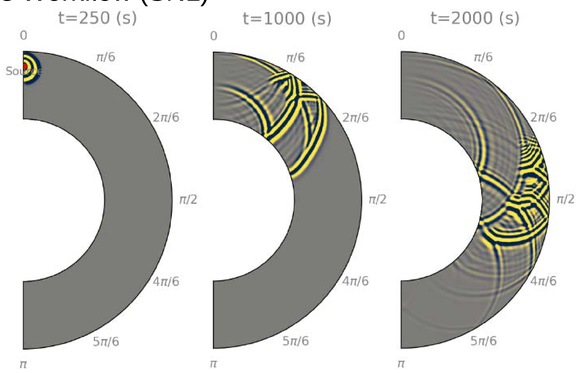
Optimizing Data Movement for GPU-Based In-Situ Workflow

Bo Zhang*, Manish Parashar*

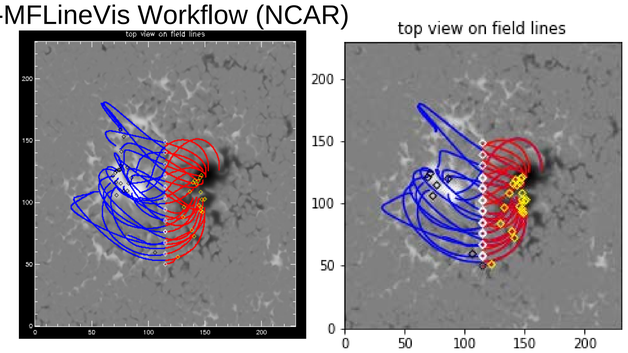
***Scientific Computing and Imaging Institute, University of Utah**

Background & Motivation

- GPUs are widely adopted by modern HPC clusters and many scientific simulations and analyses have been ported to GPUs.
- Porting existing in-situ workflow to GPUs is an ongoing effort.
- Intra-Application GPU communications, e.g. MPI, has leveraged GPUDirect RDMA for I/O optimization.
- SHAW-Vis Workflow (SNL)

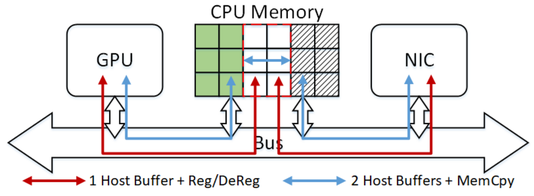


- MURaM-MFLineVis Workflow (NCAR)

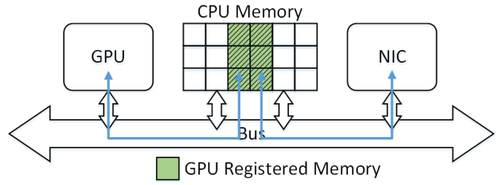


GPUDirect Technologies

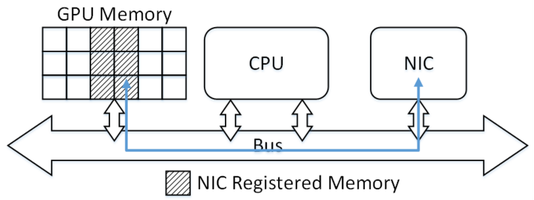
- Vanilla Solution



- GPUDirect

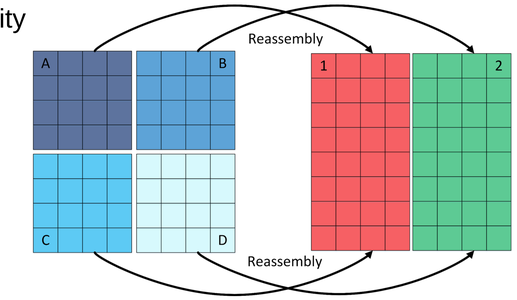


- GPUDirect RDMA

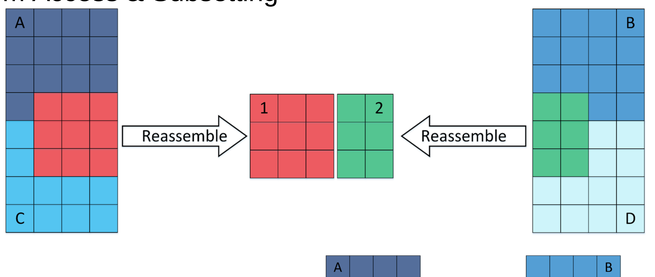


Data Object Reassembly

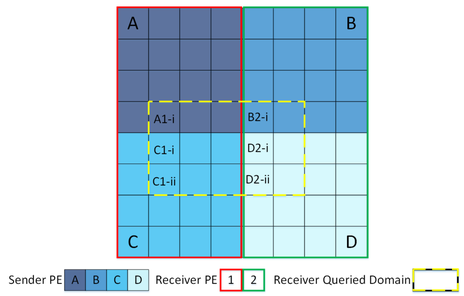
- MxN scale flexibility



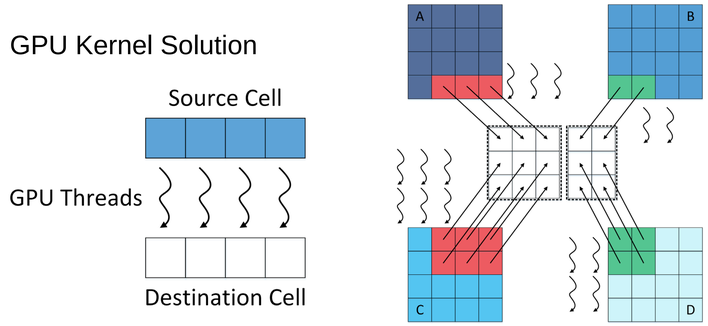
- Random Access & Subsetting



- CPU Solution

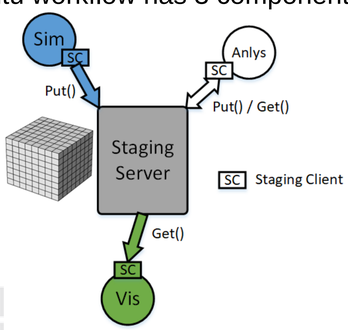


- GPU Kernel Solution

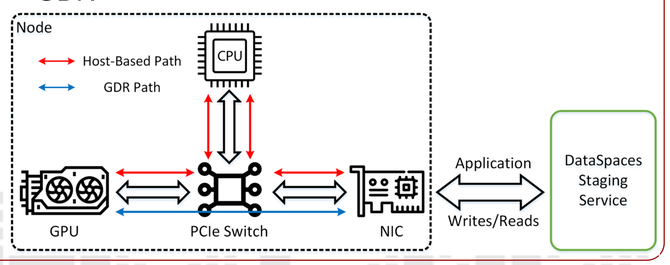


Inter-Application I/O in Staging-Based In-Situ Workflow

- Staging-Based in-situ workflow has 3 components
  - Writer
  - Reader
  - Staging Server

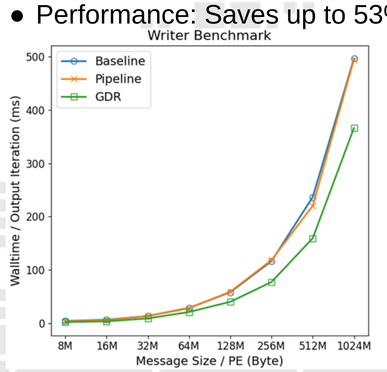


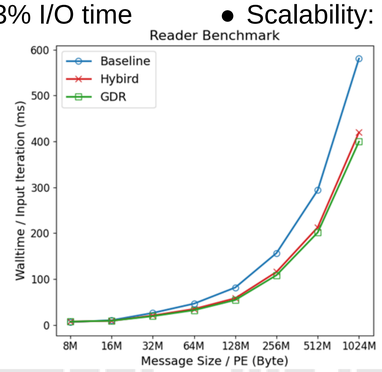
- I/O path has 2 options
  - Host-Based
  - GDR

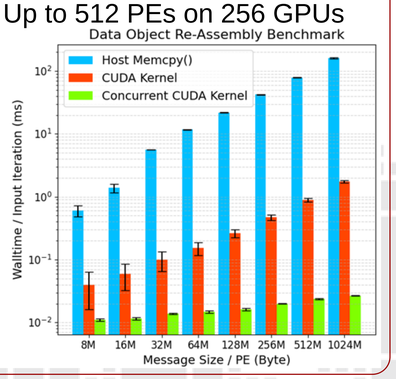


Results

- Performance: Saves up to 53% I/O time
- Scalability: Up to 512 PEs on 256 GPUs







THE UNIVERSITY OF UTAH
U.S. DEPARTMENT OF ENERGY | Office of Science
NERSC
OAK RIDGE National Laboratory
Sandia National Laboratories
TACC
SCI www.sci.utah.edu