

REVISED FOURTH EDITION

Computer Organization and Design

THE HARDWARE / SOFTWARE INTERFACE

David A. Patterson

University of California, Berkeley

John L. Hennessy

Stanford University

With contributions by

Perry Alexander
The University of Kansas

Peter J. Ashenden
Ashenden Designs Pty Ltd

Javier Bruguera
Universidade de Santiago de Compostela

Jichuan Chang
Hewlett-Packard

Matthew Farrens
University of California, Davis

David Kaeli
Northeastern University

Nicole Kaiyan
University of Adelaide

David Kirk
NVIDIA

James R. Larus
Microsoft Research

Jacob Leverich
Hewlett-Packard

Kevin Lim
Hewlett-Packard

John Nickolls
NVIDIA

John Oliver
Cal Poly, San Luis Obispo

Milos Prvulovic
Georgia Tech

Partha Ranganathan
Hewlett-Packard



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Morgan Kaufmann is an imprint of Elsevier



Contents

Preface xv



CHAPTERS

1 Computer Abstractions and Technology 2


- 1.1 Introduction 3
- 1.2 Below Your Program 10
- 1.3 Under the Covers 13
- 1.4 Performance 26
- 1.5 The Power Wall 39
- 1.6 The Sea Change: The Switch from Uniprocessors to Multiprocessors 41
- 1.7 Real Stuff: Manufacturing and Benchmarking the AMD Opteron X4 44
- 1.8 Fallacies and Pitfalls 51
- 1.9 Concluding Remarks 54
- 1.10 Historical Perspective and Further Reading 55
- 1.11 Exercises 56

2 Instructions: Language of the Computer 74



- 2.1 Introduction 76
- 2.2 Operations of the Computer Hardware 77
- 2.3 Operands of the Computer Hardware 80
- 2.4 Signed and Unsigned Numbers 87
- 2.5 Representing Instructions in the Computer 94
- 2.6 Logical Operations 102
- 2.7 Instructions for Making Decisions 105
- 2.8 Supporting Procedures in Computer Hardware 112
- 2.9 Communicating with People 122
- 2.10 MIPS Addressing for 32-Bit Immediates and Addresses 128
- 2.11 Parallelism and Instructions: Synchronization 137
- 2.12 Translating and Starting a Program 139
- 2.13 A C Sort Example to Put It All Together 149

- 2.14 Arrays versus Pointers 157
-  2.15 Advanced Material: Compiling C and Interpreting Java 161
- 2.16 Real Stuff: ARM Instructions 161
- 2.17 Real Stuff: x86 Instructions 165
- 2.18 Fallacies and Pitfalls 174
- 2.19 Concluding Remarks 176
-  2.20 Historical Perspective and Further Reading 179
- 2.21 Exercises 179

3 Arithmetic for Computers 222

- 3.1 Introduction 224
- 3.2 Addition and Subtraction 224
- 3.3 Multiplication 230
- 3.4 Division 236
- 3.5 Floating Point 242
- 3.6 Parallelism and Computer Arithmetic: Associativity 270
- 3.7 Real Stuff: Floating Point in the x86 272
- 3.8 Fallacies and Pitfalls 275
- 3.9 Concluding Remarks 280
-  3.10 Historical Perspective and Further Reading 283
- 3.11 Exercises 283

4 The Processor 298

- 4.1 Introduction 300
- 4.2 Logic Design Conventions 303
- 4.3 Building a Datapath 307
- 4.4 A Simple Implementation Scheme 316
- 4.5 An Overview of Pipelining 330
- 4.6 Pipelined Datapath and Control 344
- 4.7 Data Hazards: Forwarding versus Stalling 363
- 4.8 Control Hazards 375
- 4.9 Exceptions 384
- 4.10 Parallelism and Advanced Instruction-Level Parallelism 391
- 4.11 Real Stuff: the AMD Opteron X4 (Barcelona) Pipeline 404
-  4.12 Advanced Topic: an Introduction to Digital Design Using a Hardware Design Language to Describe and Model a Pipeline and More Pipelining Illustrations 406
- 4.13 Fallacies and Pitfalls 407
- 4.14 Concluding Remarks 408
-  4.15 Historical Perspective and Further Reading 409
- 4.16 Exercises 409

5 Large and Fast: Exploiting Memory Hierarchy 450

- 5.1 Introduction 452
- 5.2 The Basics of Caches 457
- 5.3 Measuring and Improving Cache Performance 475
- 5.4 Virtual Memory 492
- 5.5 A Common Framework for Memory Hierarchies 518
- 5.6 Virtual Machines 525
- 5.7 Using a Finite-State Machine to Control a Simple Cache 529
- 5.8 Parallelism and Memory Hierarchies: Cache Coherence 534
- 5.9 Advanced Material: Implementing Cache Controllers 538
- 5.10 Real Stuff: the AMD Opteron X4 (Barcelona) and Intel Nehalem Memory Hierarchies 539
- 5.11 Fallacies and Pitfalls 543
- 5.12 Concluding Remarks 547
- 5.13 Historical Perspective and Further Reading 548
- 5.14 Exercises 548

6 Storage and Other I/O Topics 568

- 6.1 Introduction 570
- 6.2 Dependability, Reliability, and Availability 573
- 6.3 Disk Storage 575
- 6.4 Flash Storage 580
- 6.5 Connecting Processors, Memory, and I/O Devices 582
- 6.6 Interfacing I/O Devices to the Processor, Memory, and Operating System 586
- 6.7 I/O Performance Measures: Examples from Disk and File Systems 596
- 6.8 Designing an I/O System 598
- 6.9 Parallelism and I/O: Redundant Arrays of Inexpensive Disks 599
- 6.10 Real Stuff: Sun Fire x4150 Server 606
- 6.11 Advanced Topics: Networks 612
- 6.12 Fallacies and Pitfalls 613
- 6.13 Concluding Remarks 617
- 6.14 Historical Perspective and Further Reading 618
- 6.15 Exercises 619

7 Multicores, Multiprocessors, and Clusters 630

- 7.1 Introduction 632
- 7.2 The Difficulty of Creating Parallel Processing Programs 634
- 7.3 Shared Memory Multiprocessors 638

- 7.4 Clusters and Other Message-Passing Multiprocessors 641
- 7.5 Hardware Multithreading 645
- 7.6 SISD, MIMD, SIMD, SPMD, and Vector 648
- 7.7 Introduction to Graphics Processing Units 654
- 7.8 Introduction to Multiprocessor Network Topologies 660
- 7.9 Multiprocessor Benchmarks 664
- 7.10 Roofline: A Simple Performance Model 667
- 7.11 Real Stuff: Benchmarking Four Multicores Using the Roofline Model 675
- 7.12 Fallacies and Pitfalls 684
- 7.13 Concluding Remarks 686
- 7.14 Historical Perspective and Further Reading 688
- 7.15 Exercises 688

A P P E N D I C E S

A Graphics and Computing GPUs A-2

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

B Assemblers, Linkers, and the SPIM Simulator B-2

- B.1 Introduction B-3
- B.2 Assemblers B-10
- B.3 Linkers B-18
- B.4 Loading B-19
- B.5 Memory Usage B-20
- B.6 Procedure Call Convention B-22
- B.7 Exceptions and Interrupts B-33
- B.8 Input and Output B-38
- B.9 SPIM B-40

- B.10 MIPS R2000 Assembly Language B-45
- B.11 Concluding Remarks B-81
- B.12 Exercises B-82

Index I-1

C D - R O M C O N T E N T



The Basics of Logic Design C-2

- C.1 Introduction C-3
- C.2 Gates, Truth Tables, and Logic Equations C-4
- C.3 Combinational Logic C-9
- C.4 Using a Hardware Description Language C-20
- C.5 Constructing a Basic Arithmetic Logic Unit C-26
- C.6 Faster Addition: Carry Lookahead C-38
- C.7 Clocks C-48
- C.8 Memory Elements: Flip-Flops, Latches, and Registers C-50
- C.9 Memory Elements: SRAMs and DRAMs C-58
- C.10 Finite-State Machines C-67
- C.11 Timing Methodologies C-72
- C.12 Field Programmable Devices C-78
- C.13 Concluding Remarks C-79
- C.14 Exercises C-80



Mapping Control to Hardware D-2



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



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

- E.1 Introduction E-3
- E.2 Addressing Modes and Instruction Formats E-5
- E.3 Instructions: The MIPS Core Subset E-9

Contents

- E.4 Instructions: Multimedia Extensions of the Desktop/Server RISCs E-16
 - E.5 Instructions: Digital Signal-Processing Extensions of the Embedded RISCs E-19
 - E.6 Instructions: Common Extensions to MIPS Core E-20
 - E.7 Instructions Unique to MIPS-64 E-25
 - E.8 Instructions Unique to Alpha E-27
 - E.9 Instructions Unique to SPARC v.9 E-29
 - E.10 Instructions Unique to PowerPC E-32
 - E.11 Instructions Unique to PA-RISC 2.0 E-34
 - E.12 Instructions Unique to ARM E-36
 - E.13 Instructions Unique to Thumb E-38
 - E.14 Instructions Unique to SuperH E-39
 - E.15 Instructions Unique to M32R E-40
 - E.16 Instructions Unique to MIPS-16 E-40
 - E.17 Concluding Remarks E-43
-  Glossary G-1
 -  Further Reading FR-1