



Contribution ID: 37

Type: **Short talk**

A task-based data-flow model for distributed and heterogeneous applications

Thursday 23 March 2023 14:30 (10 minutes)

Applications traditionally leverage MPI to run efficiently on HPC systems and scale up to thousands of processors. Since one decade ago, developers have also been adapting their applications to heterogeneous systems by offloading the most time-consuming computation kernels to the available GPUs. To achieve optimal performance in such applications, developers must use the non-blocking and asynchronous services provided by the MPI and GPU-offload APIs; otherwise, the application threads would waste CPU host resources waiting on the synchronous completion. But managing the asynchronicity of communications and GPU-offloading from the application is challenging, tedious, and repetitive among applications. Furthermore, overlapping computation with communication or GPU operations is even harder.

For this purpose, we present a data-flow model that allows distributed and heterogeneous applications to easily benefit from asynchronous communications and GPU-offloading operations, so they avoid dealing with low-level details, such as progress and completion checks. The idea consists of taskifying the application with standard OpenMP tasks: the computations, the asynchronous communications, and the asynchronous GPU-related operations. The tasks declare data dependencies on the data buffers they read/write to define their execution order constraints. This way, computations naturally overlap with communications and GPU operations. We provide two libraries named Task-Aware MPI and Task-Aware CUDA (or other GPU vendors) that define task-aware asynchronous services and transparently handle all those details mentioned earlier.

Our data-flow model has already shown significant benefits at both performance and programmability levels on multiple benchmarks. This short talk aims at finding collaboration opportunities for porting real-world applications or mini-applications to this model.

JLESC topic

Primary author: SALA, KEVIN (BARCELONA SUPERCOMPUTING CENTER (BSC))

Presenter: SALA, KEVIN (BARCELONA SUPERCOMPUTING CENTER (BSC))

Session Classification: Short Talks on Tasking

Track Classification: Programming languages and runtimes