



# **MPI Collective Workgroup Portland, OR, November 2009**



Torsten Hoefler

# Agenda

---

- ▶ Topological Collectives
- ▶ Dick's proposal (collectives on Cartesian topologies)
- ▶ NBC proposal (#109)
- ▶ Nonblocking collective helper functions
- ▶ Graph Topology issues (a fancy new interface?)
- ▶ MPI-2.2 Errata Item



# Topological Collectives

---

- ▶ Dynamic optimization -- all has been said, right? 😊
  - ▶ Gather, Alltoall, Reduce
- ▶ Has synchronization issues
  - ▶ thinking about “streaming” version
    - ▶ `MPI_Ineighbor_gather(..., reqs[recvcount]) + MPI_Waitany?`
    - ▶ `MPI_Ineighbor_gather(..., *req) + MPI_Cwaitany(req)?`
  - ▶ Doesn't seem to be very MPI-ish
  - ▶ might reduce usefulness of comm. schedule
- ▶ We need to find a (representative) set of applications
  - ▶ low on resources 😞
  - ▶ anybody interested (I have an implementation)?



# Dick's proposal

---

- ▶ Expose knowledge about collective communication in Cartesian communicators
  - ▶ E.g., Alltoall along lines, Bcast along columns
- ▶ I think it cannot be done with Topologies
  - ▶ How big is the advantage?
  - ▶ Would this influence process reordering?
- ▶ Does anybody (from IBM) have more details?
  
- ▶ Does anybody want to pick it up?



# NBC Ticket #109

---

- ▶ **First, a minimal change**
  - ▶ rev 6-2: page 66:41: add “blocking” before collective
  - ▶ clarifies but it not urgent
  - ▶ can be done with the MPI-2.2 changes?
  
- ▶ **Second, a (voted on) nontrivial change**
  - ▶ displs, count arrays can not be modified
  - ▶ simple fix:
    - ▶ rev 6-2: page 49:47: add “and displacement and count arrays” after “send buffers”
  - ▶ do we need an extra ticket?



# NBC MPI-2.2 changes

---

- ▶ Will be applied before the Jan. meeting
- ▶ I want (need) formal reviews (responsibilities!)
  - ▶ any volunteers?
- ▶ Do we read/vote on it again? How do we organize this?
  - ▶ I don't really need to know this right now but we should define a methodology (also for future proposals)



# Collective helper functions #168

---

- ▶ Add nonblocking versions
  - ▶ Allow nonblocking transparent libraries
  - ▶ We need for sure:
    - ▶ MPI\_Comm\_dup (for sure)
  - ▶ Maybe? Questionable imho:
    - ▶ MPI\_Comm[\_create,\_split,\_free]
    - ▶ MPI\_Intercomm[\_create,\_merge]
    - ▶ MPI\_[Dist\_]Graph\_create[\_adjacent]
    - ▶ MPI\_Cart\_create
    - ▶ MPI\_Comm\_spawn[\_multiple]
    - ▶ MPI\_Comm[\_accept,\_connect,\_disconnect]
    - ▶ MPI\_Win[\_create,\_free,\_fence]
    - ▶ +some I/O functions (but I/O is perfect 😊)



# MPI\_Cart\_create (#194/#195)

---

- ▶ See tickets!



# MPI-2 Errata

---

- ▶ See ticket #183
  - ▶ Typo MPI\_op\_free () vs. MPI\_Op\_free()
  - ▶ we can as well not put it to the errata and just fix it for MPI-3?

