


Contents


Preface xv


CHAPTERS

1 Computer Abstractions and Technology 2


- 1.1 Introduction 3
- 1.2 Eight Great Ideas in Computer Architecture 11
- 1.3 Below Your Program 13
- 1.4 Under the Covers 16
- 1.5 Technologies for Building Processors and Memory 24
- 1.6 Performance 28
- 1.7 The Power Wall 40
- 1.8 The Sea Change: The Switch from Uniprocessors to Multiprocessors 43
- 1.9 Real Stuff: Benchmarking the Intel Core i7 46
- 1.10 Fallacies and Pitfalls 49
- 1.11 Concluding Remarks 52
-  1.12 Historical Perspective and Further Reading 54
- 1.13 Exercises 54

2 Instructions: Language of the Computer 60



- 2.1 Introduction 62
- 2.2 Operations of the Computer Hardware 63
- 2.3 Operands of the Computer Hardware 67
- 2.4 Signed and Unsigned Numbers 74
- 2.5 Representing Instructions in the Computer 81
- 2.6 Logical Operations 89
- 2.7 Instructions for Making Decisions 92
- 2.8 Supporting Procedures in Computer Hardware 98
- 2.9 Communicating with People 108
- 2.10 RISC-V Addressing for Wide Immediates and Addresses 113
- 2.11 Parallelism and Instructions: Synchronization 121
- 2.12 Translating and Starting a Program 124
- 2.13 A C Sort Example to Put it All Together 133
- 2.14 Arrays versus Pointers 141
-  2.15 Advanced Material: Compiling C and Interpreting Java 144

- 2.16 Real Stuff: MIPS Instructions 145
- 2.17 Real Stuff: x86 Instructions 146
- 2.18 Real Stuff: The Rest of the RISC-V Instruction Set 155
- 2.19 Fallacies and Pitfalls 157
- 2.20 Concluding Remarks 159
-  2.21 Historical Perspective and Further Reading 162
- 2.22 Exercises 162




3**Arithmetic for Computers 172**

- 3.1 Introduction 174
- 3.2 Addition and Subtraction 174
- 3.3 Multiplication 177
- 3.4 Division 183
- 3.5 Floating Point 191
- 3.6 Parallelism and Computer Arithmetic: Subword Parallelism 216
- 3.7 Real Stuff: Streaming SIMD Extensions and Advanced Vector Extensions in x86 217
- 3.8 Going Faster: Subword Parallelism and Matrix Multiply 218
- 3.9 Fallacies and Pitfalls 222
- 3.10 Concluding Remarks 225
-  3.11 Historical Perspective and Further Reading 227
- 3.12 Exercises 227



4**The Processor 234**

- 4.1 Introduction 236
- 4.2 Logic Design Conventions 240
- 4.3 Building a Datapath 243
- 4.4 A Simple Implementation Scheme 251
- 4.5 An Overview of Pipelining 262
- 4.6 Pipelined Datapath and Control 276
- 4.7 Data Hazards: Forwarding versus Stalling 294
- 4.8 Control Hazards 307
- 4.9 Exceptions 315
- 4.10 Parallelism via Instructions 321
- 4.11 Real Stuff: The ARM Cortex-A53 and Intel Core i7 Pipelines 334
- 4.12 Going Faster: Instruction-Level Parallelism and Matrix Multiply 342
-  4.13 Advanced Topic: An Introduction to Digital Design Using a Hardware Design Language to Describe and Model a Pipeline and More Pipelining Illustrations 345
- 4.14 Fallacies and Pitfalls 345
- 4.15 Concluding Remarks 346
-  4.16 Historical Perspective and Further Reading 347
- 4.17 Exercises 347

5 Large and Fast: Exploiting Memory Hierarchy 364

- 5.1 Introduction 366
- 5.2 Memory Technologies 370
- 5.3 The Basics of Caches 375
- 5.4 Measuring and Improving Cache Performance 390
- 5.5 Dependable Memory Hierarchy 410
- 5.6 Virtual Machines 416
- 5.7 Virtual Memory 419
- 5.8 A Common Framework for Memory Hierarchy 443
- 5.9 Using a Finite-State Machine to Control a Simple Cache 449
- 5.10 Parallelism and Memory Hierarchy: Cache Coherence 454
-  5.11 Parallelism and Memory Hierarchy: Redundant Arrays of Inexpensive Disks 458
-  5.12 Advanced Material: Implementing Cache Controllers 459
- 5.13 Real Stuff: The ARM Cortex-A53 and Intel Core i7 Memory Hierarchies 459
- 5.14 Real Stuff: The Rest of the RISC-V System and Special Instructions 464
- 5.15 Going Faster: Cache Blocking and Matrix Multiply 465
- 5.16 Fallacies and Pitfalls 468
- 5.17 Concluding Remarks 472
-  5.18 Historical Perspective and Further Reading 473
- 5.19 Exercises 473

6 Parallel Processors from Client to Cloud 490

- 6.1 Introduction 492
- 6.2 The Difficulty of Creating Parallel Processing Programs 494
- 6.3 SISD, MIMD, SIMD, SPMD, and Vector 499
- 6.4 Hardware Multithreading 506
- 6.5 Multicore and Other Shared Memory Multiprocessors 509
- 6.6 Introduction to Graphics Processing Units 514
- 6.7 Clusters, Warehouse Scale Computers, and Other Message-Passing Multiprocessors 521
- 6.8 Introduction to Multiprocessor Network Topologies 526
-  6.9 Communicating to the Outside World: Cluster Networking 529
- 6.10 Multiprocessor Benchmarks and Performance Models 530
- 6.11 Real Stuff: Benchmarking and Rooflines of the Intel Core i7 960 and the NVIDIA Tesla GPU 540
- 6.12 Going Faster: Multiple Processors and Matrix Multiply 545
- 6.13 Fallacies and Pitfalls 548
- 6.14 Concluding Remarks 550
-  6.15 Historical Perspective and Further Reading 553
- 6.16 Exercises 553

APPENDIX



The Basics of Logic Design A-2

- A.1 Introduction A-3
- A.2 Gates, Truth Tables, and Logic Equations A-4
- A.3 Combinational Logic A-9
- A.4 Using a Hardware Description Language A-20
- A.5 Constructing a Basic Arithmetic Logic Unit A-26
- A.6 Faster Addition: Carry Lookahead A-37
- A.7 Clocks A-47
- A.8 Memory Elements: Flip-Flops, Latches, and Registers A-49
- A.9 Memory Elements: SRAMs and DRAMs A-57
- A.10 Finite-State Machines A-66
- A.11 Timing Methodologies A-71
- A.12 Field Programmable Devices A-77
- A.13 Concluding Remarks A-78
- A.14 Exercises A-79

Index I-1

ONLINE CONTENT



Graphics and Computing GPUs B-2

- B.1 Introduction B-3
- B.2 GPU System Architectures B-7
- B.3 Programming GPUs B-12
- B.4 Multithreaded Multiprocessor Architecture B-25
- B.5 Parallel Memory System B-36
- B.6 Floating Point Arithmetic B-41
- B.7 Real Stuff: The NVIDIA GeForce 8800 B-46
- B.8 Real Stuff: Mapping Applications to GPUs B-55
- B.9 Fallacies and Pitfalls B-72
- B.10 Concluding Remarks B-76
- B.11 Historical Perspective and Further Reading B-77




Mapping Control to Hardware C-2


- C.1 Introduction C-3
- C.2 Implementing Combinational Control Units C-4
- C.3 Implementing Finite-State Machine Control C-8
- C.4 Implementing the Next-State Function with a Sequencer C-22
- C.5 Translating a Microprogram to Hardware C-28
- C.6 Concluding Remarks C-32
- C.7 Exercises C-33



A Survey of RISC Architectures for Desktop, Server, and Embedded Computers D-2

- D.1 Introduction D-3
- D.2 Addressing Modes and Instruction Formats D-5
- D.3 Instructions: the MIPS Core Subset D-9
- D.4 Instructions: Multimedia Extensions of the Desktop/Server RISCs D-16
- D.5 Instructions: Digital Signal-Processing Extensions of the Embedded RISCs D-19
- D.6 Instructions: Common Extensions to MIPS Core D-20
- D.7 Instructions Unique to MIPS-64 D-25
- D.8 Instructions Unique to Alpha D-27
- D.9 Instructions Unique to SPARC v9 D-29
- D.10 Instructions Unique to PowerPC D-32
- D.11 Instructions Unique to PA-RISC 2.0 D-34
- D.12 Instructions Unique to ARM D-36
- D.13 Instructions Unique to Thumb D-38
- D.14 Instructions Unique to SuperH D-39
- D.15 Instructions Unique to M32R D-40
- D.16 Instructions Unique to MIPS-16 D-40
- D.17 Concluding Remarks D-43

 Glossary G-1

 Further Reading FR-1