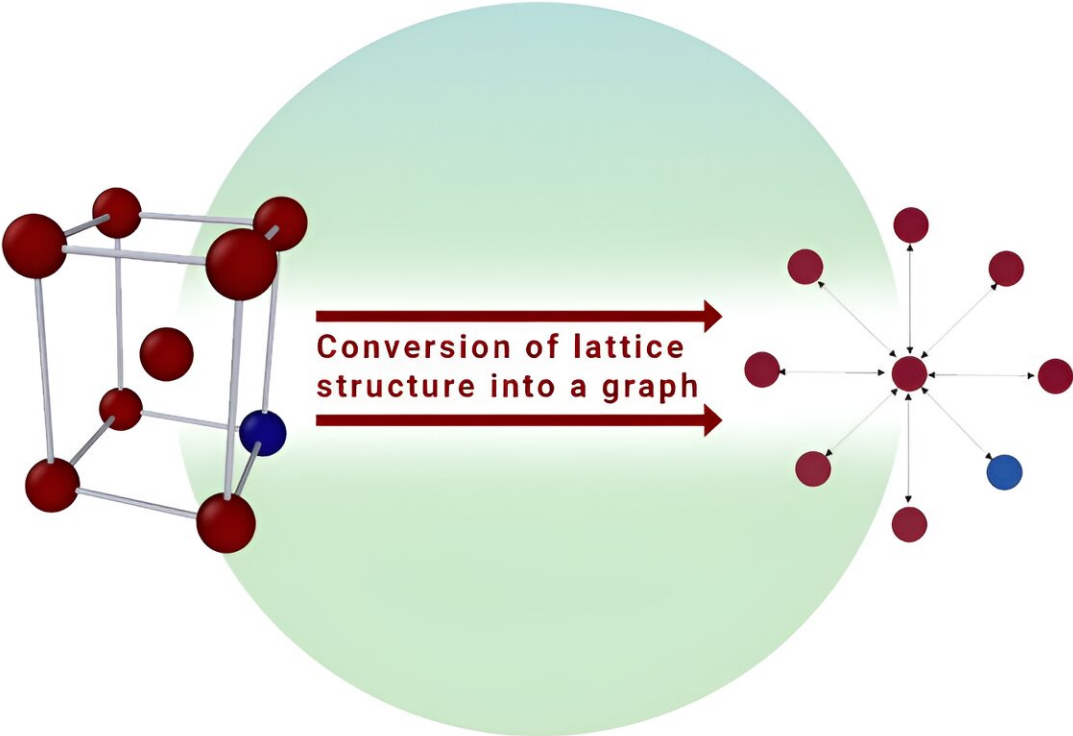


# Researchers demonstrate scalability of graph neural networks on world's most powerful computing systems

January 19 2024, by Scott Jones

---



Conversion of an atomic structure into a graph, where atoms are treating as nodes and interatomic bonds as edges. Credit: Massimiliano “Max” Lupo Pasini/ORNL, U.S. Dept. of Energy

Solving today's most complex scientific challenges often means tracing links between hundreds, thousands or even millions of variables. The larger the scientific dataset, the more complex these connections become.

With experiments generating petabytes and even exabytes of data over time, tracking the connections in processes such as drug discovery, materials development or cybersecurity can be a Herculean task.

Thankfully, with the advent of artificial intelligence, researchers can rely on graph [neural networks](#), or GNNs, to map the connections and unravel their relationships, greatly expediting time to solution and, by extension, [scientific discovery](#).

Researchers at the Department of Energy's Oak Ridge and Lawrence Berkeley national laboratories (ORNL and LBNL) are evolving GNNs to scale on the nation's most powerful computational resources, a necessary step in tackling today's data-centric scientific challenges.

ORNL's Massimiliano "Max" Lupo Pasini, Jong Youl Choi and Pei Zhang shared the multi-institutional team's findings at the Learning on Graphs 2023 conference, a virtual event that took place Nov. 27–30, 2023. Their tutorial, "Scalable graph neural network training using HPC and supercomputing facilities," illustrated how to scale GNNs on DOE's leadership-class computing systems.

Specifically, the team demonstrated the scaling of HydraGNN on the Perlmutter system at LBNL's National Energy Research Scientific Computing Center as well as the Summit and Frontier supercomputers at the Oak Ridge Leadership Computing Facility. Frontier is the world's first exascale system and is currently ranked as the most powerful computer in the world.

HydraGNN is an ORNL-branded implementation of GNN architectures that aims to produce fast and accurate predictions of material properties. It uses atomic information by abstracting the lattice structure of a solid material as a graph, where atoms are represented by nodes and metallic bonds by edges. This representation naturally incorporates information about the structure of the material, eliminating the need for computationally expensive data pre-processing required by more traditional neural networks.

"Scientific advances require discovering and designing materials with improved mechanical and thermodynamical properties, and HydraGNN is a promising surrogate model. Once trained on large volumes of first-principles data, the model can provide fast and accurate estimates of materials properties at a fraction of the computational time needed with state-of-the-art physics-based models," said Pasini, a researcher in the Computational Sciences and Engineering Division at ORNL. "The improved speed at which HydraGNN produces predictions enables a unique exploratory capability for effective materials discovery and design."

The team's tutorial was broken up into five sections. The first highlighted scientific applications that motivate the development of scalable GNN surrogate models to accelerate the study of complex physics and engineering systems. The second introduced the need to use large volumes of scientific data to scale GNNs on DOE leadership-class supercomputing facilities. The third covered HydraGNN's scalability and flexibility, which makes it portable across several DOE systems. The fourth covered examples of running HydraGNN on open-source datasets, and the fifth and final section consisted of a tutorial and closing remarks.

The tutorial was [livestreamed on YouTube](#) and is available starting at the 1:01:00 mark. The improved capabilities of HydraGNN were also

recently documented in a user manual that has been released to the public via an [ORNL Technical Report](#).

The research is part of ORNL's AI Initiative, an internal investment dedicated to ensuring safe, trustworthy and energy efficient AI in the service of scientific research and national security. Through the initiative, ORNL researchers leverage the laboratory's computing infrastructure and software capabilities to expedite times to solution and realize the potential of AI in projects of national and international importance.

For example, the initiative has helped multidisciplinary teams demonstrate that machine learning algorithms can be used to extract information from signals with low signal-to-noise ratios, develop algorithms capable of accelerating modeling and simulation with little training data and design novel biomimetic neuromorphic devices capable of detecting epileptic seizures.

"Scaling Graph Neural Networks presents unique challenges," said Prasanna Balaprakash, director of ORNL's AI Initiative. "Capable of being trained on extensive scientific datasets, these models unlock a wide array of downstream applications, particularly in the development of new materials and drug discovery. This achievement underscores our commitment to developing AI that is not only powerful but also energy-efficient and scalable, ensuring that we stay at the forefront of scientific research and national security."

Provided by Oak Ridge National Laboratory

Citation: Researchers demonstrate scalability of graph neural networks on world's most powerful computing systems (2024, January 19) retrieved 2 May 2024 from <https://techxplore.com/news/2024-01-scalability-graph-neural-networks-world.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.