SAND2022-9754C

Sandia National Laboratories

Accurate Hellmann-Feynman forces with optimized atom-centered Gaussian basis sets

Shivesh Pathak, Joshua Rackers, Ignacio Fdez. Galván, Rafael Lára Fernández, Alex J. Lee, William P. Bricker, Susi Lehtola

Machine Learned First Principles Forces (1)

- Increasing demand of accurate forces for very large systems: Proteins, Twisted Bilayer graphene
- Machine learning models can train on forces for *small pieces* and predict accurate forces on *large systems*
  - AGNI (Botu *et al.* J. Phys. Chem 2017)

- **Hang-up:** ML models for force require expensive first principles force calculations for training data

Hellmann-Feynman Theorem (2)

$$\vec{F}_I^E = \frac{-\partial\langle\Psi(\{\vec{R}\})|\hat{H}|\Psi(\{\vec{R}\})\rangle}{\partial\overrightarrow{R_I}} = \vec{F}_I^{HF} + \vec{F}_I^{Pulay}$$

$$\vec{F}_I^{HF} = -\langle\Psi(\{\vec{R}\})|\frac{\partial\hat{H}}{\partial\vec{R}_I}|\Psi(\{\vec{R}\})\rangle$$ **Depends only on ρ**

$$\vec{F}_I^{Pulay} = -\langle\Psi(\{\vec{R}\})|\hat{H}|\frac{\partial\Psi(\{\vec{R}\})}{\partial\overrightarrow{R_I}}\rangle$$

Solution: If we can reduce Pulay term, only need to train **ML density model** for forces!

Optimized Hellmann-Feynman Basis Sets: σNZHF (3)

- Rico *et al.* J Comput. Chem. (2007) for Slater-type orbitals
- For a single slater determinant wavefunction, a Gaussian basis that can reproduce its own nuclear gradients has small Pulay forces

Technique:

1. Begin with a starting basis set, which we will improve: σNZ ; N = S, D, T (Ema *et al.* 2022, ArXiV:2207.02782)
2. Compute the distance Δ_λ between the basis set space and nuclear derivative space
3. If $\Delta_\lambda < \epsilon$, a chosen value ($\epsilon = 10^{-3}$), stop
4. Else, add additional primitives to the starting basis set, return to step 2

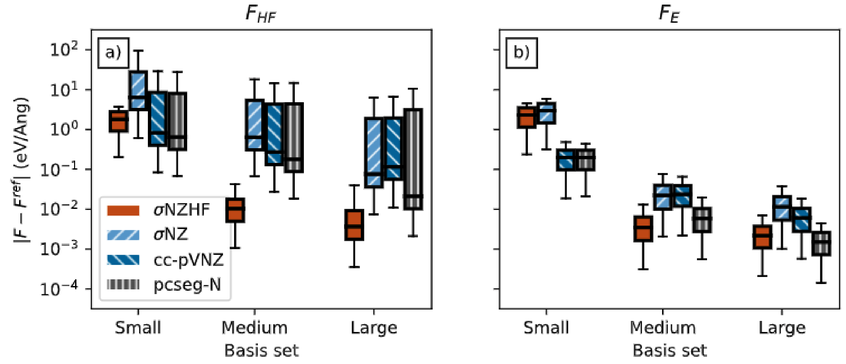
We built bases for C, N, O, F, P, S, Cl with three different composition: $\boldsymbol{\sigma SZHF, \sigma DZHF, \sigma TZHF}$

arXiv:2207.03587, Submitted to JCTC
$\sigma NZHF$ bases on Basis Set Exchange

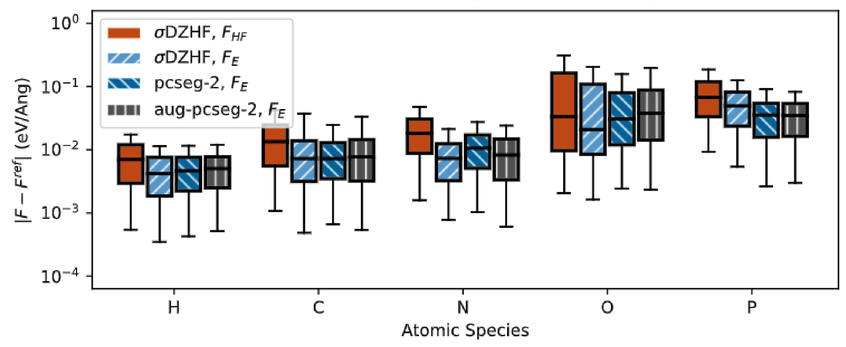
Accurate Forces on H2O Clusters and DNA Fragments (4)

- Compare analytic and HF gradients using DFT PBE0, σNZHF, (aug)-cc-pvNZ and (aug)-pcseg-N bases
- Configurations sampled from room temperature MD
  H2O: 10 molecule clusters, 50 configurations
  DNA: 35 atom fragments, 60 structures

H2O (Reference: aug-cc-pv5z, F_E)



DNA (Reference: aug-pcseg-3, F_E)



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology &Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. _____