

Contents

Listings	xv
Preface	xix
Preliminaries.....	xxiii
CHAPTER 1 Introduction	1
1.1 Think Parallel	2
1.2 Performance	4
1.3 Motivation: Pervasive Parallelism	7
1.3.1 Hardware Trends Encouraging Parallelism	7
1.3.2 Observed Historical Trends in Parallelism	11
1.3.3 Need for Explicit Parallel Programming	14
1.4 Structured Pattern-Based Programming	19
1.5 Parallel Programming Models	21
1.5.1 Desired Properties	21
1.5.2 Abstractions Instead of Mechanisms	23
1.5.3 Expression of Regular Data Parallelism	24
1.5.4 Composability	27
1.5.5 Portability of Functionality	28
1.5.6 Performance Portability	28
1.5.7 Safety, Determinism, and Maintainability	29
1.5.8 Overview of Programming Models Used	29
1.5.9 When to Use Which Model?	36
1.6 Organization of this Book	37
1.7 Summary	38
CHAPTER 2 Background	39
2.1 Vocabulary and Notation	39
2.2 Strategies	40
2.3 Mechanisms	41
2.4 Machine Models	44
2.4.1 Machine Model	44
2.4.2 Key Features for Performance	50
2.4.3 Flynn's Characterization	51
2.4.4 Evolution.....	53
2.5 Performance Theory	54
2.5.1 Latency and Throughput	55
2.5.2 Speedup, Efficiency, and Scalability.....	56

2.5.3	Power	57
2.5.4	Amdahl's Law	58
2.5.5	Gustafson-Barsis' Law	60
2.5.6	Work-Span Model	62
2.5.7	Asymptotic Complexity	65
2.5.8	Asymptotic Speedup and Efficiency	67
2.5.9	Little's Formula	67
2.6	Pitfalls	68
2.6.1	Race Conditions	68
2.6.2	Mutual Exclusion and Locks	70
2.6.3	Deadlock	72
2.6.4	Strangled Scaling	73
2.6.5	Lack of Locality	73
2.6.6	Load Imbalance	74
2.6.7	Overhead	74
2.7	Summary	75

PART I PATTERNS

CHAPTER 3	Patterns	79
3.1	Nesting Pattern	80
3.2	Structured Serial Control Flow Patterns	82
3.2.1	Sequence	82
3.2.2	Selection	84
3.2.3	Iteration	84
3.2.4	Recursion	87
3.3	Parallel Control Patterns	88
3.3.1	Fork-Join	88
3.3.2	Map	88
3.3.3	Stencil	89
3.3.4	Reduction	90
3.3.5	Scan	92
3.3.6	Recurrence	95
3.4	Serial Data Management Patterns	95
3.4.1	Random Read and Write	96
3.4.2	Stack Allocation	96
3.4.3	Heap Allocation	96
3.4.4	Closures	97
3.4.5	Objects	97

3.5	Parallel Data Management Patterns	98
3.5.1	Pack	98
3.5.2	Pipeline.....	99
3.5.3	Geometric Decomposition	100
3.5.4	Gather	101
3.5.5	Scatter	101
3.6	Other Parallel Patterns	102
3.6.1	Superscalar Sequences	102
3.6.2	Futures	102
3.6.3	Speculative Selection.....	104
3.6.4	Workpile	105
3.6.5	Search	105
3.6.6	Segmentation	105
3.6.7	Expand	106
3.6.8	Category Reduction	106
3.6.9	Term Graph Rewriting	107
3.7	Non-Deterministic Patterns	108
3.7.1	Branch and Bound.....	108
3.7.2	Transactions.....	109
3.8	Programming Model Support for Patterns	110
3.8.1	Cilk Plus	112
3.8.2	Threading Building Blocks	113
3.8.3	OpenMP.....	114
3.8.4	Array Building Blocks	115
3.8.5	OpenCL	116
3.9	Summary	118
CHAPTER 4	Map	121
4.1	Map.....	123
4.2	Scaled Vector Addition (SAXPY)	124
4.2.1	Description of the Problem.....	124
4.2.2	Serial Implementation	125
4.2.3	TBB	125
4.2.4	Cilk Plus	127
4.2.5	Cilk Plus with Array Notation	127
4.2.6	OpenMP.....	128
4.2.7	ArBB Using Vector Operations	128
4.2.8	ArBB Using Elemental Functions	129
4.2.9	OpenCL	130

4.3	Mandelbrot	131
4.3.1	Description of the Problem	131
4.3.2	Serial Implementation	132
4.3.3	TBB	132
4.3.4	Cilk Plus	132
4.3.5	Cilk Plus with Array Notations	134
4.3.6	OpenMP	134
4.3.7	ArBB	134
4.3.8	OpenCL	138
4.4	Sequence of Maps versus Map of Sequence	139
4.5	Comparison of Parallel Models	141
4.6	Related Patterns	141
4.6.1	Stencil	141
4.6.2	Workpile	142
4.6.3	Divide-and-conquer	142
4.7	Summary	143
CHAPTER 5	Collectives	145
5.1	Reduce	145
5.1.1	Reordering Computations	146
5.1.2	Vectorization	148
5.1.3	Tiling	149
5.1.4	Precision	150
5.1.5	Implementation	151
5.2	Fusing Map and Reduce	152
5.2.1	Explicit Fusion in TBB	152
5.2.2	Explicit Fusion in Cilk Plus	153
5.2.3	Automatic Fusion in ArBB	153
5.3	Dot Product	154
5.3.1	Description of the Problem	154
5.3.2	Serial Implementation	154
5.3.3	SSE Intrinsics	155
5.3.4	TBB	156
5.3.5	Cilk Plus	158
5.3.6	OpenMP	160
5.3.7	ArBB	161
5.4	Scan	162
5.4.1	Cilk Plus	164
5.4.2	TBB	165
5.4.3	ArBB	165
5.4.4	OpenMP	165
5.5	Fusing Map and Scan	166

5.6	Integration	169
5.6.1	Description of the Problem	170
5.6.2	Serial Implementation	170
5.6.3	Cilk Plus	170
5.6.4	OpenMP	170
5.6.5	TBB	172
5.6.6	ArBB	175
5.7	Summary	177
CHAPTER 6	Data Reorganization	179
6.1	Gather	180
6.1.1	General Gather	180
6.1.2	Shift	182
6.1.3	Zip	182
6.2	Scatter	182
6.2.1	Atomic Scatter	184
6.2.2	Permutation Scatter	184
6.2.3	Merge Scatter	185
6.2.4	Priority Scatter	185
6.3	Converting Scatter to Gather	186
6.4	Pack	187
6.5	Fusing Map and Pack	189
6.6	Geometric Decomposition and Partition	191
6.7	Array of Structures vs. Structures of Arrays	194
6.8	Summary	197
CHAPTER 7	Stencil and Recurrence	199
7.1	Stencil	199
7.2	Implementing Stencil with Shift	201
7.3	Tiling Stencils for Cache	202
7.4	Optimizing Stencils for Communication	203
7.5	Recurrence	204
7.6	Summary	207
CHAPTER 8	Fork-Join	209
8.1	Definition	210
8.2	Programming Model Support for Fork-Join	211
8.2.1	Cilk Plus Support for Fork-Join	212
8.2.2	TBB Support for Fork-Join	213
8.2.3	OpenMP Support for Fork-Join	214
8.3	Recursive Implementation of Map	215
8.4	Choosing Base Cases	217

8.5	Load Balancing	218
8.6	Complexity of Parallel Divide-and-Conquer	221
8.7	Karatsuba Multiplication of Polynomials	224
8.7.1	Note on Allocating Scratch Space	227
8.8	Cache Locality and Cache-Oblivious Algorithms	227
8.9	Quicksort	230
8.9.1	Cilk Quicksort	231
8.9.2	TBB Quicksort	233
8.9.3	Work and Span for Quicksort	237
8.10	Reductions and Hyperobjects	238
8.11	Implementing Scan with Fork-Join	241
8.12	Applying Fork-Join to Recurrences	246
8.12.1	Analysis	250
8.12.2	Flat Fork-Join	251
8.13	Summary	251
CHAPTER 9	Pipeline	253
9.1	Basic Pipeline	253
9.2	Pipeline with Parallel Stages	254
9.3	Implementation of a Pipeline	255
9.4	Programming Model Support for Pipelines	257
9.4.1	Pipeline in TBB	257
9.4.2	Pipeline in Cilk Plus	258
9.5	More General Topologies	261
9.6	Mandatory versus Optional Parallelism	261
9.7	Summary	262
PART II	EXAMPLES	
CHAPTER 10	Forward Seismic Simulation	265
10.1	Background	265
10.2	Stencil Computation	266
10.3	Impact of Caches on Arithmetic Intensity	267
10.4	Raising Arithmetic Intensity with Space-Time Tiling	270
10.5	Cilk Plus Code	272
10.6	ArBB Implementation	275
10.7	Summary	277

CHAPTER 11	K-Means Clustering	279
11.1	Algorithm	279
11.2	K-Means with Cilk Plus	281
11.2.1	Hyperobjects	281
11.3	K-Means with TBB	285
11.4	Summary	289
CHAPTER 12	Bzip2 Data Compression	291
12.1	The Bzip2 Algorithm	291
12.2	Three-Stage Pipeline Using TBB	292
12.3	Four-Stage Pipeline Using TBB	296
12.4	Three-Stage Pipeline Using Cilk Plus	296
12.5	Summary	297
CHAPTER 13	Merge Sort	299
13.1	Parallel Merge	299
13.1.1	TBB Parallel Merge	301
13.1.2	Work and Span of Parallel Merge	301
13.2	Parallel Merge Sort	303
13.2.1	Work and Span of Merge Sort	304
13.3	Summary	305
CHAPTER 14	Sample Sort	307
14.1	Overall Structure	307
14.2	Choosing the Number of Bins	309
14.3	Binning	309
14.3.1	TBB Implementation	310
14.4	Repacking and Subsorting	310
14.5	Performance Analysis of Sample Sort	312
14.6	For C++ Experts	313
14.7	Summary	313
CHAPTER 15	Cholesky Factorization	315
15.1	Fortran Rules!	315
15.2	Recursive Cholesky Decomposition	317
15.3	Triangular Solve	318
15.4	Symmetric Rank Update	319
15.5	Where Is the Time Spent?	321
15.6	Summary	322

APPENDICES

APPENDIX A	Further Reading	325
	A.1 Parallel Algorithms and Patterns.....	325
	A.2 Computer Architecture Including Parallel Systems.....	325
	A.3 Parallel Programming Models	326
APPENDIX B	Cilk Plus	329
	B.1 Shared Design Principles with TBB	329
	B.2 Unique Characteristics	329
	B.3 Borrowing Components from TBB.....	331
	B.4 Keyword Spelling	332
	B.5 <code>cilk_for</code>	332
	B.6 <code>cilk_spawn</code> and <code>cilk_sync</code>	333
	B.7 Reducers (Hyperobjects).....	334
	B.7.1 C++ Syntax.....	335
	B.7.2 C Syntax.....	337
	B.8 Array Notation.....	338
	B.8.1 Specifying Array Sections.....	339
	B.8.2 Operations on Array Sections	340
	B.8.3 Reductions on Array Sections	341
	B.8.4 Implicit Index	342
	B.8.5 Avoid Partial Overlap of Array Sections	342
	B.9 <code>#pragma simd</code>	343
	B.10 Elemental Functions.....	344
	B.10.1 Attribute Syntax	345
	B.11 Note on C++11	345
	B.12 Notes on Setup.....	346
	B.13 History	346
	B.14 Summary.....	347
APPENDIX C	TBB	349
	C.1 Unique Characteristics	349
	C.2 Using TBB	350
	C.3 <code>parallel_for</code>	351
	C.3.1 <code>blocked_range</code>	351
	C.3.2 Partitioners	352
	C.4 <code>parallel_reduce</code>	353
	C.5 <code>parallel_deterministic_reduce</code>	354
	C.6 <code>parallel_pipeline</code>	354
	C.7 <code>parallel_invoke</code>	354

C.8 task_group	355
C.9 task.....	355
C.9.1 empty_task.....	356
C.10 atomic.....	356
C.11 enumerable_thread_specific.....	358
C.12 Notes on C++11	358
C.13 History	359
C.14 Summary.....	360
APPENDIX D C++11	361
D.1 Declaring with auto	361
D.2 Lambda Expressions	361
D.3 std::move	365
APPENDIX E Glossary	367
Bibliography	391
Index	397