

Parallel Programming

MSBD5009

Course Introduction

Course Background

- An MSBD Elective
 - Assume CSE basic programming, OS, Algorithms
 - Reference books and structured lectures
 - Parallel programming knowledge
 - Exclusion COMP5112
 - Taught last year for the first time
 - A term project: reports, presentation, code
 - No exams
 - Everyone got good grades

Course Topics

- Introduction to parallel computer architectures
- Principles of parallel algorithm design
- Shared-memory programming models
- Message passing programming models
- Data-parallel programming models for GPUs
- Case studies of parallel algorithms, systems, and applications
- Hands-on experience with writing parallel programs for tasks of interest

Parallel Computer Architectures

- Review on OS and Computer Architecture
 - The von Neumann architecture
 - Processes, multitasking, and threads
 - Modifications to the von Neumann Model
 - Caches
 - Virtual memory
 - Instruction-level parallelism
 - Hardware multithreading
- Parallel Hardware
 - SIMD systems
 - MIMD systems
 - Interconnection networks
 - Cache coherence
 - Shared-memory versus distributed-memory

Principles of parallel algorithm design

- Preliminaries
 - Decomposition, Tasks, and Dependency Graphs
 - Granularity, Concurrency, and Task-Interaction
 - Processes and Mapping
- Decomposition Techniques
- Mapping Techniques for Load Balancing
- Methods for Containing Interaction Overheads
- Parallel Algorithm Models

Message passing programming models

- Principles of Message-Passing Programming
- Building Blocks: Send and Receive Operations
- MPI: the Message Passing Interface
- Collective Communication and Computation Operations
 - Gather, Scatter, Prefix, Reduction, Broadcast, Barrier, and so on

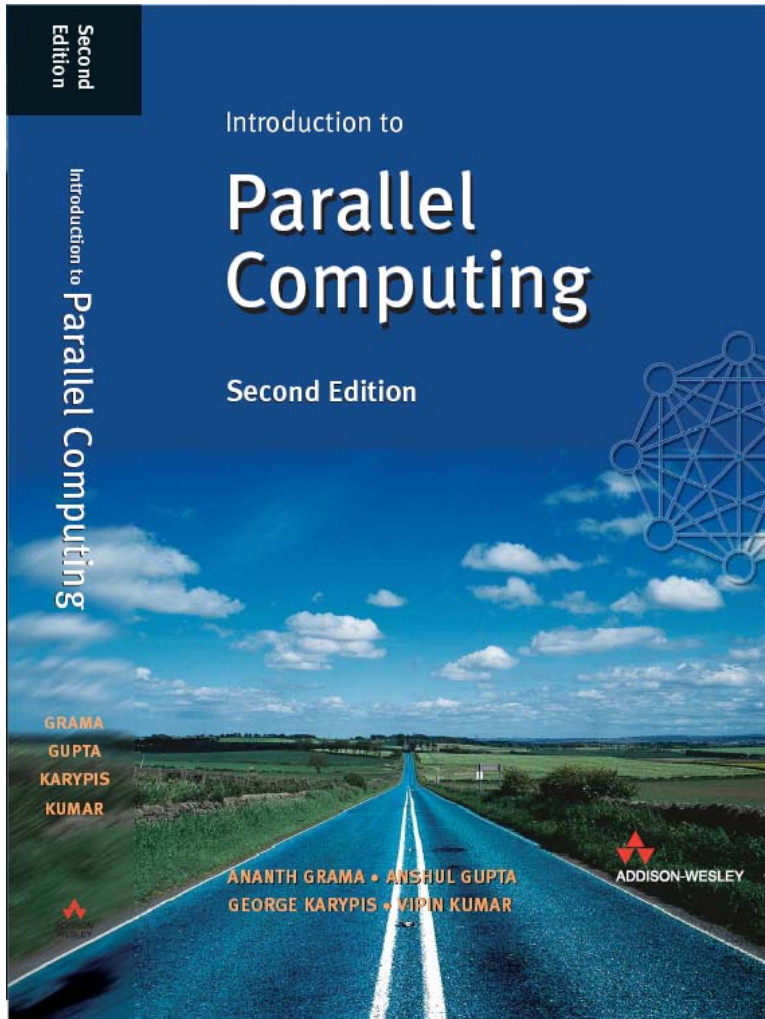
Shared-memory programming models

- Pthreads
 - Critical sections, busy-waiting, mutexes
 - Producer-Consumer Synchronization and Semaphores
 - Barriers and Condition Variables
 - Read-Write Locks
 - Caches, Cache Coherence, and False Sharing
 - Thread safety
- OpenMP

Data-parallel programming for GPUs

- CUDA C Language APIs
- CUDA Execution Model
- CUDA Memories
- Performance Considerations
- Parallel Patterns
 - Gather, Scatter, Reduction, Prefix Scan, and so on
- Case Studies

Reference Book 1



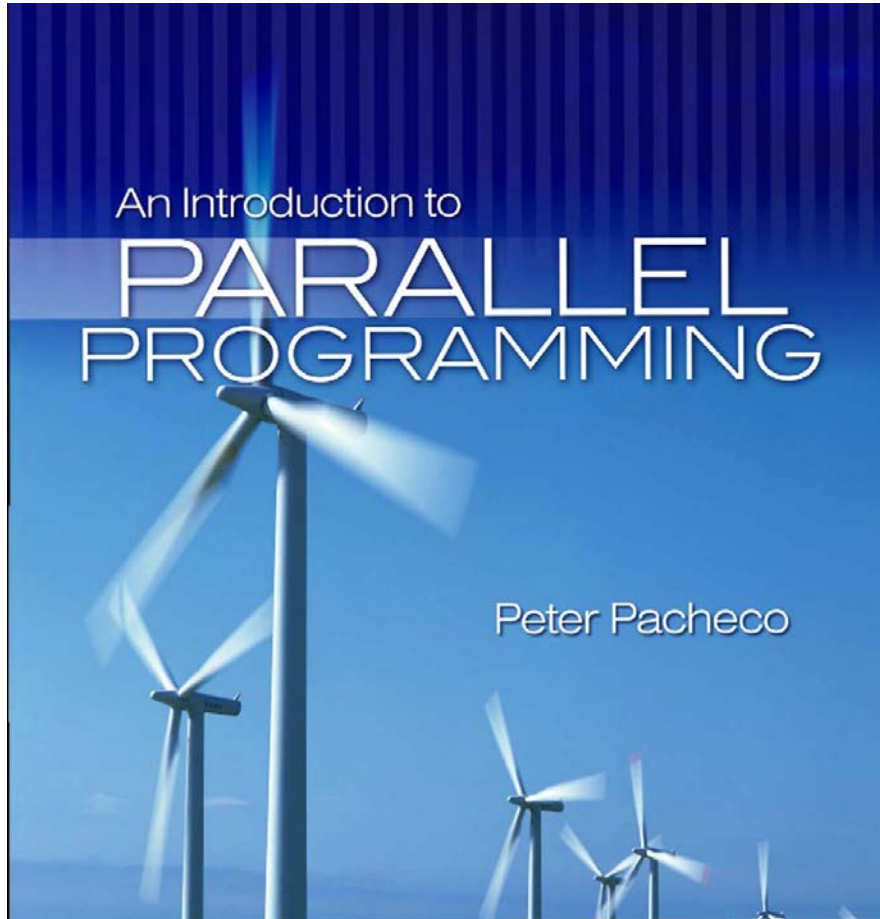
**Introduction to Parallel Computing
2nd edition**

By Ananth Grama, Anshul Gupta, George Karypis, Vipin Kumar.

Addison Wesley, 2003.

<http://www-users.cs.umn.edu/~karypis/parbook/>

Reference Book 2



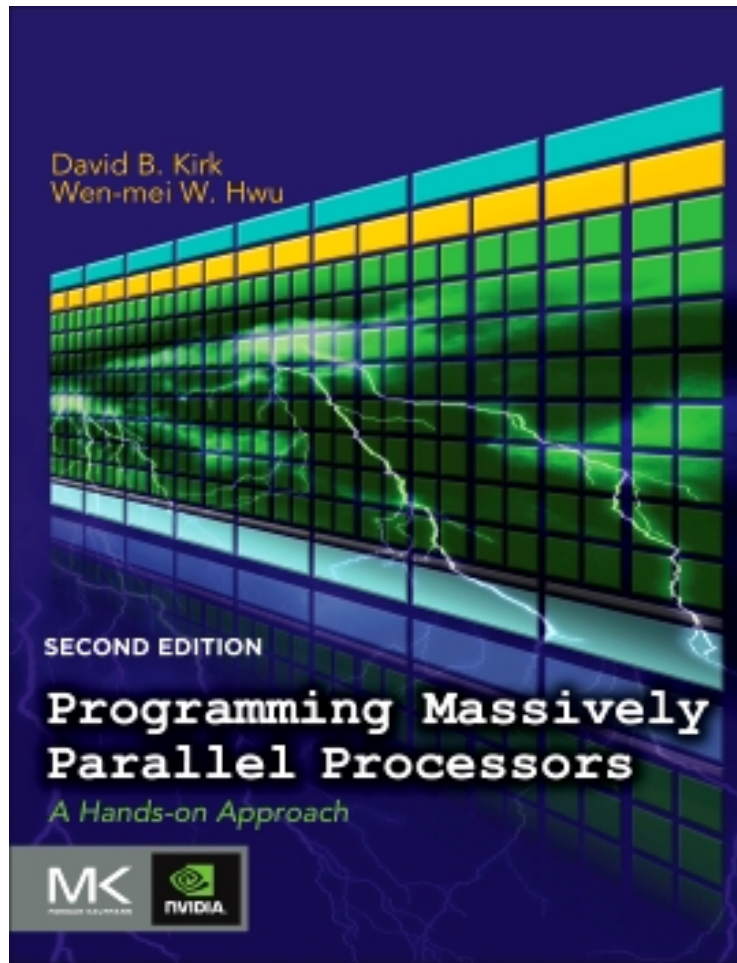
An Introduction to Parallel Programming

By Peter Pacheco

Morgan Kaufmann, 2011

<https://www.cs.usfca.edu/~peter/ipp/>

Reference Book 3



**Programming Massively
Parallel Processors:
A Hands-on Approach
2nd Edition
Author(s) : Kirk & Hwu
2013
Morgan Kaufmann**

Full-text Online Access:

<http://library.ust.hk/cgi/db/24x7.pl?id/51033>

Lecture Time and Venue

- 12 Lectures on Saturdays 3-5:50pm
 - Sept 2, 9, 16 @ Rm 2464, Lift 25-26
 - Sept 23 @ Rm 4621, Lift 31-32
 - Sept 30 – Nov 25 @ Rm 2464, Lift 25-26(No class on Oct 28 due to public holidays)

Workload & Assessment

- Tentative plan
 - Four programming assignments 50%
 - Week 4, 7, 9, 12 on MPI, Pthreads, OpenMP, CUDA
 - All assignments on a single topic (e.g., shortest path)
 - Sequential version program given (a few hundred lines of code)
 - Parallel program skeleton given
 - Your task is to fill in a few parallel components (tens of lines of code)
 - One final exam (Dec 9 Saturday 3-6pm) 50%
 - Programming: fill in code, similar to assignments
 - Short answer questions on concepts from course material

Lab Facilities

CSE Teaching Lab 2 (4214)

- Hardware: Lenovo ThinkCenter M93
 - Intel Core i3-4150 CPU, 4GB RAM, 500GB Harddisk, 21.5" LCD Monitor, Inno GeForce GTX960
- Software
 - CentOS 7.2 (x86_64 architecture)
 - CUDA Development Kit 7.0, under /usr/local/cuda
 - GCC 4.8.5 (GCC 5.3.0 available under /usr/local, as gcc5 and g++5; GCC6.2.0 available under /usr/local, as gcc6 and g++6)
 - OpenSSH

http://cssystem.cse.ust.hk/Facilities/hkust_only/labsetup/lab2.html