Researchers bring innovative AI and simulation tools to the COVID-19 battlefront

September 1 2021, by Anna Marie Tomczyk and John Spizzirri

ALCF's AI- and simulation-enabled extension of the Theta supercomputer, ThetaGPU. Credit: Argonne National Laboratory

In its on-going campaign to reveal the inner workings of the Sar-CoV-2 virus, the U.S. Department of Energy's (DOE) Argonne National Laboratory is leading efforts to couple artificial intelligence (AI) and cutting-edge simulation workflows to better understand biological observations and accelerate drug discovery.

Argonne collaborated with academic and commercial research partners to achieve near real-time feedback between simulation and AI approaches to understand how two proteins in the SARS-CoV-2 viral genome, nsp10 and nsp16, interact to help the virus replicate and elude the host's immune system.

The team achieved this milestone by coupling two distinct hardware platforms: Cerebras CS-1, a processor-packed silicon wafer deep learning accelerator; and ThetaGPU, an AI- and simulation-enabled extension of the Theta supercomputer, housed at the Argonne Leadership Computing Facility, a DOE Office of Science User Facility.

To enable this capability, the team developed Stream-AI-MD, a novel application of the AI method called deep learning to drive adaptive molecular dynamics (MD) simulations in a streaming manner. Data from simulations is streamed from ThetaGPU onto the Cerebras CS-1 platform to simultaneously analyze how the two proteins interact.

Cerebras CS-1 is a processor-packed silicon wafer deep learning accelerator. Credit: Argonne National Laboratory

"This needs to be done at a scale that is unprecedented since the data generation and AI components have to run side-by-side," said Argonne computational biologist Arvind Ramanathan, a member of the research team. "The idea is, if one machine is good at doing MD simulations and another is very good at AI, then why not couple the two to produce a much larger system that offers more throughput with AI," explained Ramanathan.

One of the AI techniques that they're using is called a variational autoencoder, which learns to capture the most essential information from MD simulations. The size of the simulation data sets is reduced in a way to make it easier for researchers to understand the dynamics occurring in the simulation.

By running their deep learning component on Cerebras CS-1, they can identify binding pockets—tiny spaces that might develop during the formation of the two proteins—that can be targeted for small-molecule drug design.

These workflows will ultimately enable drug discoveries that treat both the SARS-CoV-2 virus and other diseases, when the physical processes underlying specific biological functions are characterized, said Ramanathan. And while the study currently does not focus on vaccines, the development of more complex models could lead to vaccine design.

"This iterative workflow of supporting streaming AI and MD techniques on emerging hardware platforms will pave the way for advancing our

knowledge of how proteins function," said Ramanathan. "In the context of the SARS-CoV-2 virus, a fundamental understanding of molecular processes, such as the nsp16-nsp10 interaction, is important if we want to design drugs that can stop the virus in its path."

The research was published in the proceedings from the Platform for Advanced Scientific Computing Conference (PASC '21), July 5–9, 2021, Geneva, Switzerland. ACM, New York, NY, USA.

Provided by Argonne National Laboratory