Enabling Portability and GPU Direct on DCA++

Science

A team of ORNL researchers has used the DCA++ application, a popular code for predicting the performance of quantum materials, to verify two performance-enhancing strategies. First, the team improved the Quantum Monte Carlo solver, a tool common across the DOE application landscape, and built a portability layer to Frontier, the OLCF's first exascale system to launch in 2021. Second, the team demonstrated that GPU Direct, a feature available on the NVIDIA Volta GPUs that populate ORNL's Summit, can overcome the lost performance that haunts "fat node" systems, which relay communications between GPUs through CPUs.

Impact

The scientific impact of the research is two-fold:

- The demonstration of direct GPU-to-GPU communication proved that GPU Direct can be used to enhance code performance while maintaining the accuracy of the science. The performance benefit outweighs the resource required (developer's time, code modifications).
- The increased efficiency enables more accurate modeling of quantum materials.
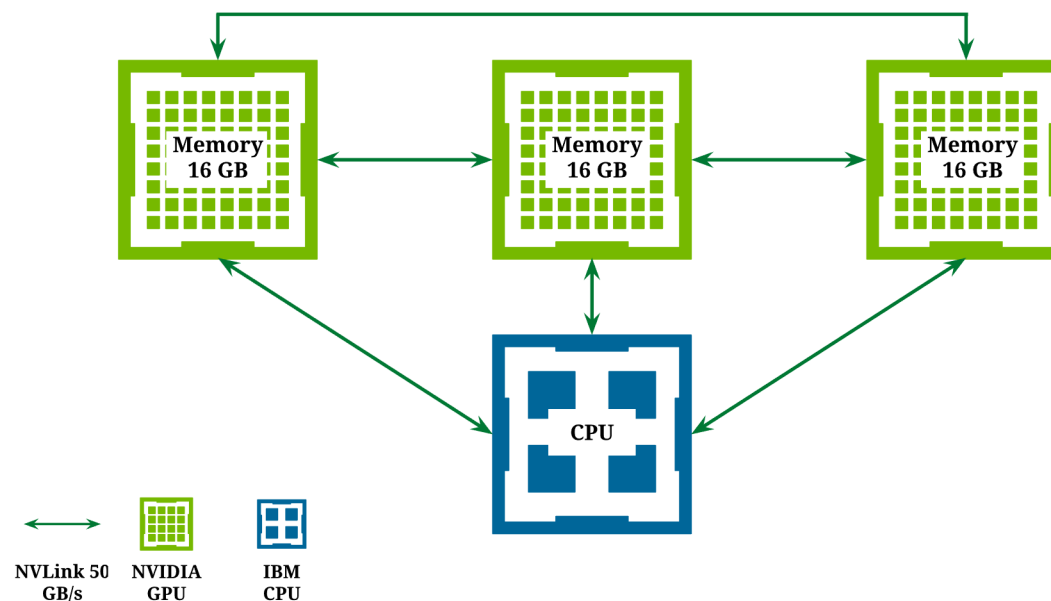

Figure: The hardware layout of half of a single summit node. This work focused on exploiting the complexity of 2 CPU and 6 GPU-node architecture to its maximum capability by further using the NVLink (fast network connection between GPUs on Summit).

OAK RIDGE
National Laboratory