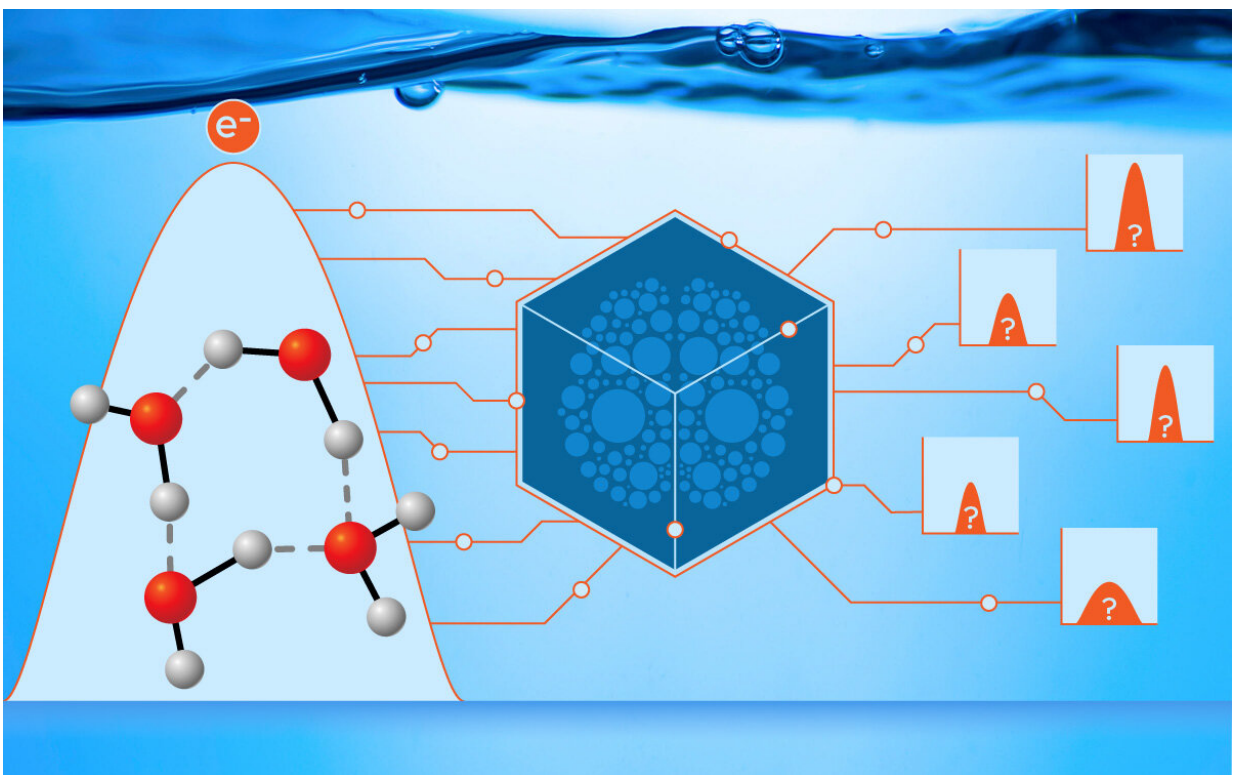


Researchers peer into water clusters database, train network to predict energy landscapes

October 14 2020, by Allan Brettman



PNNL researchers used 500,000 water clusters from a recently developed database of over 5 million water cluster minima to train a neural network relying on the mathematical power of graph theory—a collection of nodes and links representing molecular structure—to decipher structural patterns of the aggregation of water molecules. Credit: Nathan Johnson | PNNL

Machine learning algorithms, the basis of neural networks, are opening doors to new discoveries—or at least offering tantalizing clues—one massive database at a time. Case in point: Pacific Northwest National Laboratory (PNNL) researchers delved deeply into modeling the interactions among water molecules, finding information about hydrogen bonds and structural patterns while plowing a path using—you guessed it—deep learning.

"Neural networks are a way for the computer to automatically learn different properties of systems or data," said PNNL data scientist, Jenna Pope. "In this case, the [neural network](#) learns the energy of different [water](#) cluster networks based on previous data."

PNNL researchers used 500,000 [water clusters](#) from a recently developed database of over 5 million water cluster minima to train a neural network relying on the mathematical power of graph theory—a collection of nodes and links representing molecular structure—to decipher structural patterns of the aggregation of [water molecules](#). Working with the molecular, gaseous form of water, they paid particular attention to the relation between hydrogen bonding and energy relative to the most stable structure.

"That's the holy grail," said Pope. "Right now, it takes a lot of effort to develop an accurate interaction potential for water. But with neural networks, the eventual goal is to have the networks learn, from a large data set, what is the energy of a network based on its underlying [molecular structure](#)."

After sizing up 500,000 water clusters, the neural network in the PNNL-led study was trained in the various ways water molecules interact with each other. The data set theoretically could have included the whole database of 5 million water networks. But for practical reasons it didn't.

"Using the whole database to train the neural network would have taken too much computational time," said Pope. Training the deep neural networks to examine those 500,000 water clusters—just one-tenth of the full database—took more than two and half days using four state-of-the-art computers with sizable graphics processing unit (GPU) performance, working around the clock.

Part of a bigger picture

Neural networks have been around for decades. Greater processing power of GPU chips over the past 10 years, however, has sharply advanced the capability of these networks, also associated with the phrase "deep learning." Even with such promise, training neural networks is an expensive computational challenge. And as accurate as they might be, neural networks are often criticized as black boxes that offer no information about why they are giving the answer they do.

The U.S. Department of Energy's (DOE's) Exascale Computing Project (ECP) was launched in 2016 to explore the most intractable supercomputing problems, including the refinement of neural networks. In 2018, ECP spawned the ExaLearn Co-Design Center, focusing on machine learning technologies. PNNL is among eight national laboratories taking part in the ExaLearn project. James Ang, PNNL's chief scientist for computing in Physical and Computational Sciences, leads the Laboratory's participation.

Database close to home

One of ExaLearn's major goals is to develop artificial intelligence technologies that can design new chemical structures by learning from massive data sets. Research led by Sutanay Choudhury, a PNNL computer scientist, tapped into the massive water clusters database

developed at the PNNL-Richland campus by Sotiris Xantheas, a PNNL Laboratory fellow. Xantheas, known in chemical physics for his research in intermolecular interactions in aqueous ionic clusters, is a co-author on the neural networks study published in the special issue "Machine Learning Meets Chemical Physics" of the *Journal of Chemical Physics*.

"Several macroscopic properties of water have been attributed to its fleeting hydrogen bonding network, which consists of a dynamic network of bonds that break and reform in a fraction of a second at room temperature," said Xantheas, whose database work was supported by DOE's Office of Science, Basic Energy Sciences program, Chemical Sciences, Geosciences, and Biosciences Division. "Water clusters provide a testbed for probing this fleeting hydrogen bonding network by understanding the structure—energy relation of the different hydrogen bonding arrangements."

PNNL's researchers had a strategy to decipher this particular black box. They used [graph theory](#)—a branch of mathematics that studies how things are connected in a network—to represent, in graphic form, molecules and their polygon substructures. The graph-theoretical descriptors the team devised provided several insights into the water clusters' makeup.

"What we have done," said Pope, "is provide additional analysis after the network is trained. The analysis quantifies different structural properties of these water cluster networks. Then we can compare them to our predicted neural network and, in subsequent data set examinations, improve the network's understanding."

Water has a neural network future

The study's findings provide a foundation for analysis of water clusters' structural patterns in more complex hydrogen-bonded networks, such as

liquid water and ice.

"If you were able to train a neural network," said Pope, "that neural network would be able to do computational chemistry on larger systems. And then you could make similar insights in computational chemistry about chemical structure or hydrogen bonding or the molecules' response to temperature changes. Those are among the goals of this research."

More information: Avijit Rakshit et al. Atlas of putative minima and low-lying energy networks of water clusters $n = 3-25$, *The Journal of Chemical Physics* (2019). [DOI: 10.1063/1.5128378](https://doi.org/10.1063/1.5128378)

Provided by Pacific Northwest National Laboratory

Citation: Researchers peer into water clusters database, train network to predict energy landscapes (2020, October 14) retrieved 29 April 2024 from <https://phys.org/news/2020-10-peer-clusters-database-network-energy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.