Despite no changes to the computation, we saw performance...

- Runs with UPC++ and CUDA
- Use of UPC++ memory kinds for GPU memory management
- Regular 3D halo
- Unordered_map
- External storage of key metadata at the target.
- Current MPI implementations call remote procedure call on any remote part of the graph. Approx. 100% reduction.
- Results on NERSC Cori Haswell (3.68-edge Friendster):
  - Both UPC++ versions competitive with (or better than) best MPI versions up to at least 4,096 processes.

Integration efforts with NWChemEx (WS 2.2.1.02):
- Ported TAMM code base from Global Arrays to UPC++.
- TAMM implements distributed memory data store and compute for NWChemEx.
- UPC++ performance comparable to G2 code (+10.5% run time).
- Current work-in-progress includes:
  - Merging UPC++ communication into the main TAMM repository.
  - Evaluating use of `dxt_array` in TAMM.
  - Work toward UPC++ RPC in dynamically loaded libraries.

Benefits:
- Use of RPC simplifies distributed data-structure design.
- Argument passing, remote queue management, and progress engine are factored out of the application code.
- Efficient weak scaling to 512 nodes (54 core) on Cori Xeon Phi.

Support for IPA targeting GPU Memory – UPC++ and Legion/Realm Benchmarks

- Support for Nvidia GPUDirect RDMA (Nov. 2020):
  - Removes host CPU and memory bottlenecks from GPU transfers to/from GPU memories (see diagram).
  - Works with Nvidia GPUs + Mellanox NICs.
- Support for AMD ROCm RDMA (Sep. 2021):
  - Same benefits with AMD GPUs + Mellanox NICs.
- Demonstrated over HPE SlingShot-10 on OLCF’s Spock.
- Massively parallel initial GPU programs are each the subject of future work.
- Comparisons of UPC++ to MPI RMA in CUDA-Aware IBM Spectrum MPI show UPC++ saturating more quickly to the peak (top-right plot).
- Realm is the low-level runtime for the Legion Programming System (WBS 2.3.1.08):
  - Communication services originally implemented over GPU memory.
  - New communications backend (Dec 2020) embraces capabilities specific to GASNet-EX.
  - Most notable new capabilities are support for GPU RAM and for the HPE SlingShot network.
- UPC++ consistently meets or beats the performance of MPI.

For additional details see [PAW-ATM21].

References:
- IFPS19: doi.org/10.25344/S4V88H
- UPC++ 2020.11.0 vs. IBM Spectrum MPI 10.3.1.2 on OLCF Summit

GASNet-EX: communications middleware to support exascale clients

- One-sided communication – Remote Memory Access (RMA)
- Active Messages (AMs) – a form of remote procedure call
- Implemented over native APIs of all networks of interest to DOE
- Provides communication for several programming models including:
  - UPC++ (see left half of this poster)
  - Legion (WBS 2.3.1.08)
  - Chapel (from HPE, non-ECP)
- Backwards compatibility for the dozens of GASNet-1 clients.

Major features of GASNet-EX developed under ECP funding:
- "Immediate mode" injection to avoid stalls due to back-pressure
- Explicit handling of local completion (source buffer lifetime)
- Enhanced AM interfaces to reduce buffer copies between layers
- Vector-indexed-Driven for non-contiguous point-to-point RMA
- Remote Atomics, implemented with NIC offload where available
- Subset teams and non-blocking collectives
- RMA directly to/from device memory on supported hardware
  - Includes Nvidia and AMD GPUs

GASNet-EX is implemented over native APIs of all networks of interest to DOE.
- Interoperable with MPI, OpenMP, CUDA, etc.
- Runs on systems from laptops to supercomputers
- Provides building blocks to construct irregular data structures
- Provides communication for several programming models including:
  - UPC++
  - Legion (WBS 2.3.1.08)
  - Chapel (from HPE, non-ECP)
- Backwards compatibility for the dozens of GASNet-1 clients.

GASNet-EX Support for HPE Slingshot

- Demonstrated over HPE Slingshot
- EX library and its tests.
- EX's Nvidia GPU support in Realm:
  - Dual-Rail in TAMM
  - Spock – 2.3.1.14 – EMR
  - Current center default MPI version and Intel MPI Benchmarks.

Strong scaling of a Kokkos-based heat-conduction example, comparing UPC++ and CUDA-aware IBM Spectrum MPI for regular 3D halo-exchange to/from GPU buffers.

Support for REMA targeting GPU Memory – UPC++ and Legion/Realm Benchmarks

- Support for Nvidia GPUDirect RDMA (Nov. 2020):
  - Removes host CPU and memory bottlenecks from GPU transfers to/from GPU memories (see diagram).
  - Works with Nvidia GPUs + Mellanox NICs.
- Support for AMD ROCm RDMA (Sep. 2021):
  - Same benefits with AMD GPUs + Mellanox NICs.
- Demonstrated over HPE SlingShot-10 on OLCE’s Spock.
- Massively parallel initial GPU programs are each the subject of future work.
- Comparisons of UPC++ to MPI RMA in CUDA-Aware IBM Spectrum MPI show UPC++ saturating more quickly to the peak (top-right plot).
- Realm is the low-level runtime for the Legion Programming System (WBS 2.3.1.08):