

Olexandr Isayev<sup>1</sup>

<sup>1</sup>Department of Chemistry, Carnegie Mellon University

June 11, 2020

## Abstract

# 1 Referee Report

Paper by Folmsbee and Hutchison is a great example of reproducible benchmark paper and well suited for IJQC special issue. Both codes and data are available on GitHub. This paper compares the accuracy of various computational methods to evaluate single point energies of molecular conformers. Authors used DLPNO-CCSD(T) as a reference level of theory and benchmarked small-molecule force fields, semiempirical, DFT and several emerging machine learning (ML) techniques. This paper provides computational chemists with a substantial body of high accuracy data. Overall this paper could serve a solid practical guideline for applying approximate computational methods to a problem of conformer search.

However, I identified several problems to be addressed before this paper could be accepted for publication. Specific comments (not in a particular order) are below. The beauty of the availability of the code to review allowed me to run and reproduce some of the results of this paper. As one of the developers of ANI ML potential, I naturally tested our method first. Overall I applaud authors for advocating open science and open data.

1. All ANI models were fitted to wB97x DFT functional data \*minus\* D dispersion term. This is done because dispersion is an analytical ad hoc correction. The intention that at the run time dispersion should be added back. D3 could be easily computed with ASED3 code referenced in our GitHub.

2. ANI timings are simply wrong. Therefore the TOC and Figure 4 are misleading. ANI timing is at least 100 times faster. The author's script is re-loading all python dependencies and compiles the neural network model for every conformer. This takes 2.45 out of 2.5 seconds of the run. Even with sequential energy evaluation on a CPU, it should be around 0.05s for the 2x model and probably ~0.025s for 1x/ccx. Additionally, the model could be pre-compiled with JIT and embedded into applications for even faster runs. Overall our code is native GPU and also naturally batch evaluation with multidimensional tensors. Therefore the recommended use is to load all conformers and evaluate them at once. Theoretically, all conformers for the same molecule could be computed for 0.1 seconds. I am happy to share my scripts. Most of them already are in our GitHub repos anyway.

3. Authors write:

*In this work, in order to expand our range of computational methods, we only consider the relative single point energies from the same set of density-functional optimized geometries, comparing multiple current methods to a high-quality coupled cluster baseline.*

I think there is a fundamental flaw of logic here, that ultimately hurts the value of this paper. In practical research settings where conformed sampling is used, there is no access to 3D geometries obtained with

high-level QM methods. Therefore, I think the meaningful comparison would be conformed energies with geometries obtained by respective approximate methods.

4. A comparison between BOB/BAT/BATTY and ANI is also one-sided. BOB models are just molecular scorers, they just give you a number. In contrast, ANI and force fields are true automatic potentials with forces and analytic hessian. We can do geometry minimization, MD, etc.

5. There is also a small pesky bug in the authors' scripts. They use different conversion factors au to kcal/mole in different places, therefore some of the energies are inaccurate to ~0.2 kcal.