



PARTITIONED COMMUNICATION AND MESSAGE AGGREGATION

Author: W. Pepper Marts

PROBLEM STATEMENT

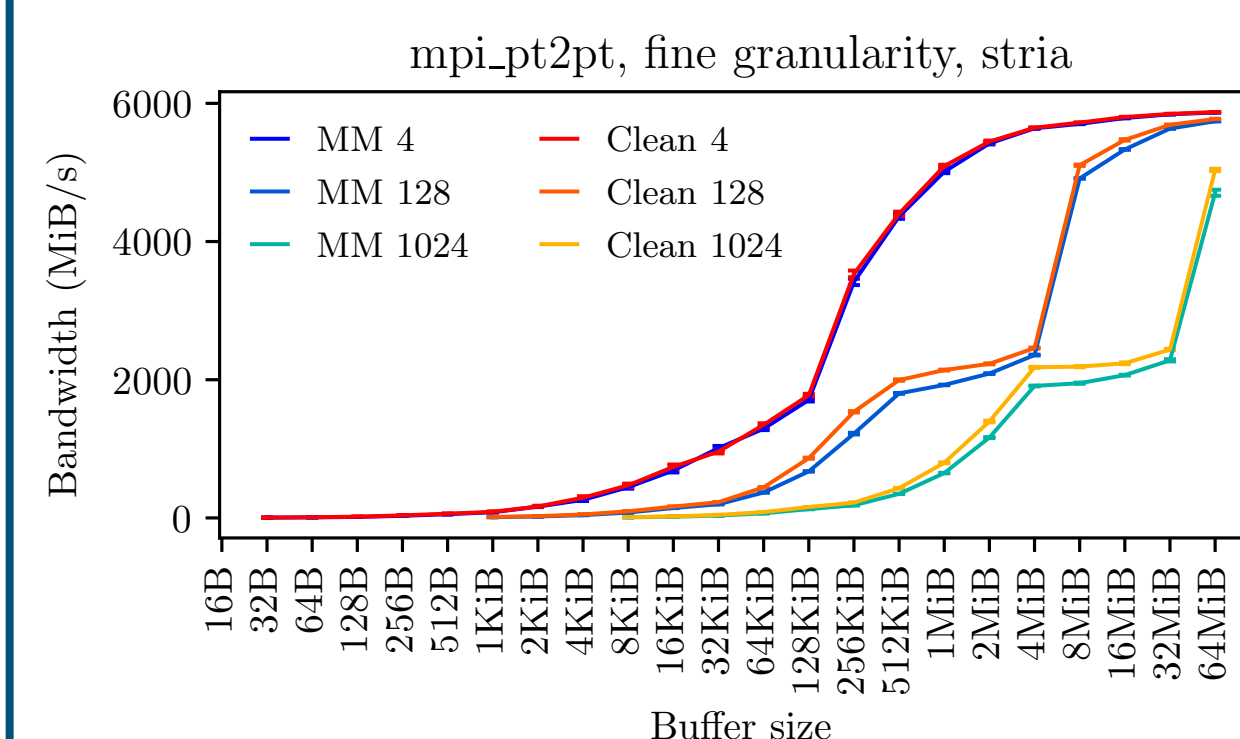
There is interest in threaded, early-bird communication both from the middleware and application communities.

- What are the potential performance benefits?
- What applications can achieve these benefits?
- What modifications are necessary?

MESSAGE AGGREGATION

The tuning of messaging granularity is necessary for performant fine grained application communication.

- Messages too Big → No early bird perceived bandwidth bump
- Messages too small → Lower bandwidth/hardware utilization



Bandwidth Analysis:

- Results are from a fine grained ping-pong test.
- Clean corresponds to a hard-coded equivalent
- Legend number are number of partitions.
- Overhead is very low.
- Network utilization very sensitive to granularity configuration.

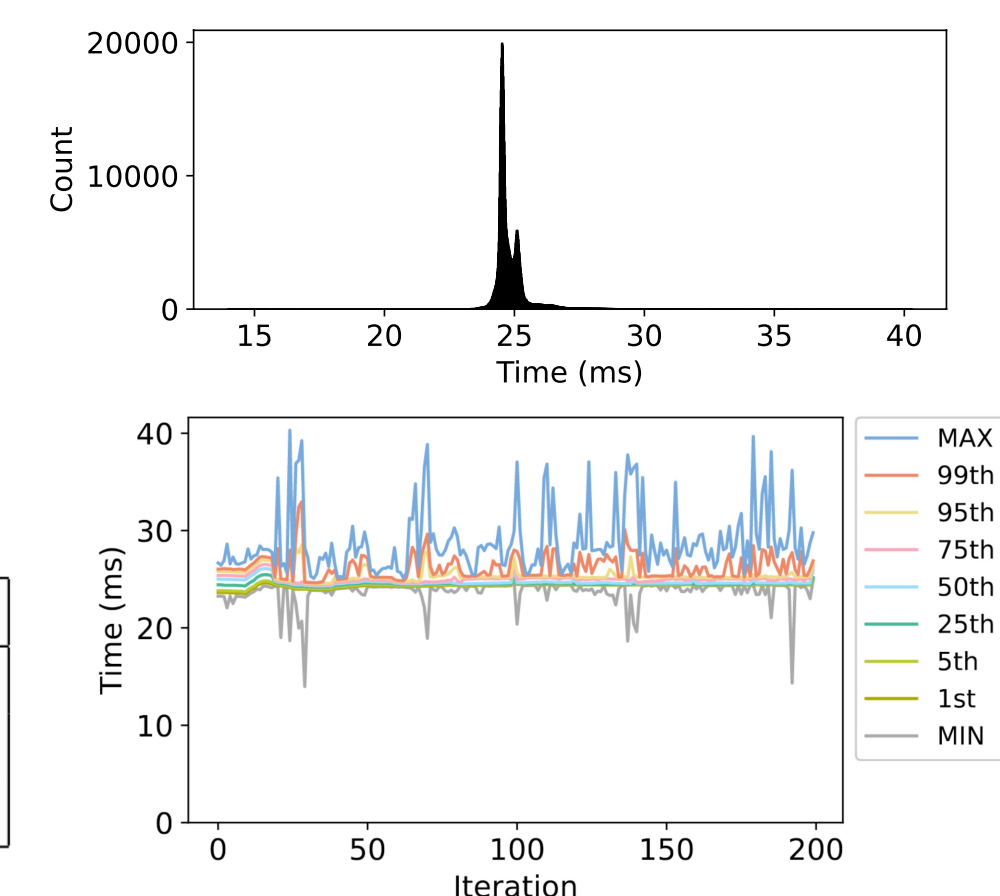
THREAD TIMING

Partitioned communication facilitates overlapping communication and computation. The performance gains are limited by reclaimable time.

Results:

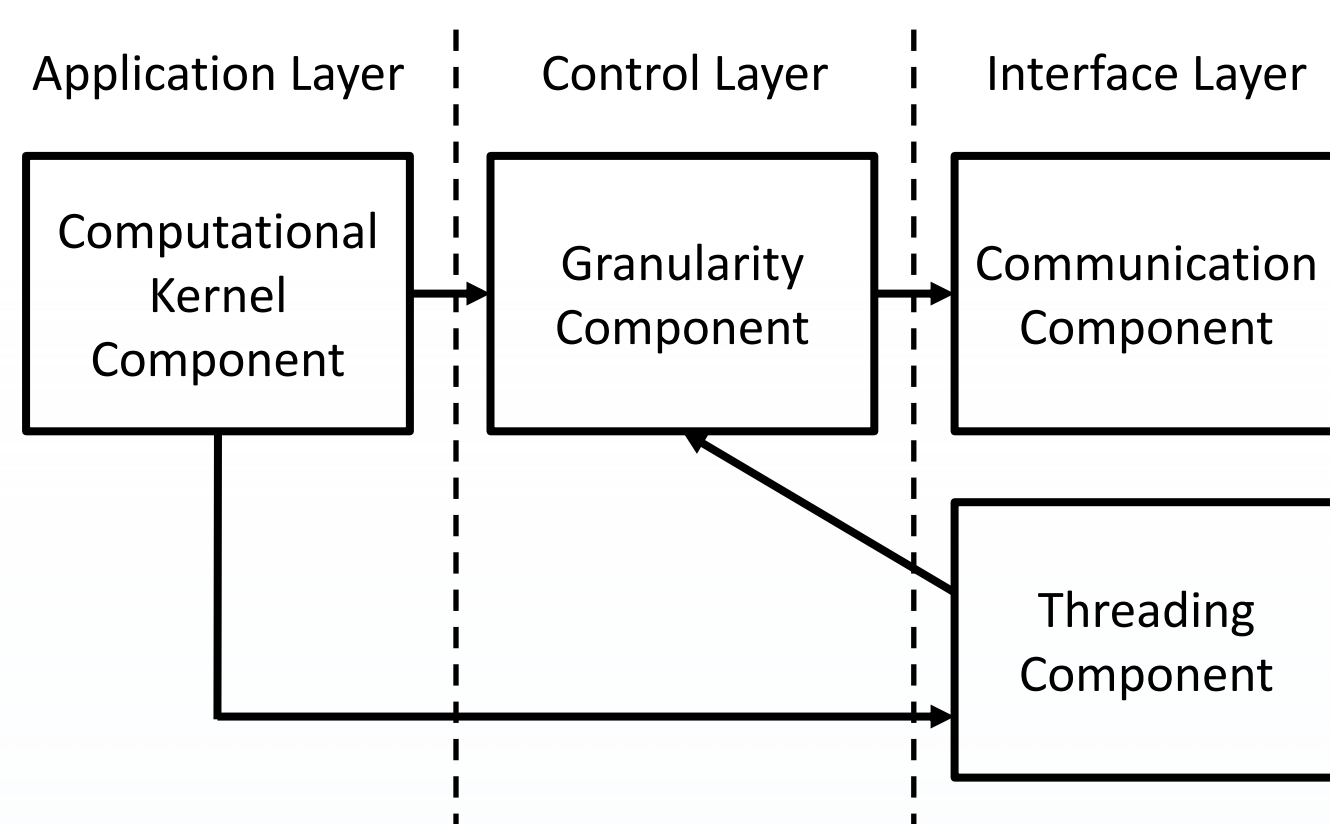
- Thread timings for miniFE, miniMD, miniQMC.
- Normalcy is common, not guaranteed.
- Laggard threads are common, but wont occur in most iterations.
- Distribution varies across application times step.

Test	MiniFE	MiniMD	MiniQMC
D'Agostino	3%	77%	95%
Shapiro-Wilk	< 1%	74%	96%
Anderson-Darling	< 1%	76%	96%



W. Pepper Marts, Matthew G. F. Dosanjh, Whit Schonbein, Scott Levy, and Patrick G. Bridges, "Measuring thread timing to assess the feasibility of early-bird message delivery," CoRR, vol. abs/2304.11122, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2304.11122>

MODULAR TESTING FRAMEWORK



MiniMod Framework:

- Runtime Configurable
- Compartmentalized Effort
- Low Overhead

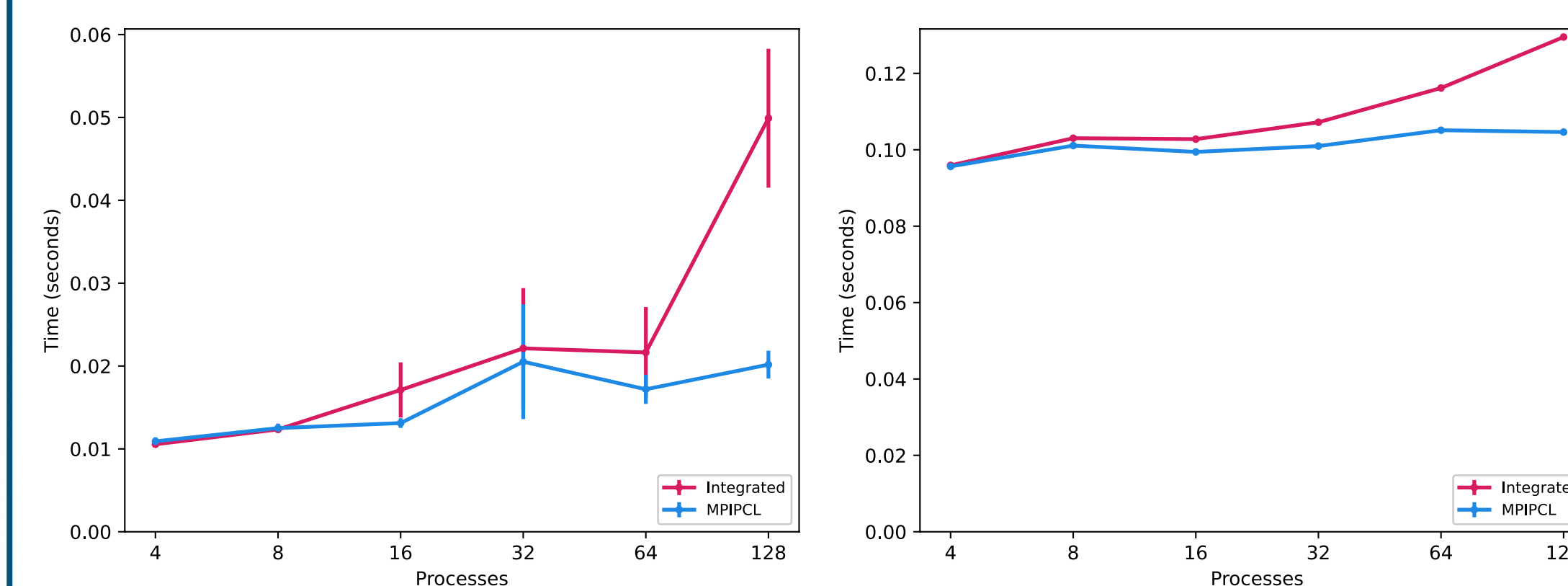
Novel Features:

- Fine-grained communication
- MPI Partitioned Comm.
- Message aggregation
- Threaded offloading

MiniMod allows the runtime configuration of an application's communication and threading.

W Pepper Marts, Matthew GF Dosanjh, Scott Levy, Whit Schonbein, Ryan E Grant, and Patrick G Bridges. MiniMod: A modular miniapplication benchmarking framework for HPC. In 2021 IEEE International Conference on Cluster Computing (CLUSTER), pages 12–22. IEEE, 2021.

PARTITIONED COMMUNICATION

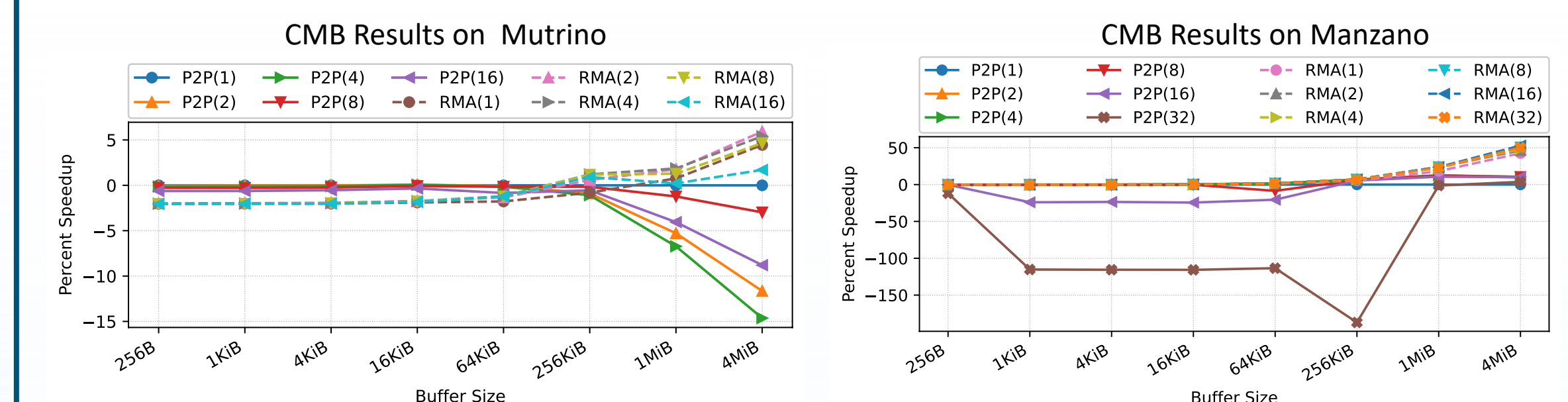


Implementations are being refined and optimized. Spec compliant implementations are not capable of true persistence and do not offer substantive benefit over other MPI interfaces.

W. Pepper Marts, Andrew Worley, Prema Soundarajan, Derek Schafer, Matthew G. F. Dosanjh, Ryan E. Grant, Purushotham V. Bangalore, Anthony Skjellum, and Sheikh Ghafoor. 2023. Design of a portable implementation of partitioned point-to-point communication primitives. *Concurrency and Computation: Practice and Experience* (2023). DOI:<https://doi.org/10.1002/cpe.7655>

CONFIGURABLE BENCHMARK

The Configurable Messaging Benchmark (CMB) is designed to replicate the communication access patterns of fine-grained applications.



Percent speedup for fine-grained binning relative to BSP MPI message passing for application profiles with laggard thread arrival distributions. Each line corresponds to a fine-grained implementation. Results shown use a 4% laggard delay parameter and 27-point stencils.