

Optimizing the Performance of Parallel Applications on Emerging Supercomputers

Sixth Workshop on Communication, I/O, and Storage at
Scale on Next-Generation Platforms

Amanda Bienz
June 26, 2026

Overall Method

- Targeted microbenchmarks for each new architecture / communication library
- Profile application to determine bottlenecks
- Develop benchmark-informed optimizations for these bottlenecks

Background: Many-core Supercomputers

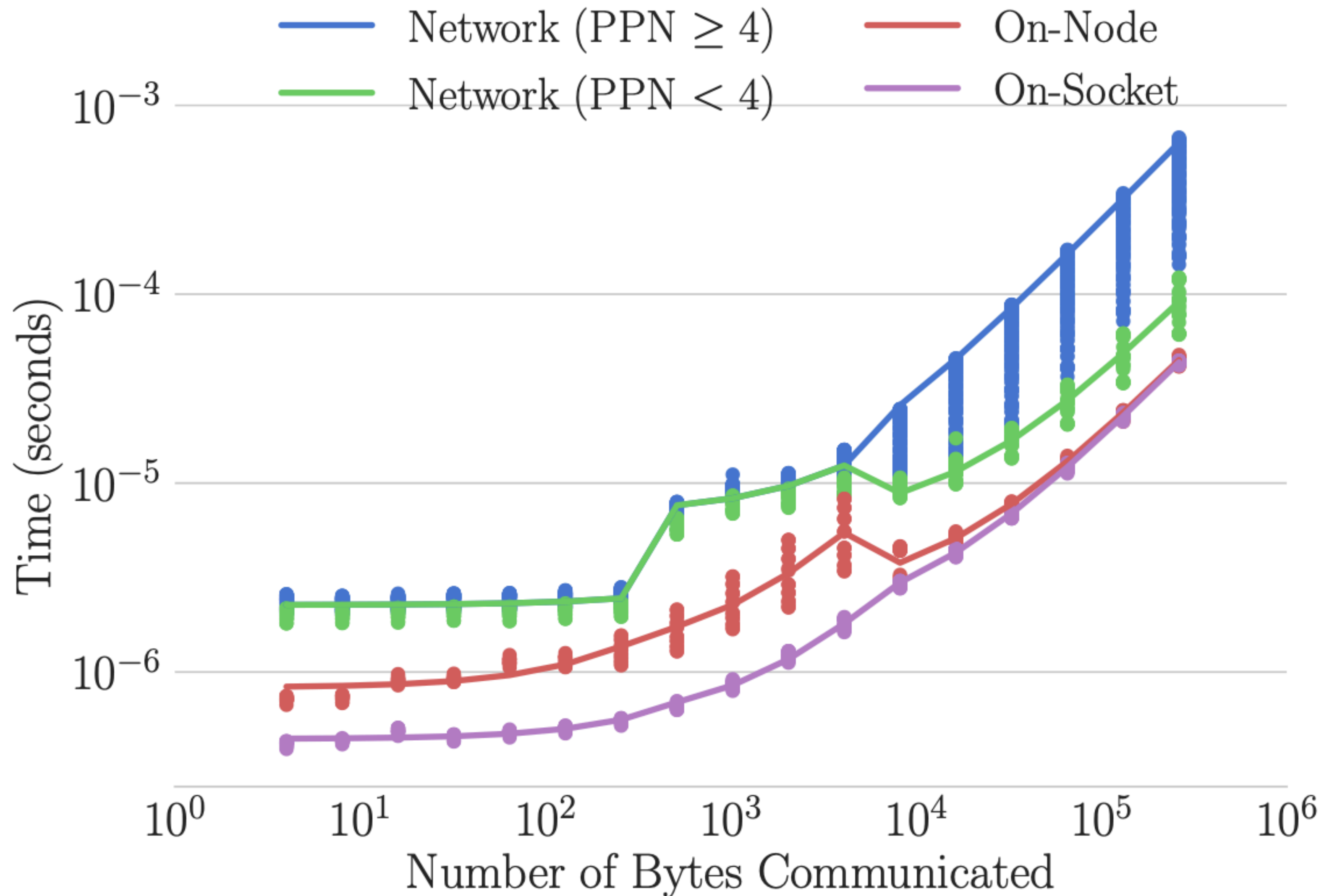
Blue Waters

- 16 cores per node
- 2 Sockets per node
- 2 NUMA regions per node
- 1 NIC per node
- CrayMPICH

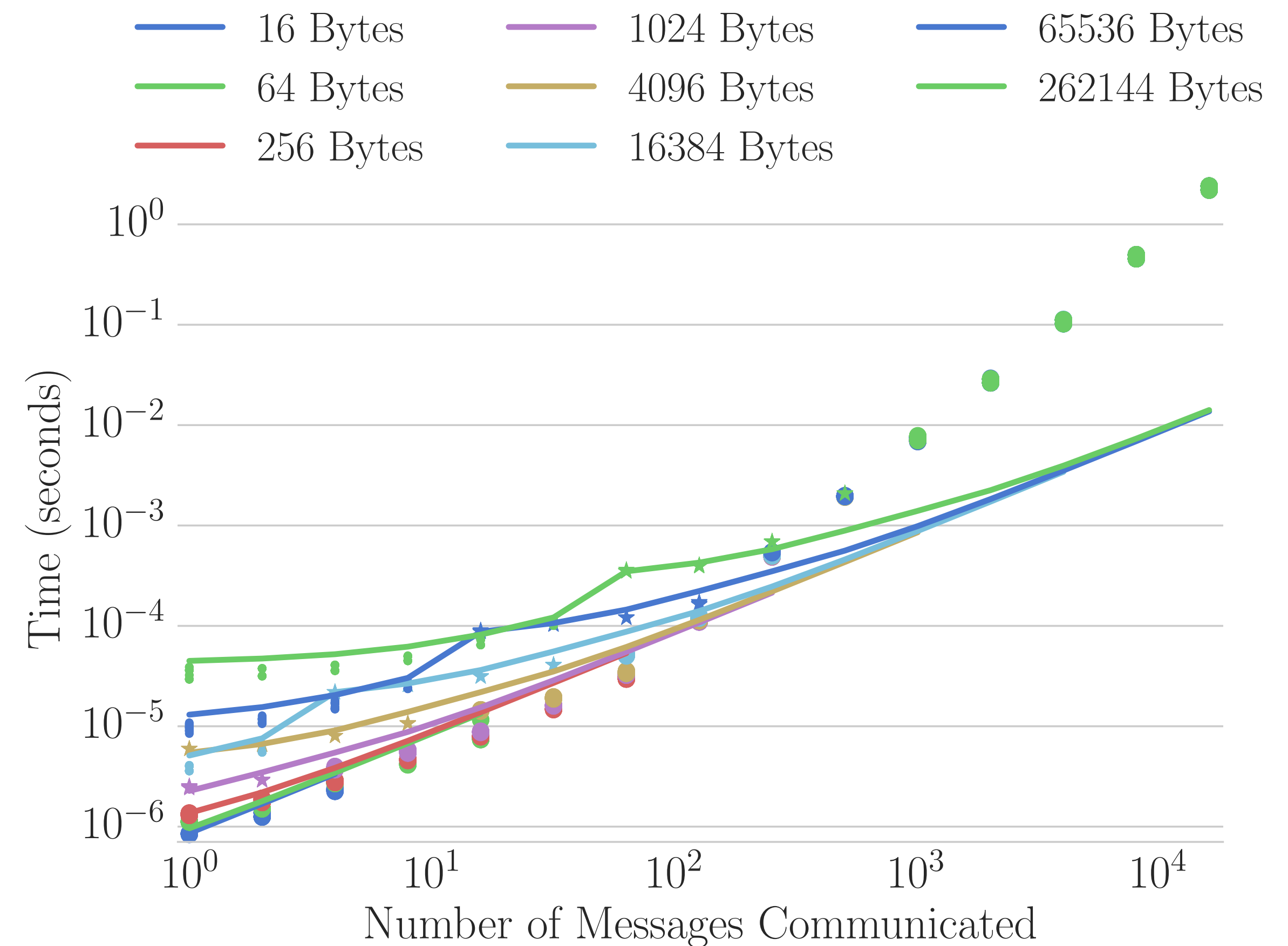
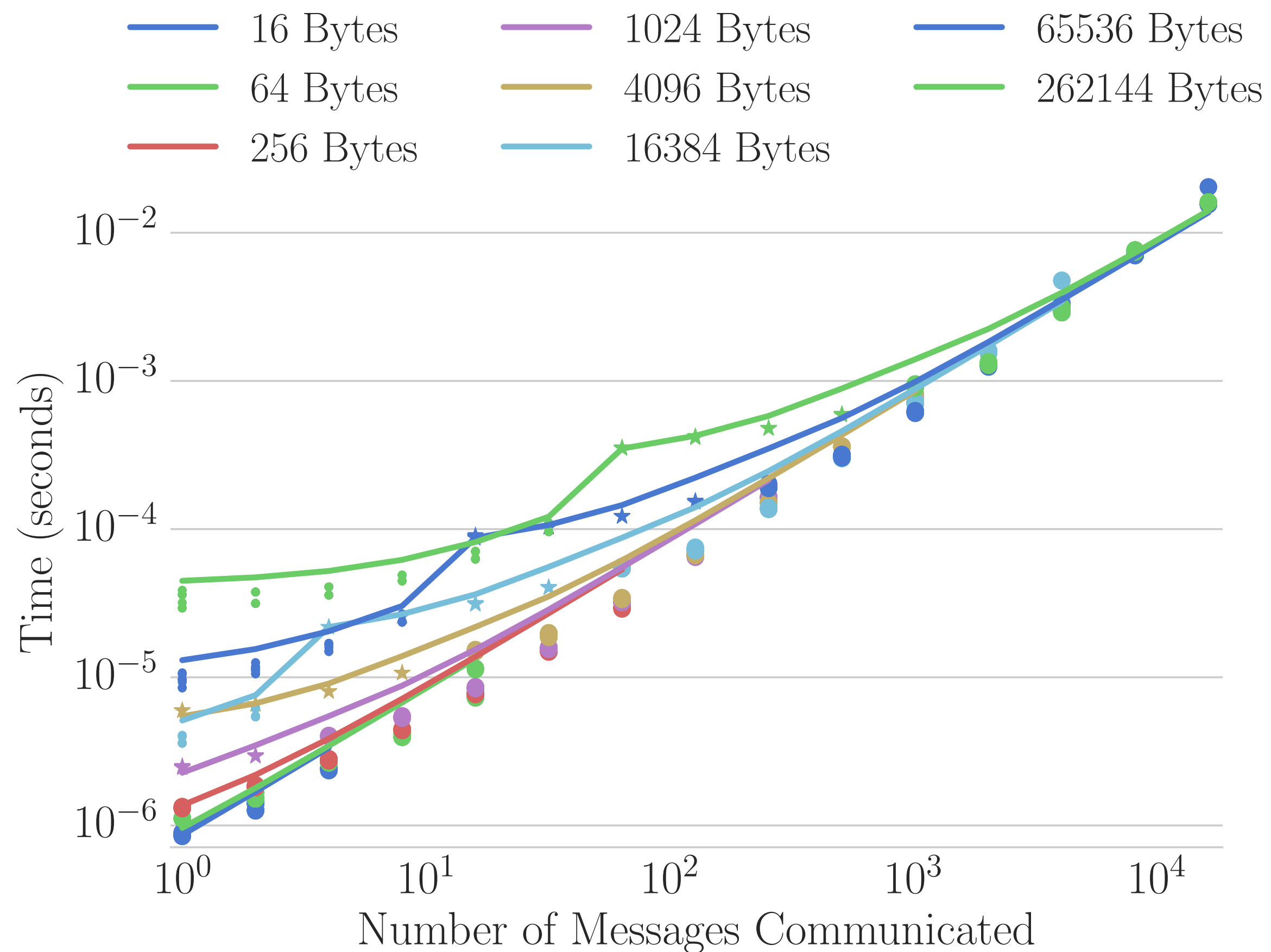


ncsa.illinois.edu

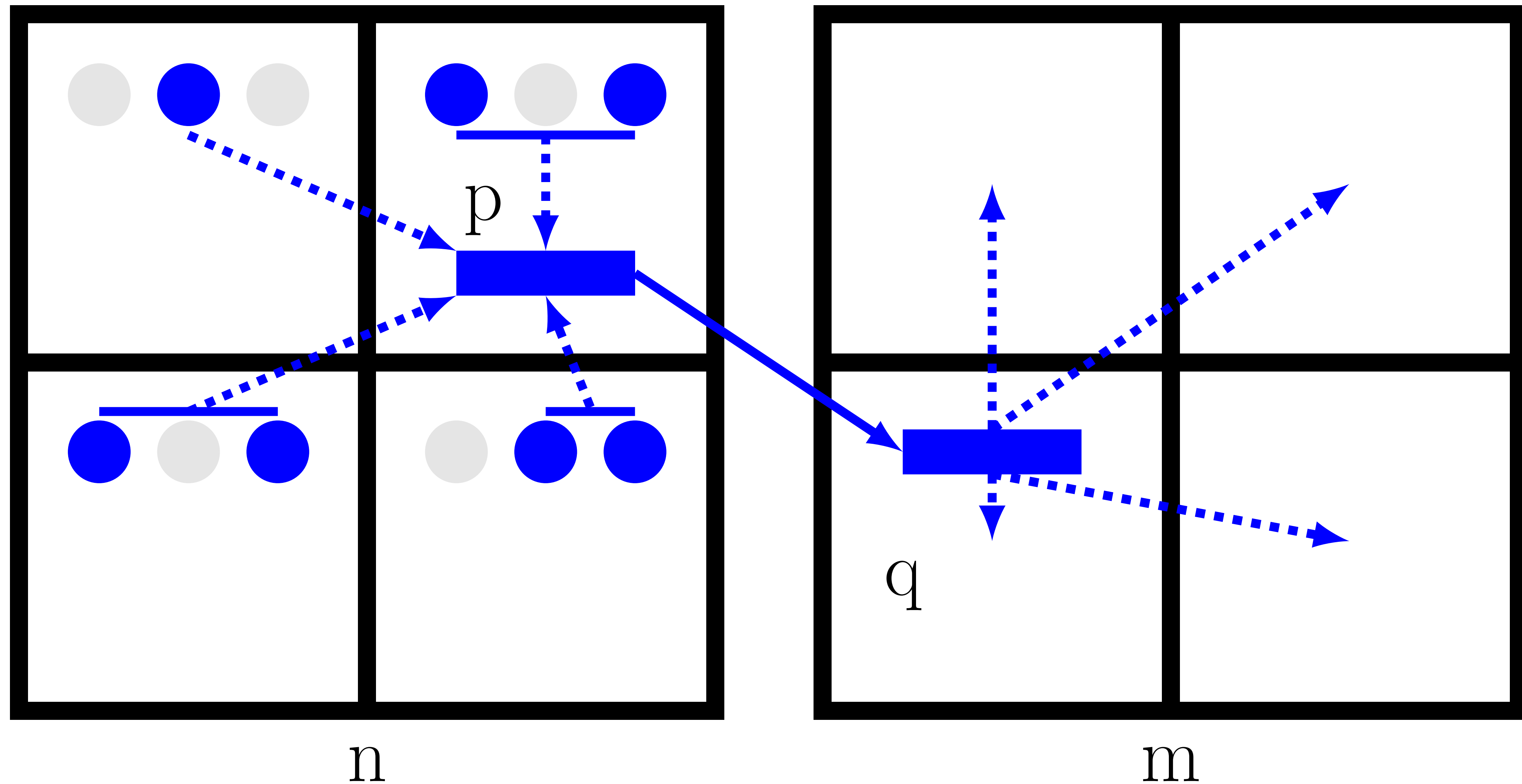
Ping-Pong Latencies



Multiple Ping-Pong Latencies

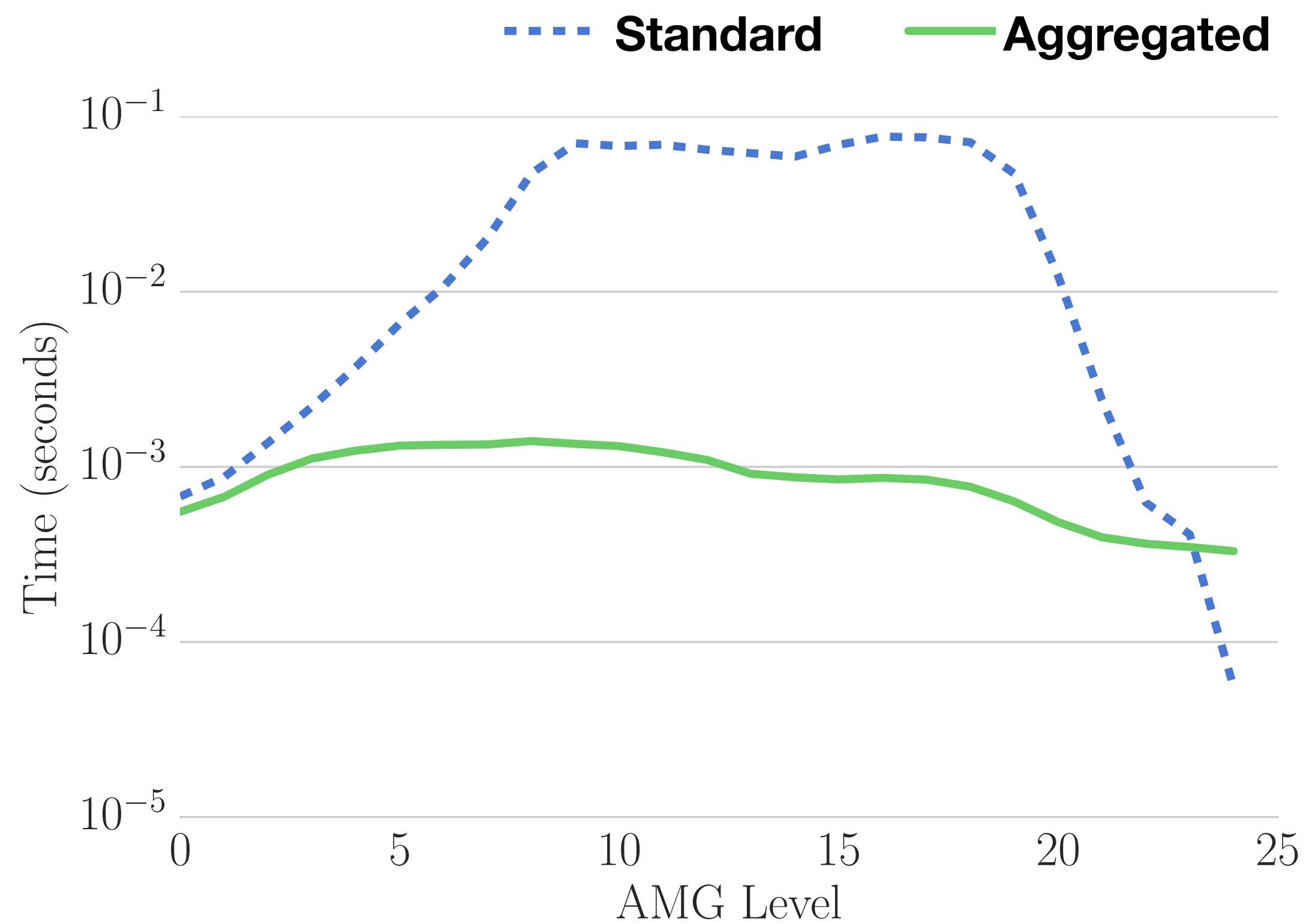


Locality-Aware Aggregation

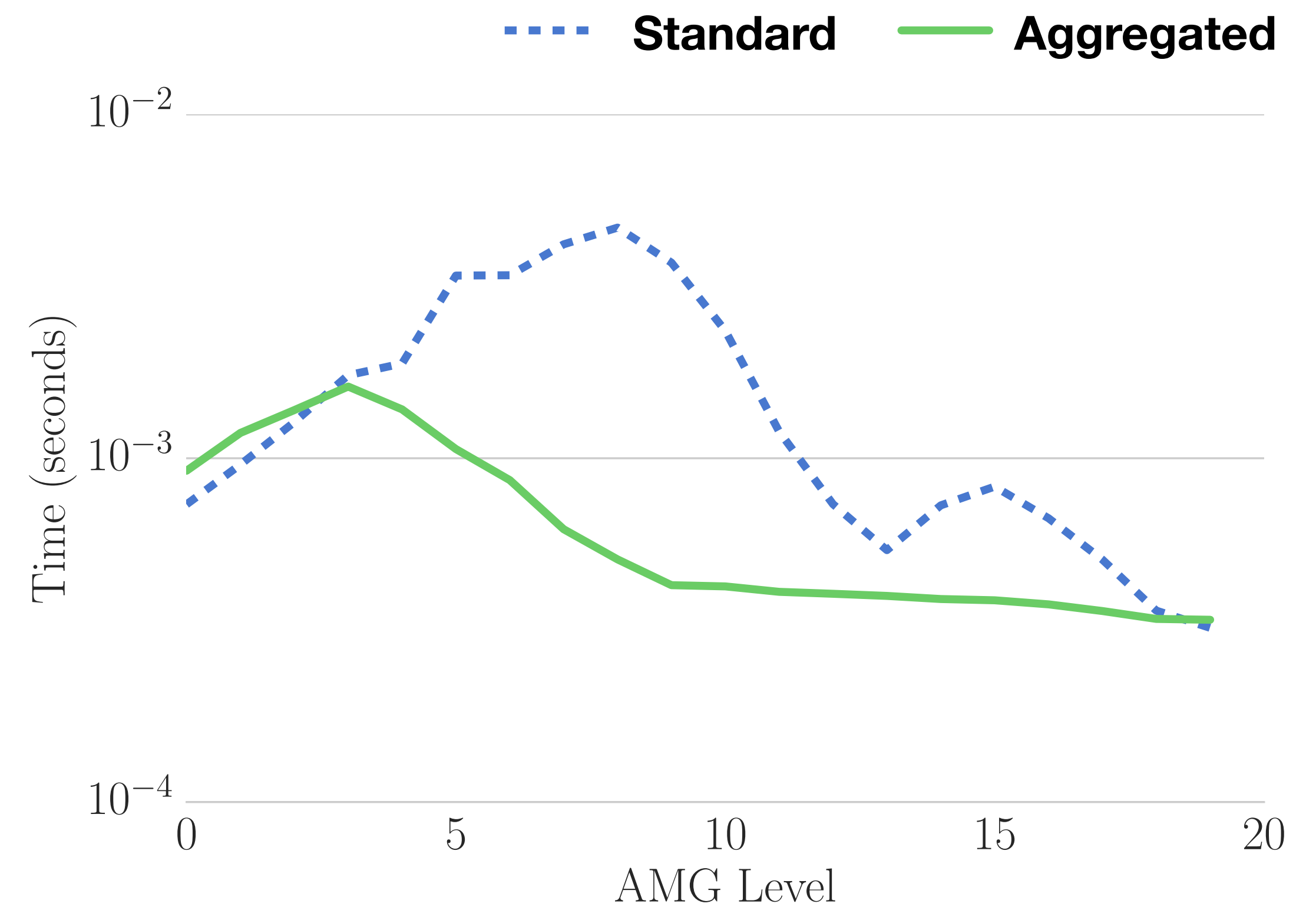


Sparse Matrix-Vector Multiplication

Time spent performing a SpMV on each level of the hierarchy with 32,768 processes

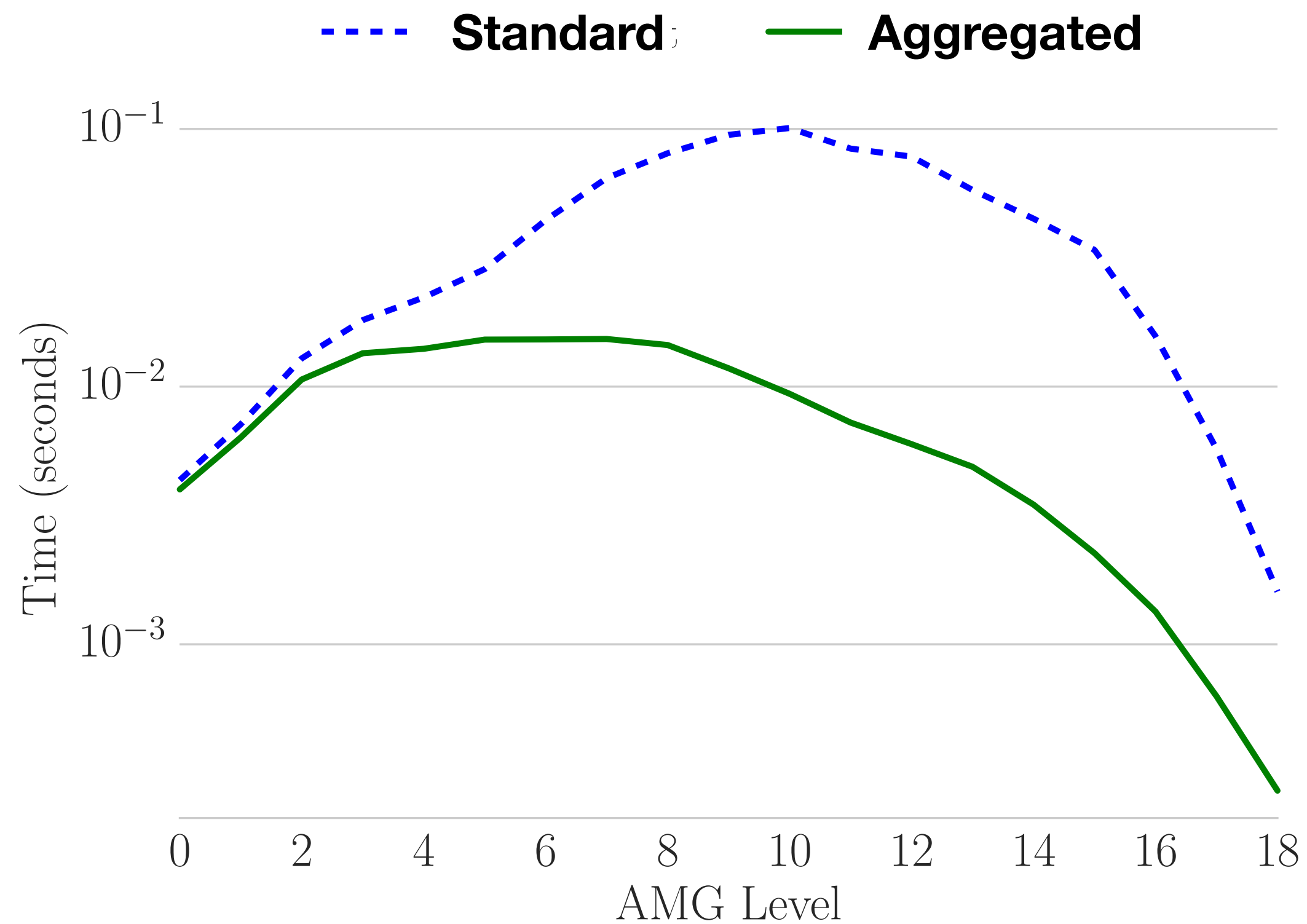


Linear Elasticity (MFEM)

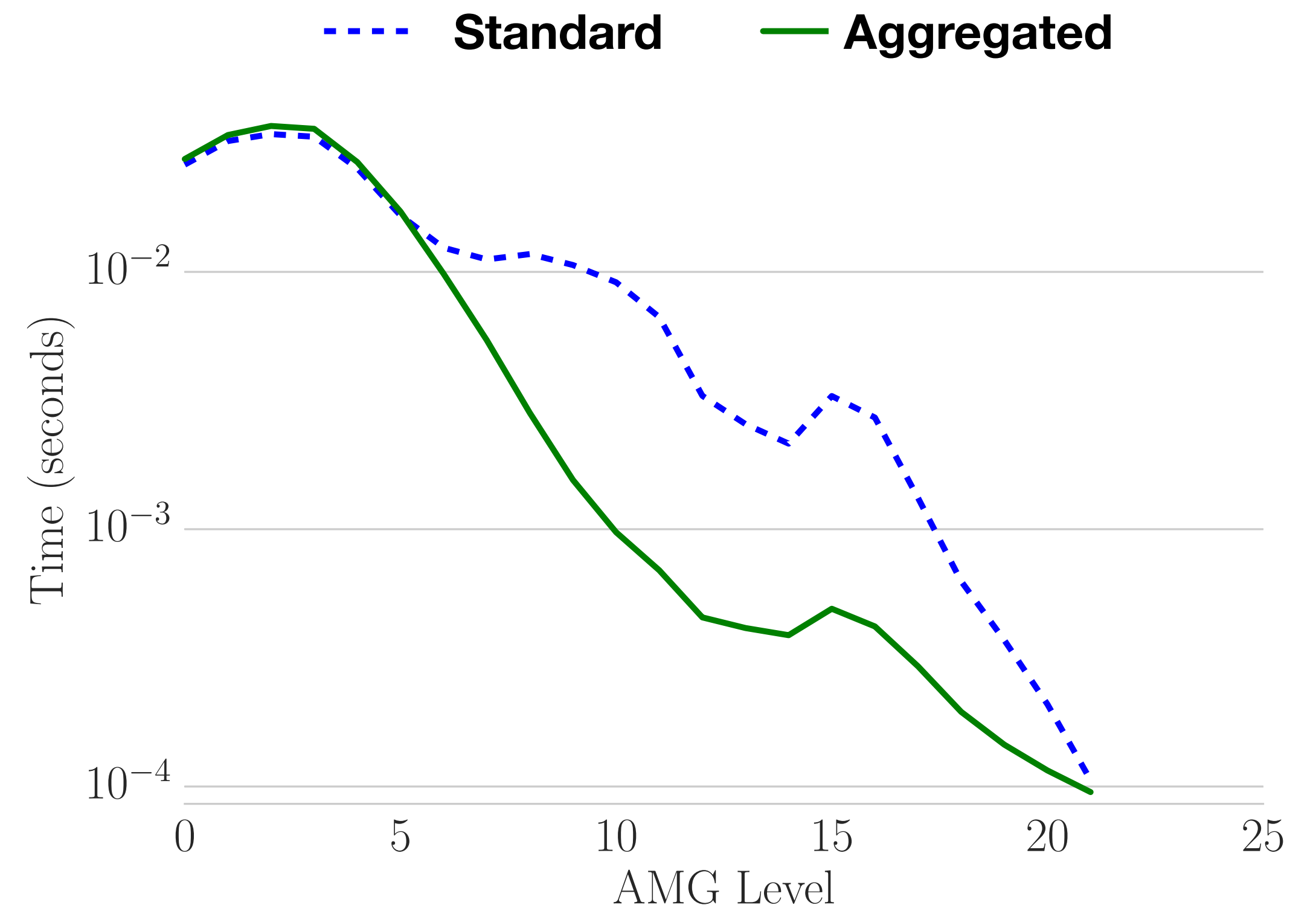


2D Rotated Anisotropic Diffusion

Sparse Matrix-Matrix Multiplication

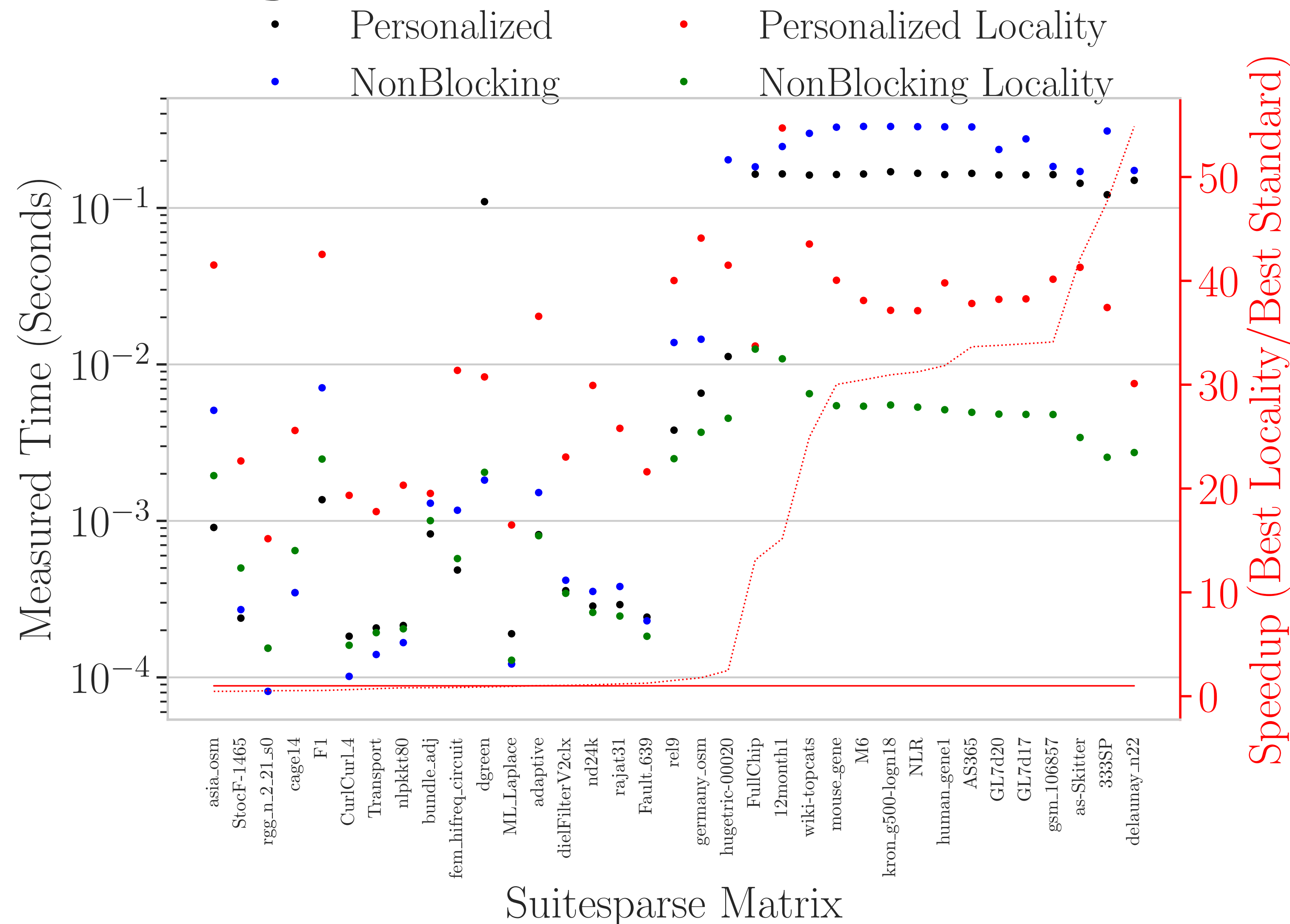


Linear Elasticity (MFEM)



2D Rotated Anisotropic Diffusion

Sparse Dynamic Data Exchange



Geyko et al, "A More Scalable Sparse Dynamic Data Exchange"

Background: Heterogeneous Supercomputers

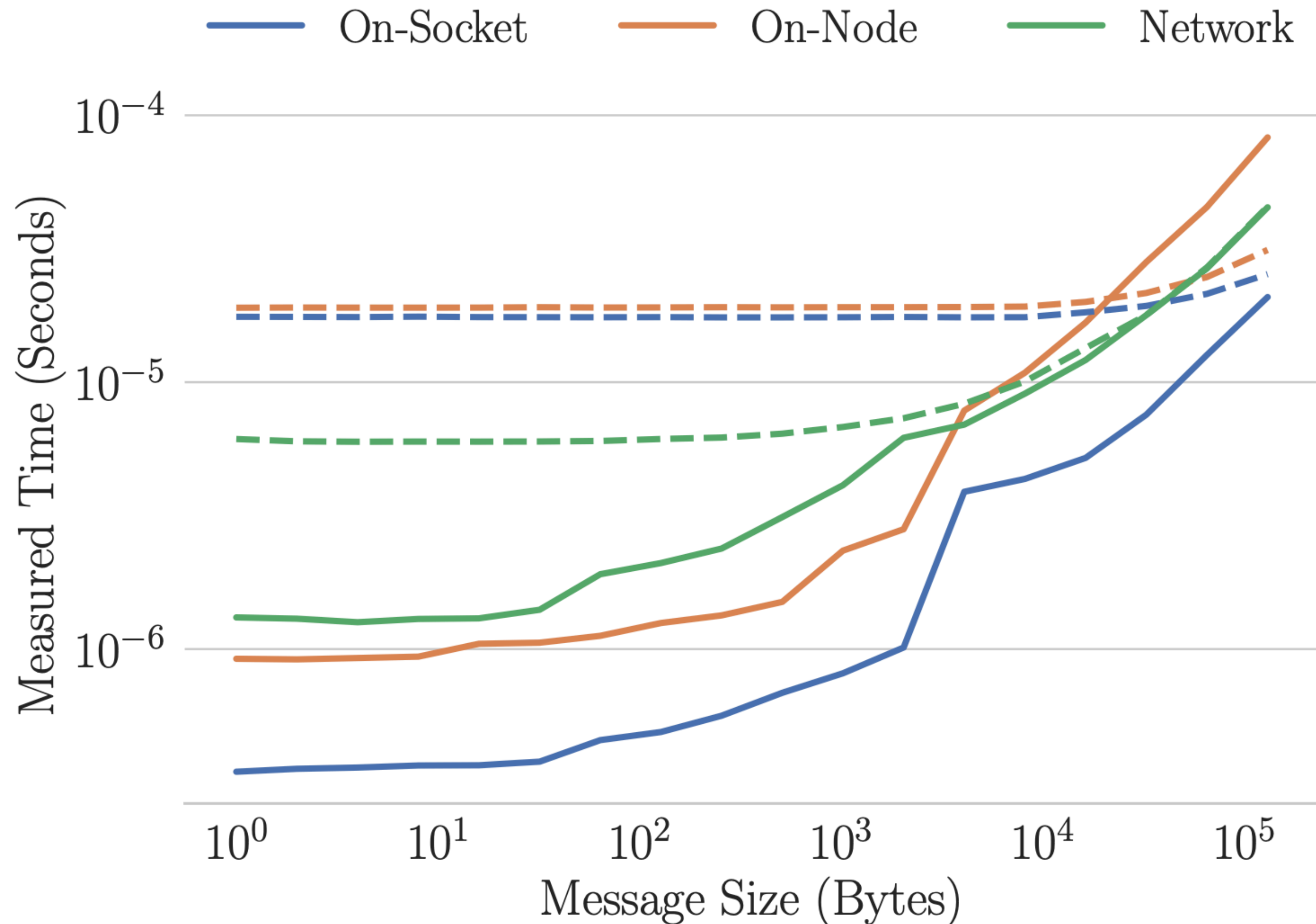
Lassen (IBM Power9)

- 44* cores per node
- 2 Sockets per node
- 2 NUMA regions per node
- 2 NICs per node
- SpectrumMPI



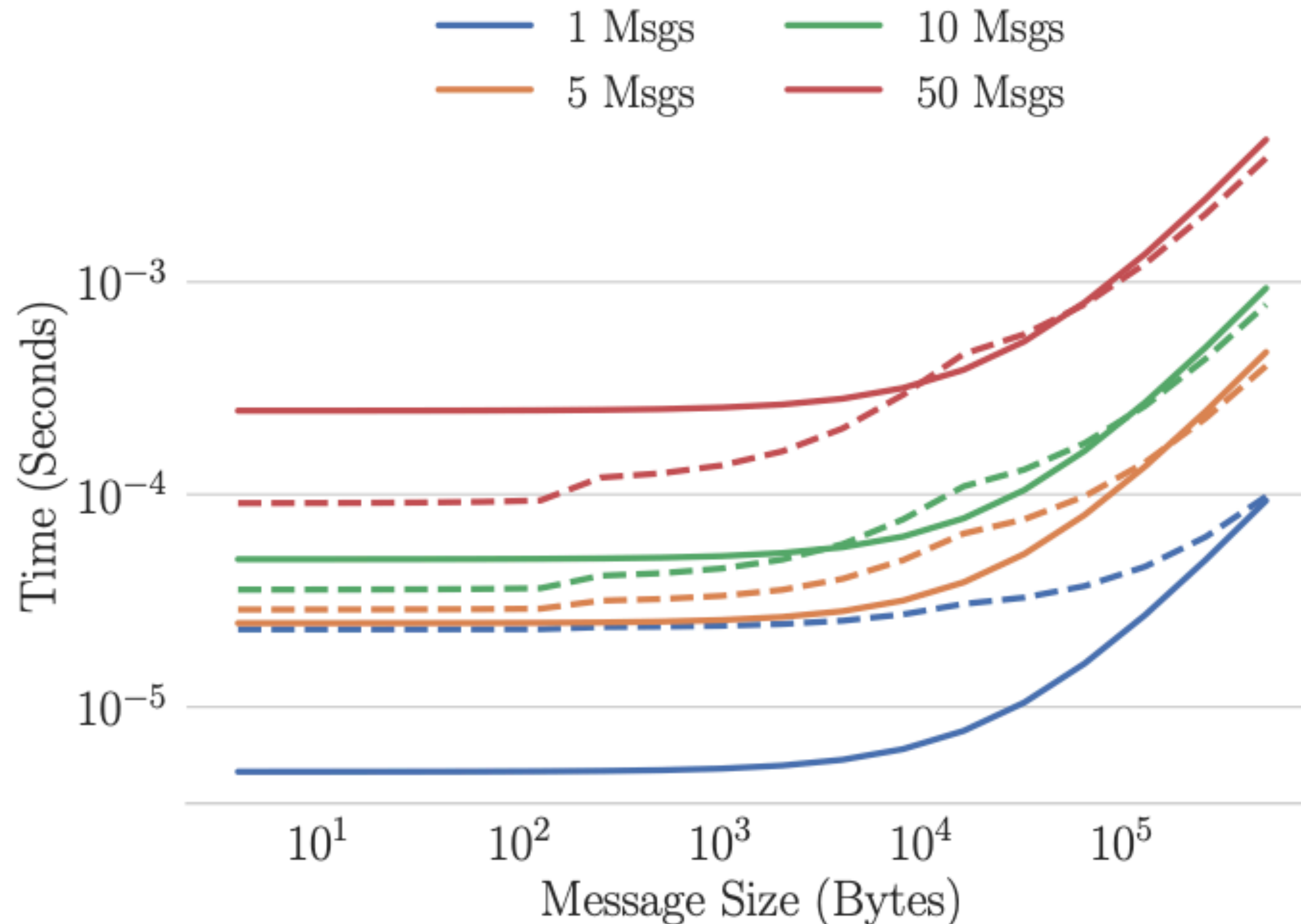
<https://hpc.llnl.gov/hardware/compute-platforms/sierra>

Ping-Pong Latencies



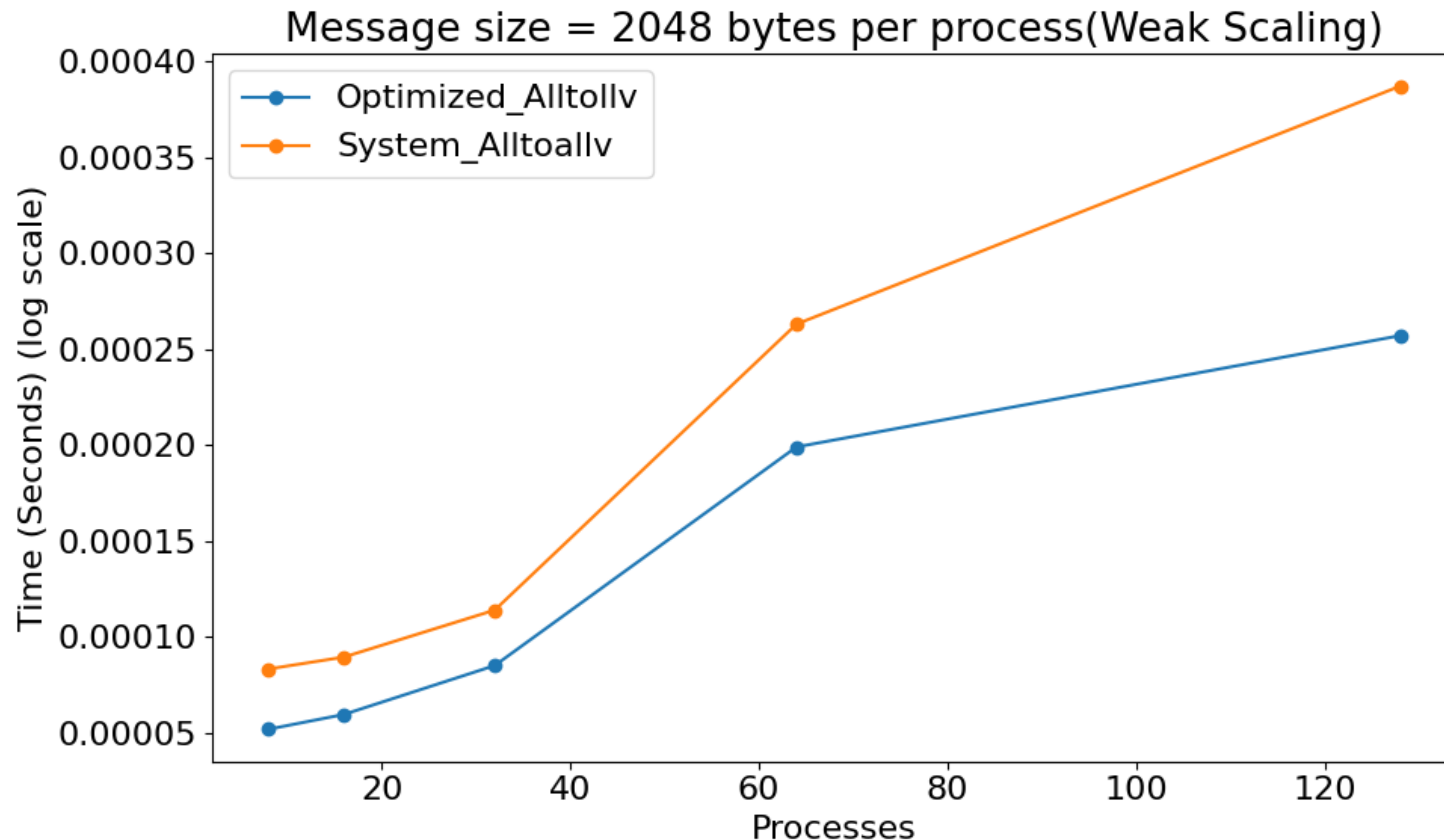
- Dashed: inter-GPU
- Solid: inter-CPU

Multiple Ping-Pong Latencies



- Dashed: manually copy to CPU
- Solid: GPUDirect

Copy-to-CPU Alltoallv



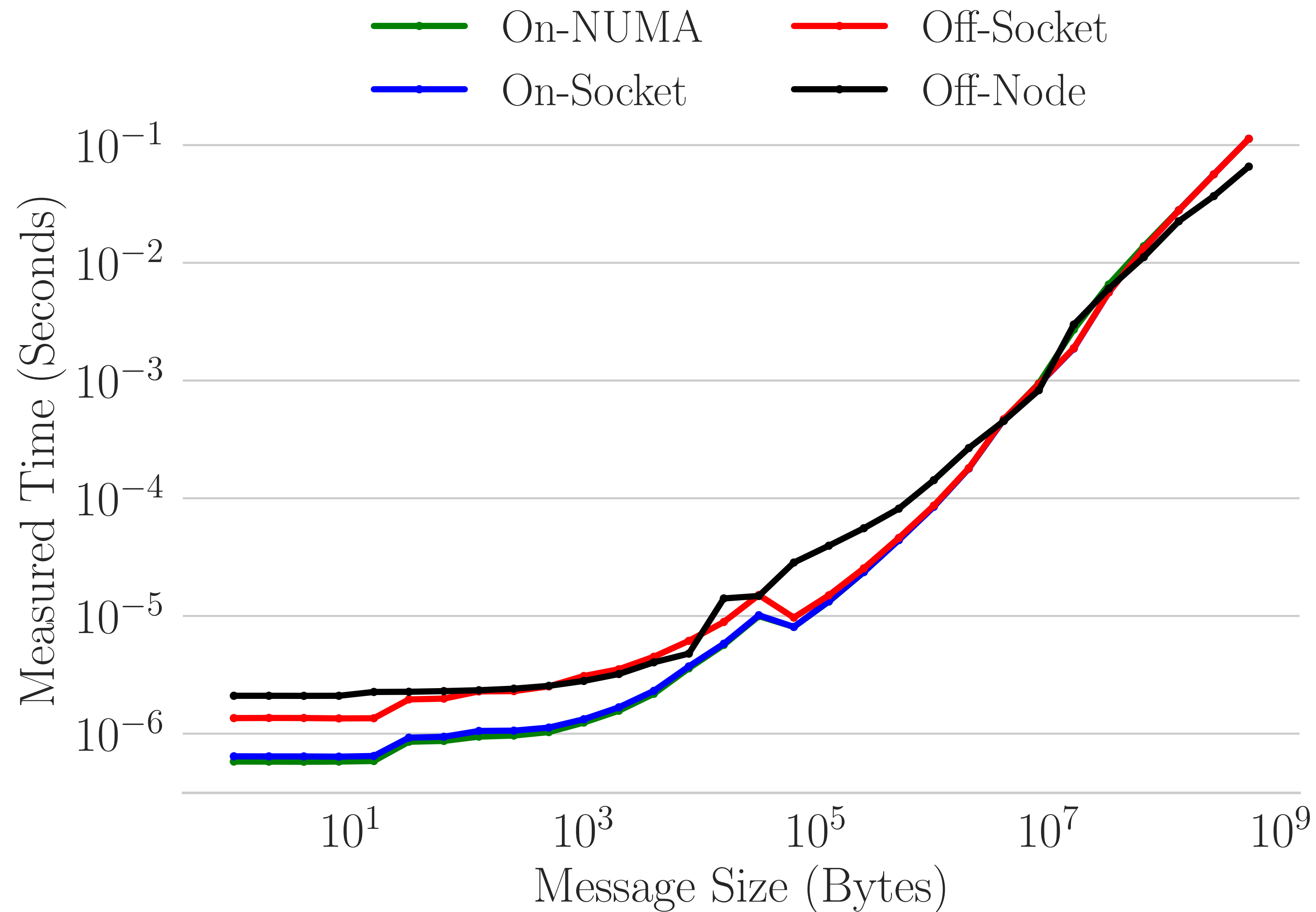
Dane (Sapphire Rapids)

- 112 cores per node
- 2 Sockets per node
- 8 NUMA regions per node
- 2 NICs per node
- OpenMPI

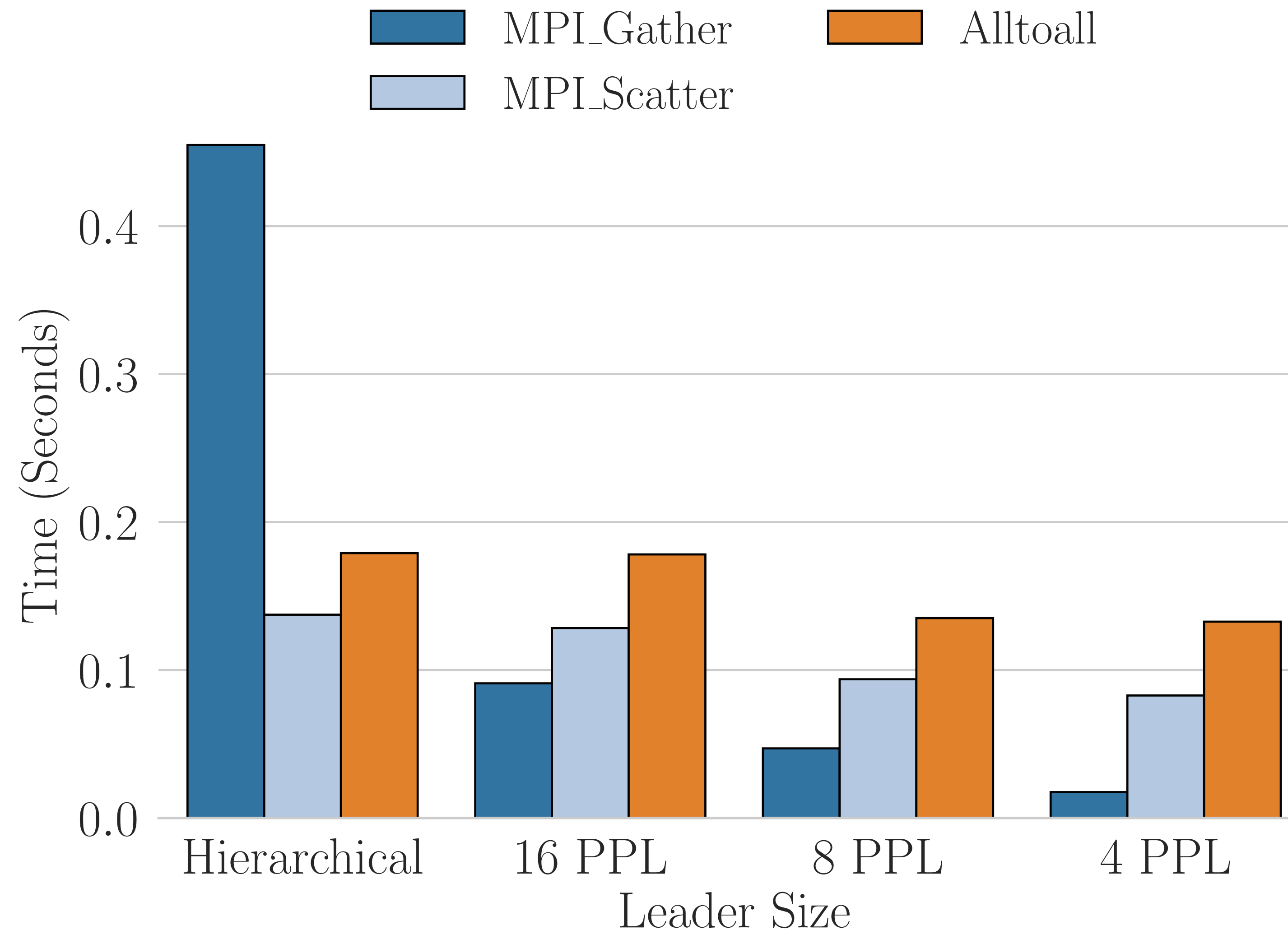


<https://hpc.llnl.gov/hardware/compute-platforms/dane>

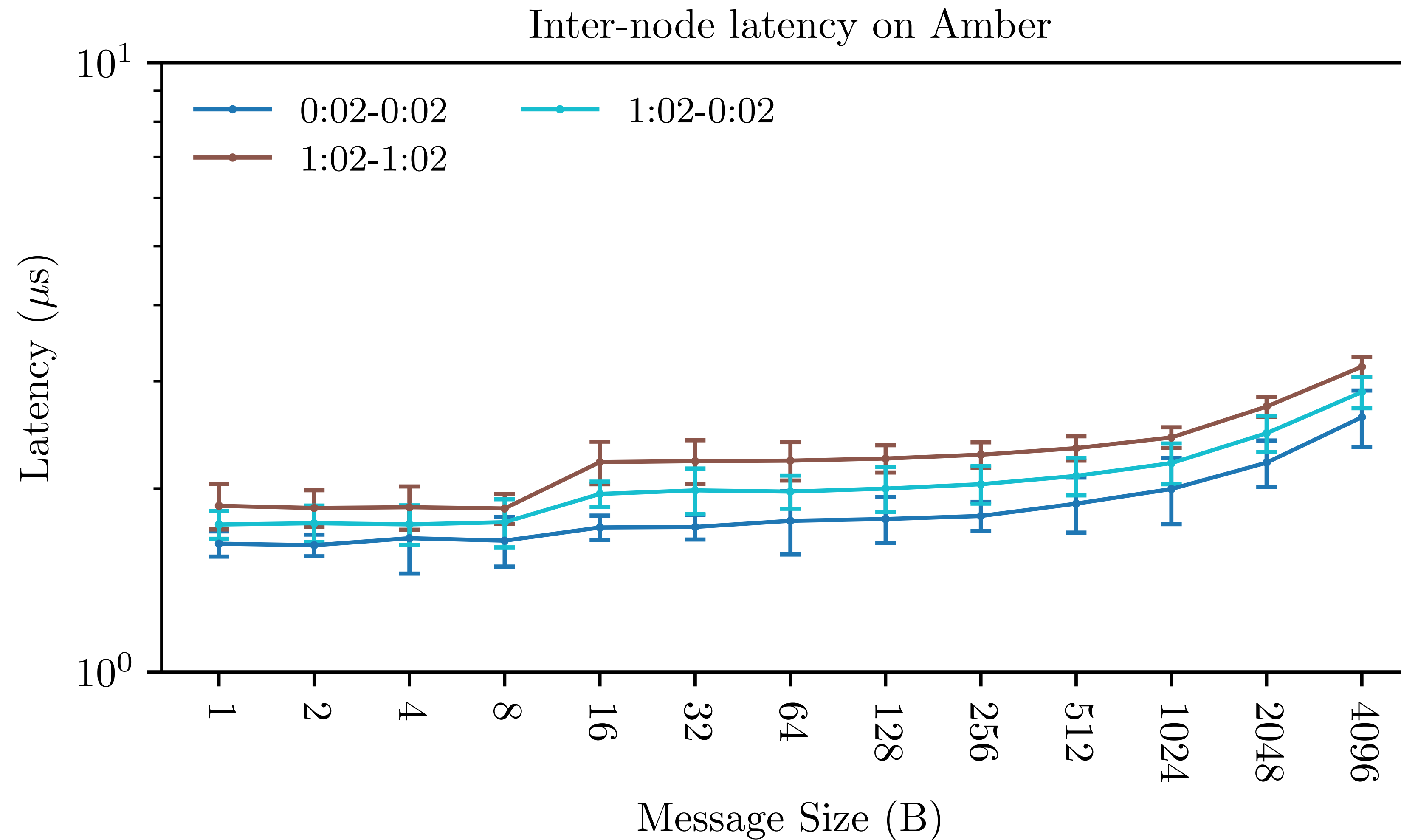
Ping-Pong Latencies



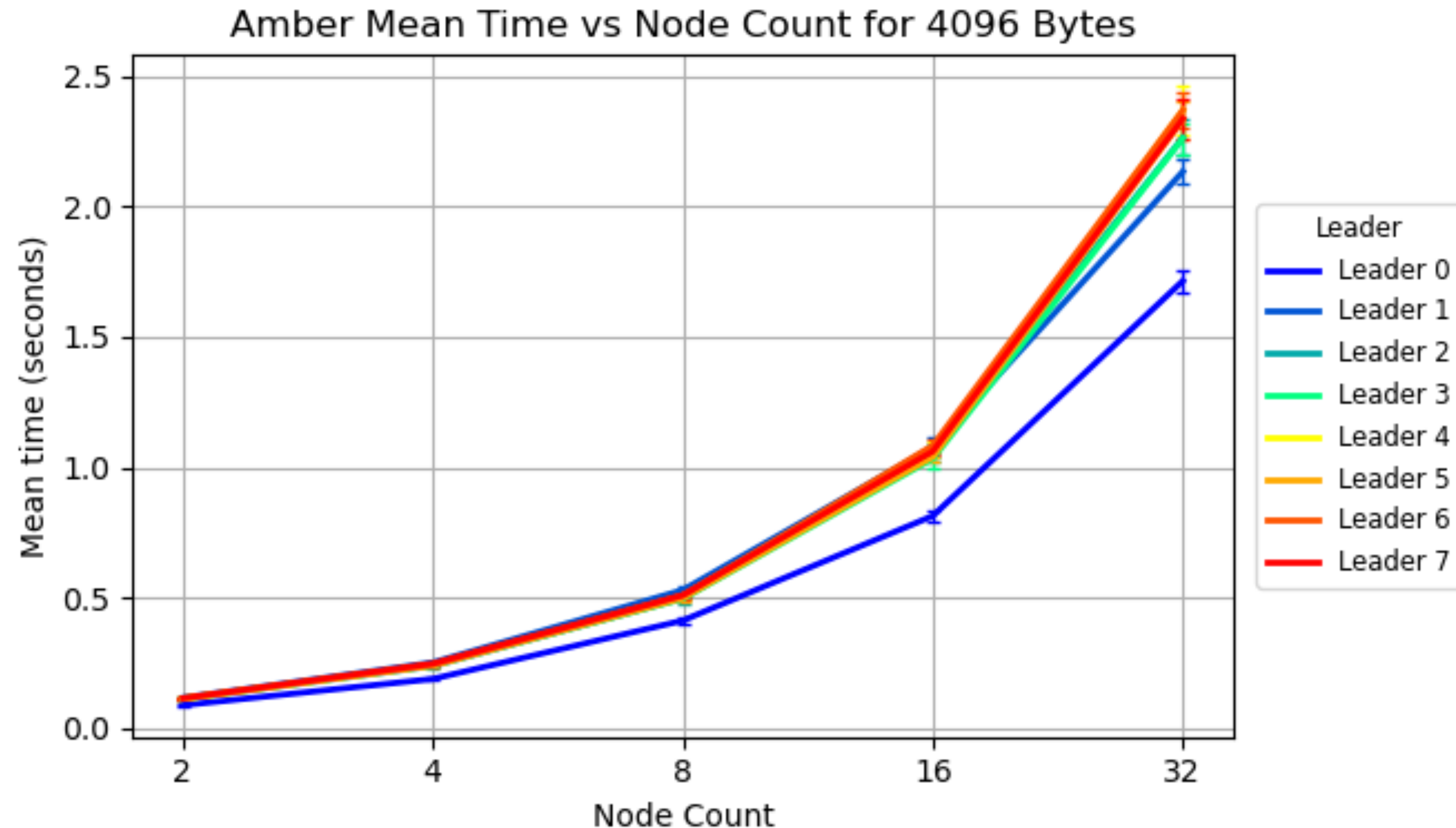
Impact of Inter-NUMA Communication



Per-Socket Performance



Impact of NIC Placement

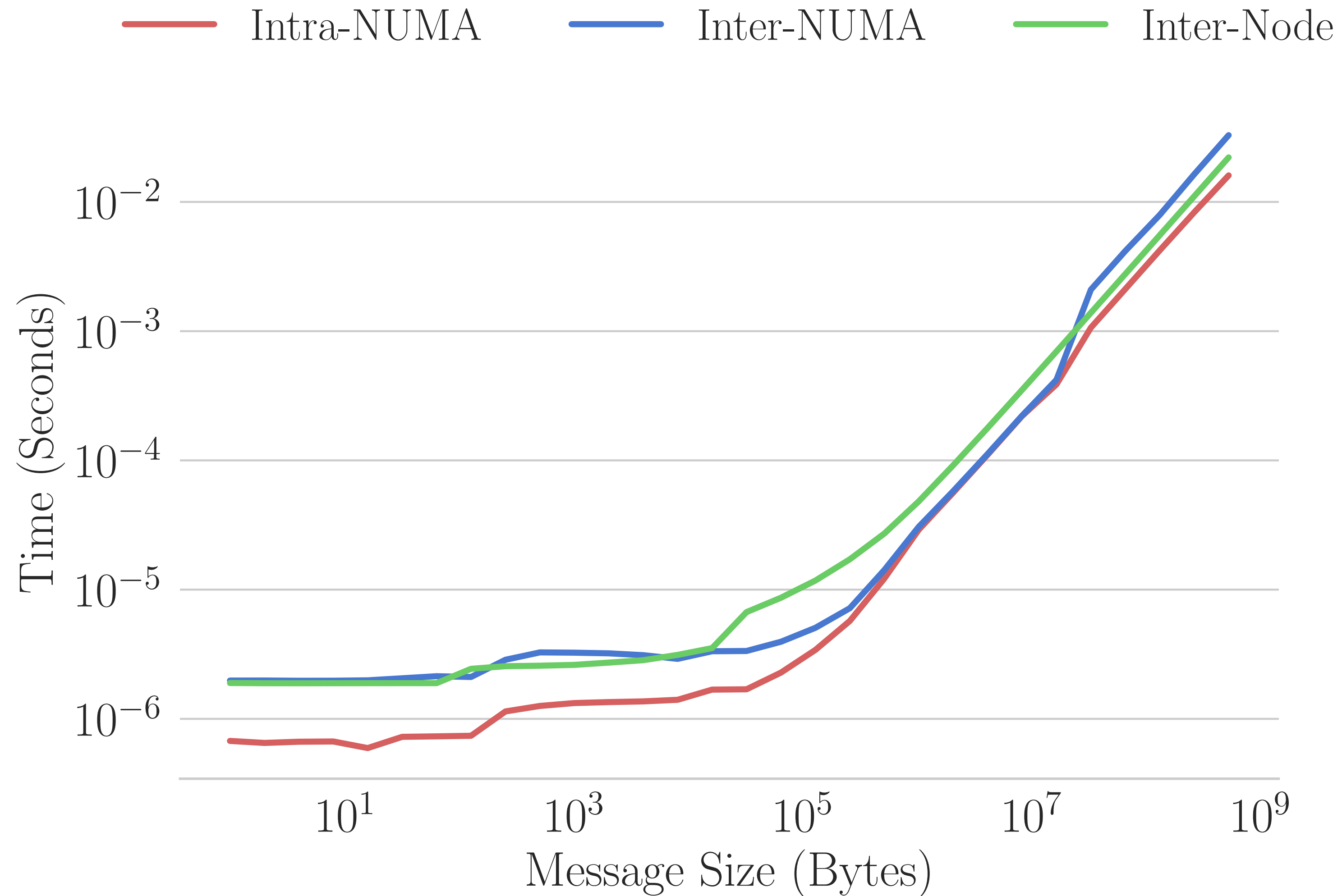


Tuolumne (El Capitan System)

- 4 NUMA regions per node, each containing:
 - 1 AMD MI300A APU
 - 1 GPU, 6 GPU dies
 - 24* CPU cores
 - 1 Network Interface Card (NIC)
- CrayMPI, HIP/ROCm/RCCL

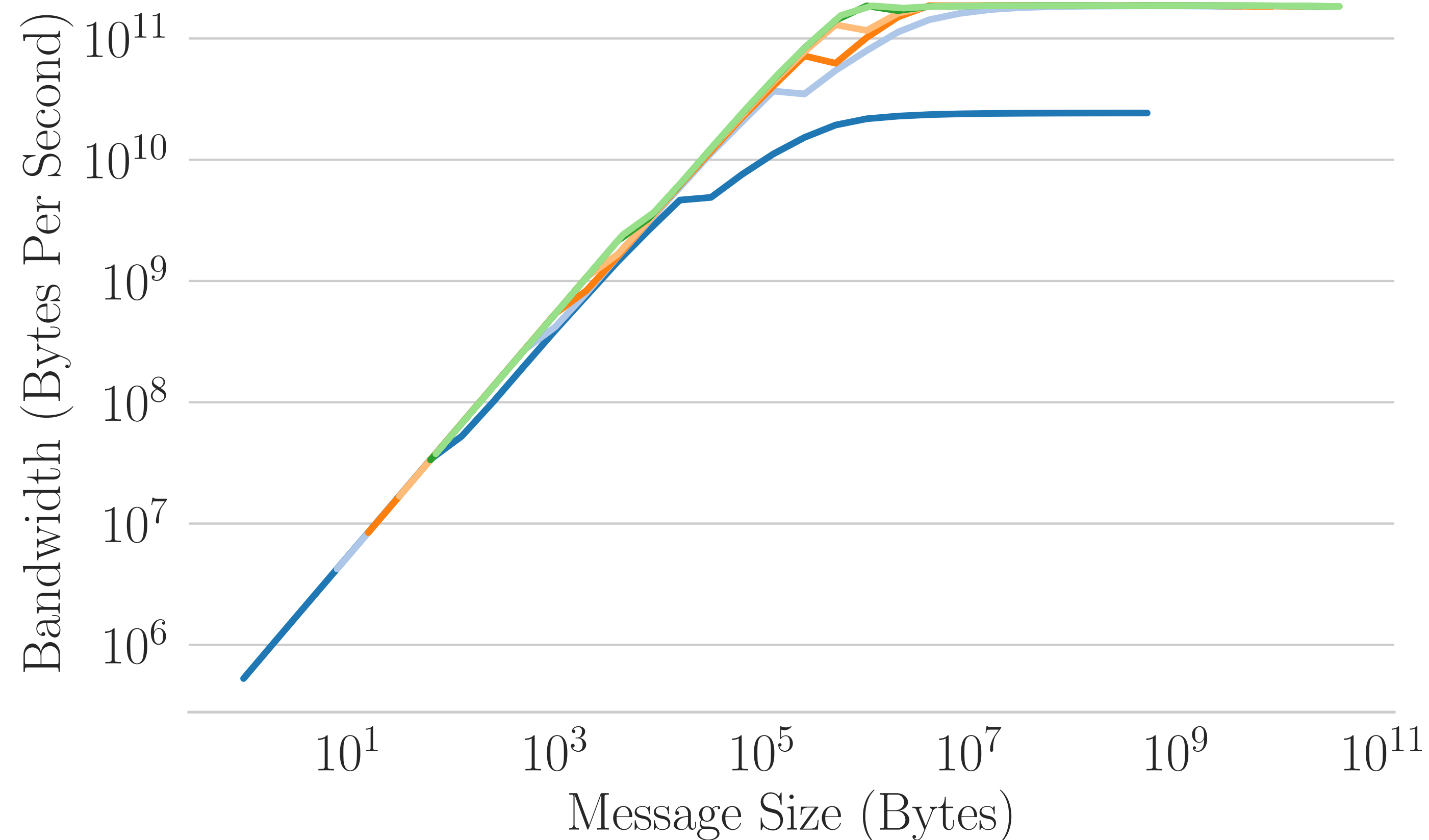


Inter-CPU Ping Pong

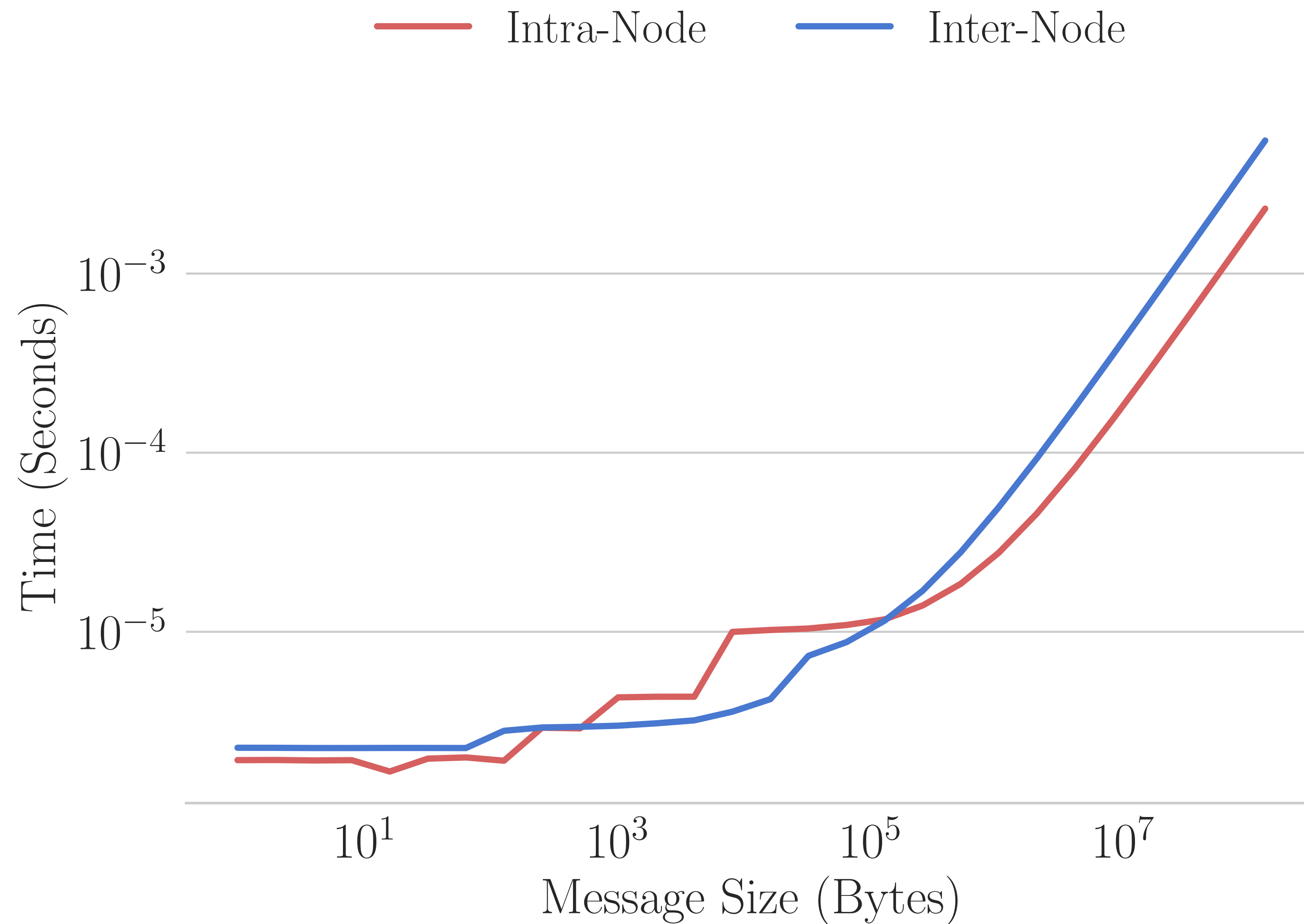


Inter-Node Bandwidth, Even NUMAs

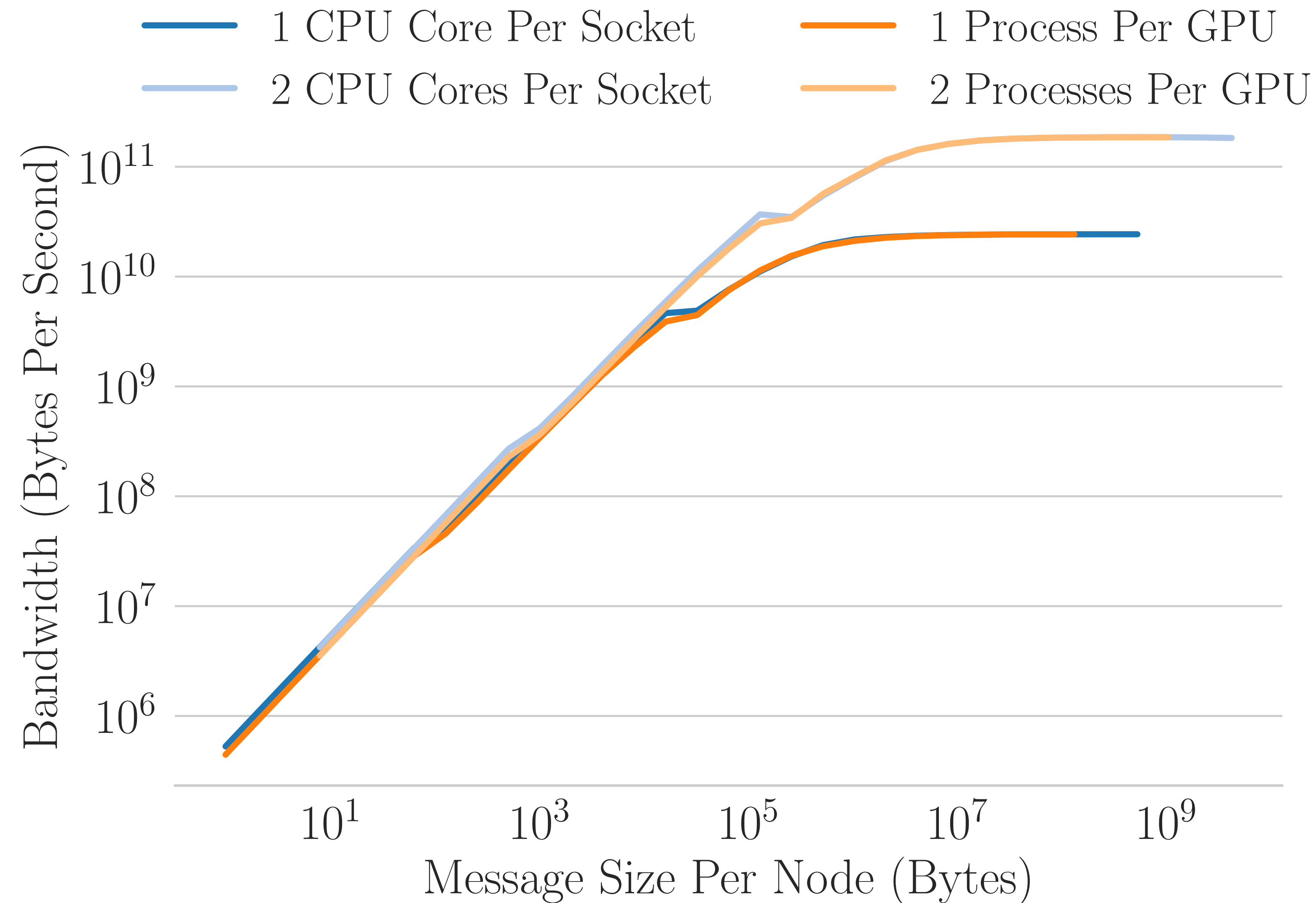
- 1 CPU Core Per Socket
- 2 CPU Cores Per Socket
- 4 CPU Cores Per Socket
- 8 CPU Cores Per Socket
- 16 CPU Cores Per Socket
- 18 CPU Cores Per Socket



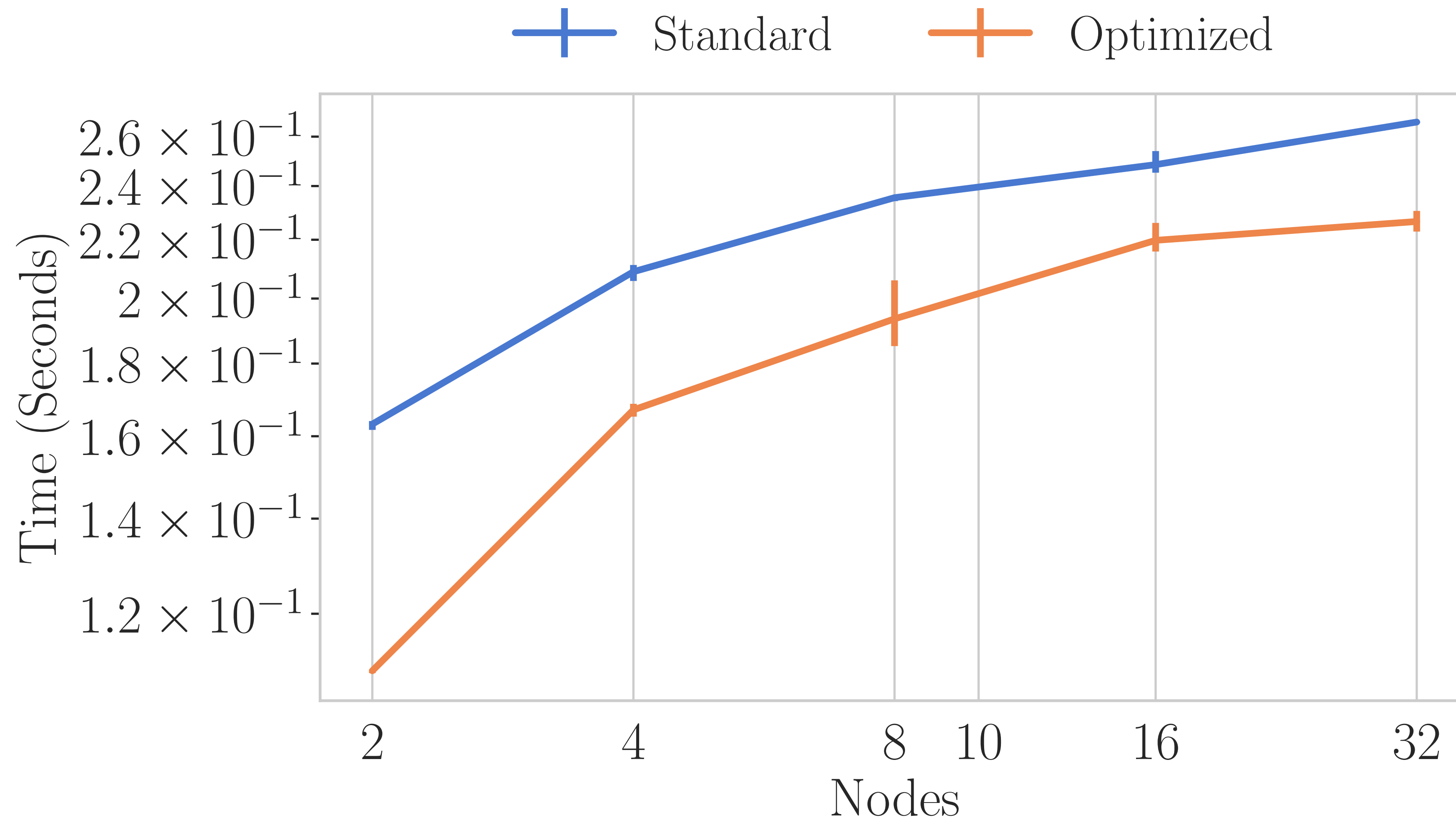
Inter-GPU Ping Pong



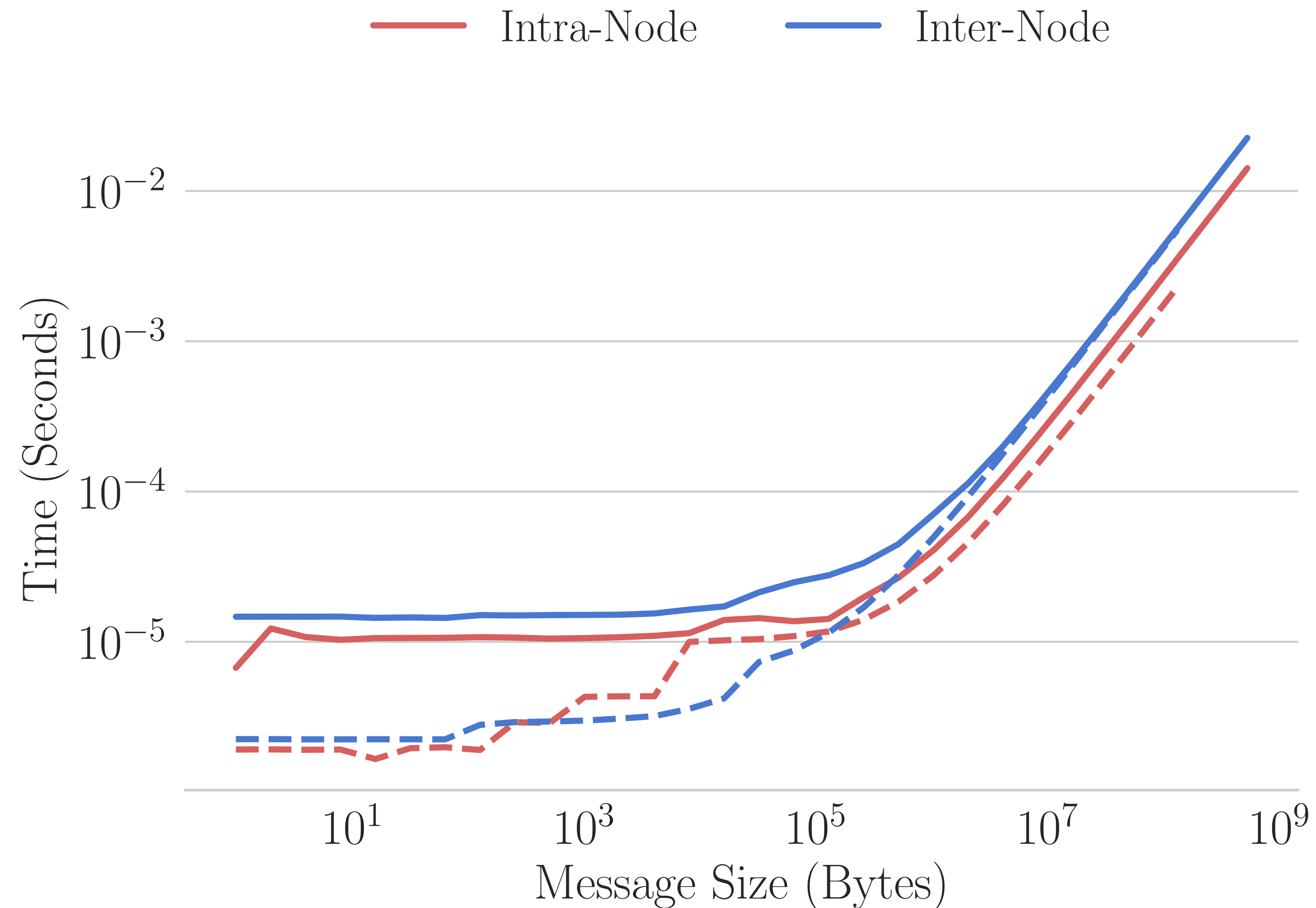
Inter-Node Bandwidth



Optimizing Large Allreduces

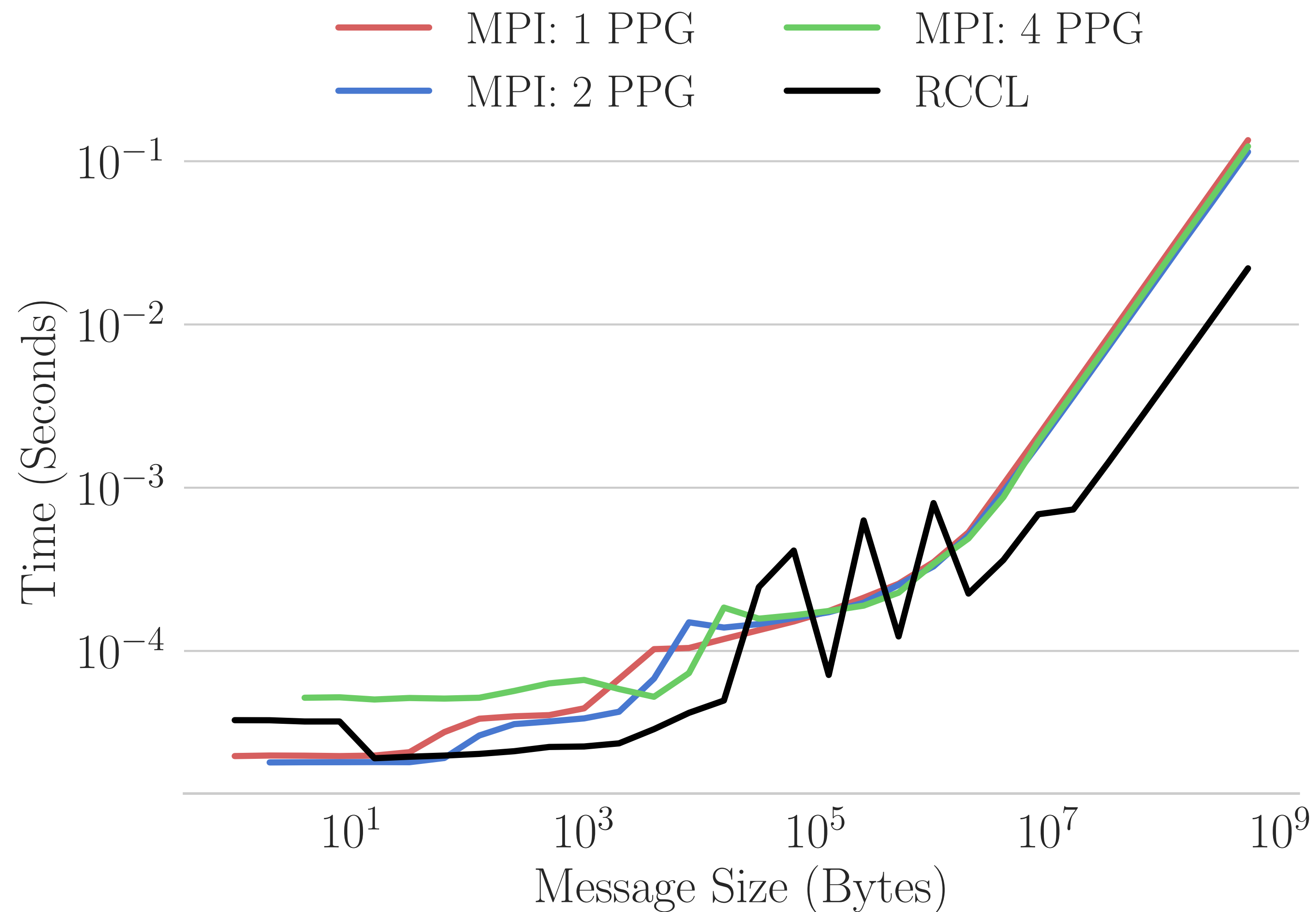


How Does RCCL Compare?

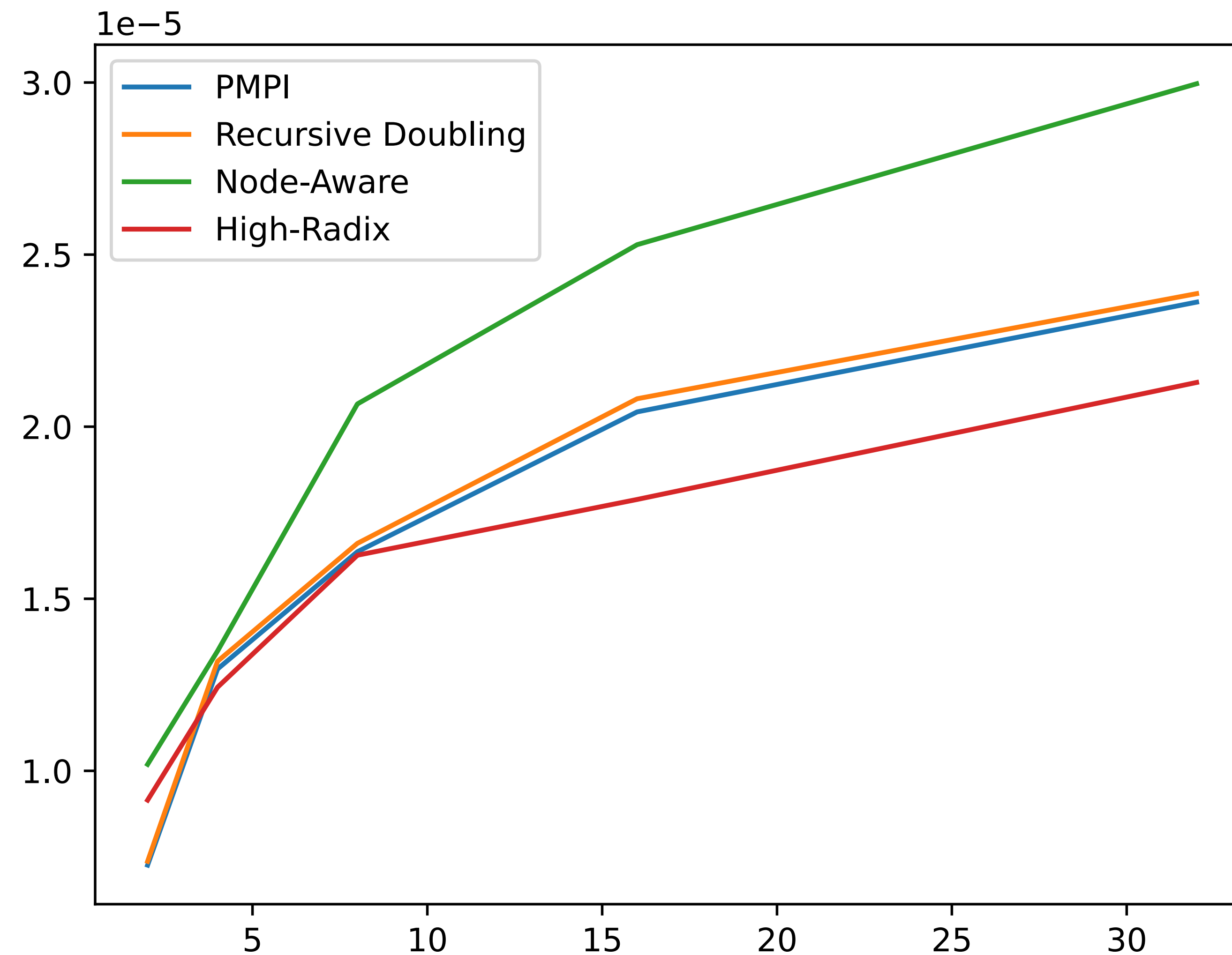


- Dashed: MPI
- Solid: RCCL

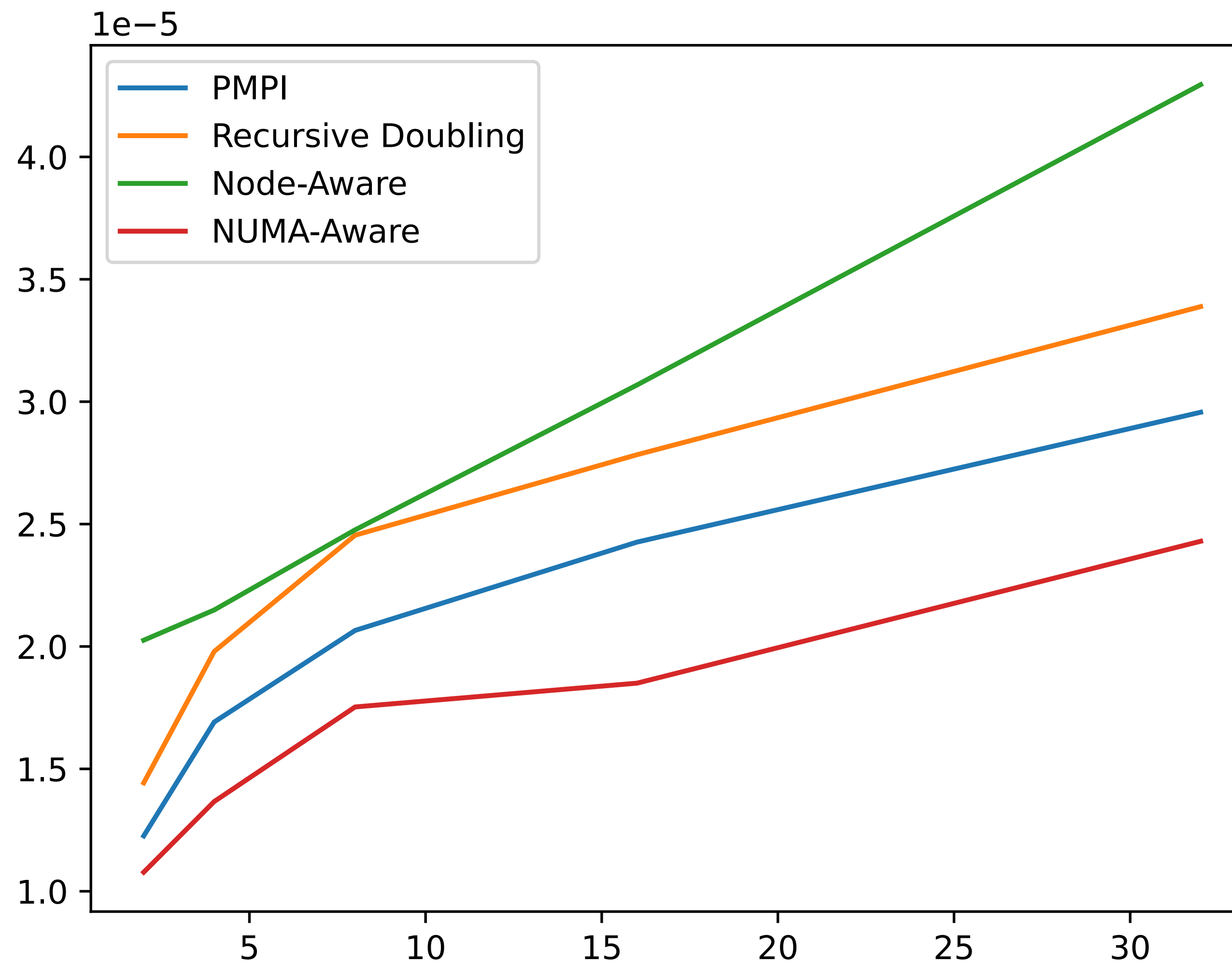
Allreduce: MPI vs RCCL



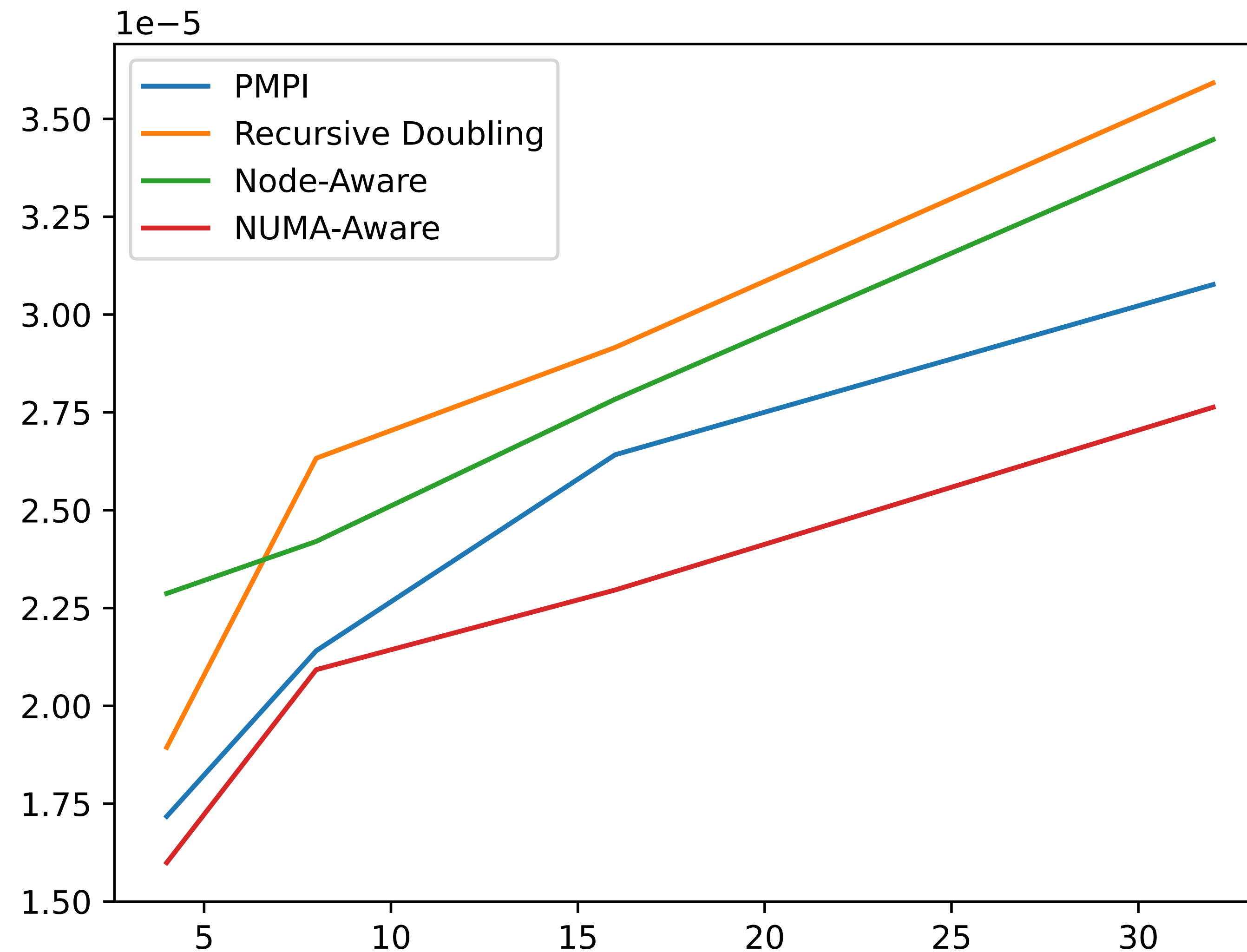
Optimizing Small Allreduces



Optimizing Small Allreduces



Optimizing Small Allreduces



MPI Advance

- <https://github.com/mpi-advance/mpi-advance>

 **mpi-advance**



 **Overview**

 Repositories **12**

 Projects

 Packages

 Teams

 People **5**

 Security and quality

 Insights

 Settings



MPI Advance Project

Better MPI Extensions for a Better HPC World

 **17** followers  United States of America

Unfollow

Popular repositories

[locality_aware](#)

Public

Collective and Neighbor Collective Optimizations and Extensions

 C++  13  18

[MPIPCL](#)

Public

Library implementation of MPI-4 Partitioned Communication

 C  8

[mpi-subset](#)

Public

Compact subset of C++ MPL messaging library.

 C++  7  1

[stream-triggering](#)

Public

An MPI Advance library exploring various stream triggering APIs

 C++  2

 **View as: Public** ▼

You are viewing the README and pinned repositories as a public user.

You can [create a README file](#) or [pin repositories](#) visible to anyone.

[Get started with tasks](#) that most successful organizations complete.

Discussions

Set up discussions to engage with your community!

[Turn on discussions](#)

Thanks!

Questions?

This work was performed with partial support from the National Science Foundation under Grant No. CCF-2338077 and CNS-2450092 and the U.S. Department of Energy's National Nuclear Security Administration (NNSA) under the Predictive Science Academic Alliance Program (PSAAP III \& IV) Awards DE-NA0003966 and DE-NA0004267.