RUNNING PLASCOM2 ON BLUE WATERS WITH OPENMP OFFLOADING

Research Challenge
Programming heterogeneous systems is a challenging task, as few programming models support executing code on accelerator devices, leading to the use of specialized solutions such as CUDA, OpenCL, Legion, or Kokkos. This provides the best performance in many cases, as they can often offer support for special device features and good code generation for specific device types. For PlasCom2, the researchers needed an approach which could use the existing code (C++ and FORTRAN) on several accelerator types without having separate implementations for different device types.

Methods & Codes
Based on the offloading support available in recent OpenMP versions (≥4.5), the researchers developed HybridOMP, which is a library to support concurrent execution on host and accelerator devices. For PlasCom2, HybridOMP measures the relative performance of the host and accelerator at startup and determines the best work distribution based on these data.

Why Blue Waters
Blue Waters was essential to the research by providing a stable, high-performance platform with easy access to modern accelerators. HybridOMP and its integration into PlasCom2 could be developed directly on Blue Waters to evaluate and compare different implementation possibilities on a real system.

Results & Impact
OpenMP offloading enables simple and efficient execution of a single code base on different types of devices, with minimal changes to existing code. HybridOMP builds on top of OpenMP offloading to add support for fully heterogeneous execution. Using HybridOMP with PlasCom2 resulted in a speedup of 2.2 times compared to CPU-only execution on a Blue Waters XK7 node. Some of the computationally intensive kernels of PlasCom2 showed a speedup of 5 times as seen in the above figure. These gains remained similar when performing a strong scaling experiment on multiple nodes.

Figure 1: Single-node results of PlasCom2 kernels on the Blue Waters system.