Collective Communication
Operations on various topologies

Murat Manguoğlu
Middle East Technical University
Computer Engineering

CENG478 – Spring 2020
one to all broadcast ↔ all to one reduction

MPI_Bcast ↔ MPI_Reduce

• Linear array
• Ring
• 2D Mesh
• 3D Mesh
• Hypercube

On the board
all to all broadcast ↔ all to all reduction

MPI_Allgather ↔ MPI_Reduce_scatter

- Ring
- Linear array
- 2D Mesh
- Hypercube

On the board
All reduce

MPI_Allreduce

- Ring
- Linear array
- 2D Mesh
- Hypercube

On the board
Prefix Sum

MPI_Scan

Example: [0,1,2,3,4,5,6,7] → [0,1,3,6,10,15,21,28]

Figure 4.13 Computing prefix sums on an eight-node hypercube. At each node, square brackets show the local prefix sum accumulated in the result buffer and parentheses enclose the contents of the outgoing message buffer for the next step.

Algorithm 4.9 Prefix sums on a $d$-dimensional hypercube.

1. procedure PREFIX_SUMS_HCUBE($my\_id$, $my\_number$, $d$, result)
2. begin
3. $result := my\_number$;
4. $msg := result$;
5. for $i := 0$ to $d - 1$ do
6. $partner := my\_id$ XOR $2^i$;
7. send $msg$ to $partner$;
8. receive $number$ from $partner$;
9. $msg := msg + number$;
10. if ($partner < my\_id$) then $result := result + number$;
11. endfor;
12. end PREFIX_SUMS_HCUBE
scatter ↔ gather

MPI_Scatter ↔ MPI_Gather

- Linear array
- Hypercube

On the board
all to all personalized communication

MPI_Alltoall

- Ring
- Hypercube