Performance Engineering of Scientific Applications with MVAPICH and TAU using Emerging Communication Primitives

Dhabaleswar K. (DK) Panda
panda@cse.ohio-state.edu
The Ohio State University
Columbus, Ohio, USA

Sameer Shende
sameer@cs.uoregon.edu
University of Oregon
Eugene, Oregon, USA

Ahmad Abdelfattah
ahmad@icl.utk.edu
University of Tennessee
Knoxville, Tennessee, USA

Yifeng Cui
yfcui@sdsc.edu
San Diego Supercomputer Center
San Diego, California, USA

Keywords

MPI communication, Profiling tools, High performance computing,

ACM Reference Format:
Dhabaleswar K. (DK) Panda, Sameer Shende, Ahmad Abdelfattah, and Yifeng Cui. 2025. Performance Engineering of Scientific Applications with MVAPICH and TAU using Emerging Communication Primitives. In *SC'25: The International Conference for High Performance Computing, Networking, Storage, and Analysis, November 16-21, 2025, St. Louis, MO.* ACM, New York, NY, USA, 1 page. https://doi.org/XXXXXXX.XXXXXXX

1 Poster Summary

We propose a co-design approach that integrates two powerful tools – MVAPICH and TAU – to demonstrate the new possibilities for performance-guided control and optimization for two large scale applications — AWP-ODC and heFFTe. AWP-ODC is a highly scalable parallel finite-difference application with point-to-point operations that enables 3D earthquake calculations, while heFFTe is a massively parallel application that provides scalable and efficient implementations of the widely used Fast Fourier Transform using several MPI primitives. Through a deep integration between MVAPICH and TAU, the two applications can identify their performance bottlenecks on various supercomputers with different architectures. AWP-ODC and heFFTe can also act as representative real-world benchmarks to MVAPICH and TAU. We show how the co-design approach enables AWP-ODC and heFFTe to deliver better performance on cutting-edge HPC architectures. This is achieved using (1 more optimized and fine-tuned collective operations, and 2) reduced network traffic through real-time data compression. Performance engineering of AWP-ODC and heFFTe is supported by TAU profiling and tracing experiments. Profiling can highlight the load imbalance in the code by computing the time spent in barrier synchronization calls within an MPI collective operation and tracing can show the temporal variation of performance on the CPUs and GPUs along a timeline display with a flamegraph showing the levels of nesting of instrumented code regions. The innovations highlighted in this work include:

(1) Load-aware designs for MPI asynchronous communication
(2) Cross runtime coordination for MPI+X applications
(3) Partitioned point-to-point primitives for efficient communication
(4) Coordinated communication kernels on GPUs
(5) On-the-fly compression for accelerating scientific applications

The proposed compression runtime where MVAPICH-Plus is configured with ZFP delivers a 7.5x speedup over NCCL at 16MB and 4x improvement over OpenMPI at 64MB message size on the NVIDIA Grace-Hopper system Vista at TACC. On Vista, heFFTe achieves throughput improvements of 8%, 5%, 11%, and up to 24% across 8 to 64 nodes. On Frontier, where 95% parallel efficiency is observed on the AWP-ODC application on the full system, the proposed compression runtime delivers a 21× speedup over RCCL at 32MB and 4.8× improvement over Cray MPICH at 32MB message size. AWP-ODC weak scaling on OLCF Frontier, with 95% parallel efficiency on full machine scale. Its MVAPICH2-GDR enhancement improves time-to-solution performance by 17.2% on 8,192 nodes or 65,536 MI250X GCDs. TACC Vista (GH200) sees a 3.5% gain for nonlinear solver with on-the-fly compression at 256 nodes. The future work in this area focuses on updates to MVAPICH-Plus and TAU as well as AWP-ODC and HeFFTe. These updates include:

- MVAPICH: Support for adaptive persistent collective communication.
- MVAPICH: Application-aware neighborhood collective communication.
- TAU: Support persistent collective operations and communicate data on patterns for collectives.
- AWP-ODC: Support on-the-fly compression for Iwan nonlinearity.
- HeFFTe: Design FFT communication using persistent collectives.

SC'25, St. Louis, MO

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX