Contents

	Series Foreword	xiii
	Preface to the Third Edition	xv
	Preface to the Second Edition	xix
	Preface to the First Edition	xxi
1	**Background**	1
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
1.4	Evolution of Message-Passing Systems	10
1.5	The MPI Forum	11
2	**Introduction to MPI**	13
2.1	Goal	13
2.2	What Is MPI?	13
2.3	Basic MPI Concepts	14
2.4	Other Interesting Features of MPI	18
2.5	Is MPI Large or Small?	20
2.6	Decisions Left to the Implementor	21
3	**Using MPI in Simple Programs**	23
3.1	A First MPI Program	23
3.2	Running Your First MPI Program	28
3.3	A First MPI Program in C	29
3.4	Using MPI from Other Languages	29
3.5	Timing MPI Programs	31
3.6	A Self-Scheduling Example: Matrix-Vector Multiplication	32
3.7	Studying Parallel Performance	38
	3.7.1 Elementary Scalability Calculations	39

	3.7.2	Gathering Data on Program Execution	41
	3.7.3	Instrumenting a Parallel Program with MPE Logging	42
	3.7.4	Events and States	43
	3.7.5	Instrumenting the Matrix-Matrix Multiply Program	43
	3.7.6	Notes on Implementation of Logging	47
	3.7.7	Graphical Display of Logfiles	48
3.8	Using Communicators		49
3.9	Another Way of Forming New Communicators		55
3.10	A Handy Graphics Library for Parallel Programs		57
3.11	Common Errors and Misunderstandings		60
3.12	Summary of a Simple Subset of MPI		62
3.13	Application: Computational Fluid Dynamics		62
	3.13.1	Parallel Formulation	63
	3.13.2	Parallel Implementation	65
4	**Intermediate MPI**		69
4.1	The Poisson Problem		70
4.2	Topologies		73
4.3	A Code for the Poisson Problem		81
4.4	Using Nonblocking Communications		91
4.5	Synchronous Sends and "Safe" Programs		94
4.6	More on Scalability		95
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		106
4.12	Common Errors and Misunderstandings		107
4.13	Application: Nek5000/NekCEM		108
5	**Fun with Datatypes**		113

5.1 MPI Datatypes 113
5.1.1 Basic Datatypes and Concepts 113
5.1.2 Derived Datatypes 116
5.1.3 Understanding Extents 118
5.2 The N-Body Problem 119
5.2.1 Gather 120
5.2.2 Nonblocking Pipeline 124
5.2.3 Moving Particles between Processes 127
5.2.4 Sending Dynamically Allocated Data 132
5.2.5 User-Controlled Data Packing 134
5.3 Visualizing the Mandelbrot Set 136
5.3.1 Sending Arrays of Structures 144
5.4 Gaps in Datatypes 146
5.5 More on Datatypes for Structures 148
5.6 Deprecated and Removed Functions 149
5.7 Common Errors and Misunderstandings 150
5.8 Application: Cosmological Large-Scale Structure Formation 152

6 Parallel Libraries 155

6.1 Motivation 155
6.1.1 The Need for Parallel Libraries 155
6.1.2 Common Deficiencies of Early Message-Passing Systems 156
6.1.3 Review of MPI Features That Support Libraries 158
6.2 A First MPI Library 161
6.3 Linear Algebra on Grids 170
6.3.1 Mappings and Logical Grids 170
6.3.2 Vectors and Matrices 175
6.3.3 Components of a Parallel Library 177
6.4 The LINPACK Benchmark in MPI 179
6.5 Strategies for Library Building 183
6.6 Examples of Libraries 184
6.7 Application: Nuclear Green's Function Monte Carlo 185

7	**Other Features of MPI**	189
7.1	Working with Global Data	189
7.1.1	Shared Memory, Global Data, and Distributed Memory	189
7.1.2	A Counter Example	190
7.1.3	The Shared Counter Using Polling Instead of an Extra Process	193
7.1.4	Fairness in Message Passing	196
7.1.5	Exploiting Request-Response Message Patterns	198
7.2	Advanced Collective Operations	201
7.2.1	Data Movement	201
7.2.2	Collective Computation	201
7.2.3	Common Errors and Misunderstandings	206
7.3	Intercommunicators	208
7.4	Heterogeneous Computing	216
7.5	Hybrid Programming with MPI and OpenMP	217
7.6	The MPI Profiling Interface	218
7.6.1	Finding Buffering Problems	221
7.6.2	Finding Load Imbalances	223
7.6.3	Mechanics of Using the Profiling Interface	223
7.7	Error Handling	226
7.7.1	Error Handlers	226
7.7.2	Example of Error Handling	229
7.7.3	User-Defined Error Handlers	229
7.7.4	Terminating MPI Programs	232
7.7.5	Common Errors and Misunderstandings	232
7.8	The MPI Environment	234
7.8.1	Processor Name	236
7.8.2	Is MPI Initialized?	236
7.9	Determining the Version of MPI	237
7.10	Other Functions in MPI	239
7.11	Application: No-Core Configuration Interaction Calculations in Nuclear Physics	240

8 Understanding How MPI Implementations Work 245
8.1 Introduction 245
8.1.1 Sending Data 245
8.1.2 Receiving Data 246
8.1.3 Rendezvous Protocol 246
8.1.4 Matching Protocols to MPI's Send Modes 247
8.1.5 Performance Implications 248
8.1.6 Alternative MPI Implementation Strategies 249
8.1.7 Tuning MPI Implementations 249
8.2 How Difficult Is MPI to Implement? 249
8.3 Device Capabilities and the MPI Library Definition 250
8.4 Reliability of Data Transfer 251

9 Comparing MPI with Sockets 253
9.1 Process Startup and Shutdown 255
9.2 Handling Faults 257

10 Wait! There's More! 259
10.1 Beyond MPI-1 259
10.2 Using Advanced MPI 260
10.3 Will There Be an MPI-4? 261
10.4 Beyond Message Passing Altogether 261
10.5 Final Words 262

Glossary of Selected Terms 263

A The MPE Multiprocessing Environment 273
A.1 MPE Logging 273
A.2 MPE Graphics 275
A.3 MPE Helpers 276

B MPI Resources Online 279

C Language Details 281
C.1 Arrays in C and Fortran 281
C.1.1 Column and Row Major Ordering 281
C.1.2 Meshes vs. Matrices 281
C.1.3 Higher Dimensional Arrays 282
C.2 Aliasing 285

References 287

Subject Index 301

Function and Term Index 305