Assessing Conformer Energies:
Machine Learning vs Conventional Quantum Chemistry

Dakota Folmsbee and Geoffrey Hutchison
Department of Chemistry, University of Pittsburgh

Funding: NSF CHE-1800435

Motivation

Understanding and predicting thermodynamically accessible conformations is a key task as the number of possible minima increase exponentially for each rotatable bond. This becomes expensive for wavefunction and density functional first principles methods leading to a desire of a fast and accurate method, such as machine learning (ML), to be used as a surrogate for these time-consuming calculations.

Methodology

- Conformer dataset consists of ~6000 DLPNO-CCSD(T) single points of ~700 small molecules with up to 10 conformers each containing atoms of H, C, N, O, F, P, S, Cl, and/or Br.
- Evaluated 39 methods including dispersion and non-dispersion corrected DFT and normalized ML methods.

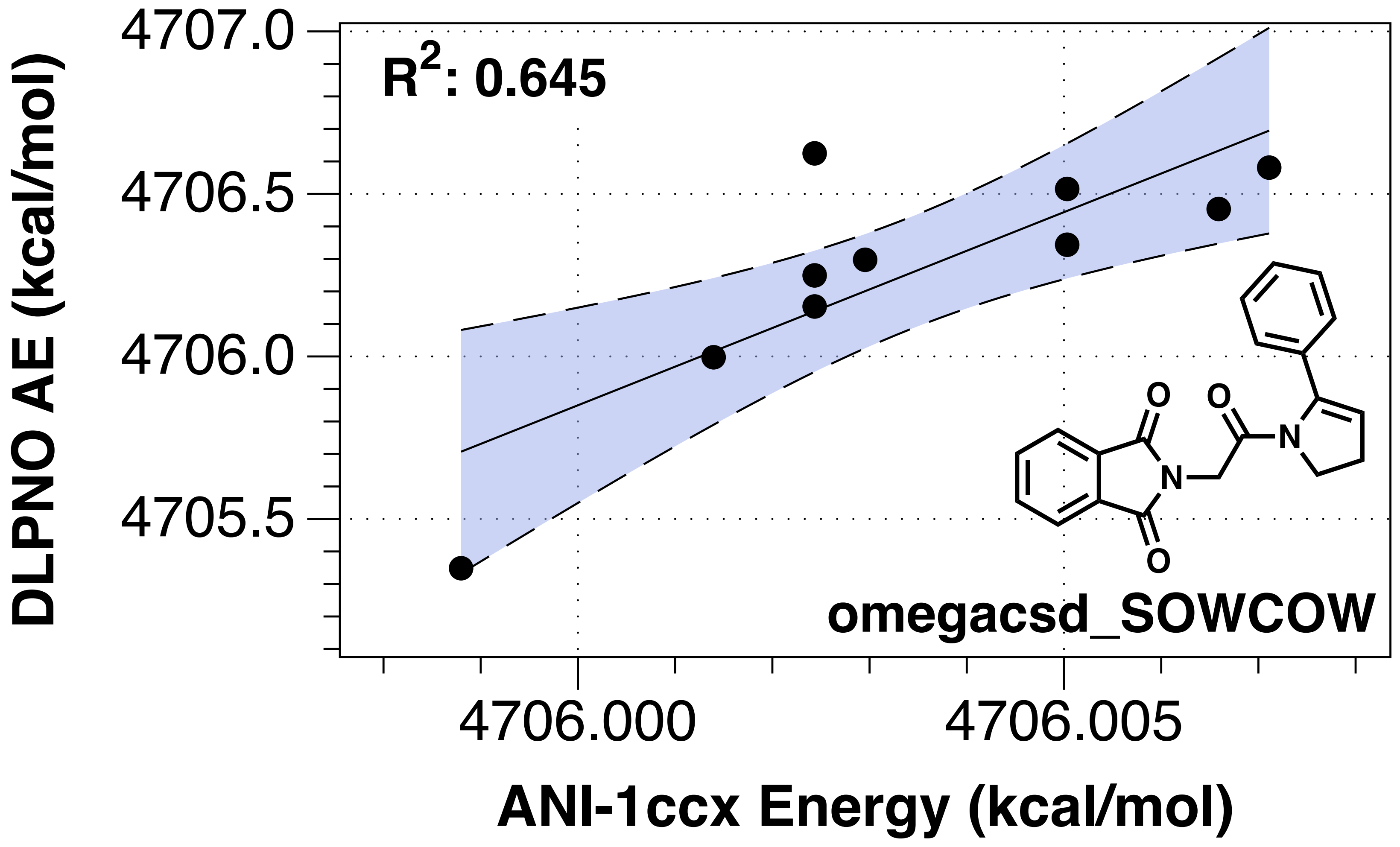
Performance

- RI-MP2 and DFT dispersion methods perform best at the cost of time while ML and semiempirical methods look to be an even ground between accuracy and time.
- Most methods are dominated by random errors as measured with R^2.
- A few force field and ML methods show increased MARE from systematic errors.

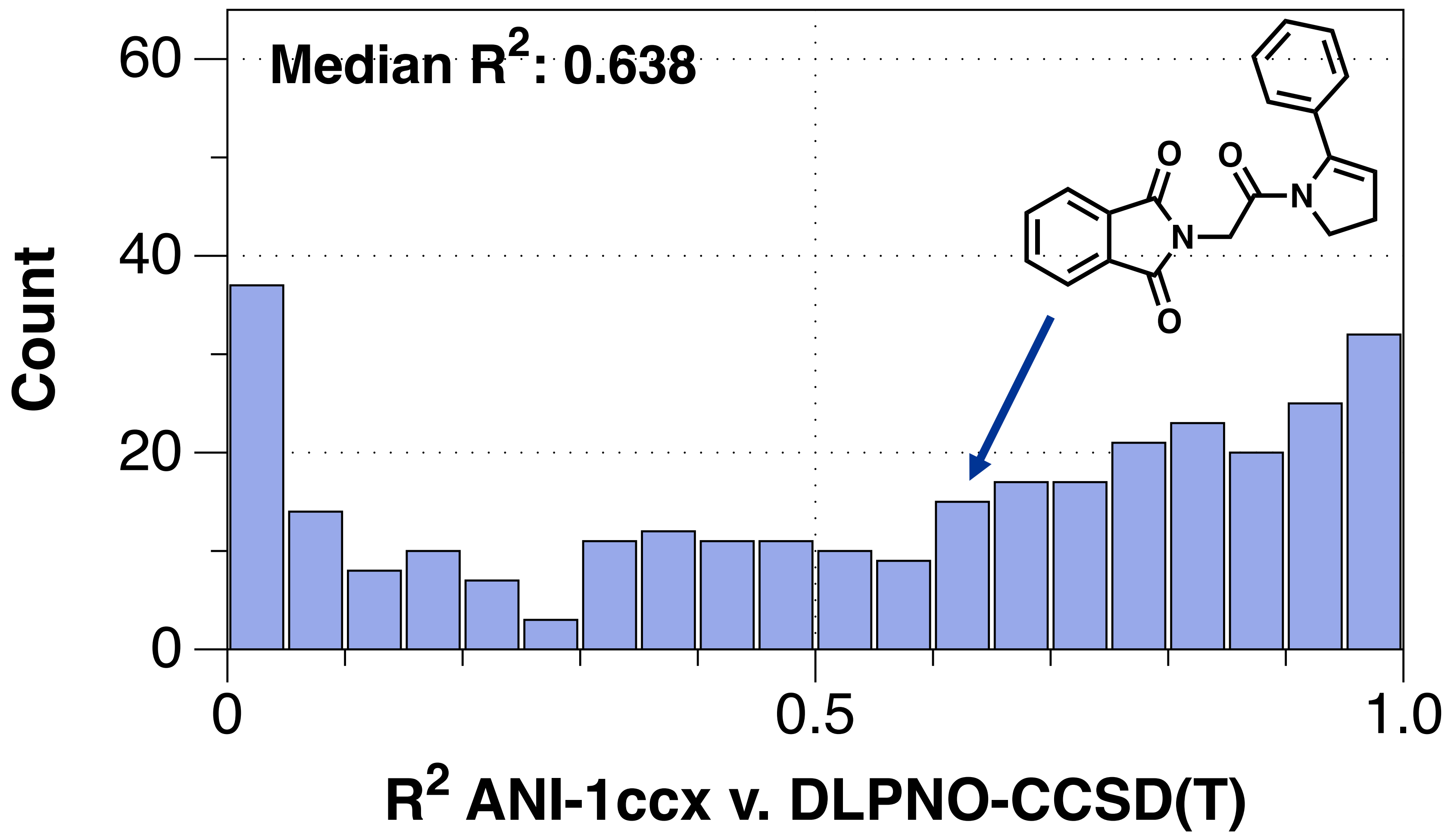
Thanks

This research was supported by the NSF and the University of Pittsburgh Center for Research Computing. We also thank Olexandr Isayev and Justin Smith for access to the ANI-2x model.

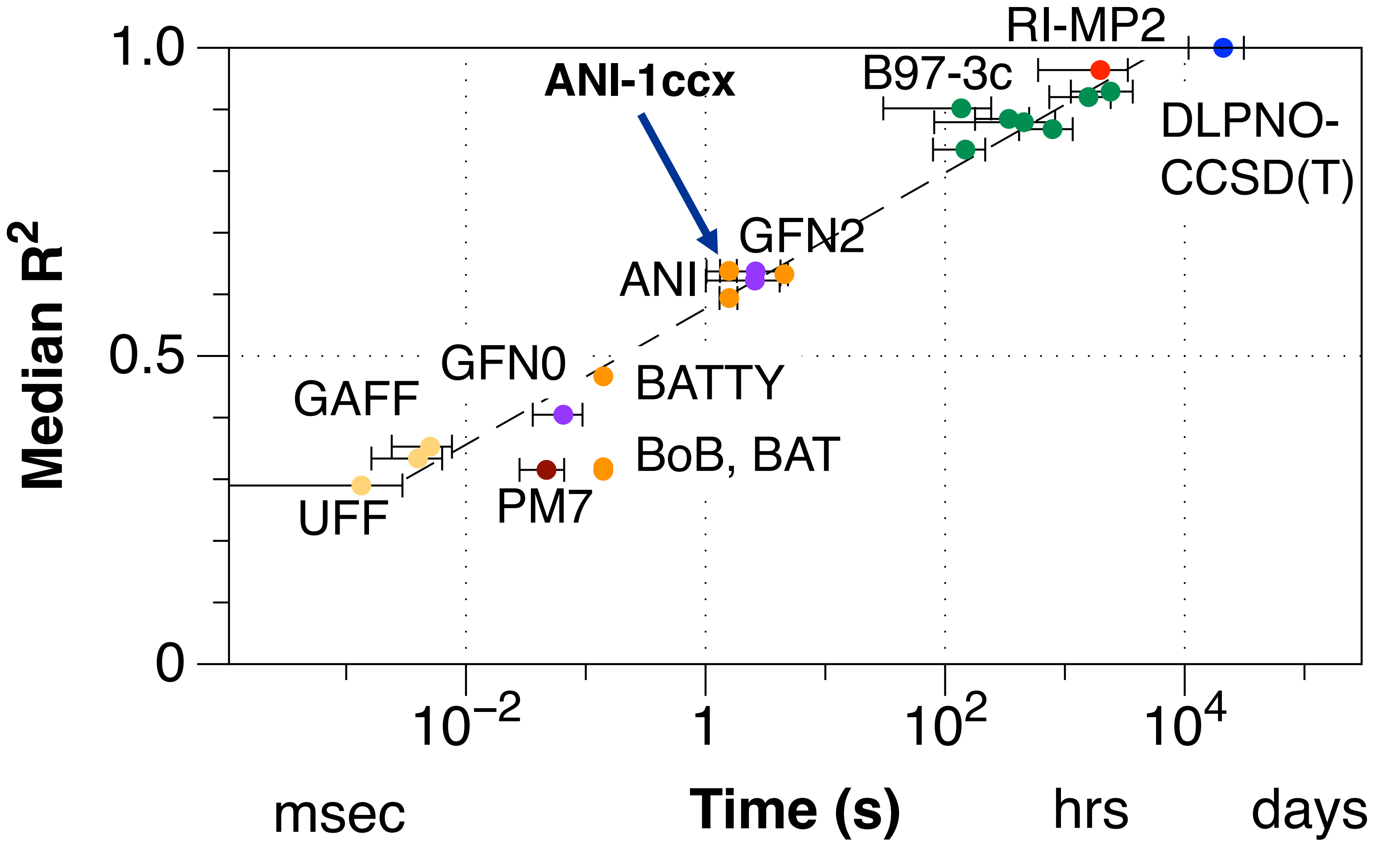
Across 10 Conformers



Across 700 Molecules



Across 39 Methods



*Single core CPU times

ML methods show promise with comparable accuracy to semiempirical methods and can be accelerated on GPUs. We expect these methods to provide increasing accuracy.

University of Pittsburgh

Take a picture to download the full paper