

COMPUTER  
ARCHITECTURE  
A  
QUANTITATIVE  
APPROACH



SECOND EDITION



JOHN L HENNESSY  
&  
DAVID A PATTERSON

2-004-340-1



2-004-340-1

# Computer Architecture A Quantitative Approach

Second Edition

David A. Patterson  
UNIVERSITY OF CALIFORNIA AT BERKELEY

John L. Hennessy  
STANFORD UNIVERSITY

With a Contribution by  
David Goldberg  
Xerox Palo Alto Research Center



MORGAN KAUFMANN PUBLISHERS, INC.  
SAN FRANCISCO, CALIFORNIA

# Contents

**Foreword** vii

**Preface** xiii

**Acknowledgments** xix

**1**

## **Fundamentals of Computer Design**

1.1	Introduction	1
1.2	The Task of a Computer Designer	3
1.3	Technology and Computer Usage Trends	6
1.4	Cost and Trends in Cost	8
1.5	Measuring and Reporting Performance	18
1.6	Quantitative Principles of Computer Design	29
1.7	Putting It All Together: The Concept of Memory Hierarchy	39
1.8	Fallacies and Pitfalls	44
1.9	Concluding Remarks	51
1.10	Historical Perspective and References	53
	Exercises	60

**2**

## **Instruction Set Principles and Examples**

2.1	Introduction	69
2.2	Classifying Instruction Set Architectures	70
2.3	Memory Addressing	73
2.4	Operations in the Instruction Set	80
2.5	Type and Size of Operands	85
2.6	Encoding an Instruction Set	87
2.7	Crosscutting Issues: The Role of Compilers	89
2.8	Putting It All Together: The DLX Architecture	96
2.9	Fallacies and Pitfalls	108
2.10	Concluding Remarks	111
2.11	Historical Perspective and References	112
	Exercises	118

**3**

## **Pipelining**

3.1	What Is Pipelining?	125
3.2	The Basic Pipeline for DLX	132
3.3	The Major Hurdle of Pipelining—Pipeline Hazards	139
3.4	Data Hazards	146
3.5	Control Hazards	161
3.6	What Makes Pipelining Hard to Implement?	178
3.7	Extending the DLX Pipeline to Handle Multicycle Operations	187
3.8	Crosscutting Issues: Instruction Set Design and Pipelining	199
3.9	Putting It All Together: The MIPS R4000 Pipeline	201

## 4

3.10	Fallacies and Pitfalls	209
3.11	Concluding Remarks	211
3.12	Historical Perspective and References	212
	Exercises	214

**Advanced Pipelining and Instruction-Level Parallelism**

4.1	Instruction-Level Parallelism: Concepts and Challenges	221
4.2	Overcoming Data Hazards with Dynamic Scheduling	240
4.3	Reducing Branch Penalties with Dynamic Hardware Prediction	262
4.4	Taking Advantage of More ILP with Multiple Issue	278
4.5	Compiler Support for Exploiting ILP	289
4.6	Hardware Support for Extracting More Parallelism	299
4.7	Studies of ILP	317
4.8	Putting It All Together: The PowerPC 620	335
4.9	Fallacies and Pitfalls	349
4.10	Concluding Remarks	352
4.11	Historical Perspective and References	354
	Exercises	362

## 5

**Memory-Hierarchy Design**

5.1	Introduction	373
5.2	The ABCs of Caches	375
5.3	Reducing Cache Misses	390
5.4	Reducing Cache Miss Penalty	411
5.5	Reducing Hit Time	422
5.6	Main Memory	427
5.7	Virtual Memory	439
5.8	Protection and Examples of Virtual Memory	447
5.9	Crosscutting Issues in the Design of Memory Hierarchies	457
5.10	Putting It All Together: The Alpha AXP 21064 Memory Hierarchy	461
5.11	Fallacies and Pitfalls	466
5.12	Concluding Remarks	471
5.13	Historical Perspective and References	472
	Exercises	476

## 6

**Storage Systems**

6.1	Introduction	485
6.2	Types of Storage Devices	486
6.3	Buses—Connecting I/O Devices to CPU/Memory	496
6.4	I/O Performance Measures	504
6.5	Reliability, Availability, and RAID	521
6.6	Crosscutting Issues: Interfacing to an Operating System	525
6.7	Designing an I/O System	528
6.8	Putting It All Together: UNIX File System Performance	539
6.9	Fallacies and Pitfalls	548
6.10	Concluding Remarks	553
6.11	Historical Perspective and References	553
	Exercises	557

## 7

**Interconnection Networks**

7.1	Introduction	563
7.2	A Simple Network	565
7.3	Connecting the Interconnection Network to the Computer	573

7.4	Interconnection Network Media	576
7.5	Connecting More Than Two Computers	579
7.6	Practical Issues for Commercial Interconnection Networks	597
7.7	Examples of Interconnection Networks	601
7.8	Crosscutting Issues for Interconnection Networks	605
7.9	Internetworking	608
7.10	Putting It All Together: An ATM Network of Workstations	613
7.11	Fallacies and Pitfalls	622
7.12	Concluding Remarks	625
7.13	Historical Perspective and References	626
	Exercises	629

### **Multiprocessors**

8.1	Introduction	635
8.2	Characteristics of Application Domains	647
8.3	Centralized Shared-Memory Architectures	654
8.4	Distributed Shared-Memory Architectures	677
8.5	Synchronization	694
8.6	Models of Memory Consistency	708
8.7	Crosscutting Issues	721
8.8	Putting It All Together: The SGI Challenge Multiprocessor	728
8.9	Fallacies and Pitfalls	734
8.10	Concluding Remarks	740
8.11	Historical Perspective and References	745
	Exercises	755

### **Appendix A: Computer Arithmetic**

by DAVID GOLDBERG

Xerox Palo Alto Research Center

A.1	Introduction	A-1
A.2	Basic Techniques of Integer Arithmetic	A-2
A.3	Floating Point	A-13
A.4	Floating-Point Multiplication	A-17
A.5	Floating-Point Addition	A-22
A.6	Division and Remainder	A-28
A.7	More on Floating-Point Arithmetic	A-34
A.8	Speeding Up Integer Addition	A-38
A.9	Speeding Up Integer Multiplication and Division	A-46
A.10	Putting It All Together	A-61
A.11	Fallacies and Pitfalls	A-65
A.12	Historical Perspective and References	A-66
	Exercises	A-72

### **Appendix B: Vector Processors**

B.1	Why Vector Processors?	B-1
B.2	Basic Vector Architecture	B-3
B.3	Two Real-World Issues: Vector Length and Stride	B-15
B.4	Effectiveness of Compiler Vectorization	B-22
B.5	Enhancing Vector Performance	B-23
B.6	Putting It All Together: Performance of Vector Processors	B-29
B.7	Fallacies and Pitfalls	B-35

B.8	Concluding Remarks	B-37
B.9	Historical Perspective and References	B-38
	Exercises	B-43

## Appendix C: Survey of RISC Architectures

C.1	Introduction	C-1
C.2	Addressing Modes and Instruction Formats	C-3
C.3	Instructions: The DLX Subset	C-5
C.4	Instructions: Common Extensions to DLX	C-9
C.5	Instructions Unique to MIPS	C-13
C.6	Instructions Unique to SPARC	C-15
C.7	Instructions Unique to PowerPC	C-18
C.8	Instructions Unique to PA-RISC	C-19
C.9	Concluding Remarks	C-22
C.10	References	C-25

## Appendix D: An Alternative to RISC: The Intel 80x86

D.1	Introduction	D-1
D.2	80x86 Registers and Data Addressing Modes	D-2
D.3	80x86 Integer Operations	D-6
D.4	80x86 Floating-Point Operations	D-9
D.5	80x86 Instruction Encoding	D-11
D.6	Putting It All Together: Measurements of instruction Set Usage	D-15
D.7	Concluding Remarks	D-22
D.8	Historical Perspective and References	D-23

## Appendix E: Implementing Coherence Protocols

E.1	Implementation Issues for the Snooping Coherence Protocol	E-1
E.2	Implementation Issues in the Distributed Directory Protocol	E-6
	Exercises	E-12

<b>References</b>	R-1
-------------------	-----

<b>Index</b>	I-1
--------------	-----