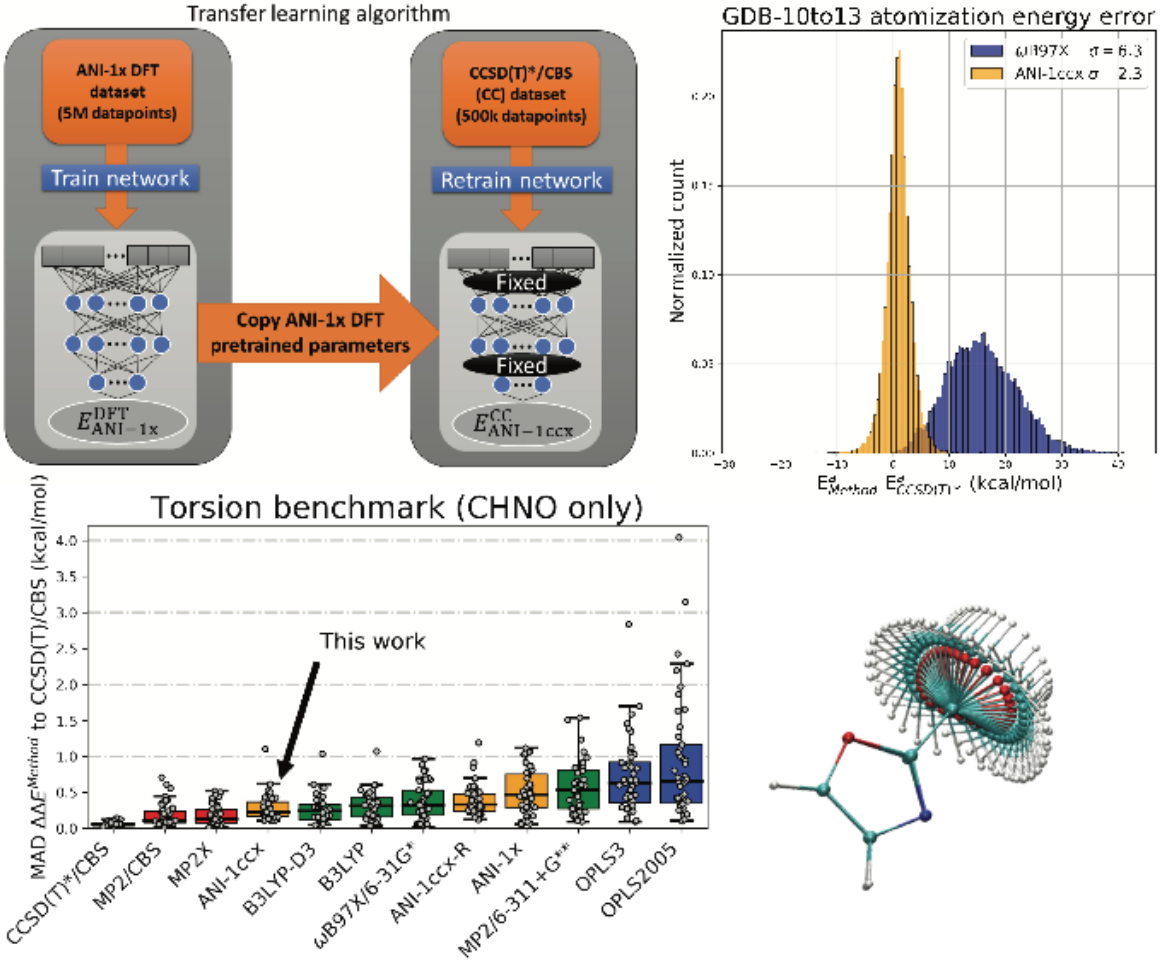
June 2019

## Outsmarting quantum chemistry through transfer learning

This work demonstrates empirical potentials based on Deep Neural Networks can surpass the accuracy of Density Functional Theory (DFT) by using Transfer Learning: A potential is first trained to reproduce a large quantity of DFT data, and then fine-tuned by partially retraining to ultra-high-fidelity Coupled-Cluster calculations. The result is the best-to-date empirical model of small organic molecules – far more accurate than traditional Force Fields, and far faster than ab-initio simulation.




Left-Top: A neural network potential is trained to density functional theory (DFT) data, then retrained to a smaller data set of highly accurate coupled cluster calculations. Right-Top: The transfer learned ANI-1ccx model predicted atomization energies vs. reference coupled cluster

with an error 3x lower than the original DFT method. This test set is built from molecules larger than those included in the training data set, which shows the models extensibility to larger molecules. Bottom: The ANI-1ccx potential out performs state-of-the-art force fields and DFT methods at the prediction of torsion profiles for small druglike molecules.

**The Science**

Los Alamos National Laboratory collaborates with the University of North Carolina at Chapel Hill and University of Florida to build the world's first coupled cluster accurate general-purpose neural network-based atomistic potential for organic molecules. This potential, dubbed ANI-1ccx, can compute energies and forces at many orders of magnitude faster than coupled cluster methods. These ML potentials scale linearly with the number of atoms rather than $O(N^5)$ or higher for coupled cluster methods. These models are developed by training a neural-network to predict density function theory (DFT) energies for 5 million diverse molecules, then retraining the model to a smaller dataset of 500k highly accurate coupled cluster energies. The transfer learning approach shows significantly better accuracy than training to the coupled cluster data alone, without using transfer learning. Three test cases are used to benchmark the resulting ANI-1ccx potential. The first compares atomization energy, absolute energy, and forces from accurate *ab initio* methods. The second compares ANI-1ccx vs. DFT on reaction and isomerization energy benchmarks. The third explores how well the resulting potential performs on a small druglike molecule torsion profile benchmark. In each case, the ANI-1ccx potential outperforms industry standard DFT methods and force fields while maintaining a high level of computational efficiency.

**The Impact**

Empirical potentials are key for enabling computational research into molecules and materials, as *ab-initio* atomistic simulation is often too expensive – especially so for the purposes of screening the immense space of possible chemical compounds. An accurate, transferable, and computationally efficient method for obtaining potential energies and forces will revolutionize many areas of research in chemistry and physics. Making these methods as general as possible expands the range of applicability of the resulting models. ANI-1ccx is accurate, general, and fast enough to be deployed in long time scale simulations or high throughput studies of millions of molecules. The ANI-1ccx potential promises to advance the capabilities of researchers in many fields, including drug development, reactive chemistry, and protein simulation. Further, the techniques developed in this work can be employed to improve the accuracy of machine learning-based potentials in future studies of metal alloys and detonation physics.

**Summary**

Understanding and explaining the atomic level physical phenomena that drive the behavior of bulk materials is paramount in modern science. Existing methods for accomplishing these goals

are either too slow or not accurate enough to provide meaningful insight. Machine learning-based atomistic potentials provide highly *accurate and efficient* predictions of energies and forces. The transfer learning methods introduced in this work provide a path towards fitting general-purpose atomistic potentials with the accuracy of high level coupled cluster calculations. The resulting potential outperforms the industry standard DFT methods and all atom force fields while retaining computational linear scaling in number of atoms.

**LANL contacts**

Justin Smith just@lanl.gov
Benjamin Nebgen bnebgen@lanl.gov
Nick Lubbers nlubbers@lanl.gov
Kipton Barros kbarros@lanl.gov
Sergei Tretiak serg@lanl.gov

**Funding**

J.S.S. thanks the University of Florida for the graduate student fellowship and the Los Alamos National Laboratory (LANL) Center for Non-linear Studies for resources and hospitality. R.Z. acknowledges support by National Science Foundation (NSF) grants 1456638 and 1338192. We gratefully acknowledge the support and hardware donation from NVIDIA Corporation and express our special gratitude to Mark Berger. The authors acknowledge support of the U.S. Department of Energy (DOE) through the LANL LDRD Program. This work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. DOE Office of Science. We also acknowledge the LANL Institutional Computing (IC) program and ACL data team for providing computational resources. O.I. acknowledges support from DOD-ONR (N00014-16-1-2311) and Eshelman Institute for Innovation award. The authors acknowledge Extreme Science and Engineering Discovery Environment (XSEDE) award DMR110088, which is supported by National Science Foundation grant number ACI-1053575. This research in part was done using resources provided by the Open Science Grid which is supported by the National Science Foundation award 1148698, and the U.S. DOE Office of Science. AER thanks NSF CHE-1802831 and OI thanks NSF CHE-1802789.

**Publications**

J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, A. E. Roitberg, "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning" *Nature Communications* **10,** 2903 (2019) **LA-UR-18-25687**