Advanced MPI Programming

Tutorial at SC14, November 2014

Latest slides and code examples are available at

www.mcs.anl.gov/~thakur/sc14-mpi-tutorial

Pavan Balaji
Argonne National Laboratory
Email: balaji@mcs.anl.gov
Web: www.mcs.anl.gov/~balaji

William Gropp
University of Illinois, Urbana-Champaign
Email: wgropp@illinois.edu
Web: www.cs.illinois.edu/~wgropp

Torsten Hoefler
ETH Zurich
Email: htor@inf.ethz.ch
Web: http://htor.inf.ethz.ch/

Rajeev Thakur
Argonne National Laboratory
Email: thakur@mcs.anl.gov
Web: www.mcs.anl.gov/~thakur

ETH Zürich

Outline

Morning

- Introduction
  - MPI-1, MPI-2, MPI-3
- Running example: 2D stencil code
  - Simple point-to-point version
- Derived datatypes
  - Use in 2D stencil code
- One-sided communication
  - Basics and new features in MPI-3
  - Use in 2D stencil code
  - Advanced topics
    - Global address space communication

Afternoon

- MPI and Threads
  - Thread safety specification in MPI
  - How it enables hybrid programming
  - Hybrid (MPI + shared memory) version of 2D stencil code
- Nonblocking collectives
  - Parallel FFT example
- Process topologies
  - 2D stencil example
- Neighborhood collectives
  - 2D stencil example
- Recent efforts of the MPI Forum
- Conclusions

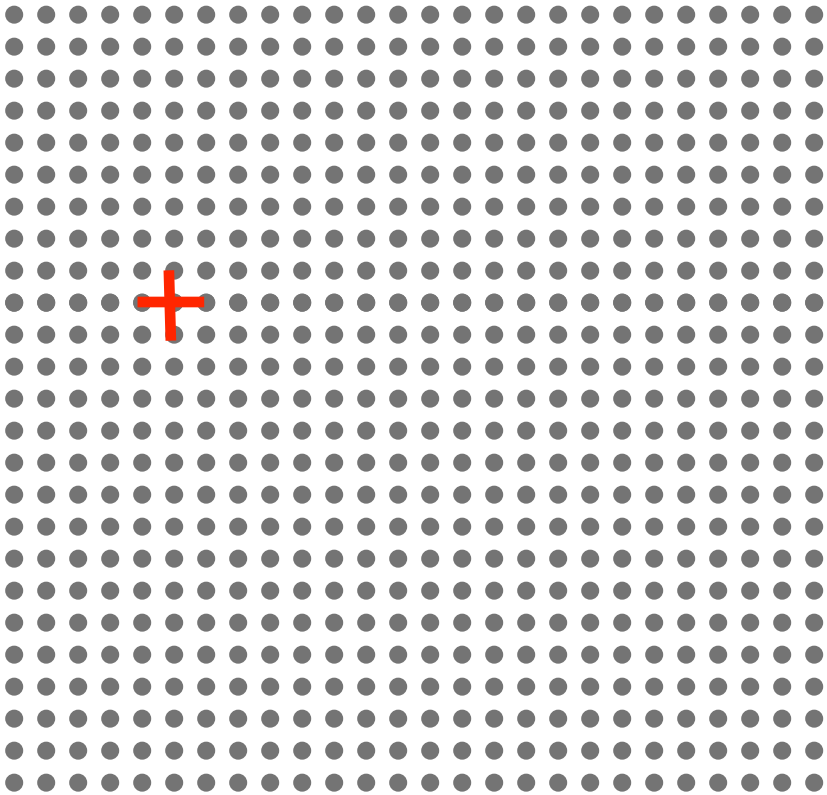
Poisson Problem

- To approximate the solution of the Poisson Problem $\nabla^2 u = f$ on the unit square, with u defined on the boundaries of the domain (Dirichlet boundary conditions), this simple 2nd order difference scheme is often used:
  - $(U(x+h,y) - 2U(x,y) + U(x-h,y)) / h^2 +$
    $(U(x,y+h) - 2U(x,y) + U(x,y-h)) / h^2 = f(x,y)$
    - Where the solution U is approximated on a discrete grid of points x=0, h, 2h, 3h, … , (1/h)h=1, y=0, h, 2h, 3h, … 1.
    - To simplify the notation, U(ih,jh) is denoted U_{ij}

- This is defined on a discrete mesh of points (x,y) = (ih,jh), for a mesh spacing "h"

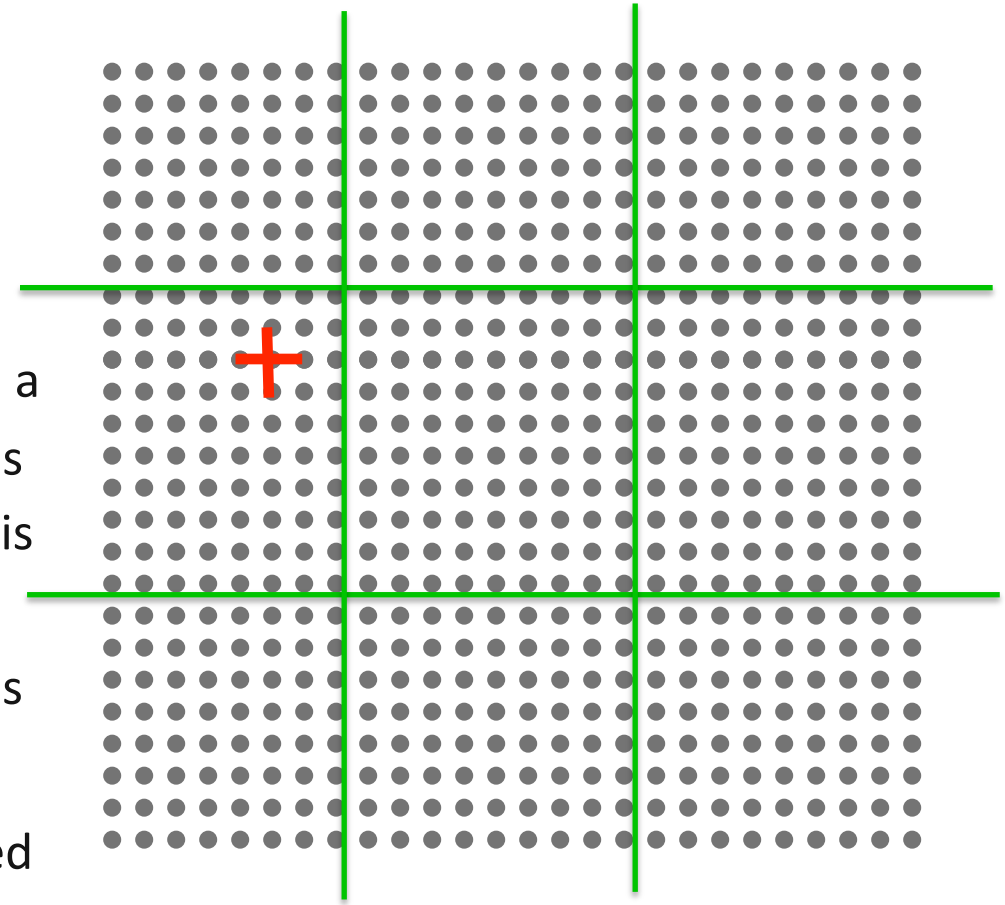
The Global Data Structure

- Each circle is a mesh point
- Difference equation evaluated at each point involves the four neighbors
- The red “plus” is called the method’s stencil
- Good numerical algorithms form a matrix equation Au=f; solving this requires computing Bv, where B is a matrix derived from A. These evaluations involve computations with the neighbors on the mesh.

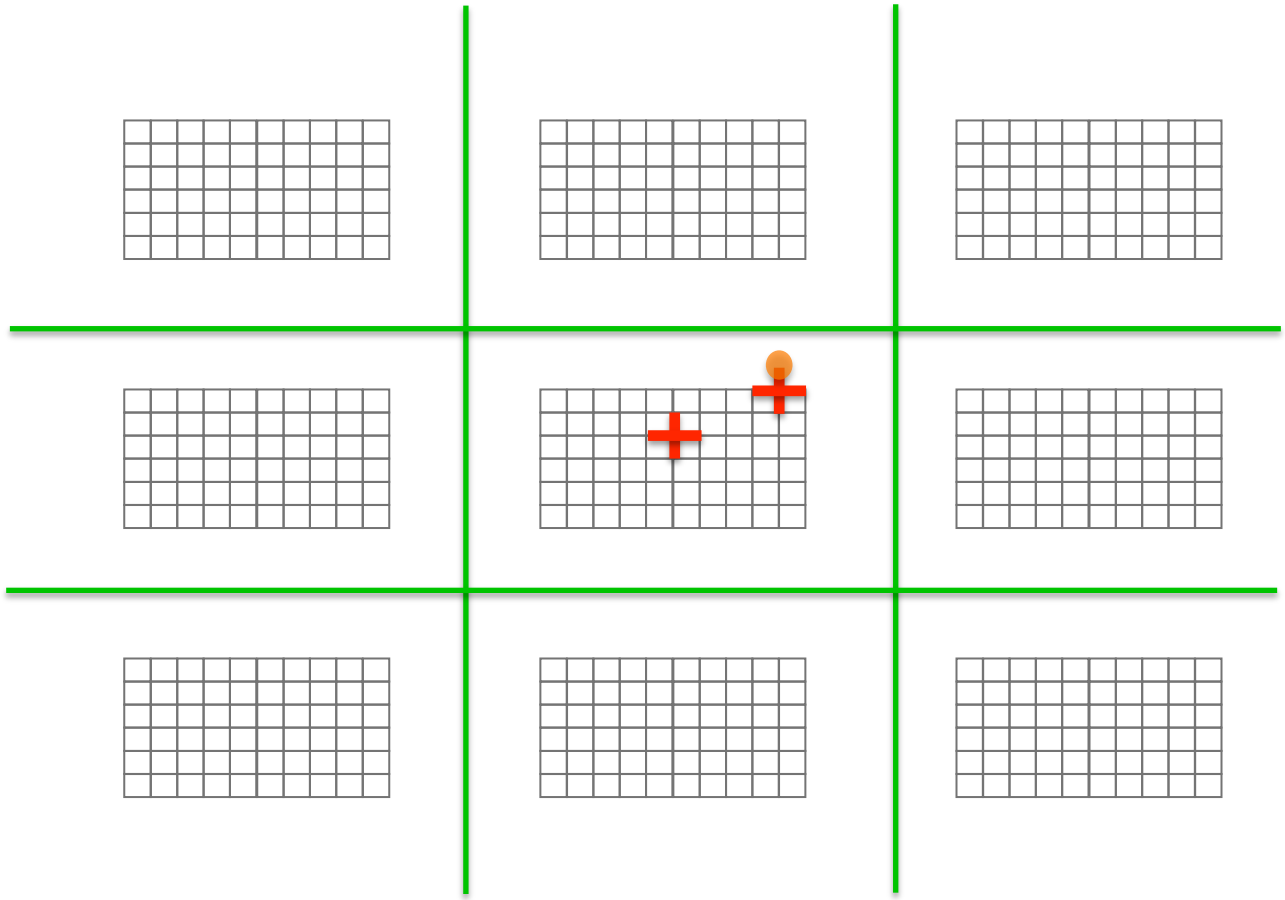


The Global Data Structure

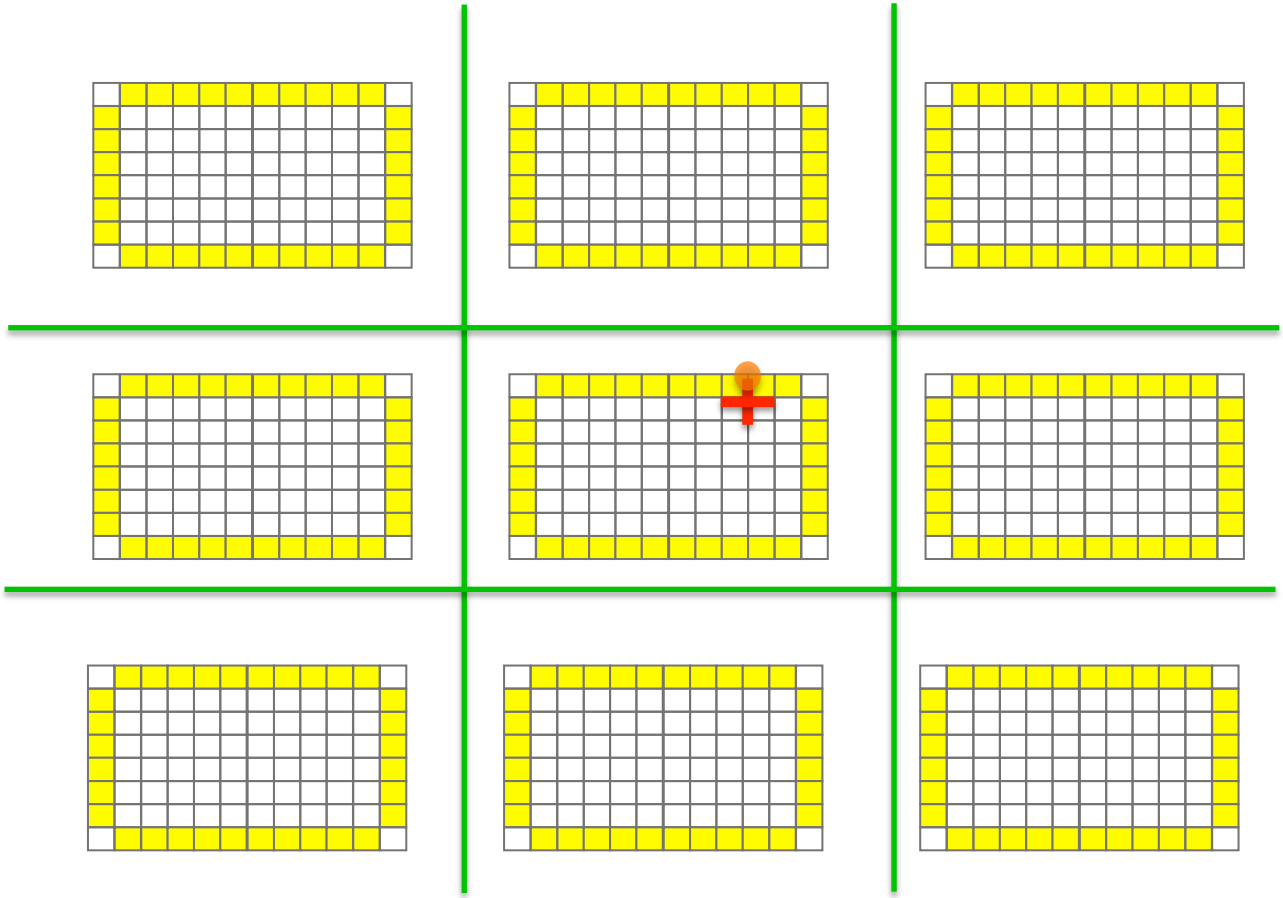
- Each circle is a mesh point
- Difference equation evaluated at each point involves the four neighbors
- The red "plus" is called the method's stencil
- Good numerical algorithms form a matrix equation Au=f; solving this requires computing Bv, where B is a matrix derived from A. These evaluations involve computations with the neighbors on the mesh.
- Decompose mesh into equal sized (work) pieces



Necessary Data Transfers

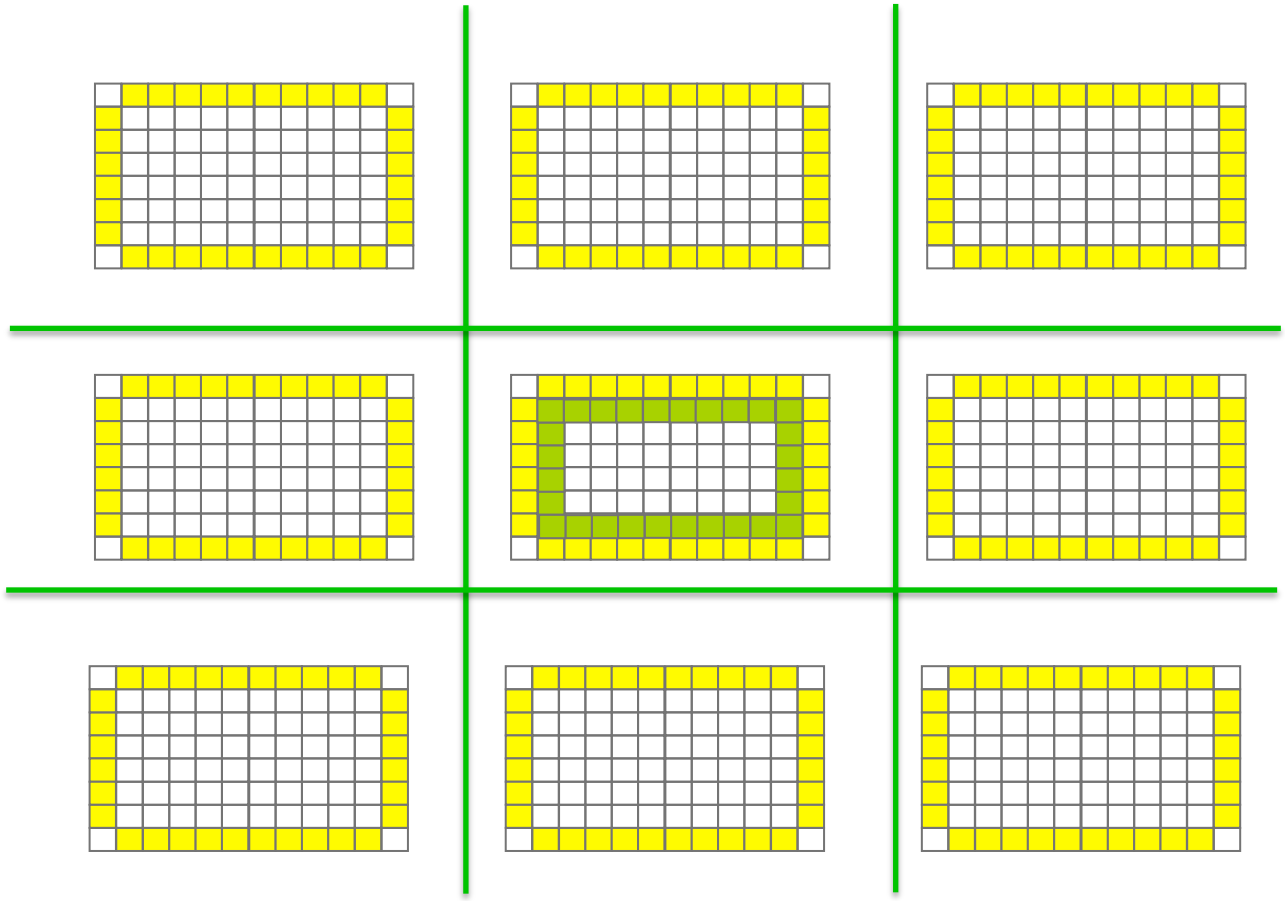


Necessary Data Transfers



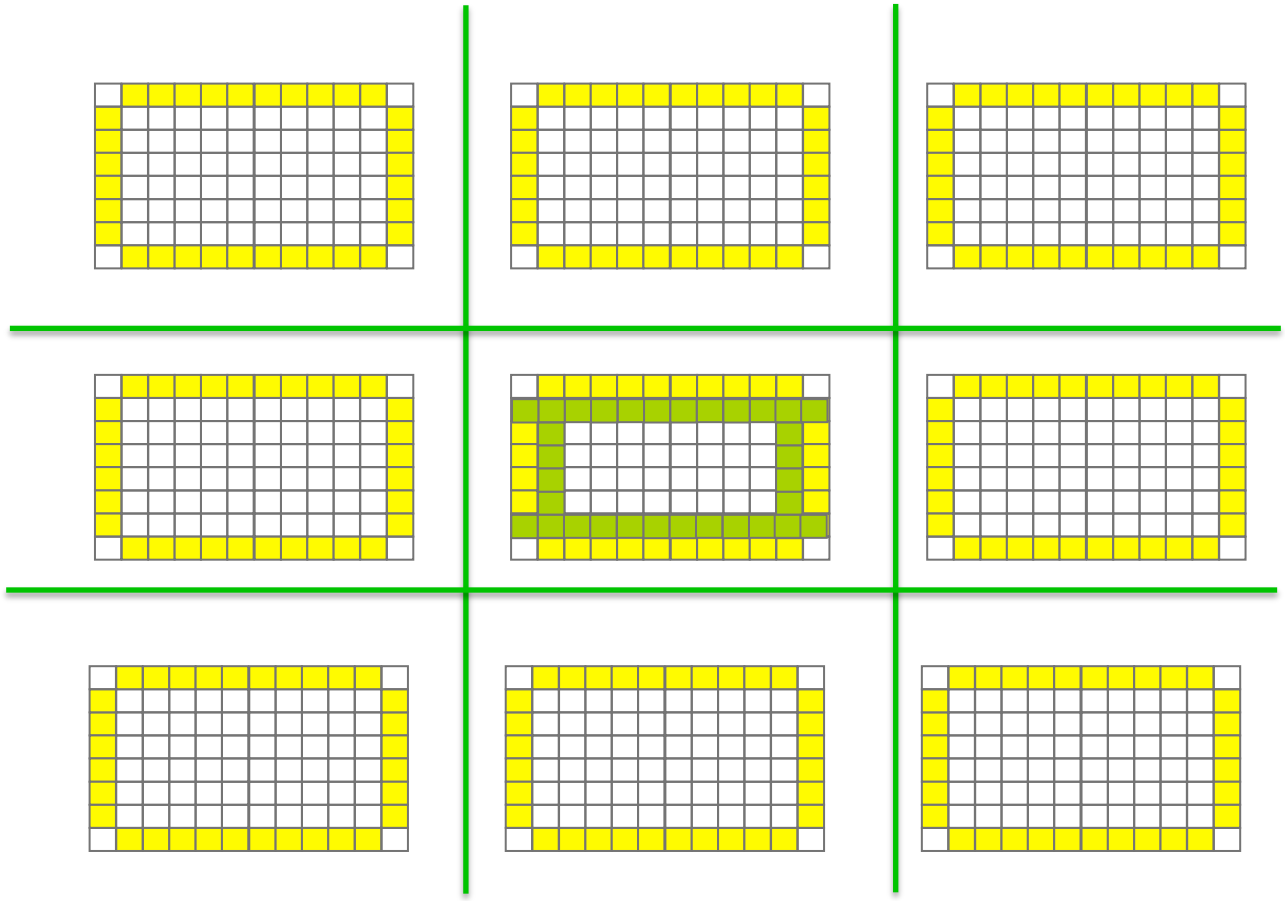
Necessary Data Transfers

- Provide access to remote data through a *halo* exchange (5 point stencil)



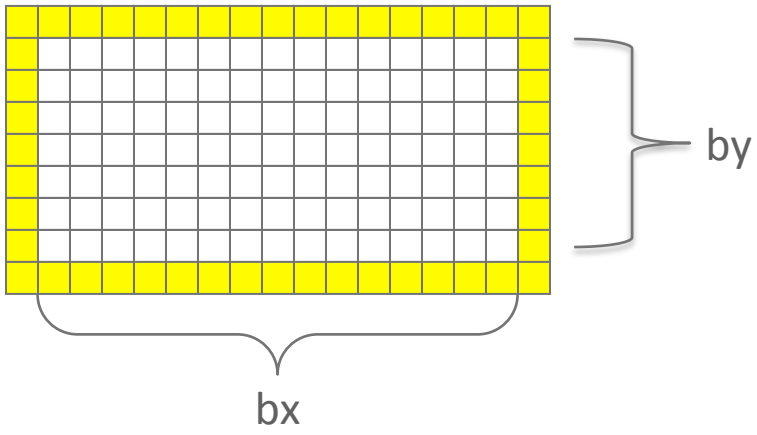
Necessary Data Transfers

- Provide access to remote data through a *halo* exchange (9 point with trick)



The Local Data Structure

- Each process has its local "patch" of the global array
  - "bx" and "by" are the sizes of the local array
  - Always allocate a halo around the patch
  - Array allocated of size (bx+2)x(by+2)



2D Stencil Code Walkthrough

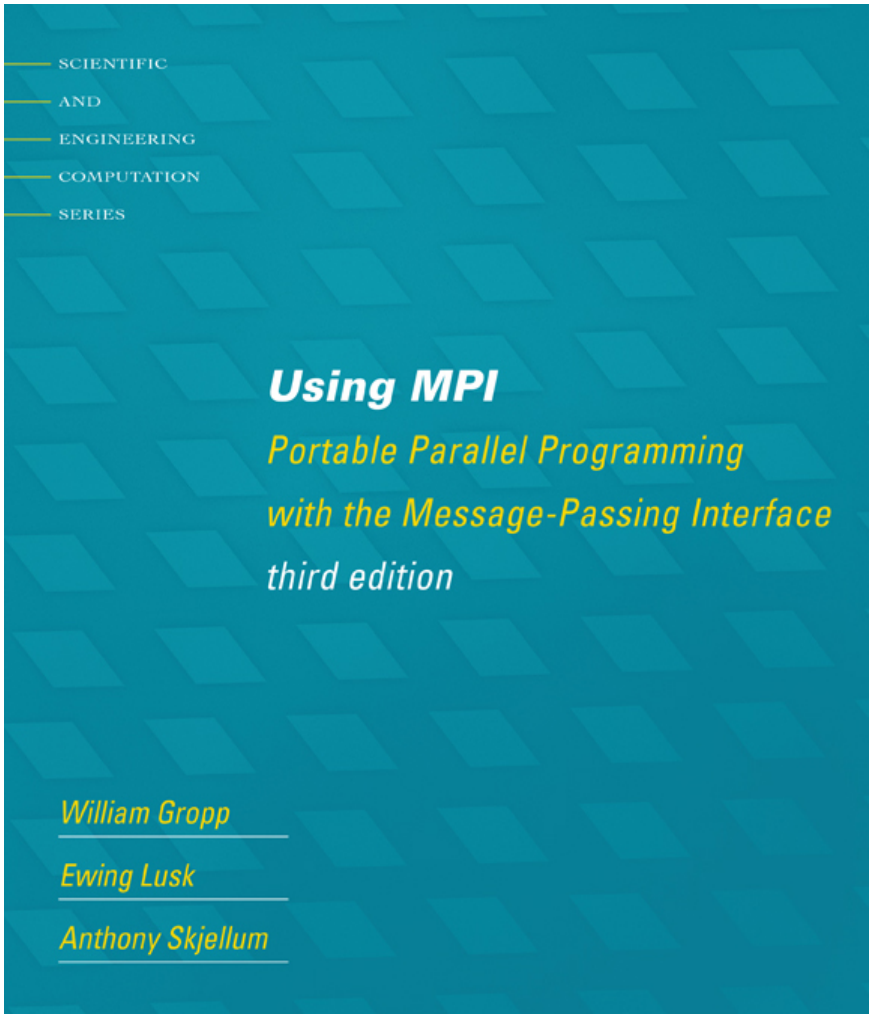
- Code can be downloaded from
  www.mcs.anl.gov/~thakur/sc14-mpi-tutorial

Web Pointers

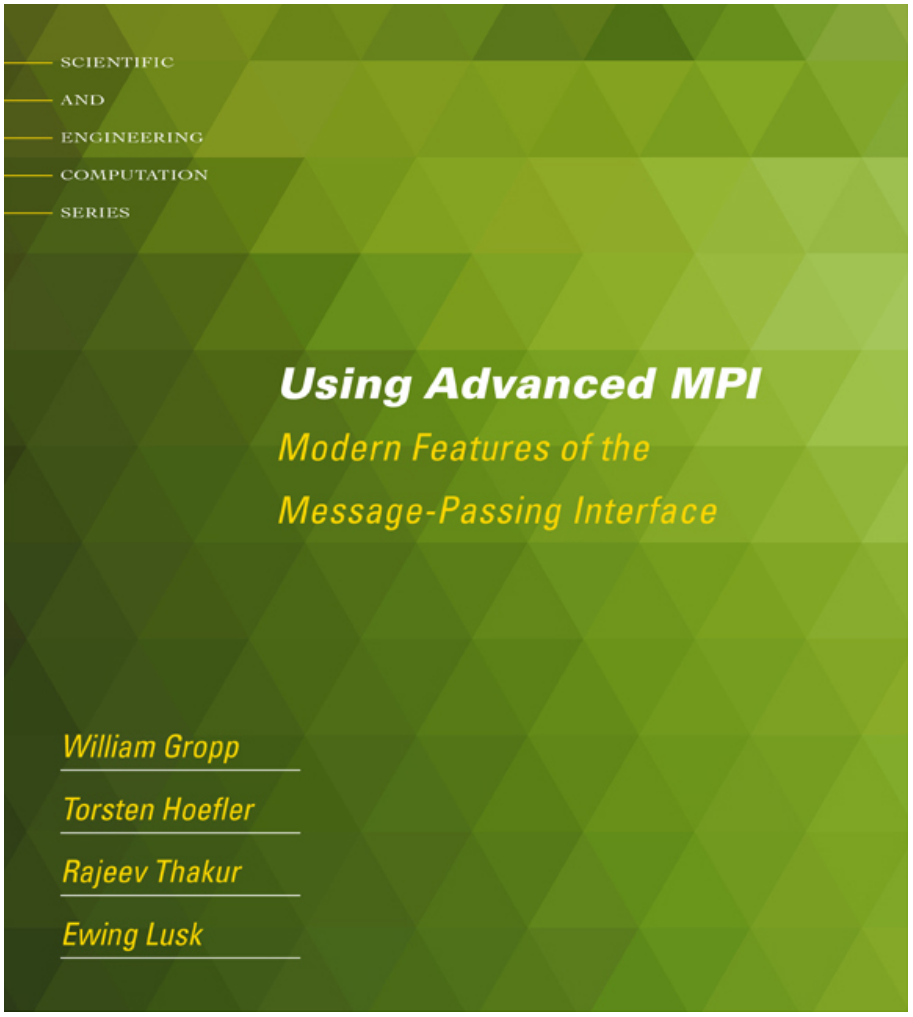
- MPI standard : http://www.mpi-forum.org/docs/docs.html
- MPI Forum : http://www.mpi-forum.org/

- MPI implementations:
  - MPICH : http://www.mpich.org
  - MVAPICH : http://mvapich.cse.ohio-state.edu/
  - Intel MPI: http://software.intel.com/en-us/intel-mpi-library/
  - Microsoft MPI: www.microsoft.com/en-us/download/details.aspx?id=39961
  - Open MPI : http://www.open-mpi.org/
  - IBM MPI, Cray MPI, HP MPI, TH MPI, …
- Several MPI tutorials can be found on the web

New Tutorial Books on MPI



Basic MPI



Advanced MPI, including MPI-3