

# Advanced MPI Programming

Tutorial at SC14, November 2014

Latest slides and code examples are available at

[www.mcs.anl.gov/~thakur/sc14-mpi-tutorial](http://www.mcs.anl.gov/~thakur/sc14-mpi-tutorial)

**Pavan Balaji**

Argonne National Laboratory

Email: [balaji@mcs.anl.gov](mailto:balaji@mcs.anl.gov)

Web: [www.mcs.anl.gov/~balaji](http://www.mcs.anl.gov/~balaji)

**William Gropp**

University of Illinois, Urbana-Champaign

Email: [wgropp@illinois.edu](mailto:wgropp@illinois.edu)

Web: [www.cs.illinois.edu/~wgropp](http://www.cs.illinois.edu/~wgropp)

**Torsten Hoefler**

ETH Zurich

Email: [htor@inf.ethz.ch](mailto:htor@inf.ethz.ch)

Web: <http://htor.inf.ethz.ch/>

**Rajeev Thakur**

Argonne National Laboratory

Email: [thakur@mcs.anl.gov](mailto:thakur@mcs.anl.gov)

Web: [www.mcs.anl.gov/~thakur](http://www.mcs.anl.gov/~thakur)



# 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

# MPI-1

- MPI is a message-passing library interface standard.
  - Specification, not implementation
  - Library, not a language
- MPI-1 supports the classical message-passing programming model: basic point-to-point communication, collectives, datatypes, etc
- MPI-1 was defined (1994) by a broadly based group of parallel computer vendors, computer scientists, and applications developers.
  - 2-year intensive process
- Implementations appeared quickly and now MPI is taken for granted as vendor-supported software on any parallel machine.
- Free, portable implementations exist for clusters and other environments (MPICH, Open MPI)

# MPI-2

- Same process of definition by MPI Forum
- MPI-2 is an extension of MPI
  - Extends the message-passing model.
    - Parallel I/O
    - Remote memory operations (one-sided)
    - Dynamic process management
  - Adds other functionality
    - C++ and Fortran 90 bindings
      - similar to original C and Fortran-77 bindings
    - External interfaces
    - Language interoperability
    - MPI interaction with threads

# Timeline of the MPI Standard

- MPI-1 (1994), presented at SC'93
  - Basic point-to-point communication, collectives, datatypes, etc
- MPI-2 (1997)
  - Added parallel I/O, Remote Memory Access (one-sided operations), dynamic processes, thread support, C++ bindings, ...
- ----- Stable for 10 years -----
- MPI-2.1 (2008)
  - Minor clarifications and bug fixes to MPI-2
- MPI-2.2 (2009)
  - Small updates and additions to MPI 2.1
- MPI-3 (2012)
  - Major new features and additions to MPI

# Overview of New Features in MPI-3

- Major new features
  - Nonblocking collectives
  - Neighborhood collectives
  - Improved one-sided communication interface
  - Tools interface
  - Fortran 2008 bindings
- Other new features
  - Matching Probe and Recv for thread-safe probe and receive
  - Noncollective communicator creation function
  - “const” correct C bindings
  - Comm\_split\_type function
  - Nonblocking Comm\_dup
  - Type\_create\_hindexed\_block function
- C++ bindings removed
- Previously deprecated functions removed

# Status of MPI-3 Implementations (\*)

	MPICH	MVAPICH	Open MPI	Cray MPI	Tianhe MPI	Intel MPI	IBM BG/Q MPI <sup>1</sup>	IBM PE MPICH <sup>2</sup>	IBM Platform	SGI MPI	Fujitsu MPI	MS MPI
NB collectives	✓	✓	✓	✓	✓	✓	✓	Q4 '14	✓	✓	✓	*
Neighborhood collectives	✓	✓	✓	✓	✓	✓	✓	Q4 '14	Q3 '15	✓	Q2 '15	
RMA	✓	✓	✓	✓	✓	✓	✓	Q4 '14	Q3 '15	✓	Q2 '15	
Shared memory	✓	✓	✓	✓	✓	✓	✓	Q4 '14	Q3 '15	✓	Q2 '15	✓
Tools Interface	✓	✓	✓	✓	✓	✓	✓ <sup>3</sup>	Q4 '14	Q3 '15	✓	Q2 '15	*
Non-collective comm. create	✓	✓	✓	✓	✓	✓	✓	Q4 '14	Q3 '15	✓	Q2 '15	
F08 Bindings	✓	✓	✓	Q4 '14	✓	Q4 '14	✓	Q4 '14	Q3 '15	✓	Q2 '15	
New Datatypes	✓	✓	✓	✓	✓	✓	✓	Q4 '14	Q3 '15	✓	Q2 '15	*
Large Counts	✓	✓	✓	✓	✓	✓	✓	Q4 '14	Q3 '15	✓	Q2 '15	*
Matched Probe	✓	✓	✓	✓	✓	✓	✓	Q4 '14	Q3 '15	✓	✓	*

Release dates are estimates and are subject to change at any time.

Empty cells indicate no *publicly announced* plan to implement/support that feature.

<sup>1</sup> Open source, but unsupported

<sup>2</sup> Beta release

<sup>3</sup> No MPI\_T variables exposed

\* Under development

(\*) Platform-specific restrictions might apply for all supported features



## Important considerations while using MPI

- All parallelism is explicit: the programmer is responsible for correctly identifying parallelism and implementing parallel algorithms using MPI constructs



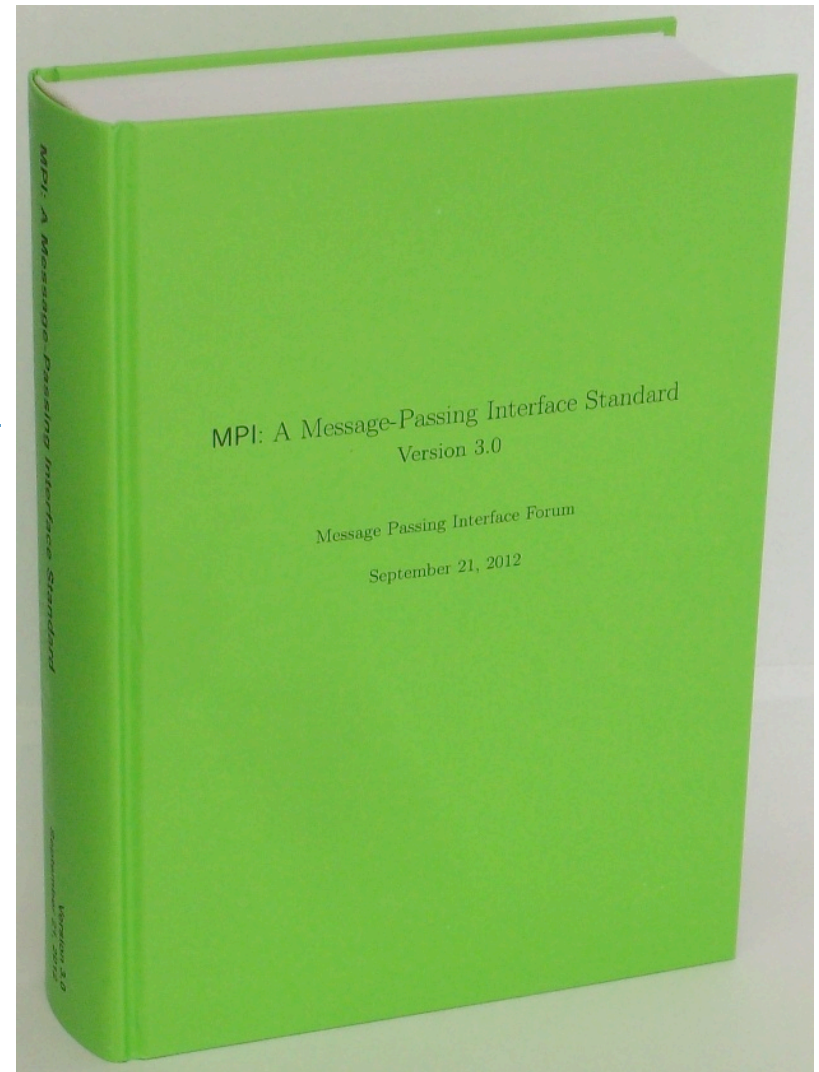
# Web Pointers

- 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](http://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

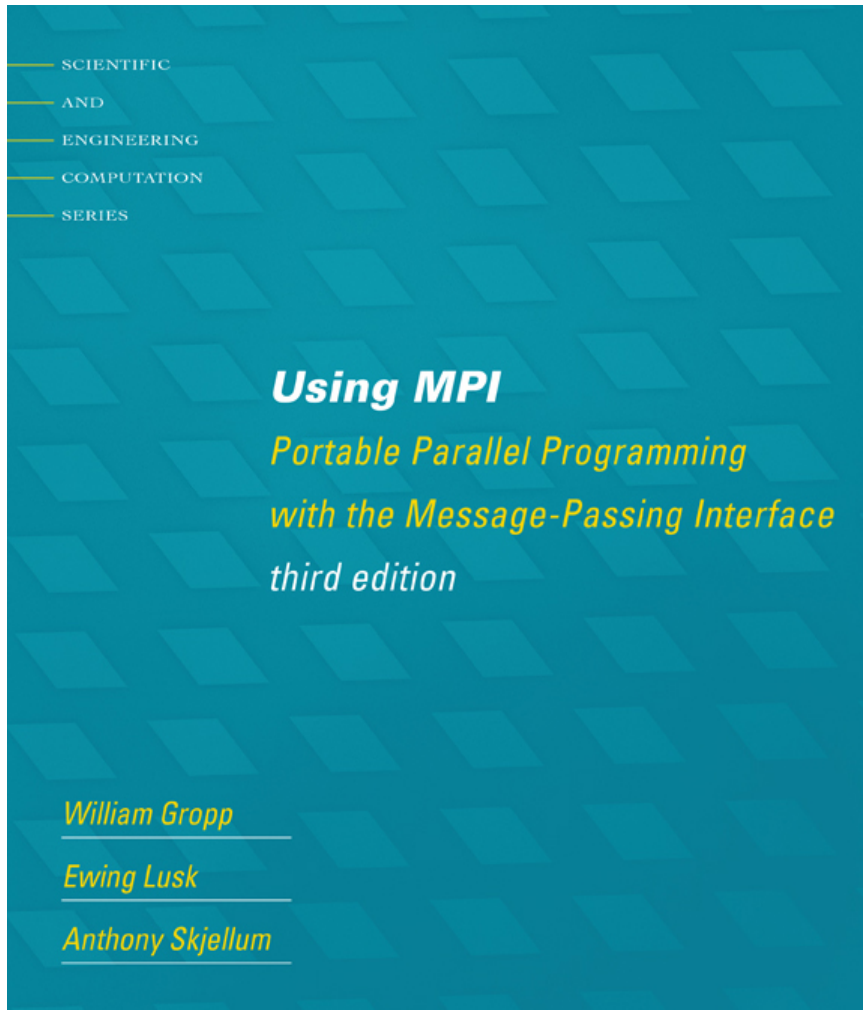
# Latest MPI 3.0 Standard in Book Form

Available from amazon.com

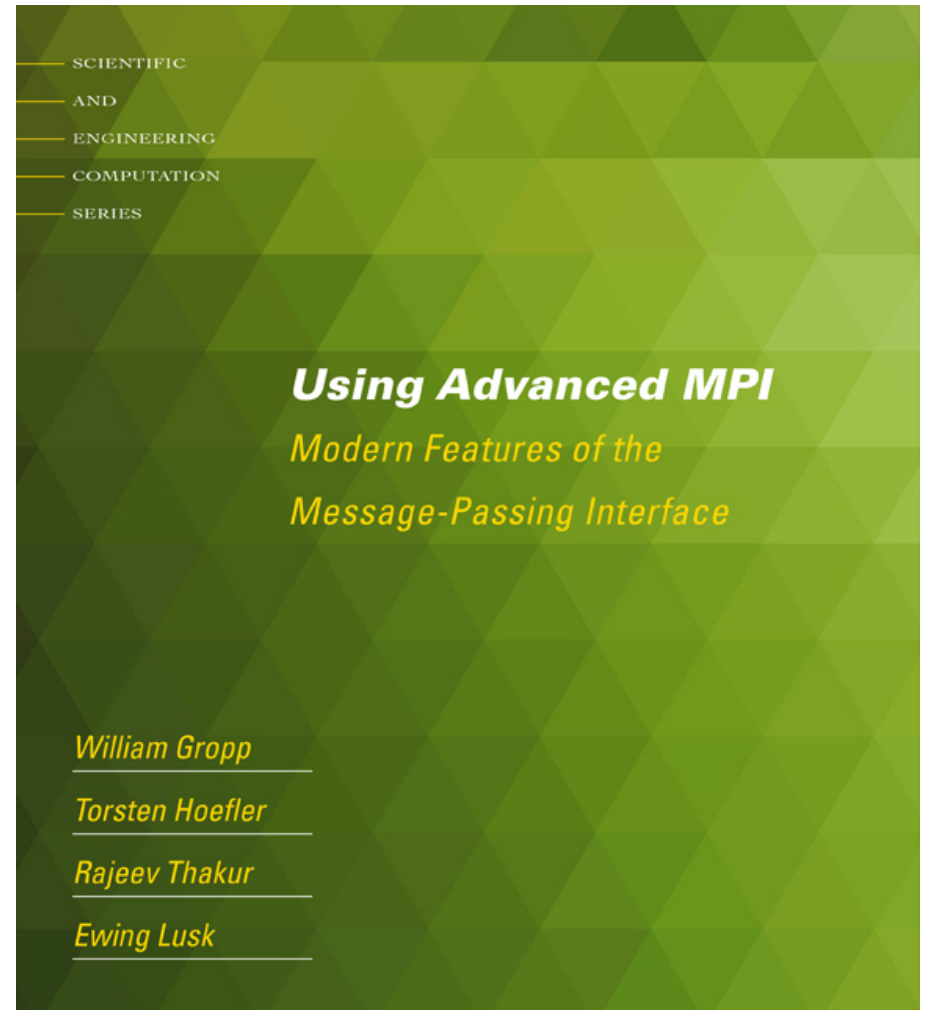
<http://www.amazon.com/dp/B002TM5BQK/>



# New Tutorial Books on MPI



**Basic MPI**



**Advanced MPI, including MPI-3**