Acknowledgments xvii
Preface xix
About the Authors xxv

Chapter 1  Introduction to Interconnection Networks 1
1.1 Three Questions About Interconnection Networks 2
1.2 Uses of Interconnection Networks 4
  1.2.1 Processor-Memory Interconnect 5
  1.2.2 I/O Interconnect 8
  1.2.3 Packet Switching Fabric 11
1.3 Network Basics 13
  1.3.1 Topology 13
  1.3.2 Routing 16
  1.3.3 Flow Control 17
1.4 History 21
1.5 Organization of this Book 23

Chapter 2  A Simple Interconnection Network 25
2.1 Network Specifications and Constraints 25
2.2 Topology 27
2.3 Routing 31
2.4 Flow Control 32
2.5 Router Design 33
2.6 Performance Analysis 36
2.7 Exercises 42
## Chapter 3  Topology Basics  

### 3.1 Nomenclature
- 3.1.1 Channels and Nodes  46
- 3.1.2 Direct and Indirect Networks  47
- 3.1.3 Cuts and Bisections  48
- 3.1.4 Paths  48
- 3.1.5 Symmetry  49

### 3.2 Traffic Patterns

### 3.3 Performance
- 3.3.1 Throughput and Maximum Channel Load  51
- 3.3.2 Latency  55
- 3.3.3 Path Diversity  57

### 3.4 Packaging Cost

### 3.5 Case Study: The SGI Origin 2000

### 3.6 Bibliographic Notes

### 3.7 Exercises

## Chapter 4  Butterfly Networks

### 4.1 The Structure of Butterfly Networks

### 4.2 Isomorphic Butterflies

### 4.3 Performance and Packaging Cost

### 4.4 Path Diversity and Extra Stages

### 4.5 Case Study: The BBN Butterfly

### 4.6 Bibliographic Notes

### 4.7 Exercises

## Chapter 5  Torus Networks

### 5.1 The Structure of Torus Networks

### 5.2 Performance
- 5.2.1 Throughput  92
- 5.2.2 Latency  95
- 5.2.3 Path Diversity  96

### 5.3 Building Mesh and Torus Networks

### 5.4 Express Cubes

### 5.5 Case Study: The MIT J-Machine

### 5.6 Bibliographic Notes

### 5.7 Exercises
Chapter 6  Non-Blocking Networks  111

6.1  Non-Blocking vs. Non-Interfering Networks  112
6.2  Crossbar Networks  112
6.3  Clos Networks  116
   6.3.1  Structure and Properties of Clos Networks  116
   6.3.2  Unicast Routing on Strictly Non-Blocking Clos Networks  118
   6.3.3  Unicast Routing on Rearrangeable Clos Networks  122
   6.3.4  Routing Clos Networks Using Matrix Decomposition  126
   6.3.5  Multicast Routing on Clos Networks  128
   6.3.6  Clos Networks with More Than Three Stages  133
6.4  Beneš Networks  134
6.5  Sorting Networks  135
6.6  Case Study: The Velio VC2002 (Zeus) Grooming Switch  137
6.7  Bibliographic Notes  142
6.8  Exercises  142

Chapter 7  Slicing and Dicing  145

7.1  Concentrators and Distributors  146
   7.1.1  Concentrators  146
   7.1.2  Distributors  148
7.2  Slicing and Dicing  149
   7.2.1  Bit Slicing  149
   7.2.2  Dimension Slicing  151
   7.2.3  Channel Slicing  152
7.3  Slicing Multistage Networks  153
7.4  Case Study: Bit Slicing in the Tiny Tera  155
7.5  Bibliographic Notes  157
7.6  Exercises  157

Chapter 8  Routing Basics  159

8.1  A Routing Example  160
8.2  Taxonomy of Routing Algorithms  162
8.3  The Routing Relation  163
8.4  Deterministic Routing  164
   8.4.1  Destination-Tag Routing in Butterfly Networks  165
   8.4.2  Dimension-Order Routing in Cube Networks  166
# Contents

8.5 Case Study: Dimension-Order Routing in the Cray T3D 168
8.6 Bibliographic Notes 170
8.7 Exercises 171

**Chapter 9** Oblivious Routing 173

9.1 Valiant's Randomized Routing Algorithm 174
  9.1.1 Valiant's Algorithm on Torus Topologies 174
  9.1.2 Valiant's Algorithm on Indirect Networks 175

9.2 Minimal Oblivious Routing 176
  9.2.1 Minimal Oblivious Routing on a Folded Clos (Fat Tree) 176
  9.2.2 Minimal Oblivious Routing on a Torus 178

9.3 Load-Balanced Oblivious Routing 180
9.4 Analysis of Oblivious Routing 180
9.5 Case Study: Oblivious Routing in the Avici Terabit Switch Router (TSR) 183
9.6 Bibliographic Notes 186
9.7 Exercises 187

**Chapter 10** Adaptive Routing 189

10.1 Adaptive Routing Basics 189
10.2 Minimal Adaptive Routing 192
10.3 Fully Adaptive Routing 193
10.4 Load-Balanced Adaptive Routing 195
10.5 Search-Based Routing 196
10.6 Case Study: Adaptive Routing in the Thinking Machines CM-5 196
10.7 Bibliographic Notes 201
10.8 Exercises 201

**Chapter 11** Routing Mechanics 203

11.1 Table-Based Routing 203
  11.1.1 Source Routing 204
  11.1.2 Node-Table Routing 208

11.2 Algorithmic Routing 211
11.3 Case Study: Oblivious Source Routing in the IBM Vulcan Network 212
14.3 Adaptive Routing 272
   14.3.1 Routing Subfunctions and Extended Dependences 272
   14.3.2 Duato's Protocol for Deadlock-Free Adaptive Algorithms 276

14.4 Deadlock Recovery 277
   14.4.1 Regressive Recovery 278
   14.4.2 Progressive Recovery 278

14.5 Livelock 279

14.6 Case Study: Deadlock Avoidance in the Cray T3E 279

14.7 Bibliographic Notes 281

14.8 Exercises 282

Chapter 15 Quality of Service 285

15.1 Service Classes and Service Contracts 285

15.2 Burstiness and Network Delays 287
   15.2.1 \((a, p)\) Regulated Flows 287
   15.2.2 Calculating Delays 288

15.3 Implementation of Guaranteed Services 290
   15.3.1 Aggregate Resource Allocation 291
   15.3.2 Resource Reservation 292

15.4 Implementation of Best-Effort Services 294
   15.4.1 Latency Fairness 294
   15.4.2 Throughput Fairness 296

15.5 Separation of Resources 297
   15.5.1 Tree Saturation 297
   15.5.2 Non-interfering Networks 299

15.6 Case Study: ATM Service Classes 299

15.7 Case Study: Virtual Networks in the Avici TSR 300

15.8 Bibliographic Notes 302

15.9 Exercises 303

Chapter 16 Router Architecture 305

16.1 Basic Router Architecture 305
   16.1.1 Block Diagram 305
   16.1.2 The Router Pipeline 308

16.2 Stalls 310

16.3 Closing the Loop with Credits 312

16.4 Reallocating a Channel 313

16.5 Speculation and Lookahead 316
16.6 Flit and Credit Encoding 319
16.7 Case Study: The Alpha 21364 Router 321
16.8 Bibliographic Notes 324
16.9 Exercises 324

Chapter 17 Router Datapath Components 325
17.1 Input Buffer Organization 325
  17.1.1 Buffer Partitioning 326
  17.1.2 Input Buffer Data Structures 328
  17.1.3 Input Buffer Allocation 333
17.2 Switches 334
  17.2.1 Bus Switches 335
  17.2.2 Crossbar Switches 338
  17.2.3 Network Switches 342
17.3 Output Organization 343
17.4 Case Study: The Datapath of the IBM Colony Router 344
17.5 Bibliographic Notes 347
17.6 Exercises 348

Chapter 18 Arbitration 349
18.1 Arbitration Timing 349
18.2 Fairness 351
18.3 Fixed Priority Arbiter 352
18.4 Variable Priority Iterative Arbiters 354
  18.4.1 Oblivious Arbiters 354
  18.4.2 Round-Robin Arbiter 355
  18.4.3 Grant-Hold Circuit 355
  18.4.4 Weighted Round-Robin Arbiter 357
18.5 Matrix Arbiter 358
18.6 Queuing Arbiter 360
18.7 Exercises 362

Chapter 19 Allocation 363
19.1 Representations 363
19.2 Exact Algorithms 366
19.3 Separable Allocators 367
  19.3.1 Parallel Iterative Matching 371
  19.3.2 iSLIP 371
  19.3.3 Lonely Output Allocator 372
19.4 Wavefront Allocator 373
19.5 Incremental vs. Batch Allocation 376
19.6 Multistage Allocation 378
19.7 Performance of Allocators 380
19.8 Case Study: The Tiny Tera Allocator 383
19.9 Bibliographic Notes 385
19.10 Exercises 386

Chapter 20  Network Interfaces 389
20.1 Processor-Network Interface 390
  20.1.1 Two-Register Interface 391
  20.1.2 Register-Mapped Interface 392
  20.1.3 Descriptor-Based Interface 393
  20.1.4 Message Reception 393
20.2 Shared-Memory Interface 394
  20.2.1 Processor-Network Interface 395
  20.2.2 Cache Coherence 397
  20.2.3 Memory-Network Interface 398
20.3 Line-Fabric Interface 400
20.4 Case Study: The MIT M-Machine Network Interface 403
20.5 Bibliographic Notes 407
20.6 Exercises 408

Chapter 21  Error Control 411
21.1 Know Thy Enemy: Failure Modes and Fault Models 411
21.2 The Error Control Process: Detection, Containment, and Recovery 414
21.3 Link Level Error Control 415
  21.3.1 Link Monitoring 415
  21.3.2 Link-Level Retransmission 416
  21.3.3 Channel Reconfiguration, Degradation, and Shutdown 419
21.4 Router Error Control 421
21.5 Network-Level Error Control 422
21.6 End-to-end Error Control 423
21.7 Bibliographic Notes 423
21.8 Exercises 424
### Chapter 25 Simulation Examples

#### 25.1 Routing
- 25.1.1 Latency 496
- 25.1.2 Throughput Distributions 499

#### 25.2 Flow Control Performance
- 25.2.1 Virtual Channels 500
- 25.2.2 Network Size 502
- 25.2.3 Injection Processes 503
- 25.2.4 Prioritization 505
- 25.2.5 Stability 507

#### 25.3 Fault Tolerance 508

### Appendix A Nomenclature 511
### Appendix B Glossary 515
### Appendix C Network Simulator 521
### Bibliography 523
### Index 539