

— SCIENTIFIC  
— AND  
— ENGINEERING  
— COMPUTATION  
— SERIES

# ***Using MPI***

*Portable Parallel Programming  
with the Message-Passing Interface  
second edition*

*William Gropp*

---

*Ewing Lusk*

---

*Anthony Skjellum*

---

© 1999 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in L<sup>A</sup>T<sub>E</sub>X by the authors and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Gropp, William.

Using MPI : portable parallel programming with the message-passing interface / William Gropp, Ewing Lusk, Anthony Skjellum.—2nd ed.

p. cm.—(Scientific and engineering computation)

Includes bibliographical references and index.

ISBN 0-262-57134-X (set : pbk. : alk. paper)—0-262-57132-3

(v. 1 : pbk. : alk. paper)

1. Parallel programming (Computer science). 2. Parallel computers—programming. 3. Computer interfaces. I. Lusk, Ewing. II. Skjellum, Anthony. III. Title. IV. Series.

QA76.642.G76 1999

005.2'75—dc21

99-16613

CIP

10 9 8 7 6 5 4 3

## Contents

Series Foreword	xiii
Preface to the Second Edition	xv
Preface to the First Edition	xix
<b>1 Background</b>	<b>1</b>
1.1 Why Parallel Computing?	1
1.2 Obstacles to Progress	2
1.3 Why Message Passing?	3
1.3.1 Parallel Computational Models	3
1.3.2 Advantages of the Message-Passing Model	9
<a href="#">1.4 Evolution of Message-Passing Systems</a>	<a href="#">10</a>
1.5 The MPI Forum	11
<b>2 Introduction to MPI</b>	<b>13</b>
2.1 Goal	13
2.2 What Is MPI?	13
<a href="#">2.3 Basic MPI Concepts</a>	<a href="#">14</a>
<a href="#">2.4 Other Interesting Features of MPI</a>	<a href="#">18</a>
2.5 Is MPI Large or Small?	20
2.6 Decisions Left to the Implementor	21
<b>3 Using MPI in Simple Programs</b>	<b>23</b>
<a href="#">3.1 A First MPI Program</a>	<a href="#">23</a>
3.2 Running Your First MPI Program	28
3.3 A First MPI Program in C	29
3.4 A First MPI Program in C++	29
<a href="#">3.5 Timing MPI Programs</a>	<a href="#">34</a>
3.6 A Self-Scheduling Example: Matrix-Vector Multiplication	35
<a href="#">3.7 Studying Parallel Performance</a>	<a href="#">43</a>
<a href="#">3.7.1 Elementary Scalability Calculations</a>	<a href="#">43</a>
<a href="#">3.7.2 Gathering Data on Program Execution</a>	<a href="#">45</a>
3.7.3 Instrumenting a Parallel Program with MPE Logging	46

3.7.4	Events and States	47
3.7.5	Instrumenting the Matrix-Matrix Multiply Program	47
3.7.6	<a href="#">Notes on Implementation of Logging</a>	49
3.7.7	<a href="#">Examining Logfiles with Upshot</a>	52
3.8	Using Communicators	53
3.9	<a href="#">Another Way of Forming New Communicators</a>	59
3.10	<a href="#">A Handy Graphics Library for Parallel Programs</a>	62
3.11	<a href="#">Common Errors and Misunderstandings</a>	64
3.12	Application: Quantum Monte Carlo Calculations in Nuclear Physics	66
3.13	Summary of a Simple Subset of MPI	67
<b>4</b>	<b><a href="#">Intermediate MPI</a></b>	<b>69</b>
4.1	The Poisson Problem	70
4.2	<a href="#">Topologies</a>	73
4.3	A Code for the Poisson Problem	81
4.4	Using Nonblocking Communications	93
4.5	Synchronous Sends and “Safe” Programs	96
4.6	More on Scalability	96
4.7	Jacobi with a 2-D Decomposition	98
4.8	An MPI Derived Datatype	100
4.9	Overlapping Communication and Computation	101
4.10	More on Timing Programs	105
4.11	Three Dimensions	107
4.12	Common Errors and Misunderstandings	107
4.13	Application: Simulating Vortex Evolution in Superconducting Materials	109
<b>5</b>	<b>Advanced Message Passing in MPI</b>	<b>111</b>
5.1	MPI Datatypes	111
5.1.1	Basic Datatypes and Concepts	111
5.1.2	Derived Datatypes	114

5.1.3	Understanding Extents	117
5.2	The N-Body Problem	117
5.2.1	Gather	118
5.2.2	Nonblocking Pipeline	123
5.2.3	Moving Particles between Processes	124
5.2.4	Sending Dynamically Allocated Data	132
5.2.5	User-Controlled Data Packing	134
5.3	Visualizing the Mandelbrot Set	138
5.3.1	Sending Arrays of Structures	145
5.4	Gaps in Datatypes	146
5.4.1	MPI-2 Functions for Manipulating Extents	148
5.5	New MPI-2 Datatype Routines	150
5.6	More on Datatypes for Structures	152
5.7	Deprecated Functions	154
5.8	Common Errors and Misunderstandings	156
<b>6</b>	<b>Parallel Libraries</b>	<b>157</b>
6.1	Motivation	157
6.1.1	The Need for Parallel Libraries	157
6.1.2	Common Deficiencies of Previous Message-Passing Systems	158
6.1.3	Review of MPI Features That Support Libraries	160
6.2	A First MPI Library	163
6.2.1	MPI-2 Attribute-Caching Routines	172
6.2.2	A C++ Alternative to <code>MPI_Comm_dup</code>	172
6.3	Linear Algebra on Grids	177
6.3.1	Mappings and Logical Grids	178
6.3.2	Vectors and Matrices	181
6.3.3	Components of a Parallel Library	185
6.4	The LINPACK Benchmark in MPI	189
6.5	Strategies for Library Building	190
6.6	Examples of Libraries	192

<b>7</b>	<b>Other Features of MPI</b>	<b>195</b>
7.1	Simulating Shared-Memory Operations	195
7.1.1	Shared vs. Distributed Memory	195
7.1.2	A Counter Example	196
7.1.3	The Shared Counter Using Polling instead of an Extra Process	200
7.1.4	Fairness in Message Passing	201
7.1.5	Exploiting Request-Response Message Patterns	202
7.2	Application: Full-Configuration Interaction	205
7.3	Advanced Collective Operations	206
7.3.1	Data Movement	206
7.3.2	Collective Computation	206
7.3.3	Common Errors and Misunderstandings	213
7.4	Intercommunicators	214
7.5	Heterogeneous Computing	220
7.6	The MPI Profiling Interface	222
7.6.1	Finding Buffering Problems	226
7.6.2	Finding Load Imbalances	228
7.6.3	The Mechanics of Using the Profiling Interface	228
7.7	Error Handling	229
7.7.1	Error Handlers	230
7.7.2	An Example of Error Handling	233
7.7.3	User-Defined Error Handlers	234
7.7.4	Terminating MPI Programs	237
7.7.5	MPI-2 Functions for Error Handling	239
7.8	The MPI Environment	240
7.8.1	Processor Name	242
7.8.2	Is MPI Initialized?	242
7.9	Determining the Version of MPI	243
7.10	Other Functions in MPI	245
7.11	Application: Computational Fluid Dynamics	246
7.11.1	Parallel Formulation	246

7.11.2	Parallel Implementation	248
<b>8</b>	<b>Understanding how MPI Implementations Work</b>	<b>253</b>
8.1	Introduction	253
8.1.1	Sending Data	253
8.1.2	Receiving Data	254
8.1.3	Rendezvous Protocol	254
8.1.4	Matching Protocols to MPI's Send Modes	255
8.1.5	Performance Implications	256
8.1.6	Alternative MPI Implementation Strategies	257
8.1.7	Tuning MPI Implementations	257
8.2	How Difficult Is MPI to Implement?	257
8.3	Device Capabilities and the MPI Library Definition	258
8.4	Reliability of Data Transfer	259
<b>9</b>	<b>Comparing MPI with Other Systems for Interprocess Communication</b>	<b>261</b>
9.1	Sockets	261
9.1.1	Process Startup and Shutdown	263
9.1.2	Handling Faults	265
9.2	PVM 3	266
9.2.1	The Basics	267
9.2.2	Miscellaneous Functions	268
9.2.3	Collective Operations	268
9.2.4	MPI Counterparts of Other Features	269
9.2.5	Features Not in MPI	270
9.2.6	Process Startup	270
9.2.7	MPI and PVM related tools	271
9.3	Where to Learn More	272
<b>10</b>	<b>Beyond Message Passing</b>	<b>273</b>
10.1	Dynamic Process Management	274
10.2	Threads	275

10.3	Action at a Distance	276
10.4	Parallel I/O	277
10.5	MPI-2	277
10.6	Will There Be an MPI-3?	278
10.7	Final Words	278
	<b>Glossary of Selected Terms</b>	<b>279</b>
<b>A</b>	<b>Summary of MPI-1 Routines and Their Arguments</b>	<b>289</b>
<b>B</b>	<b>The MPICH Implementation of MPI</b>	<b>329</b>
<b>C</b>	<b>The MPE Multiprocessing Environment</b>	<b>337</b>
<b>D</b>	<b>MPI Resources on the World Wide Web</b>	<b>345</b>
<b>E</b>	<b>Language Details</b>	<b>347</b>
	<a href="#">References</a>	<a href="#">353</a>
	<a href="#">Subject Index</a>	<a href="#">363</a>
	<a href="#">Function and Term Index</a>	<a href="#">367</a>