

图书基本信息

书名：<<可扩展并行计算--技术、结构与编程>>

13位ISBN编号：9787111071761

10位ISBN编号：711107176X

出版时间：1999-05

出版时间：机械工业出版社

作者：黄铠

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

作者简介

About the Authors

Kai Hwang presently holds a Chair Professor of Computer Engineering at the University of Hong Kong (HKU), while taking a research leave from the University of Southern California (USC). He has been engaged in higher education and computer research for 26 years, after earning the Ph.D. in Electrical Engineering and Computer Science from the University of California at Berkeley. His work on this book started at the USC and was mostly completed at the HKU.

An IEEE Fellow, he has published extensively in the areas of computer architecture, digital arithmetic, parallel processing, and distributed computing. He is the founding Editor-in-Chief of the Journal of Parallel and Distributed Computing. He has chaired the international conferences: ICPP86, ARITH-7, IPP96, ICAPP 96, and HPCA-4 in 1998. He has received several achievement awards for outstanding research and academic contributions to the field of parallel computing.

He has lectured worldwide and performed consulting and advisory work for US National Academy of Sciences, MIT Lincoln Laboratory, IBM Fishkill, TriTech in Singapore, Fujitsu and ETL in Japan, GMD in Germany, CERN School of Computing, and Academia Sinica in China. Presently, he leads a research group at HKU in developing an ATM-based multicomputer cluster for high-performance computing and distributed multimedia, Intranet, and Internet applications.

Zhiwei Xu is a Professor and Chief Architect at the National Center for Intelligent Computing Systems (NCIC), Chinese Academy of Sciences, Beijing, China. He is also a Honorary Research Fellow of HKU. He received a Ph.D. in Computer Engineering from the University of Southern California. He participated in the STAP/MPP benchmark projects led by Dr. Hwang at USC and HKU during the past 5 years.

He has taught at the Rutgers University and New York Polytechnic University. He has published in the areas of parallel languages, pipelined vector processing, and benchmark evaluation of massively parallel processors. Presently, he leads a design group at NCIC in building a series of cluster-based superservers in China. His current research interest lies mainly in network-based cluster computing and the software environments for parallel programming.

书籍目录

- Table of Contents
- About the Authors Iv
- Foreword xv
- Preface xvi
- Guide to Instructors/Readers xlix
- Part I Scalability and Clustering
- Chapter I Scalable Computer Platforms and Models
- I.1 Evolution of Computer Architecture
- I.1.1 Computer Generations
- I.1.2 Scalable Computer Architectures
- I.1.3 Converging System Architectures
- I.2 Dimensions of Scalability
- I.2.1 Resource Scalability
- I.2.2 Application Scalability
- I.2.3 Technology Scalability
- I.3 Parallel Computer Models
- I.3.1 Semantic Attributes
- I.3.2 Performance Attributes
- I.3.3 Abstract Machine Model
- I.3.4 Physical Machine Model
- I.4 Basic Concepts of Clustering
- I.4.1 Cluster Characteristics
- I.4.2 Architectural Comparisons
- I.4.3 Benefits and Difficulties of Clusters
- I.5 Scalable Design Principles
- I.5.1 Principle of Independence
- I.5.2 Principle of Balanced Design
- I.5.3 Design for Scalability
- I.6 Bibliographic Notes and Problems
- Chapter 2 Basics of Parallel Programming
- 2.1 Parallel Programming Overview
- 2.1.1 Why Is Parallel Programming Difficult?
- 2.1.2 Parallel Programming Environments
- 2.1.3 Parallel Programming Approaches
- 2.2 Processes, Tasks, and Threads
- 2.2.1 Definitions of an Abstract Process
- 2.2.2 Execution Mode
- 2.2.3 Address Space
- 2.2.4 Process Context
- 2.2.5 Process Descriptor
- 2.2.6 Process Control
- 2.2.7 Variations of Process
- 2.3 Parallelism Issues
- 2.3.1 Homogeneity in Processes
- 2.3.2 Static versus Dynamic Parallelism

- 2.3.3 Process Grouping.
- 2.3.4 Allocation Issues
- 2.4 Interaction/Communication Issues
 - 2.4.1 Interaction Operations
 - 2.4.2 Interaction Modes
 - 2.4.3 Interaction Patterns
 - 2.4.4 Cooperative versus Competitive Interactions
- 2.5 Semantic Issues in Parallel Programs
 - 2.5.1 Program Termination
 - 2.5.2 Determinacy of Programs
- 2.6 Bibliographic Notes and Problems
- Chapter 3 Performance Metrics and Benchmarks
 - 3.1 System and Application Benchmarks
 - 3.1.1 Micro Benchmarks
 - 3.1.2 Parallel Computing Benchmarks
 - 3.1.3 Business and TPC Benchmarks
 - 3.1.4 SPEC Benchmark Family
 - 3.2 Performance versus Cost
 - 3.2.1 Execution Time and Throughput
 - 3.2.2 Utilization and Cost-Effectiveness
 - 3.3 Basic Performance Metrics
 - 3.3.1 Workload and Speed Metrics
 - 3.3.2 Caveats in Sequential Performance
 - 3.4 Performance of Parallel Computers
 - 3.4.1 Computational Characteristics
 - 3.4.2 Parallelism and Interaction Overheads
 - 3.4.3 Overhead Quantification
 - 3.5 Performance of Parallel Programs
 - 3.5.1 Performance Metrics
 - 3.5.2 Available Parallelism in Benchmarks
 - 3.6 Scalability and Speedup Analysis
 - 3.6.1 Amdahl's Law: Fixed Problem Size
 - 3.6.2 Gustafson's Law: Fixed Time
 - 3.6.3 Sun and Ni's Law: Memory Bounding
 - 3.6.4 Isoperformance Models
 - 3.7 Bibliographic Notes and Problems
- Part II Enabling Technologies
- Chapter 4 Microprocessors as Building Blocks
 - 4.1 System Development Trends
 - 4.1.1 Advances in Hardware
 - 4.1.2 Advances in Software
 - 4.1.3 Advances in Applications
 - 4.2 Principles of Processor Design
 - 4.2.1 Basics of Instruction Pipeline
 - 4.2.2 From CISC to RISC and Beyond
 - 4.2.3 Architectural Enhancement Approaches
 - 4.3 Microprocessor Architecture Families

- 4.3.1 Major Architecture Families
- 4.3.2 Superscalar versus Superpipelined Processors
- 4.3.3 Embedded Microprocessors
- 4.4 Case Studies of Microprocessors
 - 4.4.1 Digital's Alpha 21 164 Microprocessor
 - 4.4.2 Intel Pentium Pro Processor
- 4.5 Post-RISC, Multimedia, and VLIW
 - 4.5.1 Post-RISC Processor Features
 - 4.5.2 Multimedia Extensions
 - 4.5.3 The VLIW Architecture
- 4.6 The Future of Microprocessors
 - 4.6.1 Hardware Trends and Physical Limits
 - 4.6.2 Future Workloads and Challenges ,
 - 4.6.3 Future Microprocessor Architectures
- 4.7 Bibliographic Notes and Problems
- Chapter 5 Distributed Memory and Latency Tolerance
 - 5.1 Hierarchical Memory Technology
 - 5.1.1 Characteristics of Storage Devices
 - 5.1.2 Memory Hierarchy Properties
 - 5.1.3 Memory Capacity Planning
 - 5.2 Cache Coherence Protocol
 - 5.2.1 Cache Coherency Problem
 - 5.2.2 Snoopy Coherency Protocols
 - 5.2.3 The MESI Snoopy Protocol
 - 5.3 Shared-Memory Consistency
 - 5.3.1 Memory Event Ordering
 - 5.3.2 Memory Consistency Models
 - 5.3.3 Relaxed Memory Models
 - 5.4 Distributed Cache/Memory Architecture
 - 5.4.1 NORMA, NUMA, COMA, and DSM Models
 - 5.4.2 Directory-Based Coherency Protocol
 - 5.4.3 The Stanford Dash Multiprocessor
 - 5.4.4 Directory-Based Protocol in Dash
 - 5.5 Latency Tolerance Techniques 250
 - 5.5.1 Latency Avoidance, Reduction, and Hiding
 - 5.5.2 Distributed Coherent Caches
 - 5.5.3 Data Prefetching Strategies
 - 5.5.4 Effects of Relaxed Memory Consistency
 - 5.6 Multithreaded Latency Hiding
 - 5.6.1 Multithreaded Processor Model
 - 5.6.2 Context-Switching Policies
 - 5.6.3 Combining Latency Hiding Mechanisms
 - 5.7 Bibliographic Notes and Problems
- Chapter 6 System Interconnects and Gigabit Networks
 - 6.1 Basics of Interconnection Network
 - 6.1.1 Interconnection Environments
 - 6.1.2 Network Components

- 6.1.3 Network Characteristics
- 6.1.4 Network Performance Metrics
- 6.2 Network Topologies and Properties
 - 6.2.1 Topological and Functional Properties
 - 6.2.2 Routing Schemes and Functions
 - 6.2.3 Networking Topologies
- 6.3 Buses, Crossbar, and Multistage Switches
 - 6.3.1 Multiprocessor Buses
 - 6.3.2 Crossbar Switches
 - 6.3.3 Multistage Interconnection Networks
 - 6.3.4 Comparison of Switched Interconnects
- 6.4 Gigabit Network Technologies
 - 6.4.1 Fiber Channel and FDDI Rings
 - 6.4.2 Fast Ethernet and Gigabit Ethernet
 - 6.4.3 Myrinet for SAN/LAN Construction
 - 6.4.4 HiPPI and SuperHiPPI
- 6.5 ATM Switches and Networks
 - 6.5.1 ATM Technology
 - 6.5.2 ATM Network Interfaces
 - 6.5.3 Four Layers of ATM Architecture
 - 6.5.4 ATM Internetwork Connectivity
- 6.6 Scalable Coherence Interface
 - 6.6.1 SCI Interconnects
 - 6.6.2 Implementation Issues
 - 6.6.3 SCI Coherence Protocol
- 6.7 Comparison of Network Technologies
 - 6.7.1 Standard Networks and Perspectives
 - 6.7.2 Network Performance and Applications
- 6.8 Bibliographic Notes and Problems
- Chapter 7 Threading, Synchronization, and Communication
 - 7.1 Software Multithreading
 - 7.1.1 The Thread Concept
 - 7.1.2 Threads Management
 - 7.1.3 Thread Synchronization
 - 7.2 Synchronization Mechanisms
 - 7.2.1 Atomicity versus Mutual Exclusion
 - 7.2.2 High-Level Synchronization Constructs
 - 7.2.3 Low-Level Synchronization Primitives
 - 7.2.4 Fast Locking Mechanisms
 - 7.3 The TCP/IP Communication Protocol Suite
 - 7.3.1 Features of The TCP/IP Suite
 - 7.3.2 UDP, TCP, and IP
 - 7.3.3 The Sockets Interface
 - 7.4 Fast and Efficient Communication
 - 7.4.1 Key Problems in Communication
 - 7.4.2 The Log P Communication Model
 - 7.4.3 Low-Level Communications Support

- 7.4.4 Communication Algorithms
- 7.5 Bibliographic Notes and Problems
- Part III Systems Architecture
- Chapter 8 Symmetric and CC-NUMA Multiprocessors
- 8.1 SMP and CC-NUMA Technology
- 8.1.1 Multiprocessor Architecture
- 8.1.2 Commercial SMP Servers
- 8.1.3 The Intel SHV Server Board
- 8.2 Sun Ultra Enterprise I/O System
- 8.2.1 The Ultra E-10000 Architecture
- 8.2.2 System Board Architecture
- 8.2.3 Scalability and Availability Support
- 8.2.4 Dynamic Domains and Performance
- 8.3 HP/Convex Exemplar X-Class
- 8.3.1 The Exemplar X System Architecture
- 8.3.2 Exemplar Software Environment
- 8.4 The Sequent NUMA-Q 2000
- 8.4.1 The NUMA-Q 2000 Architecture
- 8.4.2 Software Environment of NUMA-Q
- 8.4.3 Performance of the NUMA-Q
- 8.5 The SGI/Cray Origin 2000 Superserver
- 8.5.1 Design Goals of Origin 2000 Series
- 8.5.2 The Origin 2000 Architecture
- 8.5.3 The Cellular IRIX Environment
- 8.5.4 Performance of the Origin 2000
- 8.6 Comparison of CC-NUMA Architectures
- 8.7 Bibliographic Notes and Problems
- Chapter 9 Support of Clustering and Availability
- 9.1 Challenges in Clustering
- 9.1.1 Classification of Clusters
- 9.1.2 Cluster Architectures
- 9.1.3 Cluster Design Issues
- 9.2 Availability Support for Clustering
- 9.2.1 The Availability Concept
- 9.2.2 Availability Techniques
- 9.2.3 Checkpointing and Failure Recovery
- 9.3 Support for Single System Image
- 9.3.1 Single System Image Layers
- 9.3.2 Single Entry and Single File Hierarchy
- 9.3.3 Single I/O, Networking, and Memory Space
- 9.4 Single System Image in Solaris MC
- 9.4.1 Global File System
- 9.4.2 Global Process Management
- 9.4.3 Single I/O System Image
- 9.5 Job Management in Clusters
- 9.5.1 Job Management System
- 9.5.2 Survey of Job Management Systems

- 9.5.3 Load-Sharing Facility (LSF)
- 9.6 Bibliographic Notes and ProNems
- Chapter 10 Clusters of Servers and Workstations
- 10.1 Cluster Products and Research Projects
- 10.1.1 Supporting Trend of Cluster Products
- 10.1.2 Cluster of SMP Servers
- 10.1.3 Cluster Research Projects
- 10.2 Microsoft Wolfpack for NT Clusters
- 10.2.1 Microsoft Wolfpack Configurations
- 10.2.2 Hot Standby Multiserver Clusters
- 10.2.3 Active Availability Clusters
- 10.2.4 Fault-Tolerant Multiserver Cluster
- 10.3 The IBM SP System
- 10.3.1 Design Goals and Strategies
- 10.3.2 The SP2 System Architecture
- 10.3.3 I/O and Internetworking.
- 10.3.4 The SP System Software
- 10.3.5 The SP2 and Beyond
- 10.4 The Digital TruCluster
- 10.4.1 The TruCluster Architecture
- 10.4.2 The Memory Channel Interconnect
- 10.4.3 Programming the TruCluster
- 10.4.4 The TruCluster System Software
- 10.5 The Berkeley NOW Project
- 10.5.1 Active Messages for Fast Communication
- 10.5.2 GLUnix for Global Resource Management
- 10.5.3 The FSServerless Network File System
- 10.6 TreadMarks: A Software-Implemented DSM Cluster
- 10.6.1 Boundary Conditions
- 10.6.2 User Interface for DSM
- 10.6.3 Implementation Issues
- 10.7 Bibliographic Notes and Problems
- Chapter 11 MPP Architecture and Performance
- 11.1 An Overview of MPP Technology
- 11.1.1 MPP Characteristics and Issues
- 11.1.2 MPP Systems - An Overview
- 11.2 The Cray T3E System
- 11.2.1 The System Architecture of T3E
- 11.2.2 The System Software in T3E
- 11.3 New Generation of ASCI/MPPs
- 11.3.1 ASCI Scalable Design Strategy
- 11.3.2 Hardware and Software Requirements
- 11.3.3 Contracted ASCI/MPP Platforms
- 11.4 Intel/Sandia ASCI Option Red
- 11.4.1 The Option Red Architecture
- 11.4.2 Option Red System Software
- 11.5 Parallel NAS Benchmark Results

- II.5.1 The NAS Parallel Benchmarks
- II.5.2 Superstep Structure and Granulanty
- II.5.3 Memory, VO, and Communications
- 11.6 MPI and STAP Benchmark Results
- II.6.1 MPI Performance Measurements
- II.6.2 MPI Latency and Aggregate Bandwidth
- II.6.3 STAP Benchmark Evaluation of MPPs
- II.6.4 MPP Architectural Implications
- 11.7 Bibliographic Notes and Problems
- Part IV Part IV Parallel Programming
- Chapter 12 Parallel Paradigms and Programming Models
- 12.1 Paradigms and Programmability
- 12.1.1 Algorithmic Paradigms
- 12.1.2 Programmability Issues
- 12.1.3 Parallel Programming Examples
- 12.2 Parallel Programming Models
- 12.2.1 Implicit Parallelism
- 12.2.2 Explicit Parallel Modeis
- 12.2.3 ComparisonofFourModels
- 12.2.4 Other Parallel Programming Models
- 12.3 Shared-Memory ProgrammiBg
- 12.3.1 The ANSI X3H5 Shared-Memory Model
- 12.3.2 ThePOSIX Threads(Pthreads)Model
- 12.3.3 The OpenMP Standard
- 12.3.4 TheSGIPowerCModel
- 12.3.5 CII: A Structured Parallel C Language
- 12.4 Bibliographic Notes and Problems
- Chapter 13 Message-Passing Programmmg
- 13.1 The Message-Passing Paradigm
- 13.1.1 Message-Passing Libraries
- 13.1.2 Message-Passing Modes
- 13.2 Message-Passing Interface (MPI)
- 13.2.1 MPIMessages
- 13.2.2 Message Envelope in MPI
- 13.2.3 Point-to-Point Communications
- 13.2.4 Collective MPI Communications
- 13.2.5 The MP1-2 Extensions
- 13.3 Parallel Virtual Machine (PVM)
- 13.3.1 Virtual Machine Construction
- 13.3.2 Process Management in PVM
- 13.3.3 Communication with PVM
- 13.4 Bibliographic Notes and Problems
- Chapter 14 Data-Parallel Programming
- 14.1 The Data-Parallel Model
- 14.2 The Fortran 90 Approach
- 14.2.1 Parallel Array Operations
- 14.2.2 Intrinsic Functions in Fortran 90

14.3 High-Performance Fortran

14.3.1 Support for Data Parallelism

14.3.2 DataMappinginHPF

14.3.3 SummaryofFortran90andHPF

14.4 Other Data-Parallel Approaches

14.4.1 Fortran 95andFortran2001

14.4.2 ThepC++andNeslApproaches

14.5 Bibliographic Notes and Problems

Bibliography

Web Resources List

Subject Index

Author Index

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>