August 29, 2019 | By Kimberly Mann Bruch and Nancy Ambrosiano

Supercomputers Pave the Way for New Machine Learning Approach
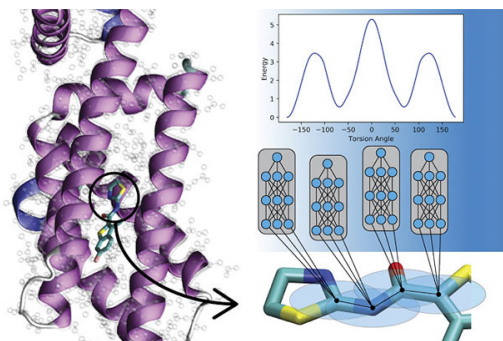
San Diego Supercomputer Center and Los Alamos National Laboratory aid in transfer learning

According to a release issued earlier this month by the Los Alamos National Laboratory (LANL), researchers have developed a machine learning approach called transfer learning that lets them model novel materials by learning from data collected about millions of other compounds. The new approach can be applied to new molecules in milliseconds, enabling research into a far greater number of compounds over much longer timescales.

The new technique, called ANI-1ccx potential, promises to advance the capabilities of researchers in many fields and improve the accuracy of machine learning-based potentials in future studies of metal alloys and detonation physics.

New deep learning models predict the interactions between atoms in organic molecules. These models, which were generated using supercomputers at the San Diego Supercomputer Center and the Los Alamos National Laboratory, help computational biologists and drug development researchers better understand and treat disease. Los Alamos National Laboratory

"Our quantum mechanical calculations to create ANI-1ccx potential were conducted over two years with time split on the *Comet* supercomputer at the San Diego Supercomputer Center and the *Badger* supercomputer at LANL," said Olexandr Isayev, paper author and a pharmacy professor at the University of North Carolina at Chapel Hill. "We chose these two supercomputers to train our neural networks as there are few machines that can run these — due to the high memory and core requirements."

Isayev and colleagues from University of Florida and LANL recently published their research in a *Nature Communications* paper called Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. The paper details how quantum

mechanical (QM) algorithms, used on classical computers, can accurately describe the mechanical motions of a compound in its operational environment.

However, QM scales very poorly with varying molecular sizes, severely limiting the scope of possible simulations. Even a slight increase in molecular size within a simulation can dramatically increase the computational burden. So practitioners often resort to using empirical information, which describes the motion of atoms in terms of classical physics and Newton's Laws, enabling simulations that scale to billions of atoms or millions of chemical compounds.

Traditionally, similar models have had to strike a tradeoff between accuracy and transferability. When the many parameters of the potential are finely tuned for one compound, the accuracy decreases on other compounds.

"This means we can now model materials and molecular dynamics billions of times faster compared to conventional quantum methods, while retaining the same level of accuracy," explained Justin Smith, LANL Physicist and Metropolis Fellow in the laboratory's Theoretical Division. Understanding how molecules move is critical to tapping their potential value for drug development, protein simulations and reactive chemistry, for example, and both quantum mechanics and experimental (empirical) methods feed into the simulations.

MEDIA CONTACT

Kimberly Mann Bruch, 858-822-3622, kbruch@ucsd.edu

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.