GPU-aware Communications in AWP-ODC that Improves the Performance of Collective Communication on GPU Clusters

Yifeng Cui, Akash Palla, Tom Zhang, Shijie Wang, Daniel Roten, Lars Koesterke, Wenyang Zhang, Philip Maechling

We implemented GPU-aware communications in the AWP-ODC software and enhanced its message-passing collective communications to improve the scalability and efficiency of this widely used, memory-bound, finite-difference, earthquake wave propagation solver. These improvements allow GPUs to communicate directly with each other without routing messages through the CPU. We measured AWP-ODC performance improvements on two leadership class supercomputers, including OLCF Frontier and TACC Vista. On the DOE OLCF Leadership Class Frontier supercomputer, which is an 8192 Node Exascale System with 65,336 AMD MI250X GCDs (Graphics Compute Die), by using our GPU-aware software on 8,192 nodes we reached 37 sustained Petaflop/s and reduced our time-to-solution by 17.2% as compared to performance without these communications improvements. We have also measured excellent AWP-ODC performance on NSF's TACC Vista supercomputer, an Arm-based NVIDIA GH200 Grace Hopper Superchip, that serves as the computing bridge between TACC current leadership class system, Frontera, and the next NSF Leadership class supercomputer, Horizon. Our poster will showcase our performance studies and how we improved AWP-ODC GPU performance characteristics. We will discuss our verification of GPU-aware development and the use of high-performance MVAPICH libraries, including on-the-fly compression, on modern GPU clusters.