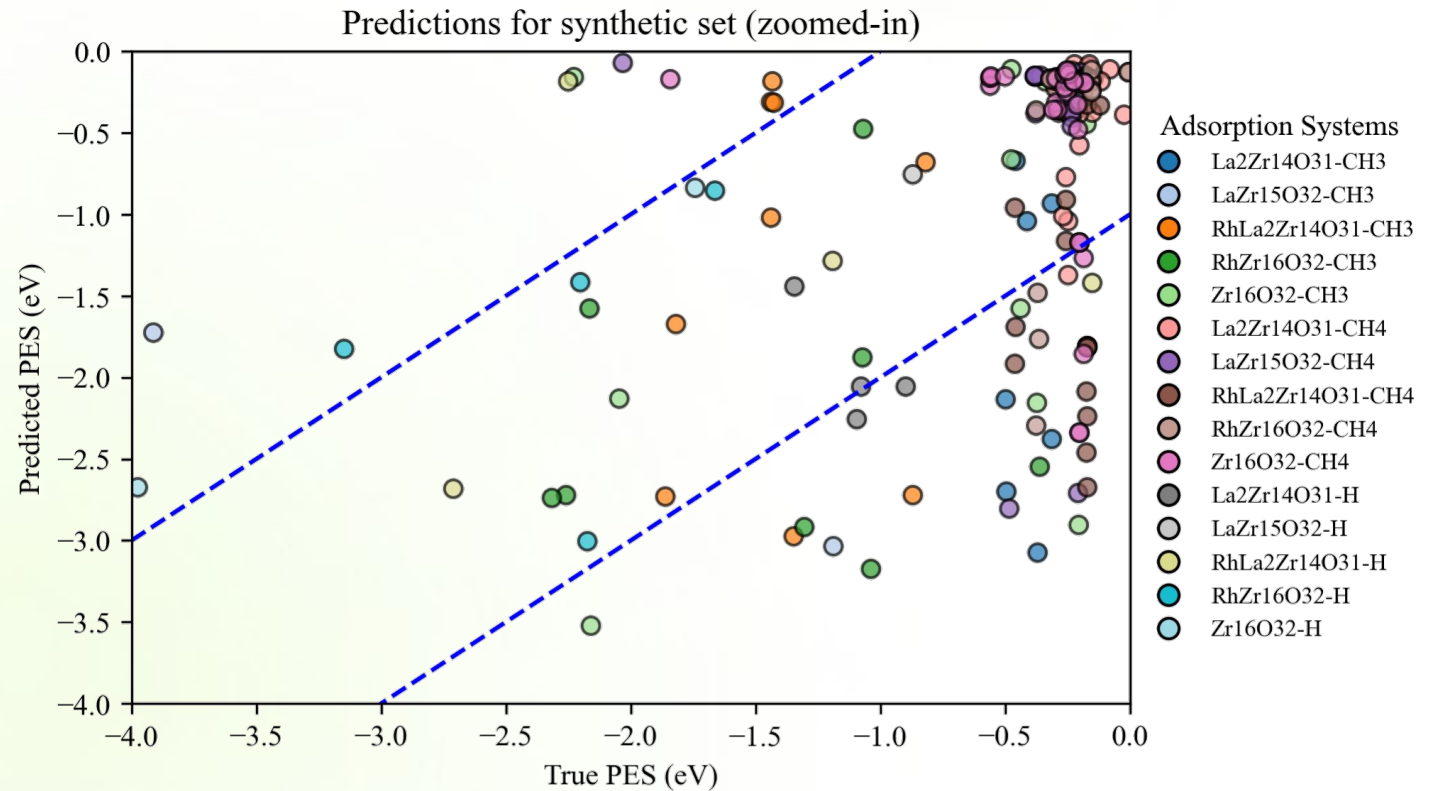
We trained an MLP on Cut-SOAP features and tested it on both a **held-out test set** and a **new synthetic dataset** of 222 systems. The latter simulates practical applications with varied orientations but familiar adsorbate-adsorbent distances.

MAE for the held-out set:
0.062 eV

MAE for the synthetic set:
0.89 eV

Time to train the network:
6:46:35

Time to predict synthetic set:
0.28 s



Pipeline overview: from training to PES prediction

The whole pipeline can be summarized as follows:

Training

Format starting data

Load FHI-Aims data, then fit a descriptor and build features. Concatenate with PES and output train/test datasets.

Train ML model

Use the formatted dataset to train an ML PES predictor with a given topology and hyperparameters.

Production

Format new data

Load systems to have their PES predicted, then generate feature vectors with the previously fitted descriptor.

Predict PES

After all the new data was converted into a table, use the trained model to predict their PES.

- It's possible to train effective ML predictors using historical datasets;
- Cut-SOAP's accuracy is close to regular SOAP's while reducing cost;
- Trained model can predict PES for systems mildly different from starting set;
- Predictions are superfast, allowing large scale operations;
- Framework has a straightforward console interface;
- We plan to submit this work and release the full framework in upcoming months.

Some icons are from flaticon.com

Thank You!

Some icons are from flaticon.com