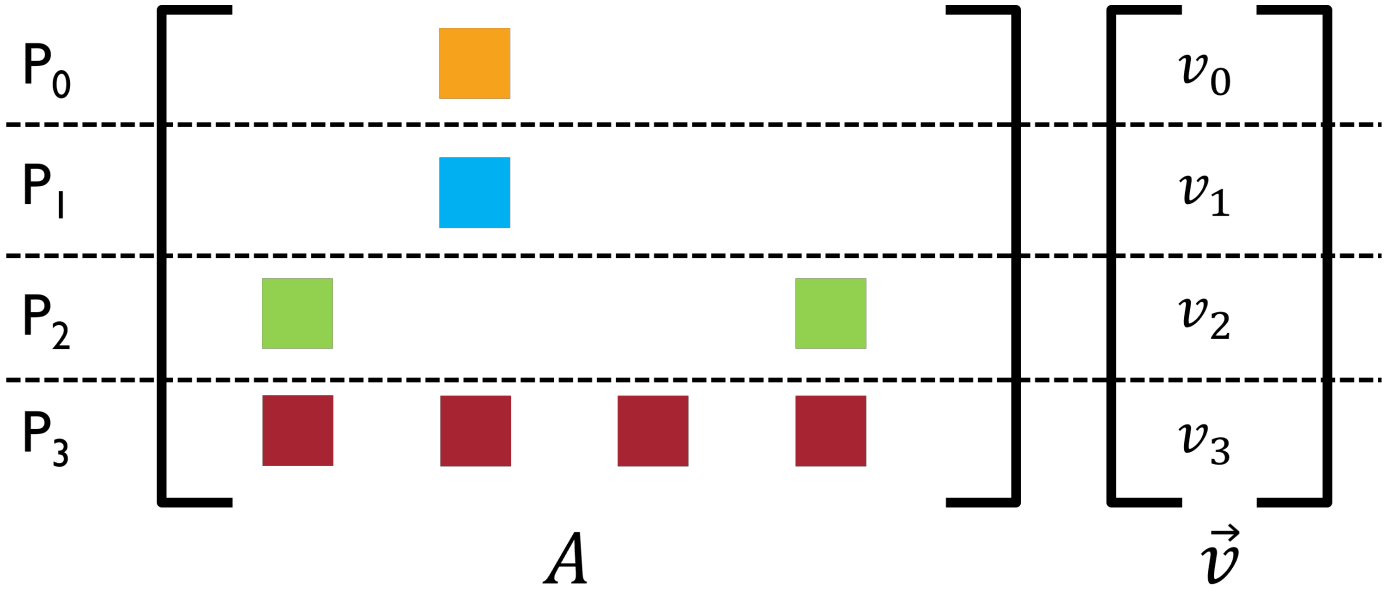
A More Scalable Sparse Dynamic Data Exchange

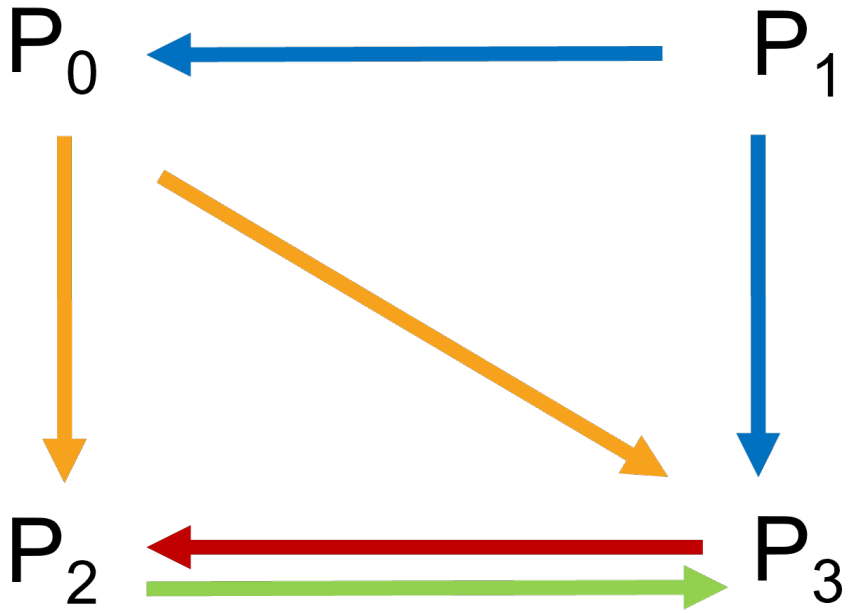
Andrew Geyko*, Gerald Collom, Derek Schafer, Patrick Bridges, Amanda Bienz

*Max Planck Institute
University of New Mexico

CUP ECS
Center for Understandable, Performant Exascale Communication Systems
THE UNIVERSITY OF NEW MEXICO

Motivating Example 1: Sparse Matrix Operations

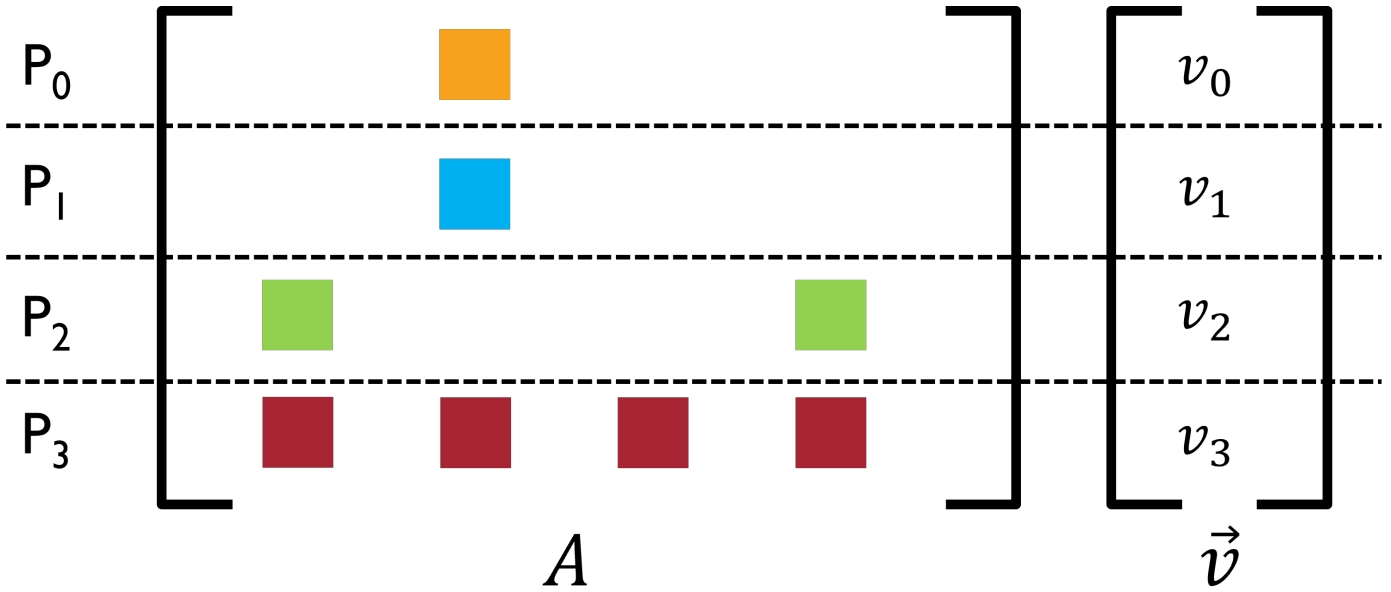


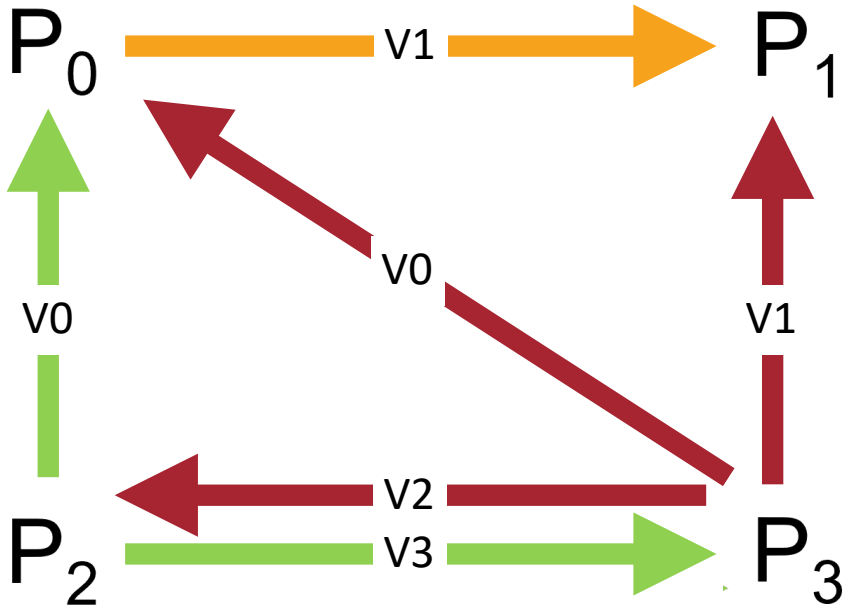


Each process *knows* processes from which it receives and what it receives from each

Each process *does not know* processes to which it sends or what it sends to each

Topology Discovery: MPIX_Alltoallv_crs

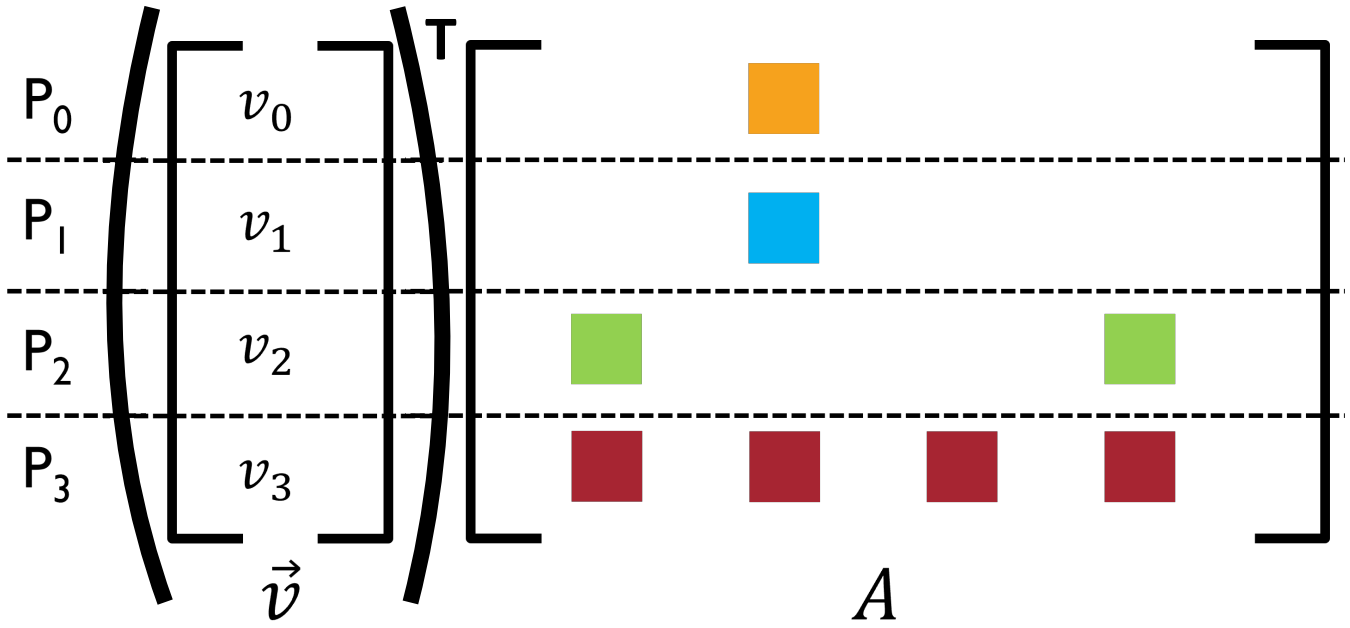




Each process sends a message to every process from which it wants to receive data

Containing all data indices it wants to receive

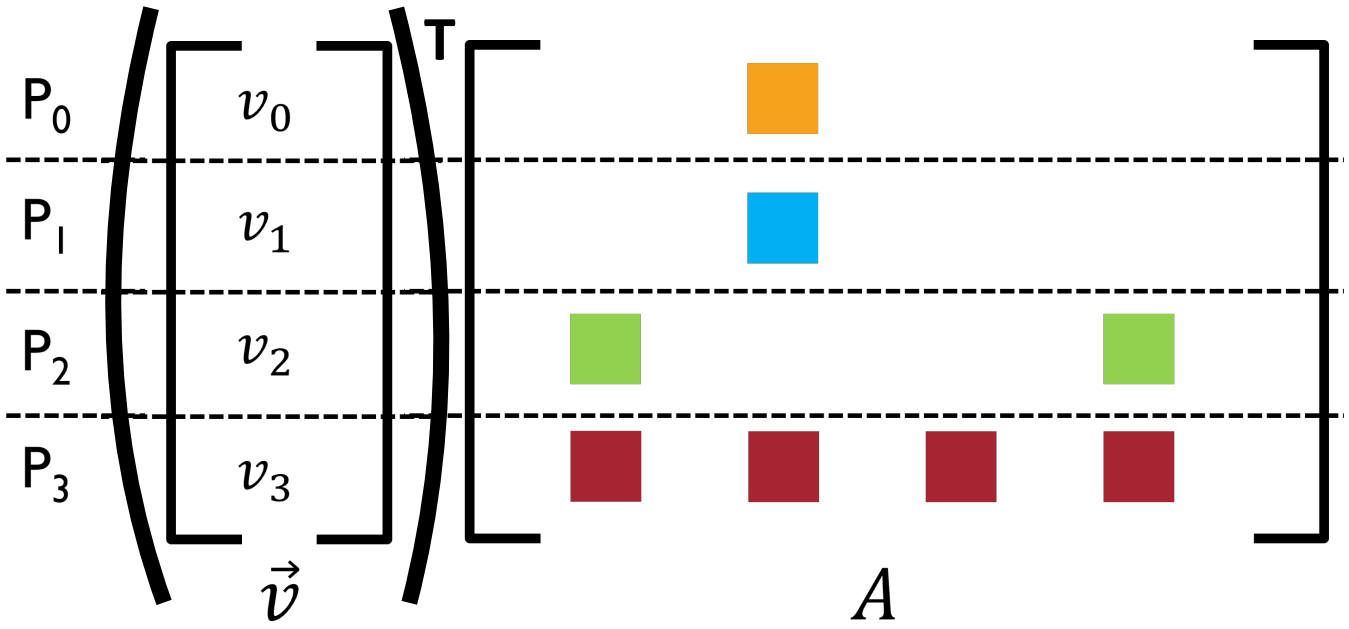
Motivating Example 2: Transpose Sparse Matrix Operations

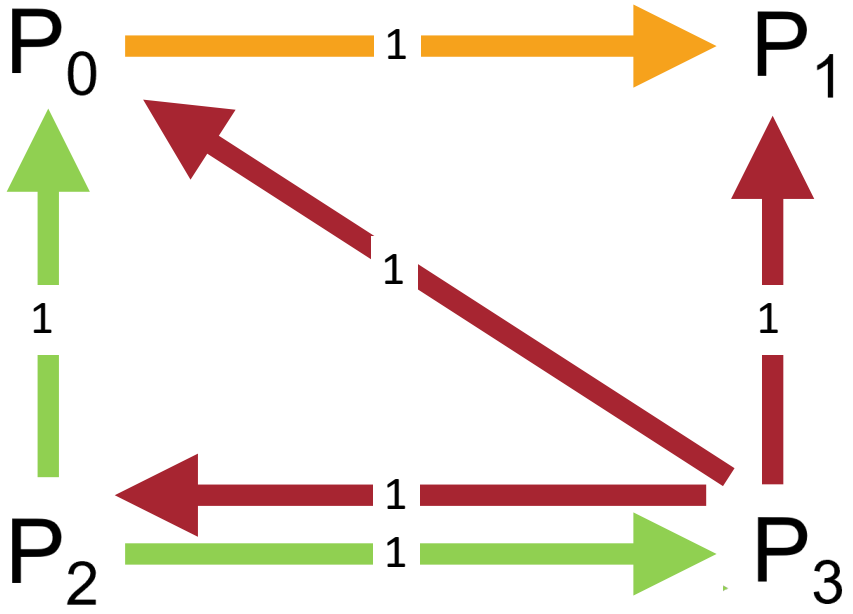


Each process *knows* processes to which it sends and what it sends to each

Each process *does not know* processes from which it receives or what it receives from each

Topology Discovery: MPIX_Alltoall_crs

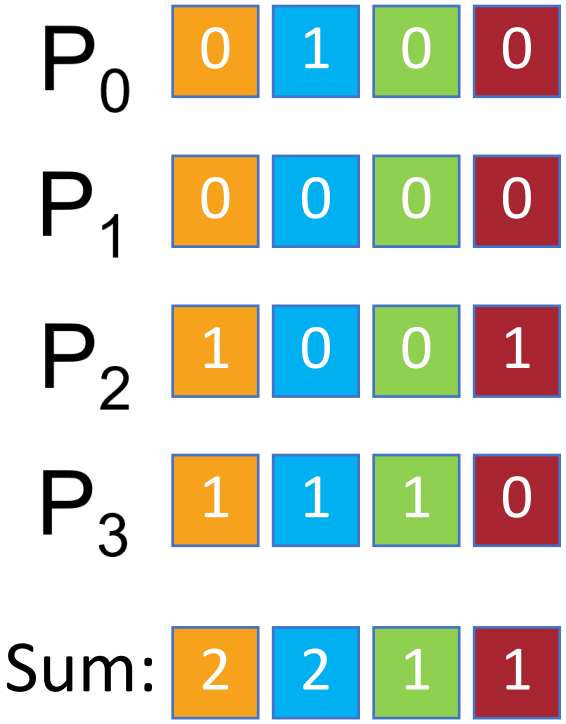




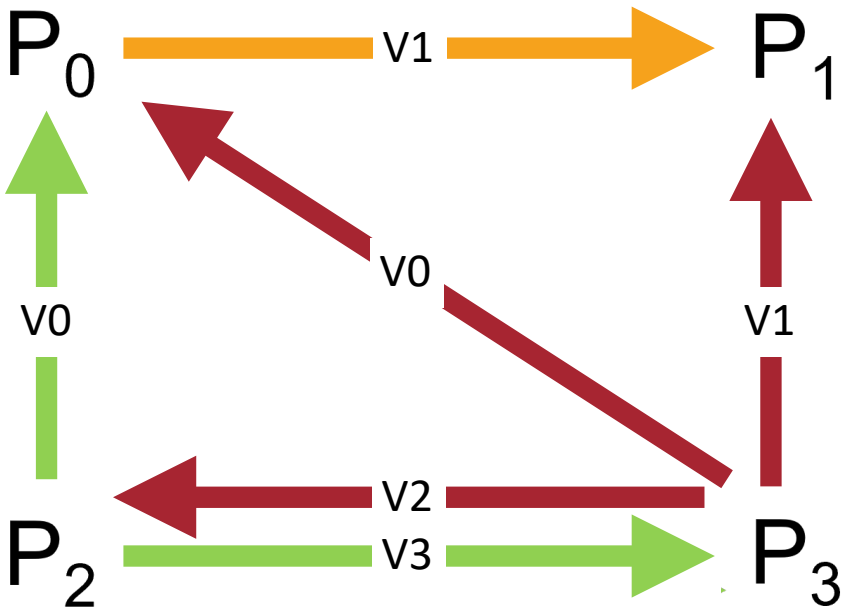
Each process sends a message to every process to which it wants to send data
Containing the number of values it will send

Standard Algorithm: Personalized

Step 1: AllReduce



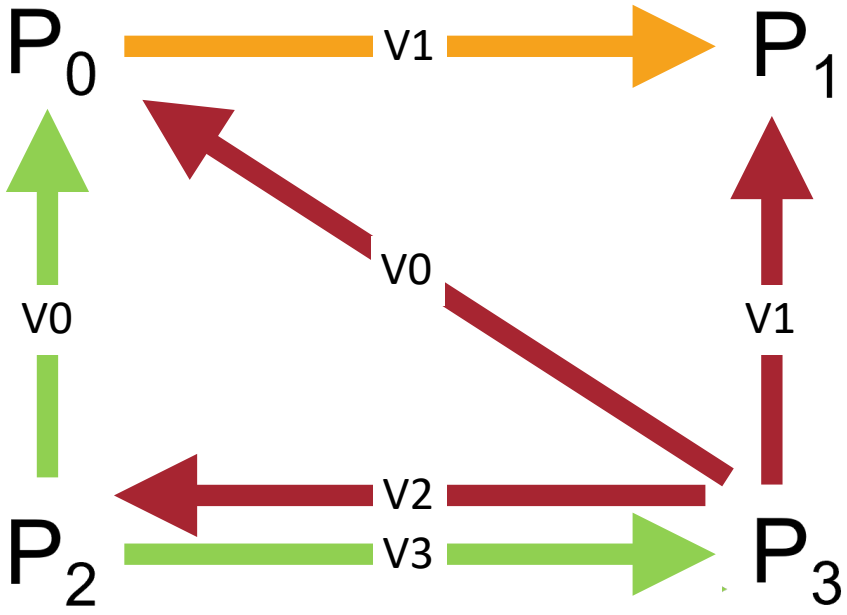
Step 2: Dynamic Exchange



MPI_Isend, MPI_Probe

Standard Algorithm: NonBlocking

Step 1: Dynamic Exchange



MPI_Issend

While sends haven't completed:

MPI_Iprobe

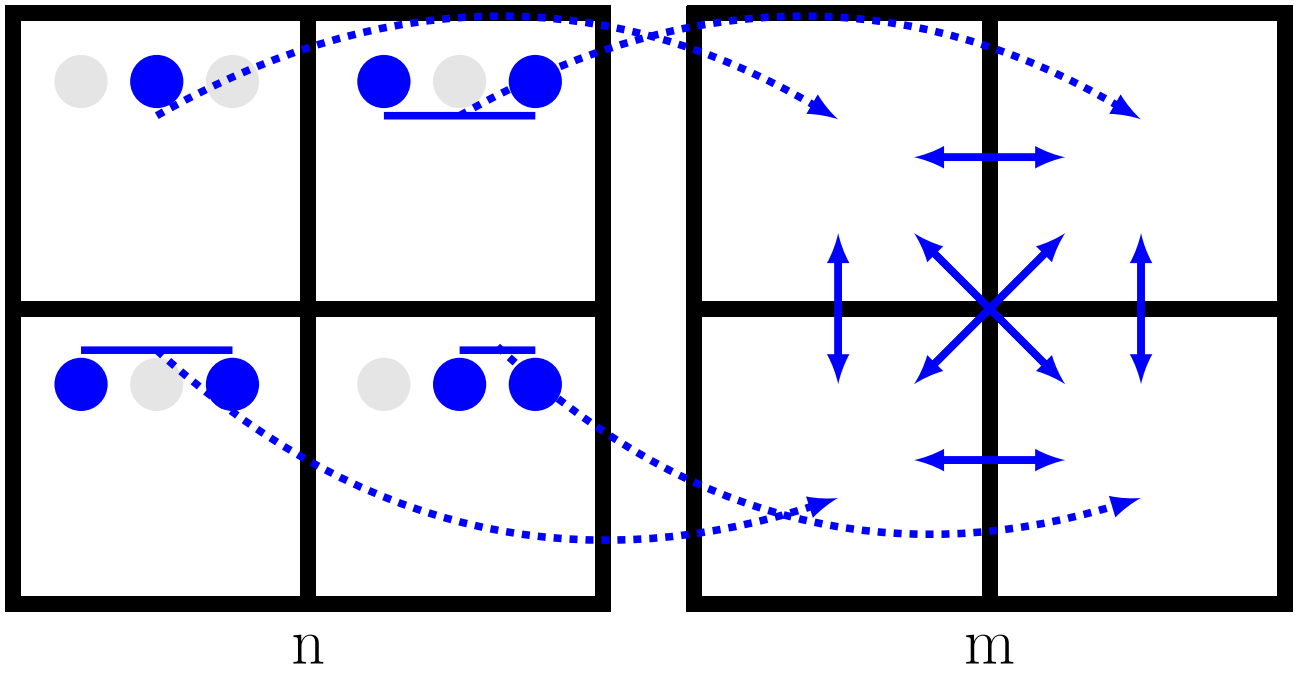
MPI_Ibarrier

While barrier hasn't completed:

MPI_Iprobe

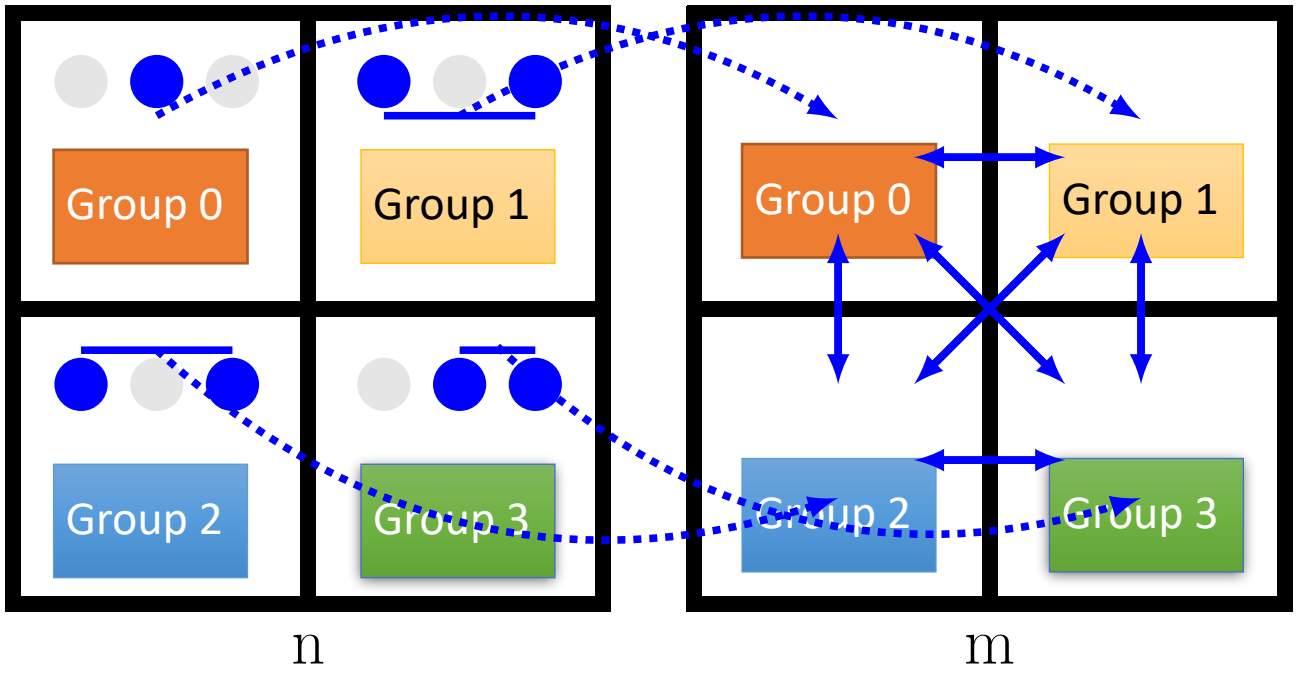
Locality-Aware Aggregation

- Previously has been used extensively in persistent point-to-point communication
- Common in collective operations
- **Novel contribution: locality-aware aggregation within dynamic exchanges**
  - **Adds metadata**



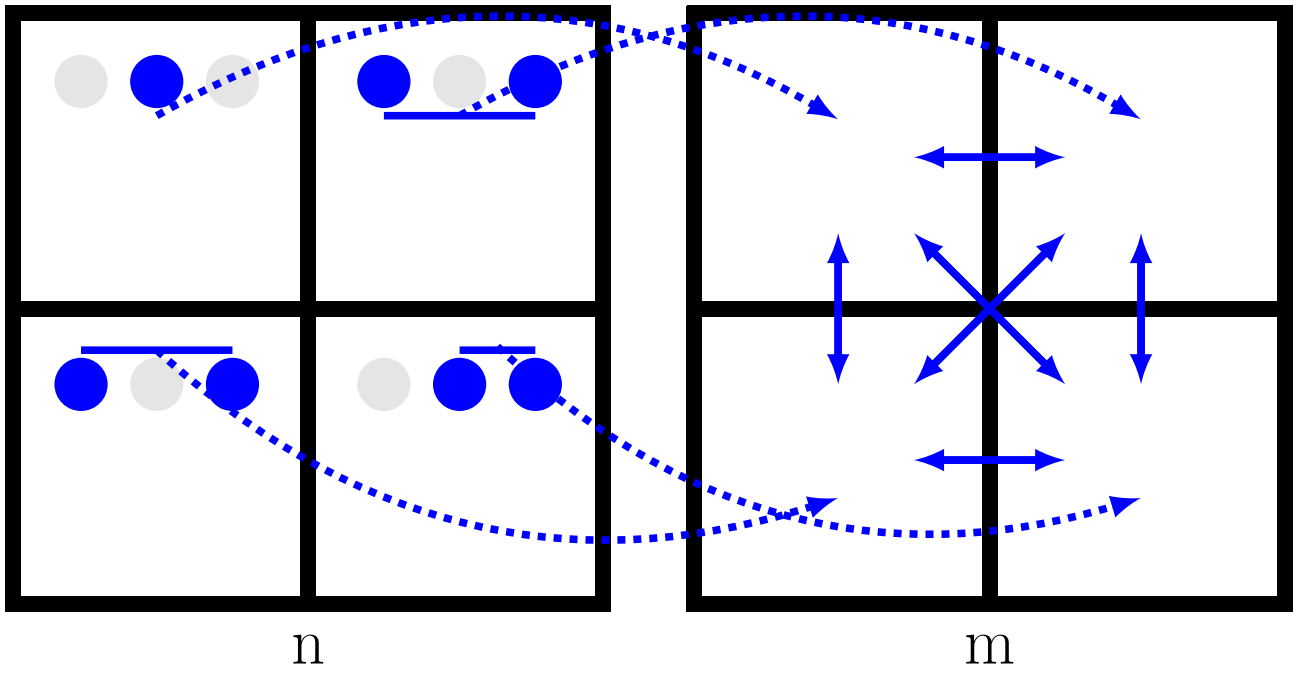
Locality-Aware Personalized

- Step 1: AllReduce among all processes in group
- Step 2: Aggregated personalized dynamic communication
  - MPI_Isend, MPI_Probe
- Step 3: Personalized dynamic exchange within region

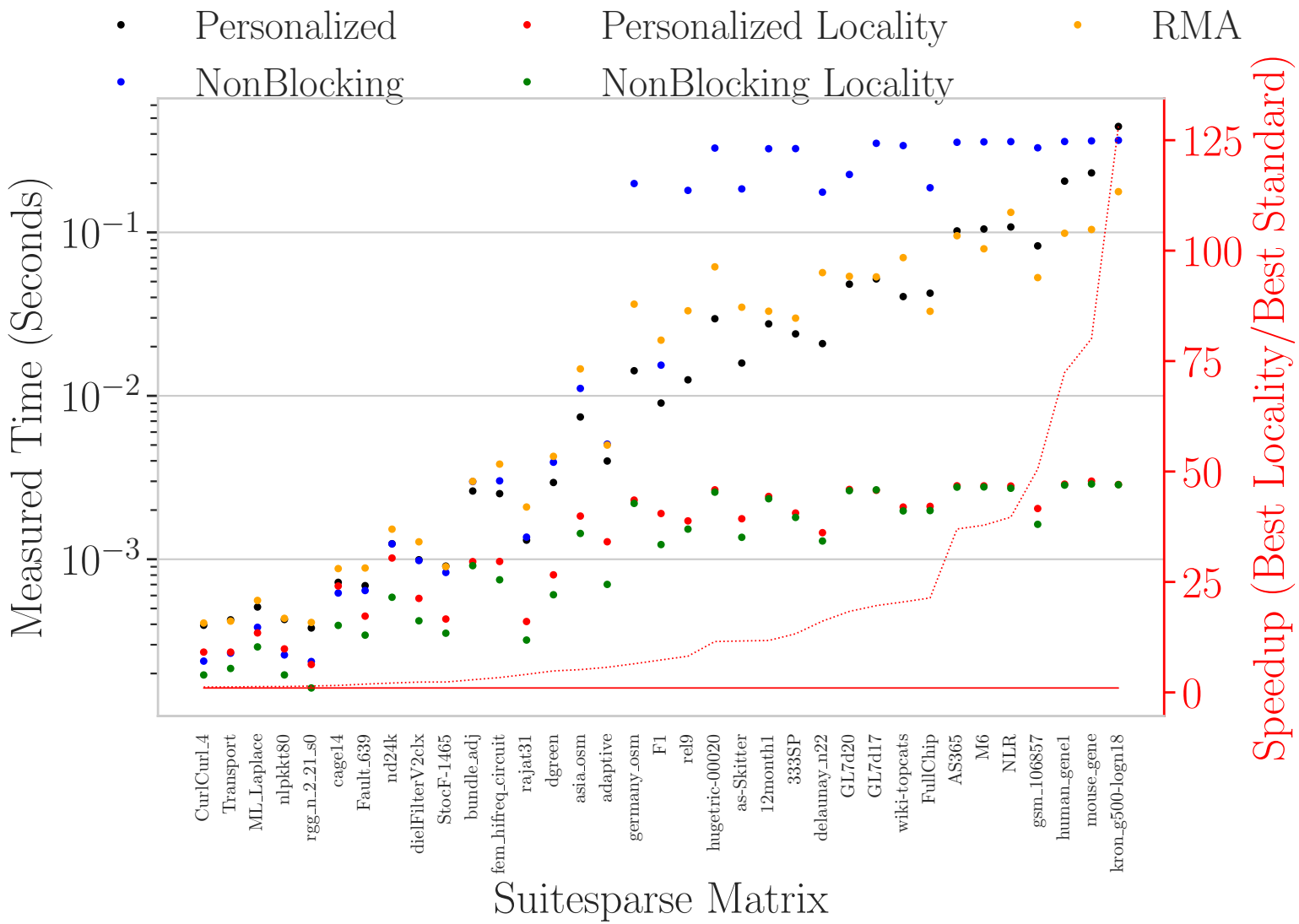


Locality-Aware NonBlocking

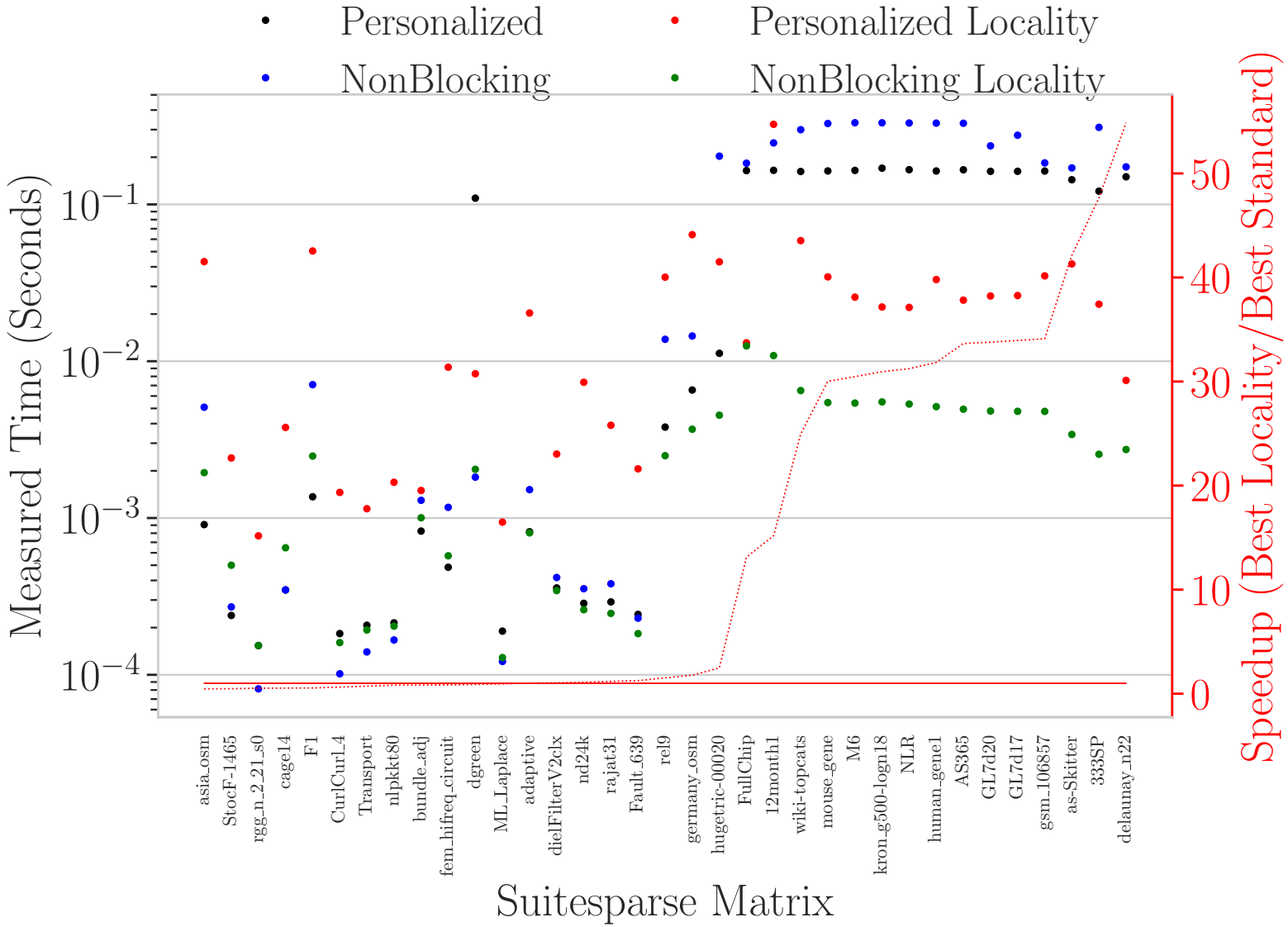
- Step 1: Aggregated nonblocking dynamic communication
  - MPI_Issend, MPI_Iprobe, etc
- Step 2: Personalized dynamic exchange within region



Locality-Aware MPIX_Alltoall_crs



Locality-Aware MPIX_Alltoallv_crs



Questions?

Email: bienz@unm.edu