„Synchronization Behavior for Collectives“

Background

- The question about the synchronization behavior for collectives came up in the previous meetings as it was mentioned and applied to neighborhood collectives. (Issue 863)
- „What exactly does MPI want to say about this for regular collectives?“

„Synchronizing“ in MPI Terms

Collective procedure:

An MPI procedure is collective if all processes in a group or groups of MPI processes need to invoke the procedure.

...

An MPI collective procedure is **synchronizing** if it will only return once all processes in the associated group or groups of MPI processes have called the appropriate matching MPI procedure.

The initiation procedures for nonblocking collective operations and the starting procedures for persistent collective operations are local and shall not be synchronizing.

All other procedures for collective operations, such as for blocking collective operations and the initialization procedures for persistent collective operations, may or may not be synchronizing.

MPI 5.0 p. 15/16 (Terms)

Collective communication chapter

Collective operations can (but are not required to) complete as soon as the caller's participation in the collective communication is finished. A blocking operation is complete as soon as the call returns. A nonblocking (immediate) call requires a separate completion call (cf. Section 3.7). The completion of a collective operation indicates that the caller is free to modify locations in the communication buffer. It does not indicate that other MPI processes in the group have completed or even started the operation (unless otherwise implied by the description of the operation). **Thus, a collective communication operation may, or may not, have the effect of synchronizing all participating MPI processes.**

MPI 5.0 p. 188 (Ch6.1, Collectives)

Rational: The statements about synchronization are made so as to allow a variety of implementations of the collective functions.

Remakr 18 in Appendix A.2

Procedure	Stages	Cpl	Loc	Blk	Op	Collective C	sq	S/W	Blocked resources and remarks
Chapter 6: Collective Communication									
MPI_BCAST, MPI_BARRIER, MPI_GATHER, MPI_GATHERV, MPI_SCATTER, MPI_SCATTERV, MPI_ALLGATHER, MPI_ALLGATHERV, MPI_ALLTOALL, MPI_ALLTOALLV, MPI_ALLTOALLW, MPI_REDUCE, MPI_ALLREDUCE, MPI_REDUCE_SCATTER_BLOCK, MPI_REDUCE_SCATTER, MPI_SCAN, MPI_EXSCAN	i-s-c-f	c+f	nl	b	b-op	C	sq	W1	) 18)

18. Based on their semantics, when called using an intra-communicator, MPI_ALLGATHER, MPI_ALLTOALL, and their V and W variants, MPI_ALLREDUCE, MPI_REDUCE_SCATTER, and MPI_REDUCE_SCATTER_BLOCK must synchronize (i.e., S1/S2 instead of W1/W2) provided that all counts and the size of all datatypes are larger than zero.

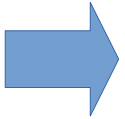
S1 = blocking synchronization, i.e., no process shall return from this procedure until all processes on the associated process group called this procedure
S2 = start-complete-synchronization, i.e., no process shall complete the associated operation until all processes on the associated process group have called the associated starting procedure
W1/W2 = the implementation is permitted to do S1 but not required to do S1/S2

MPI Forum

Zero size message behaviour in P2P

- A zero length (size) datatype may result in transfer according to:

> …
> The value returned as the count argument of MPI_GET_COUNT for a datatype of length zero **where zero bytes have been transferred** is zero. If the number of bytes transferred is greater than zero, MPI_UNDEFINED is returned.
> ...
> MPI 5.0 p. 40/41 (MPI_Get_count)



Does this P2P statement also apply to collectives and enforce synchronisation for these cases?

Questions

1) What do we intent with Remark 18

   (a) warn users about synchronization?
   (b) tell users that they do not need extra synchronisation afterwards and could be used as a „barrier“?
   (c) enforce (barrier) synchronisation by implementations?

2) Is remark 18 correct or are we missing MPI operations with this behaviour? If so is it a problem?

3) Do we want to weaken the „must synchronize“ with, e.g., „are synchronizing“ OR going even further and
   remove remark 18 for regular collectives in A.2?

Related Issues:

- **Issue 863:** Clarify synchronization behavior of neighborhood collectives
- **Issue 971:** Do All-to-All collectives *must* synchronize?