Hybrid Programming Working Group Proposals

What's this talk about?

- Introduce some of things we are planning to work on
- We haven't started meetings or formal discussions yet
  - Intend to start in January
- TRAC link:

  https://svn.mpi-forum.org/trac/mpi-forum-web/wiki/MPI3Hybrid
- Working group mailing list:
  - Please subscribe if you plan to participate
- Not soliciting votes/straw votes at this point
  - Just initial feedback on the working group

Three Proposals Currently

- Treating Threads as MPI Processes
- Dynamic Thread levels
- Thread Init/Finalize routines for thread-core mapping support
  - Could be integrated into the first proposal

Dynamic Thread Levels

- Problem: MPI specifies thread-level support at Init time
  - Even if a small fraction of the code uses THREAD_MULTIPLE, the entire code is forced to go through locks
- Performance Impact (messaging rate):
  - 2X on PROC_NULL (emulating infinitely fast networks)
  - About 20% on TCP/IP
- Proposal:
  - Add calls for MPI_Set_thread_level() to dynamically change thread-levels within the application

Dynamic Thread Levels (contd.)

- MPI_Set_thread_level(int required, int * provided)
  - Hinting mechanism only
- Relevant Issues:
  - If an implementation allows the thread-level reduction, but not increase, the application might not be able to deal with it
  - Asynchronous Progress Threads
    - Requires synchronization with the progress thread to change thread level
  - Collective Operations: Some MPI implementations use different collective operations based on the thread-level
    - Maybe add a stricter collective version as well ?

Thread Init/Finalize Routines

- Problem:
  - MPI currently does not explicitly know threads
  - Process can be mapped to different cores/SMTs
    - Thread scheduling is left to the OS
- Proposal:
  - Explicit Thread Init/Finalize Routines
  - Allow the process manager to perform intelligent mapping
  - Optional calls – application does not necessarily have to call these
- Can potentially be merged into Alexander's Thread Register/Deregister proposal