



# Persistent and Partitioned MPI for Regular Halo Exchange

Gerald Collom  
University of New Mexico

Contributions from Olga Pearce, Jason Burmark, David  
Boehme (LLNL) and Puri Bangalore (Univ. of Alabama at  
Birmingham)



# Background



- Domain decomposition for each iteration:
  - compute
  - communicate
- Regular Halo Exchange
  - Molecular Dynamics, Dislocation Dynamics, Hydrodynamics
- Supercomputer advances (GPU's) require testing new modes of communication
- Comb: communication benchmark
  - <https://github.com/LLNL/Comb>

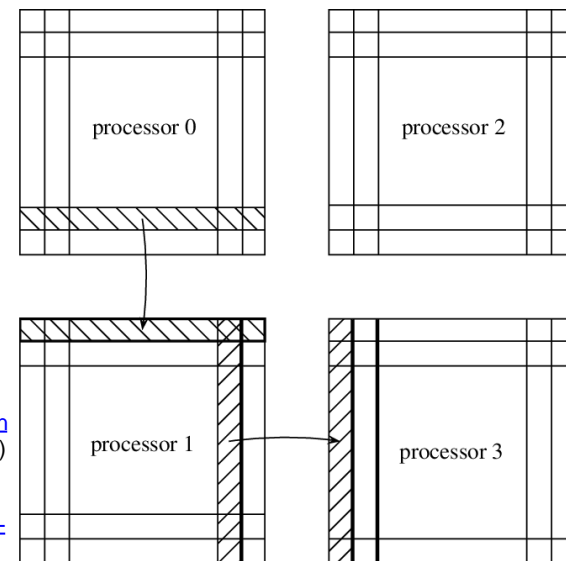
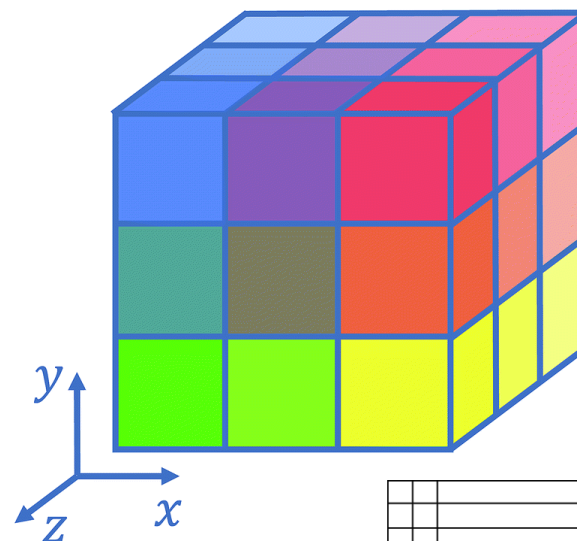


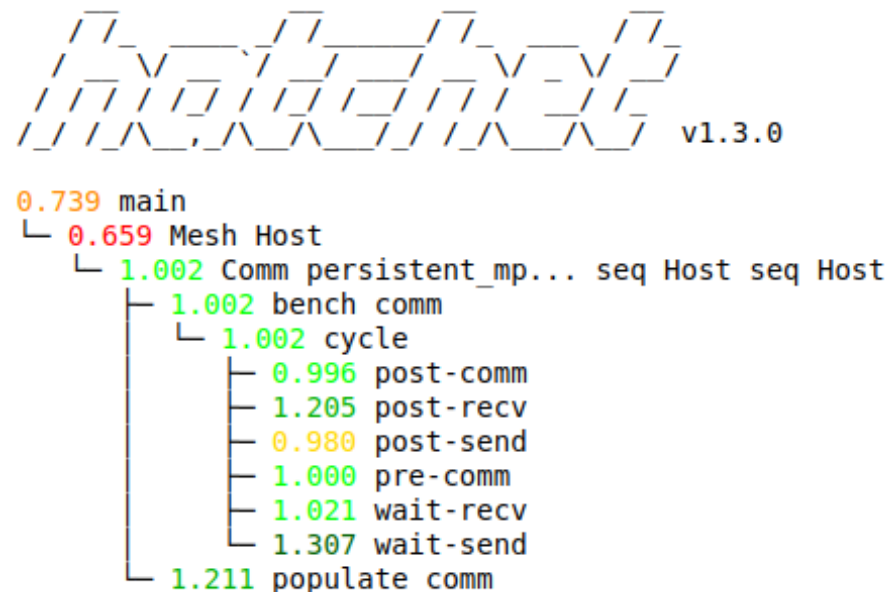
Image sources:

<https://pubs.rsc.org/en/content/articlelanding/2018/SC/C7SC04531J#divAbstract> (top)  
[https://www.researchgate.net/figure/Two-stages-of-a-halo-exchange-in-two-dimensions-Note-the-corner-is-transmitted-from\\_fig3\\_7660293](https://www.researchgate.net/figure/Two-stages-of-a-halo-exchange-in-two-dimensions-Note-the-corner-is-transmitted-from_fig3_7660293) (bottom)

# Persistent and Partitioned MPI



- Persistent MPI: reuse message arguments
  - MPI\_Send\_init, MPI\_Recv\_init
  - MPI\_Start
  - MPI\_Request\_free
  
- Partitioned MPI (MPIPCL): separate parts each sent when ready
  - MPIX\_Psend\_init, MPIX\_Precv\_init
  - MPIX\_Pready, MPIX\_Pready\_range
  - MPIX\_Request\_free



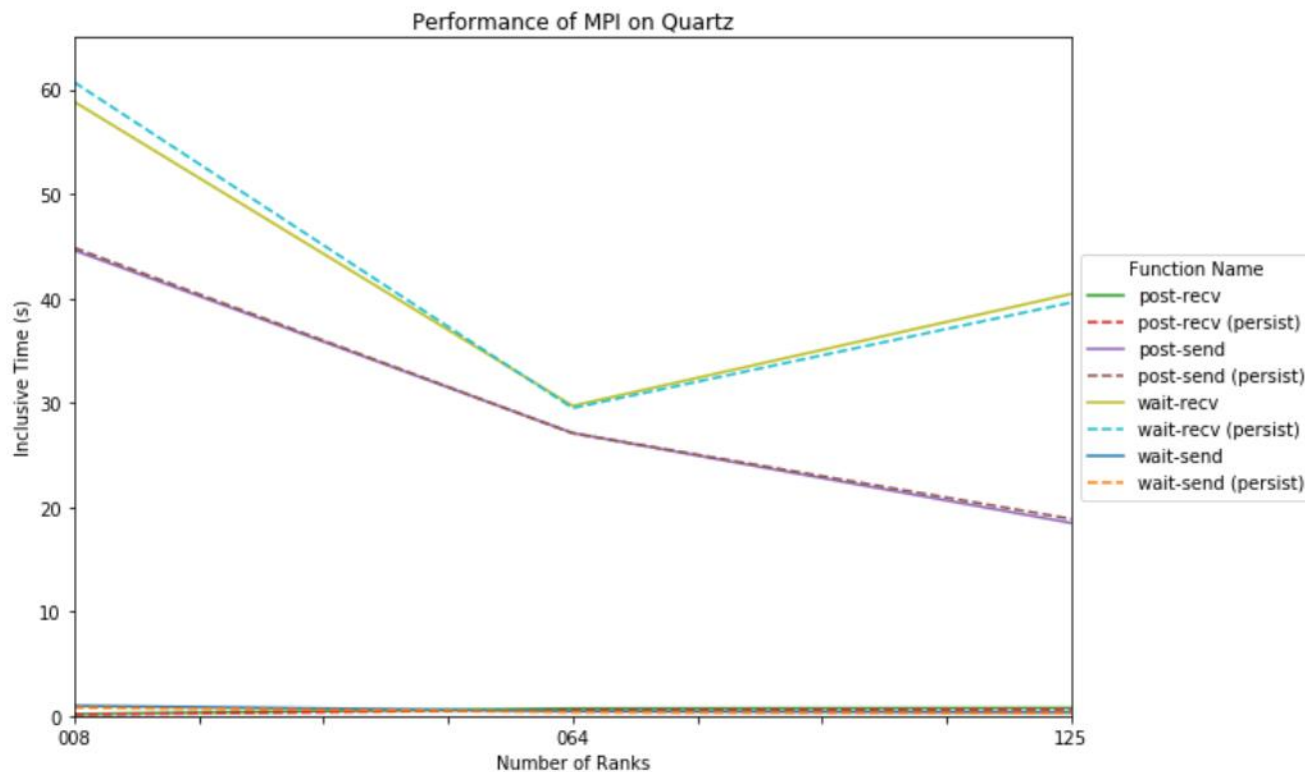
<https://github.com/hatchet/hatchet> (visualization)

<https://github.com/LLNL/Caliper> (timing)

# Results



- Successful test implementation of persistent and partitioned MPI on CPU's



## Future Work

- Performance analysis of partitioned MPI
- GPU implementations