CME 213, ME 339—Spring 2021
Introduction to parallel computing using MPI, openMP, and CUDA

Eric Darve, ICME, Stanford

"The city's central computer told you? R2D2, you know better than to trust a strange computer!" (C3PO)

Instructor

Eric Darve, ME, ICME, darve@stanford.edu

Numerical linear algebra, machine learning for mechanics and engineering, parallel computing

Teaching assistant

Chenzhuo Zhu (head TA)

Teaching assistant

Vikul Gupta

Guest lectures

Elliott Slaughter 

NVIDIA engineers 

Classes will be (mostly) pre-recorded using Panopto.

There will be a few live lectures (tutorials, guest lectures).

Recordings can be found on the class web page.

C++ refresher with Chenzhuo

This Wednesday 2:30 PM over live Zoom

Canvas

https://canvas.stanford.edu/courses/133903

Announcements, grades, office hours, zoom links

Office Hours

See Zoom tab on Canvas (sessions are not recorded)

- Monday 2–4 PM: instructor
- Tue 4–6 PM: Chenzhuo
- Wed 9–11 AM: Vikul
- Thu 4–6 PM: Chenzhuo
- Fri 9–11 AM: Vikul

Class web page

https://ericdarve.github.io/cme213-spring-2021/

Class material, homework, reading assignments, final project

Syllabus

How to get support?

Online discussion forum on Slack

https://cme213-spring-2021.slack.com/

Channel: #homework

Please be civil on the forum

No disrespectful or demeaning posts

Gradescope

https://www.gradescope.com/courses/258024

See entry code on Canvas.

- Grades and rubrics
- Regrades

Assignment weights

- Reading assignments: 25%
- Homework: 40%
- Final project: 35%

- Reading assignments: questions on the lecture videos
- Homework assignments: more in-depth programming assignments

Usually, reading assignments are due on Wednesday and homework assignments are due on Friday.

Homework submission

- PDF with answers: gradescope
- Computer code: copy your code to cardinal. A Python script is provided for the submission. Instructions are given in the homework assignment.

Policy for late assignments

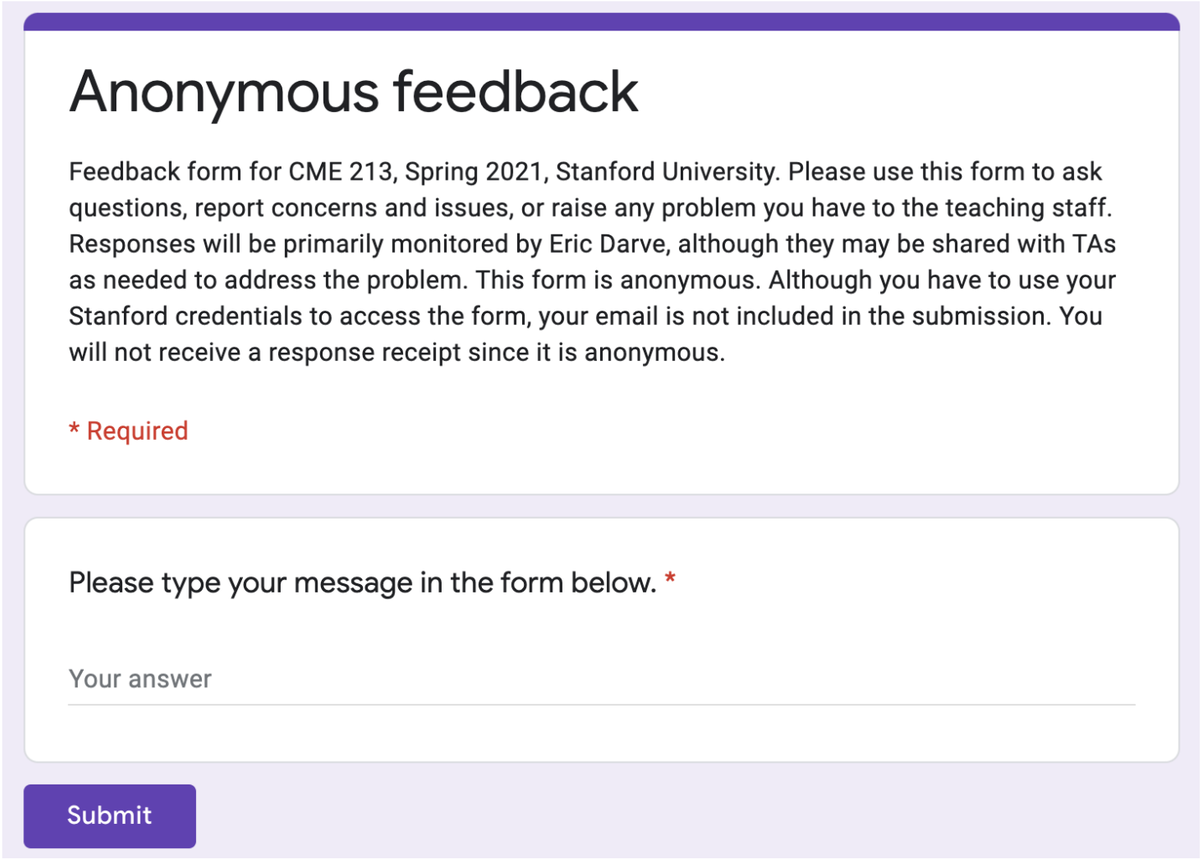
Extensions can be requested in advance for exceptional circumstances (e.g., travel, sickness, injury, COVID-related issues) and for OAE-approved accommodations.

Submissions after the deadline and late by at most two days will be accepted with a 10% penalty.

No submissions will be accepted two days after the deadline.

Anonymous feedback

Google form; see Canvas for the link



Ice breaker


AGENDA
1. Business
2. Ice Breaker
3. Ice Breaker
4. Ice Breaker
5. Ice Breaker
6. Ice Breaker
7. Ice Breaker
8. Ice Breaker
9. Ice Breaker
Why is it that nothing ever gets done at these meetings?
Reynolds

- Break into groups. Minimum: 3. Maximum: 5.
- If interested this group may become your study group for this quarter.
- You will have 10 minutes to complete the exercise.

In your group, introduce yourself: give your name, major, and degree. Then answer the following questions, one per round.

1. Would you rather travel back in time to meet your ancestors or would you rather go to the future to meet your descendants?
2. Is it ever OK to waste time?
3. Describe an experience in your life that changed your values.

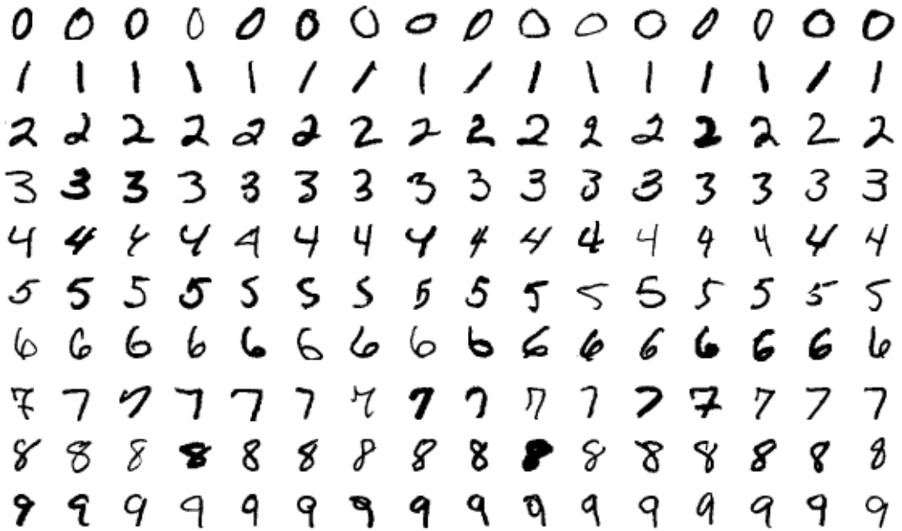
Make a note of these questions before leaving the lobby.

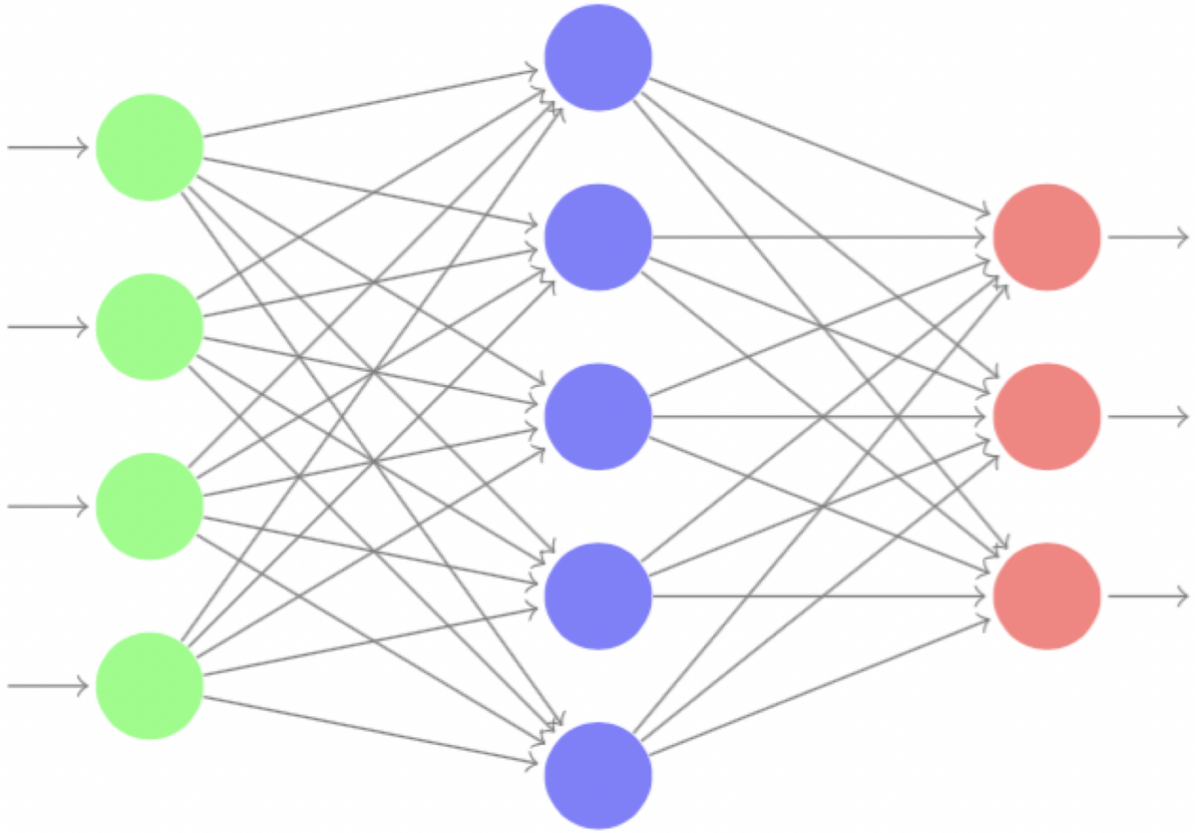
Group size: min: 3; max: 5. Time: 10 minutes

Final Project

Deep neural network to recognize hand-written digits (MNIST)

Will involve CUDA and MPI programming

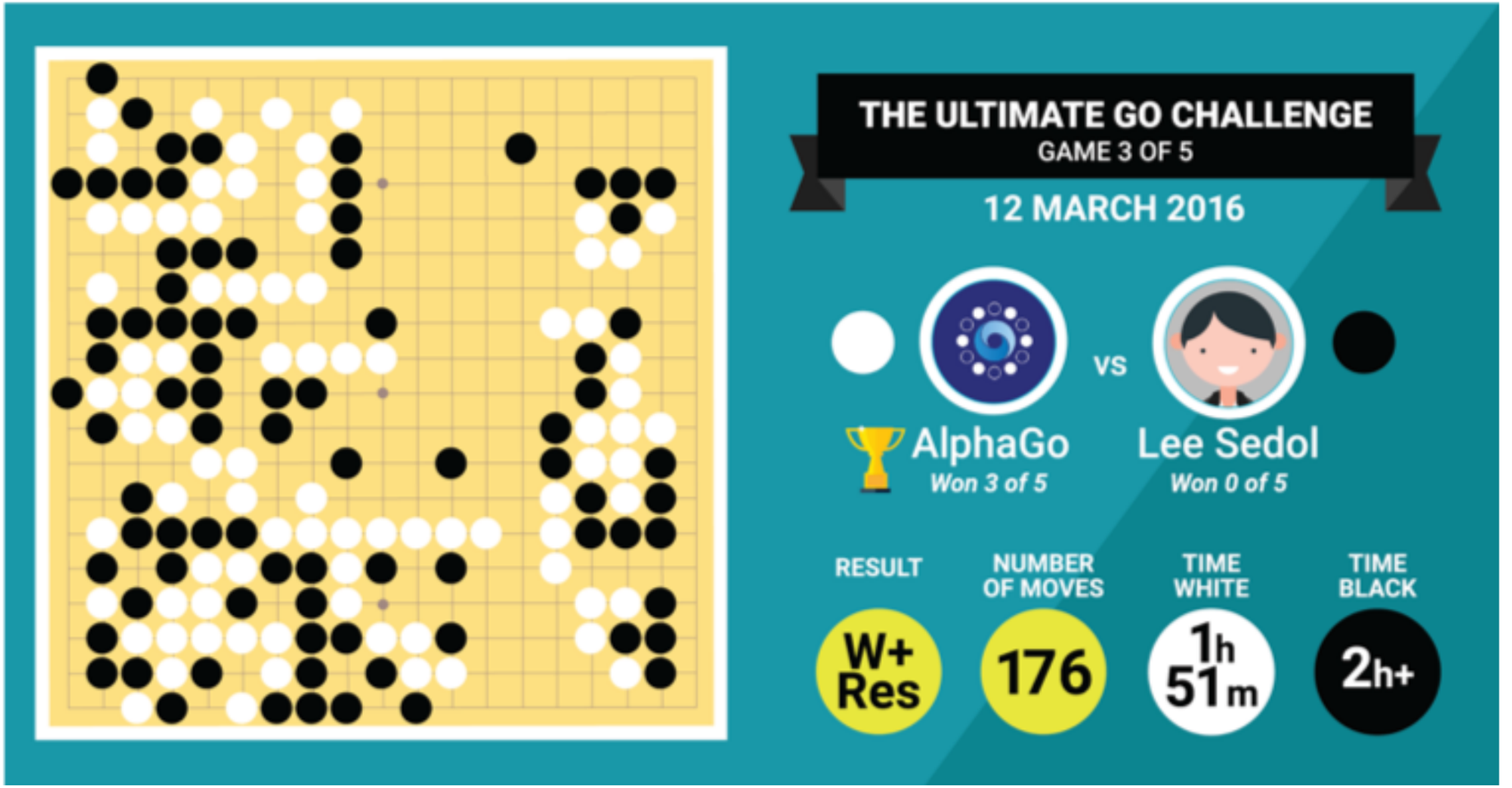




P(0)
P(1)
P(2)
P(3)
P(4)
P(5)
P(6)
P(7)
P(8)
P(9)

DNN playground

https://playground.tensorflow.org/



AlphaGo played a handful of highly inventive winning moves, several of which were so surprising they overturned hundreds of years of received wisdom.

Deep Learning relies on parallel computing

Program	Hardware
AlphaGo Fan	176 GPUs
AlphaGo Lee	48 TPUs
AlphaGo Master	Single machine with 4 TPUs
AlphaGo Zero	Single machine with 4 TPUs

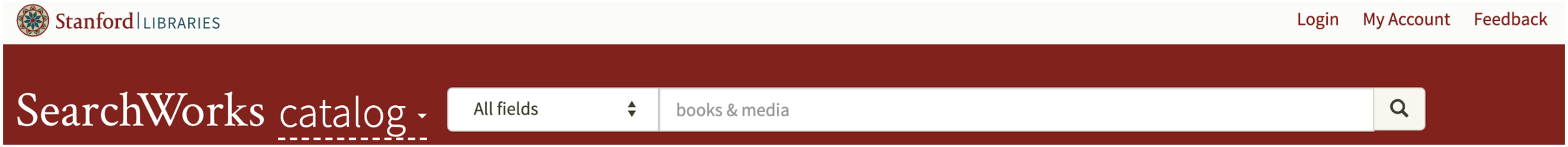
Computer access

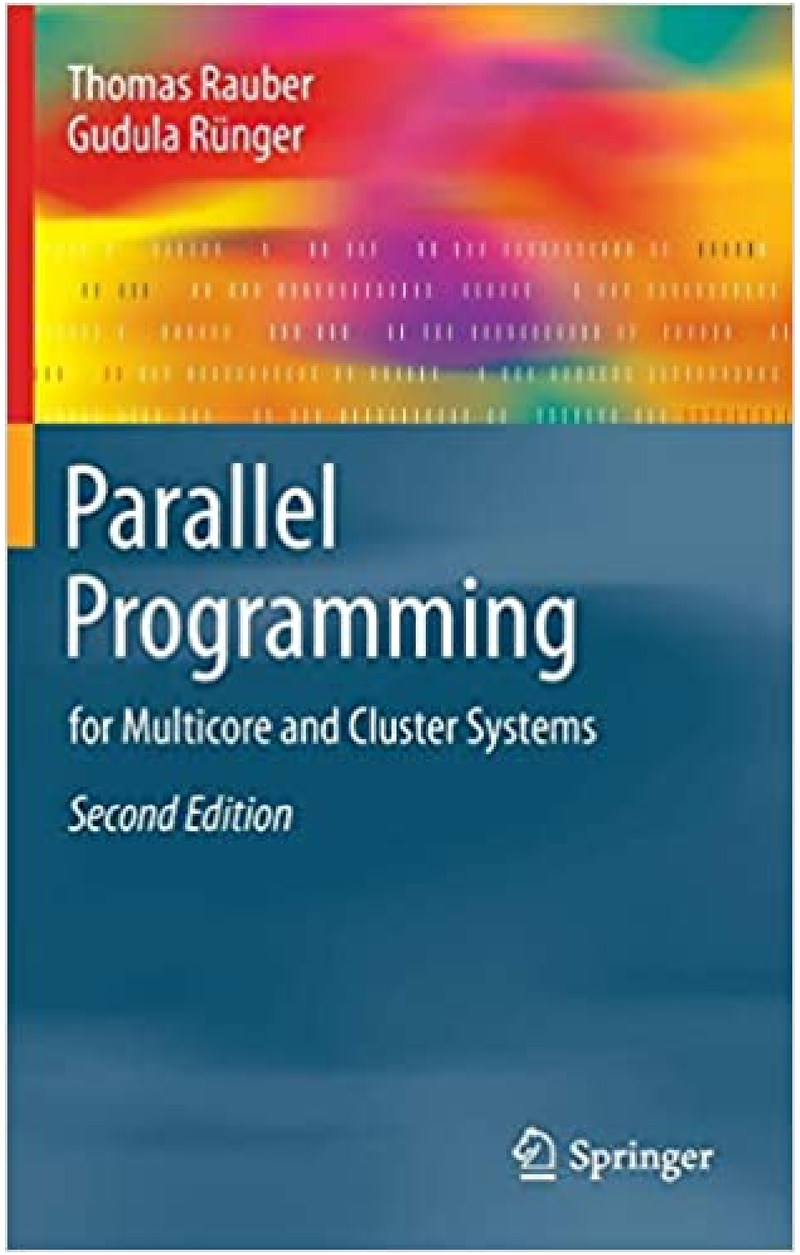
- ICME computer cluster: icme-gpu.stanford.edu
- This computer has GPUs which will be required for the GPU homework assignments and the final project.
- For the first few assignments, you can use your own computer or one of the computers on FarmShare.
- rice.stanford.edu is a good option.

Books!

Available electronically

http://searchworks.stanford.edu/

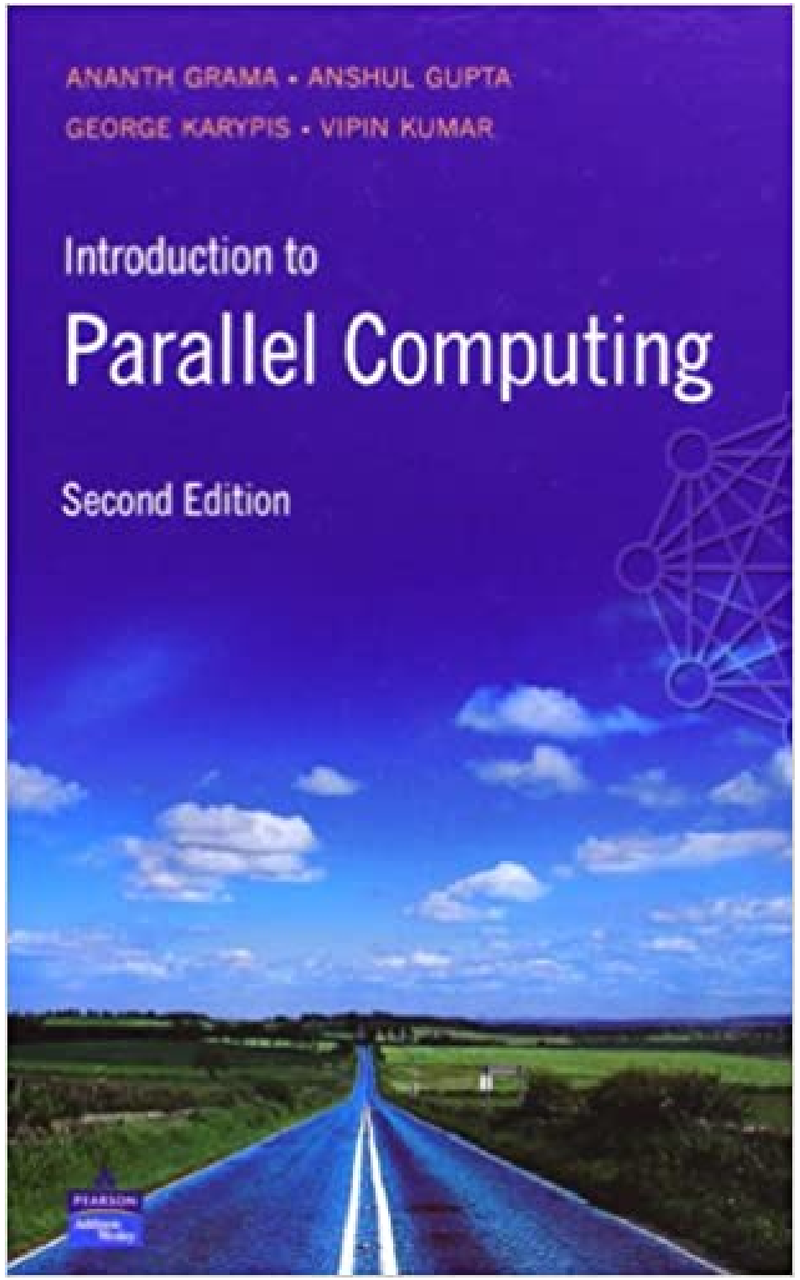




General parallel computing books

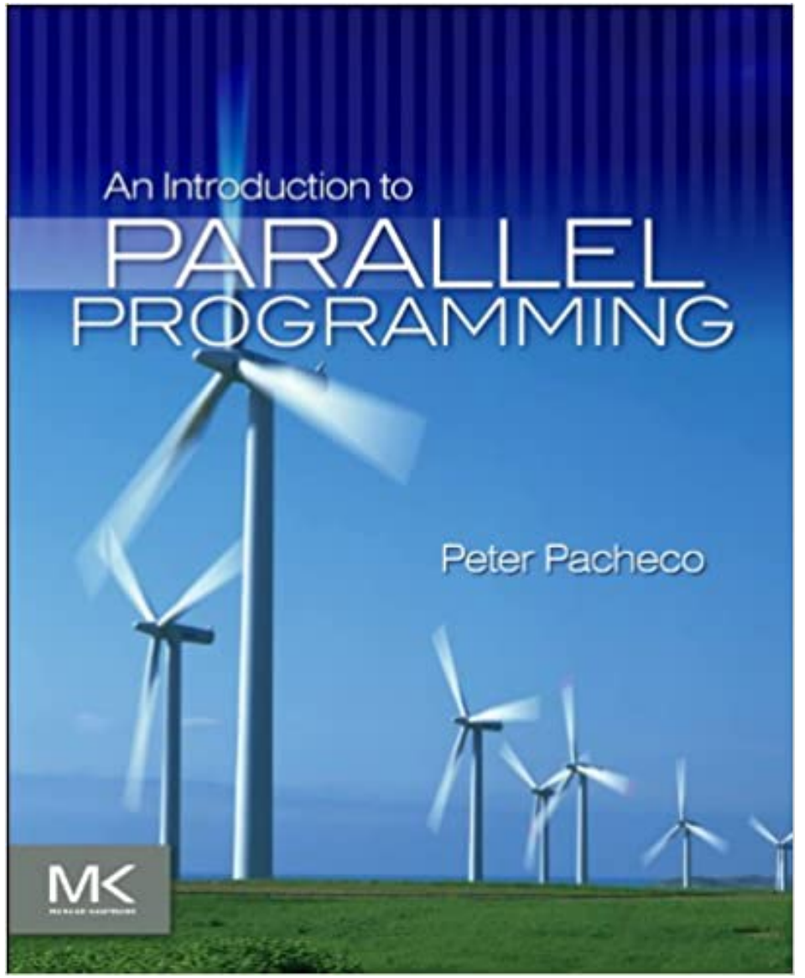
Parallel Programming for Multicore and Cluster Systems, by Rauber and Rünger

Applications focus mostly on linear algebra



Introduction to Parallel Computing, by Grama, Gupta, Karypis, Kumar

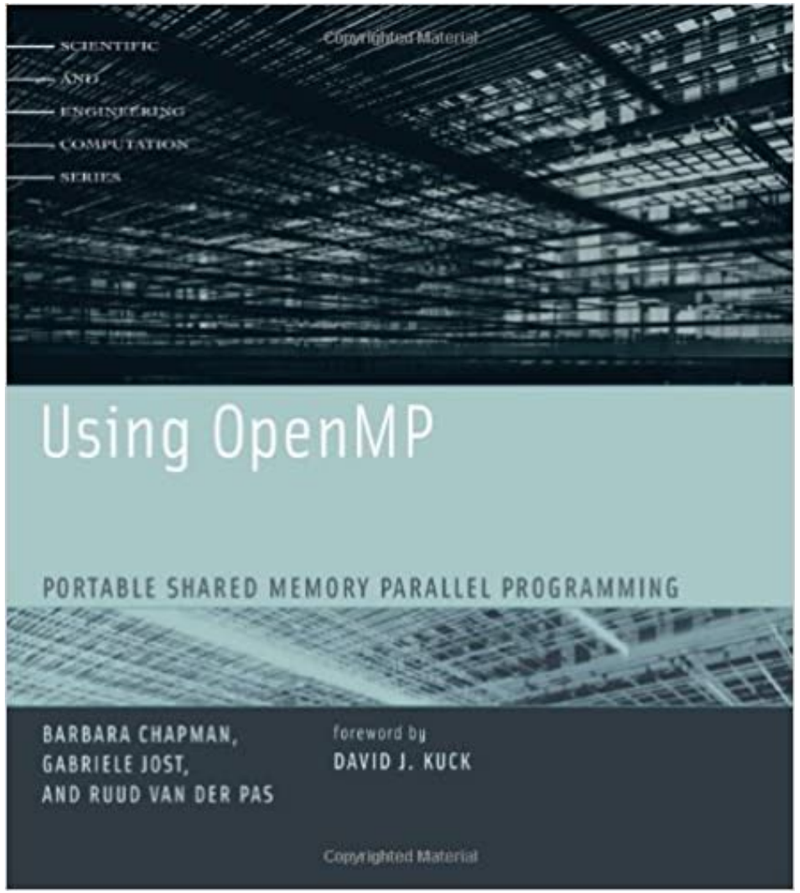
Wide range of applications from sort to FFT, linear algebra and tree search



An introduction to parallel programming, by Pacheco

More examples and less theoretical

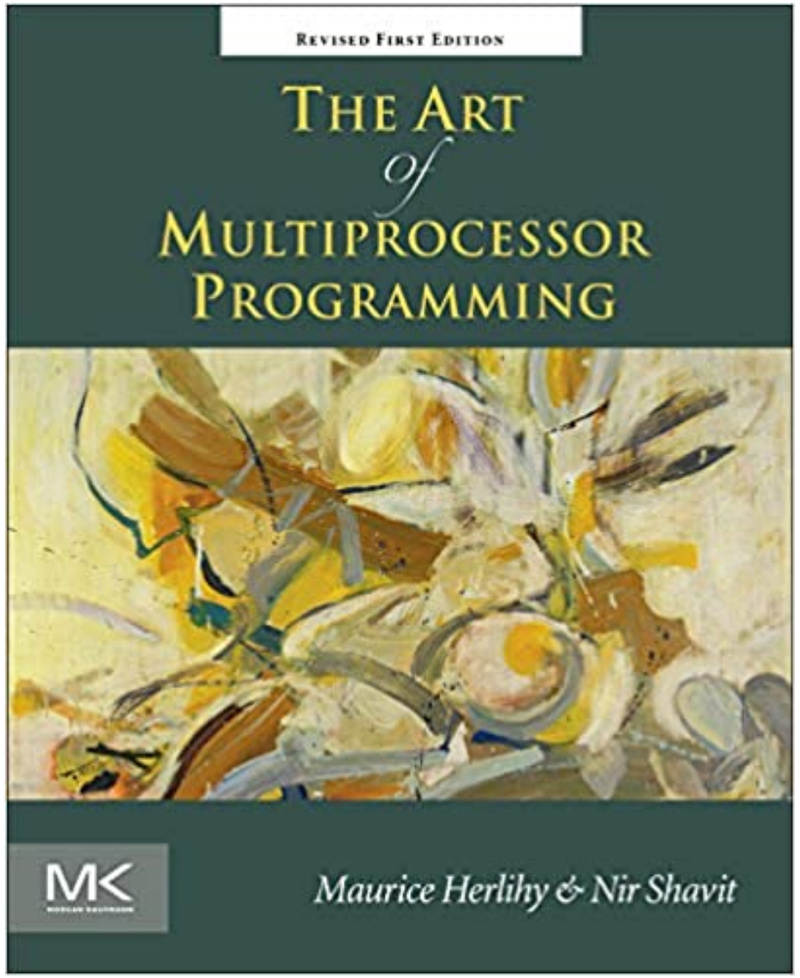
Applications include N-body codes and tree search



OpenMP and multicore books

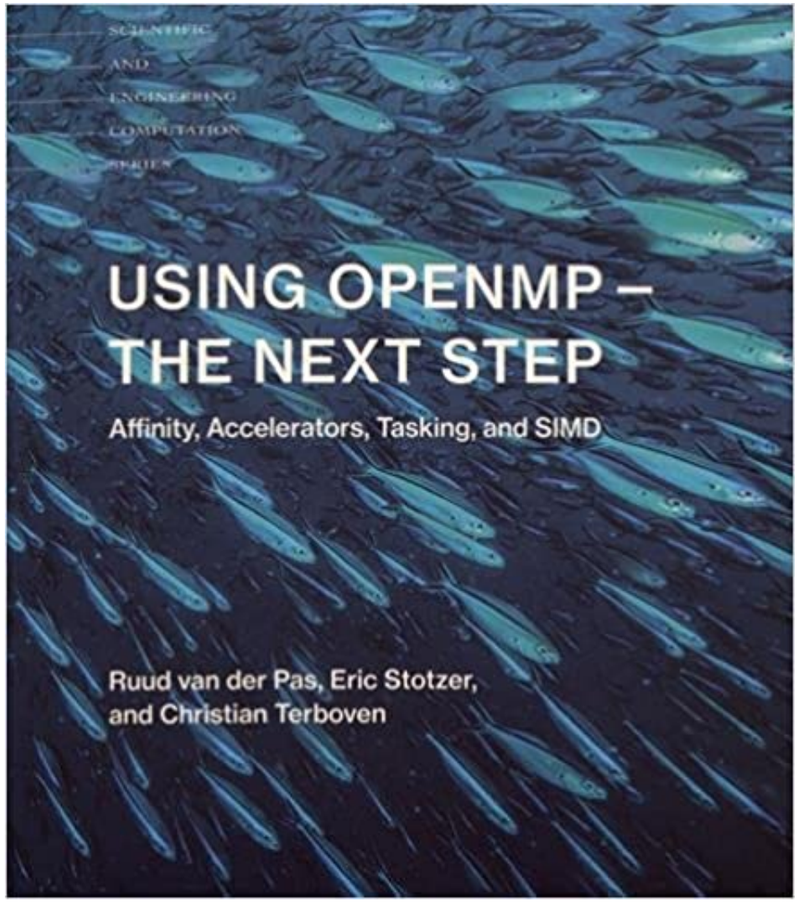
Using OpenMP: portable shared memory parallel programming, by Chapman, Jost, van der Pas

In-depth coverage of OpenMP



The art of multiprocessor programming, by Herlihy, Shavit

Specializes on advanced multicore programming



Using OpenMP—The Next Step: Affinity, Accelerators, Tasking, and SIMD, by van der Pas, Stotzer, Terbo

Covers recent extensions to OpenMP and some advanced usage

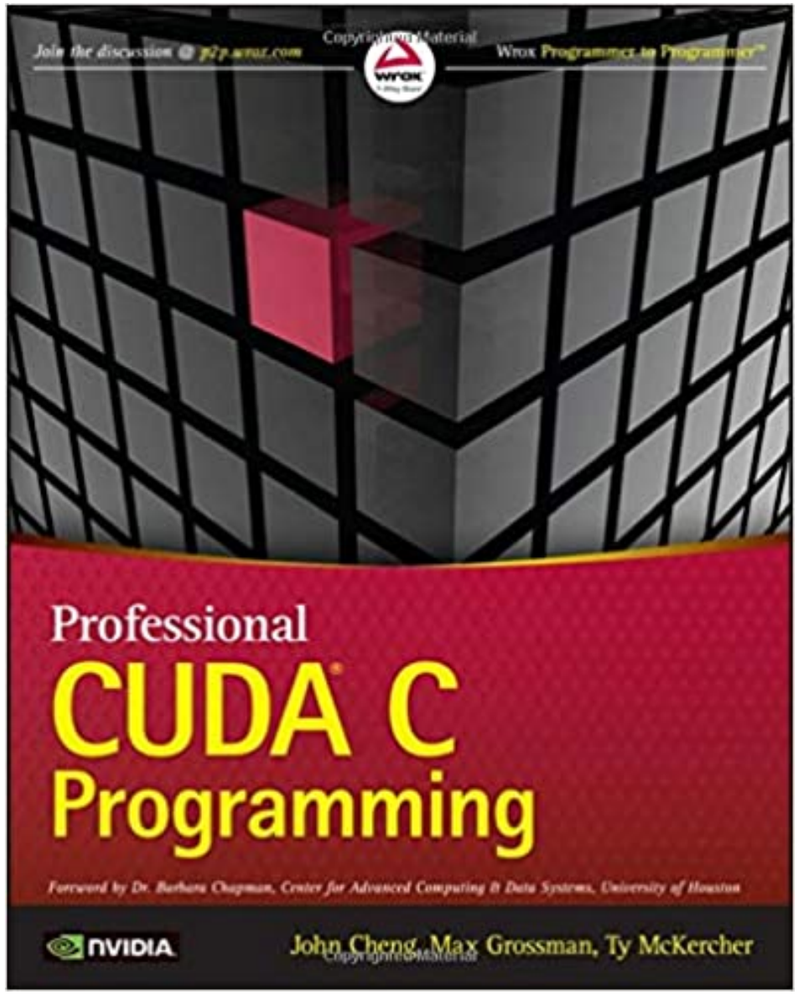
CUDA—online documentation **(preferred)**

Programming guides and API references
http://docs.nvidia.com/cuda/index.html

Teaching and learning resources from NVIDIA
https://developer.nvidia.com/cuda-education-training

Recommended reading (also on class page):

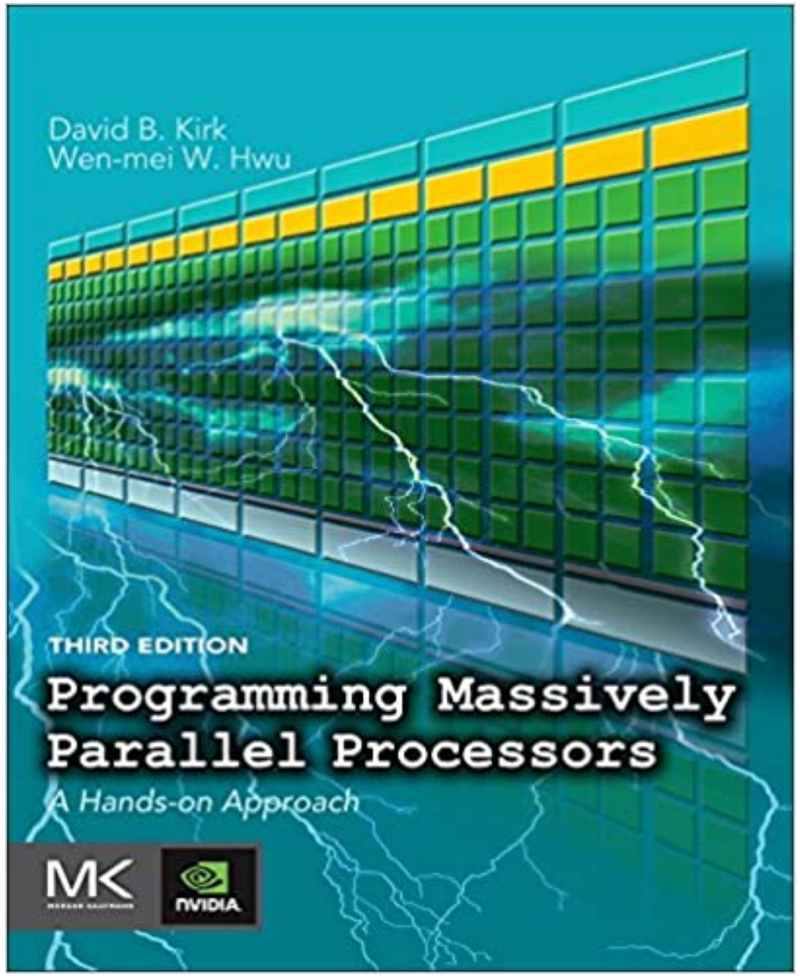
- CUDA_C_Best_Practices_Guide.pdf
- CUDA_C_Programming_Guide.pdf



CUDA books

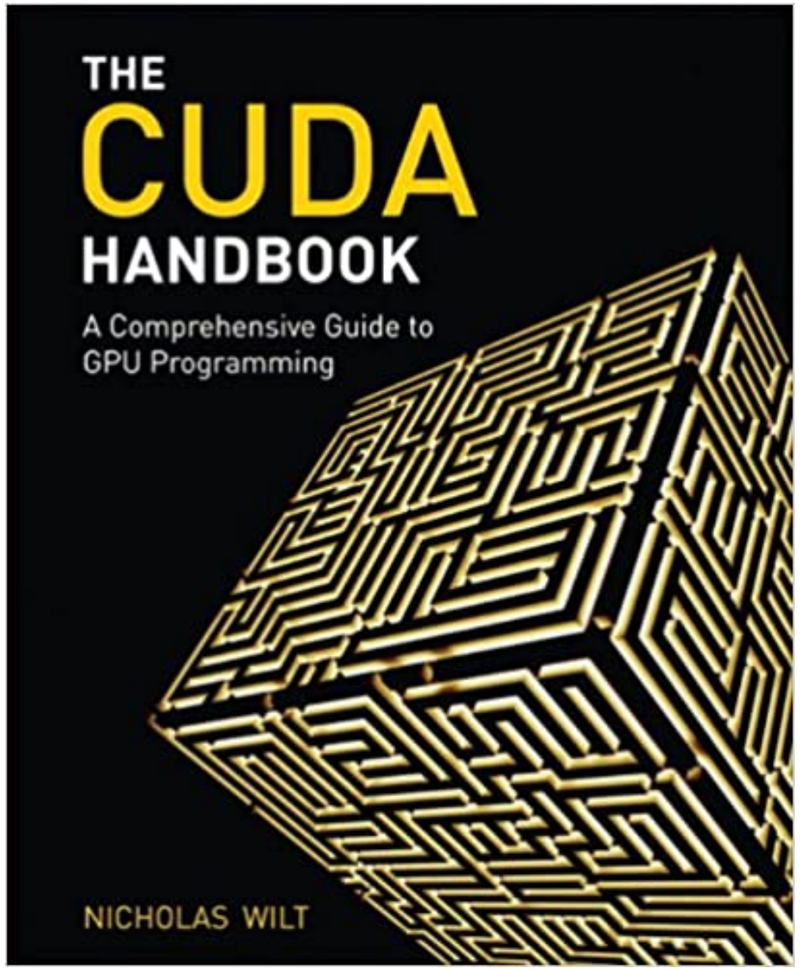
Professional CUDA C Programming, by Cheng, Grossman, McKercher

Recommended for this class; has more advanced usage like multi-GPU programming



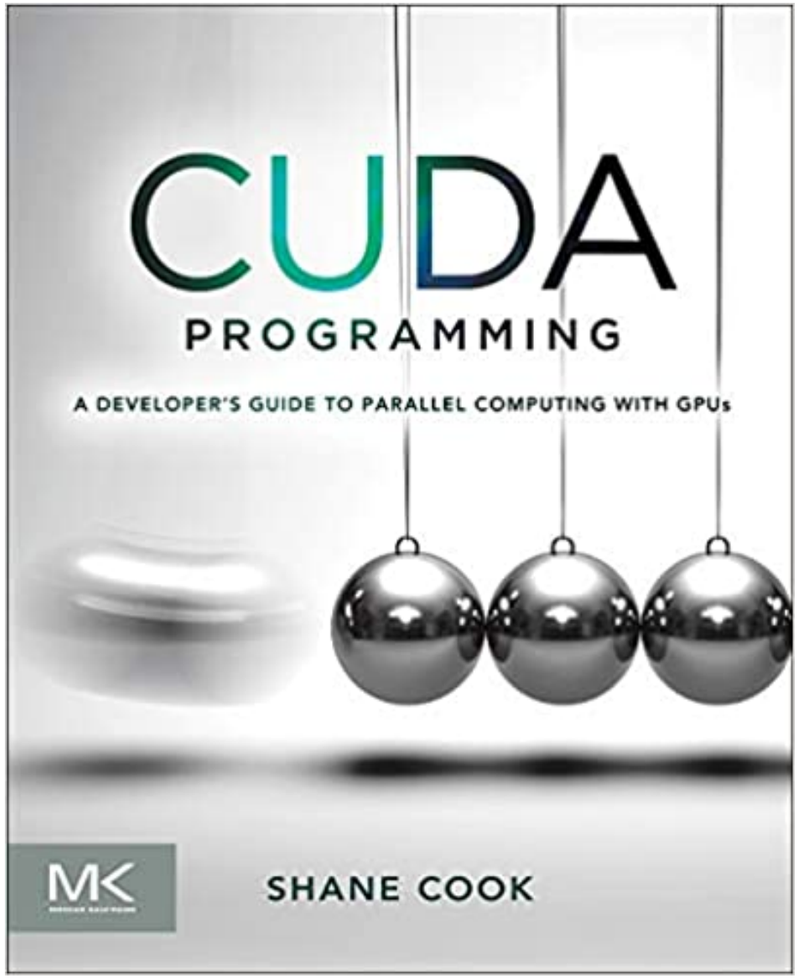
Programming Massively Parallel Processors: A Hands-on Approach, by Kirk, Hwu

In its 3rd edition now; covers a wide range of topics including numerical linear algebra, applications, parallel programming patterns



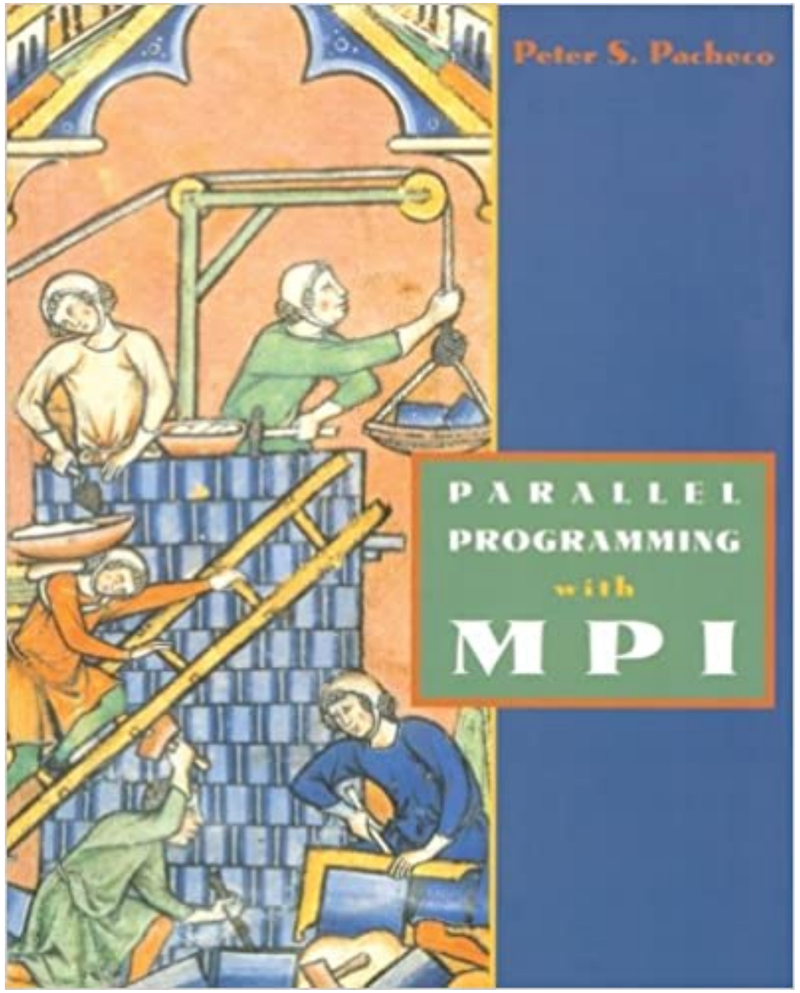
CUDA Handbook: A Comprehensive Guide to GPU Programming, by Wilt

Lots of advanced technical details on memory, streaming, the CUDA compiler, examples of CUDA optimizations



CUDA Programming: A Developer's Guide to Parallel Computing with GPUs, by Cook

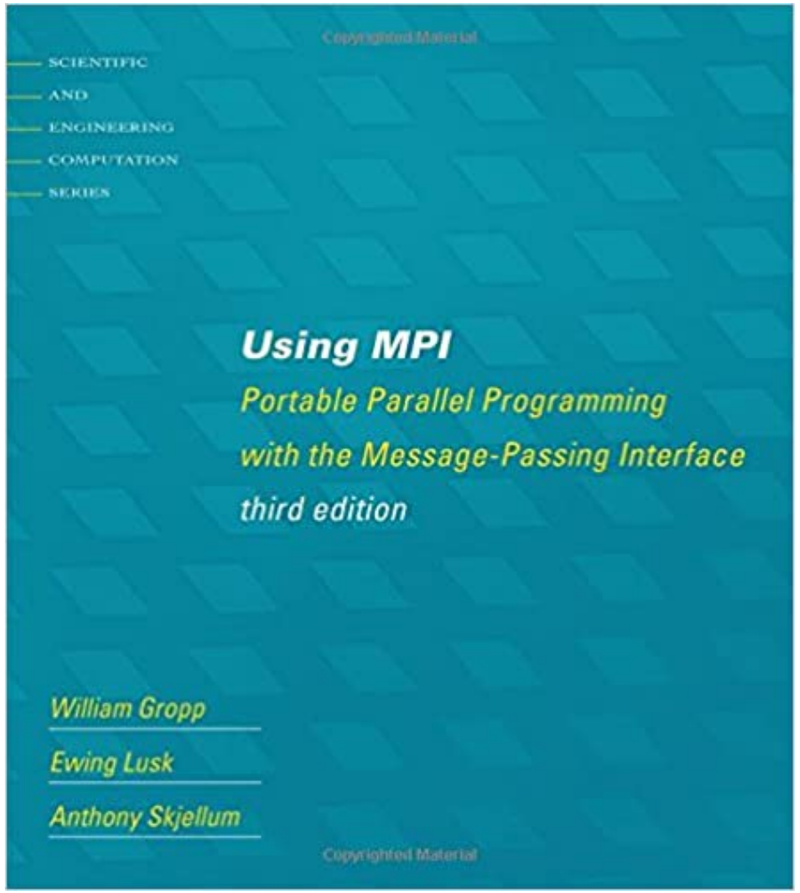
Extensive CUDA optimization guide; practical tips for debugging, memory leaks



MPI books

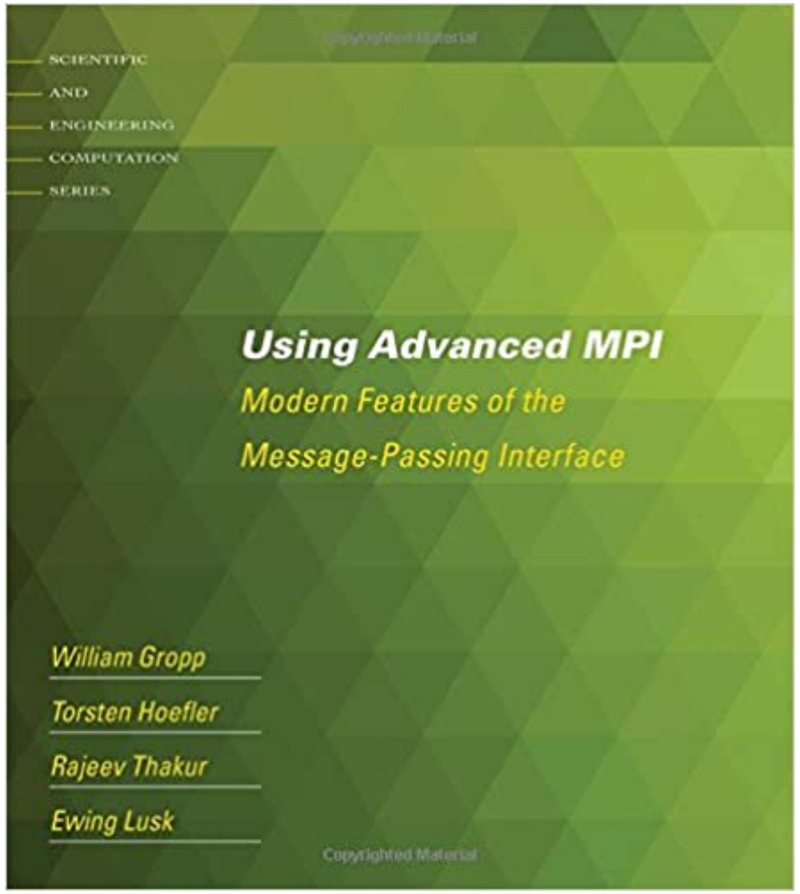
Parallel Programming with MPI, by Pacheco

Classic reference; somewhat dated at this point



Using MPI: Portable Parallel Programming with the Message-Passing Interface, by Gropp, Lusk, Skjellum

Very complete reference



Using Advanced MPI: Modern Features of the Message-Passing Interface, by Gropp, Hoefler, Thakur, Lusk

Same authors as previous entry; discusses recent and more advanced features of MPI

What this class is about

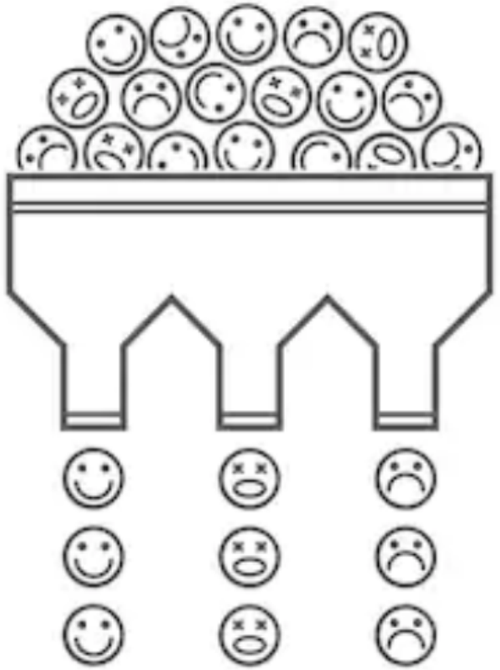
Multicore processors; Pthreads, C++ threads, OpenMP

NVIDIA graphics processors using CUDA

Computer clusters using MPI

Numerical algorithms for illustration

Sort, linear algebra, basic parallel primitives



What this class is **not** about

Parallel computer architecture

Parallel design patterns and programming models

Parallel numerical algorithms

Other related classes

- *CME 342: Parallel Methods in Numerical Analysis;* parallel algorithms
- *CS 149: Parallel Computing;* hardware, synchronization mechanisms, parallel programming models

- *EE 382A: Parallel Processors Beyond Multicore Processing;* SIMD programming, parallel sorting with sorting networks, string comparison with dynamic programming, arbitrary-precision operations with fixed-point numbers

Requirements and pre-requisites

Basic knowledge of UNIX

ssh, compilers, makefile, git

Knowledge of C and C++

Pointers, memory, templates, standard library, polymorphism

General proficiency in scientific programming

Testing, verification, and debugging

Respect for diversity

It is my intent that students from all diverse backgrounds, perspectives, and situations be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength, and benefit.

It is my intent to present materials and activities that are respectful of diversity.

I acknowledge that there is likely to be a diversity of access to resources among students and plan to support all of you as best as I can.

Please let me know of ways to improve the effectiveness of the course for you personally or for other students or student groups.

COVID! Politics! Capitol riot!

Many reasons to be stressed, anxious, and on edge.

You may experience a range of other challenges that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation.

Stanford is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. Confidential mental health services: Vaden service site

Students with Documented Disabilities

Office of Accessible Education (OAE)

563 Salvatierra Walk; 723-1066

http://oae.stanford.edu

Let us know right away!

Honor Code and Office of Community Standards

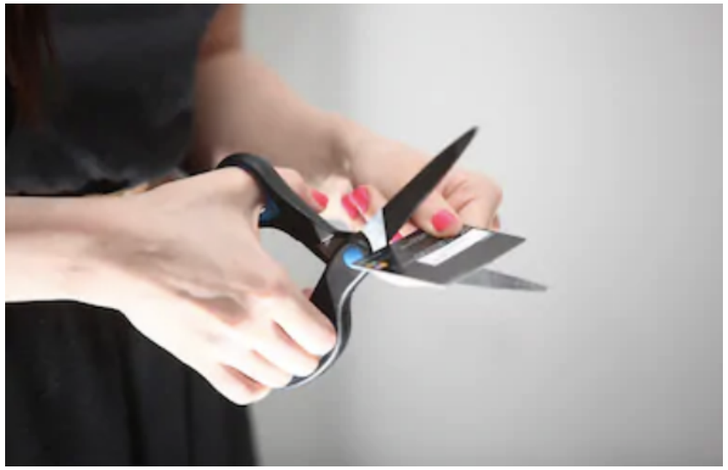
Violations include at least the following circumstances:
copying material from

- another student,
- previous year solution sets,
- solutions found on the internet

All work must be your own!
Honor Code

Do not post any solution set from this class online

If found guilty of a violation, your grade will be automatically lowered by at least one letter grade, and the instructor may decide to give you a "No Credit" grade.



Survey

Please go to Canvas to fill the online survey.