Wrap-up:
Large-Scale Computational and Data Science

CS205: Computing Foundations for Computational Science
Dr. David Sondak
Spring Term 2021

HARVARD
School of Engineering and Applied Sciences

IACS
INSTITUTE FOR APPLIED COMPUTATIONAL SCIENCE
AT HARVARD UNIVERSITY

CS205: Contents

A Practical View: From Design to Implementation

INTRO: LARGE-SCALE COMPUTATIONAL AND DATA SCIENCE

A. PARALLEL PROCESSING FUNDAMENTALS

- A.1. Parallel Processing Architectures
- A.2. Large-scale Processing on the Cloud
- A.3. Practical Aspects of Cloud Computing
- A.4. Application Parallelism
- A.5. Designing Parallel Programs

B. PARALLEL COMPUTING

- B.1. Foundations of Parallel Computing
- B.2. Performance Optimization
- B.3. Accelerated Computing
- B.4. Shared-memory Parallel Processing
- B.5. Distributed-memory Parallel Processing

C. PARALLEL DATA PROCESSING

- C.1. Foundations of Data Processing
- C.2. Batch Data Processing
- C.3. Dataflow Processing
- C4. Stream Data Processing

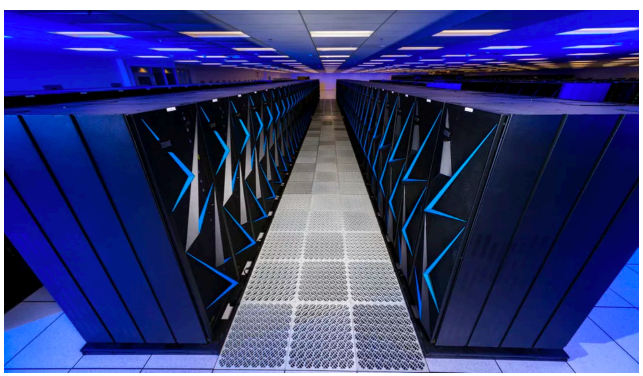
WRAP-UP: ADVANCED TOPICS

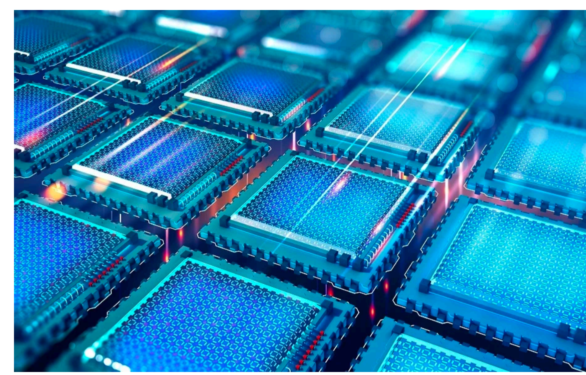
The Future of High Performance Computing

The Future of High Performance Computing

Is it obvious?

- Are more powerful computers inevitable?
  - What does more powerful mean?
- What kinds of challenges can we solve with supercomputers?
  - What issues can arise?
- Is the path forward linear?
  - New innovations may require drastic creativity





The Future of High Performance Computing

The Underlying Goal

- At it’s heart, HPC enables solutions to the world’s biggest, most important problems.
  - Healthcare
  - Education
  - Autonomous vehicles
  - Space travel and exploration
  - Climate change

- These problems are intractable with current algorithms and computing technology.

- Through a combination of algorithmic and computing innovations, we can begin to tackle and solve such problems.
  - We want good solutions and we want them fast (sometimes in real-time).

The Future of High Performance Computing

Where We Are

- We are on the cusp of the world's first *double-precision* exascale supercomputer.

- Frontier is under construction at Oak Ridge National Laboratories.
    - 1.5 EFLOP/s
    - More than 100 cabinets
    - A node contains: AMD EPYC CPU, 4 AMD Radeon Instinct GPUs
    - High speed links between CPU and GPU

https://www.olcf.ornl.gov/frontier/#3

The Future of High Performance Computing

Essential Challenges

- Major challenges going forward include:
  - Fault tolerance
  - **Energy and power**
  - Data storage and processing
  - Network interconnects
  - How to write parallel software at this scale
  - …

- Great that we can solve big problems; are we creating new ones?
  - What is are environmental / ecological impacts?
  - Can consume enormous amounts of energy.
  - Take up a lot of space and displace natural environments.
  - What steps can we take now to mitigate these risks?

HPC Energy and Power

HPC Energy and Power

How Much Energy, Really?

- Supercomputers are really, really big.

Multiple cabinets← nodes ← processors ← cores ← compute units.

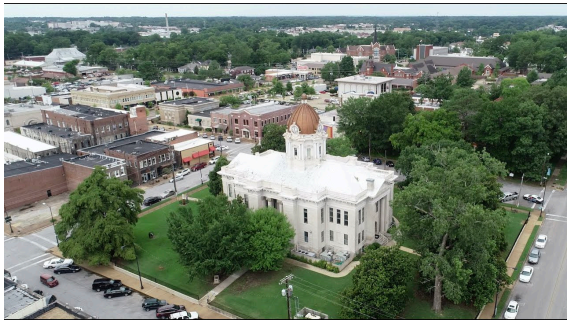
- Each little calculation costs a little bit of energy.
  - This depends on:
    - The hardware design
    - The software design
    - The algorithm design

Computer Name	Year	Peak PFLOPS	MW at Peak	PFLOPS / MW
Roadrunner	2009	1.1	2.4	0.41
Summit	2019	148	10.1	14.8

HPC Energy and Power

Comparison to a City and a Brain

- In 2009, the projected energy consumption of an exascale computer was 2000 MW!
- The Frontier supercomputer is projected to have an energy consumption of 30 MW.



	Peak performance	Energy consumption	Weight
Frontier	1.5 EFLOPS	30 MW	Hundreds of tons
Human brain	Roughly 30 PFLOPS	20 W (average)	3 lbs (average)
Tupelo, MI	N/A	17.8 MW	???

Addressing the Energy Needs

Addressing the Energy Needs

Hardware and Software and Algorithms

Hardware

- Replace copper cables with optical links
- New memory models
  - Non-volatile memory (NVM)
  - Low-voltage SRAM
- Reduce cost of data movement
  - 3D stacked memory
- Cooling technologies
  - Liquid immersion cooling systems

https://www.top500.org/lists/green500/
Supercomputing's Super Energy Needs, and What to Do About Them
The Landscape of Exascale Research: A Data-Driven Literature Analysis

Addressing the Energy Needs

Hardware and Software and Algorithms

Software

- Improve coordination between the processors
- Dynamically scale frequency, voltage, level of concurrency
- Efficiently distribute power over the nodes
  - Power-aware job scheduler

A survey on software methods to improve the energy efficiency of parallel computing
The Landscape of Exascale Research: A Data-Driven Literature Analysis

Addressing the Energy Needs

Hardware and Software and Algorithms

Algorithms

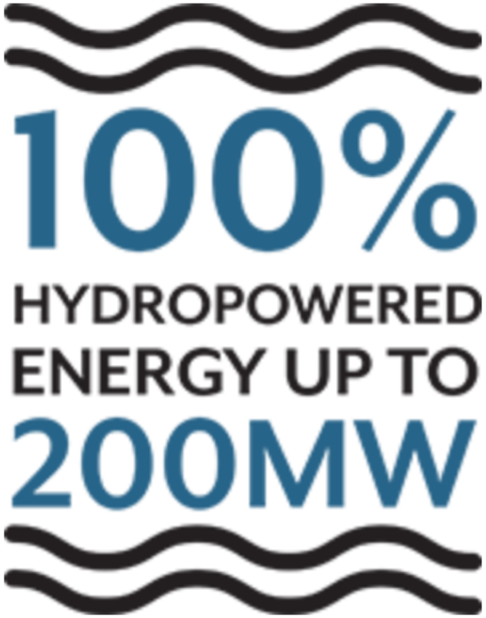
- Mixed-precision floating point calculations
- Communication-avoiding algorithms
- Approximation-based methods
  - Refine data types and the number of iterations
  - Avoid wasting memory and computing resources

A survey on software methods to improve the energy efficiency of parallel computing
The Landscape of Exascale Research: A Data-Driven Literature Analysis

Addressing the Energy Needs

Efficient Datacenters

- Develop datacenters powered from renewable energy.
- Recycle the generated heat and use in the community.



LUMI is using 100% hydropowered energy. Up to 200MWs are available. The waste heat of LUMI will produce 20 percent of the district heat of the area.

https://www.lumi-supercomputer.eu/

New Architectures on the Horizon

New Architectures on the Horizon

Neuromorphic Computing

- The brain is remarkably efficient
  - Only weighs 3 pounds on average
  - Consumes about 20 W on average
  - Rough estimate: 30 PFLOP capacity

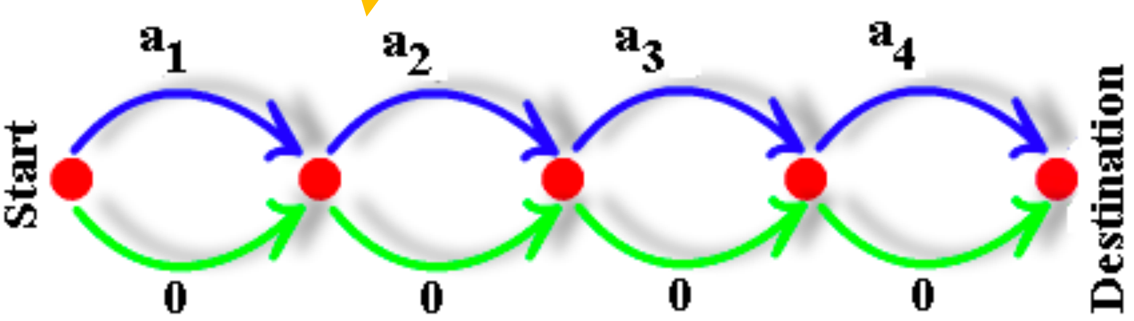
- The basic goal of neuromorphic computing:
  - Build circuits composed of physical neurons
  - These circuits are connected by physical synapses
  - These synapses have direct access to memory
  - Inspired by the brain → leads to huge gains in speed and energy efficiency

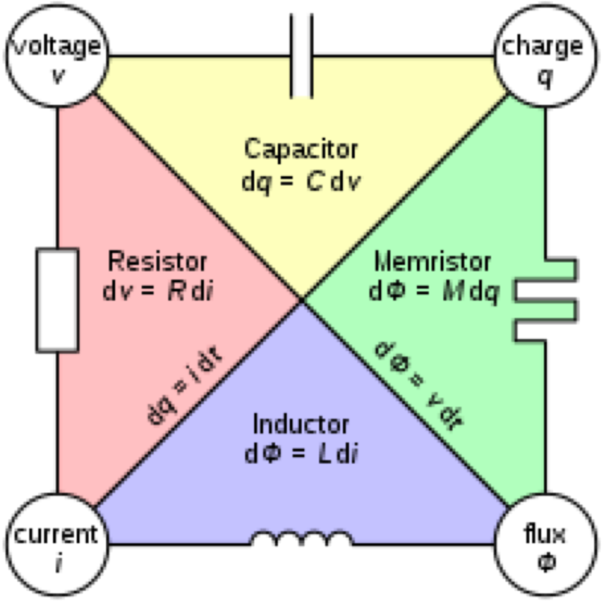
- Main challenges:
  - Current transistor technology can't reach the scales necessary
  - Current circuit technology is limited to 2D

New Architectures on the Horizon

Neuromorphic Computing: Hardware

- Can we map deep neural networks onto a chip using physics?
- Two main technologies today:
  1. Hybrid CMOS/memristive
  2. Photonic





[Physics for Neuromorphic Computing]

New Architectures on the Horizon

Neuromorphic Computing: Memristive

- Hybrid CMOS/memristive
  - Neurons are made from CMOS (transistors)
  - Information flowing through the network is electrical current
  - A synapse acts like a valve for the current
  - Memristor (memory-resistor): A nanoscale resistor that has a memory of past voltages or currents
  - **Big challenge: How to build a memristor**

- Potential gains:
  - Hundred-fold gain in energy consumption and speed compared to GPUs

New Architectures on the Horizon

Neuromorphic Computing: Photonic

- Photonic
  - Build neural networks with optical components
  - Neurons implemented by optical resonators
  - Synapses implemented by multiple interferometers or using optical waveguides
- Benefits:
  - Convey information in parallel on a single fiber
  - Build a passive neural network with very low energy consumption
- Open considerations:
  - How to shrink the size of the neurons and synapses
  - How to minimize energy cost of converting information to light

New Architectures on the Horizon

Cryogenic Computing

- Can we significantly reduce power and cooling costs of giant computers?
- Current CMOS-based systems seem to be hitting a wall in terms of energy efficiency
- New architectures based on superconducting computing may offer a solution
  - Operate a ultra-low temperatures
  - Potential for 1 PFLOP/s at 20 KW; 100 PFLOP/s at 200 KW
  - May also aid quantum computing efforts

[Cryogenic Computing Complexity (C3)]

New Architectures on the Horizon

Analog Computing

- Analog computers were used for some of the most famous engineering missions (e.g. Apollo)

- The very basic idea is to have a system that is governed by the equations you want to solve
  - Originally this meant building computers out of gears
  - Eventually electronic analog computers were developed

- They were quite difficult to work with and were eventually replaced by digital computers

Not Your Father's Analog Computer

New Architectures on the Horizon

Analog Computing

- There is a resurgence of interest in analog computers now that digital computers are confronting their own issues

- Some research using FPGAs (Field Programmable Gate Arrays) indicates that analog computers could become cost-effective

- Limitations include:
  - Accuracy
  - Can be difficult to design
  - Cost can be high for complex problems

Beyond the Carbon Footprint

Is That All?

Thinking about Things Not Seen

- A lot of thought has been given to *energy efficiency*.
  - This is good:
    - More efficient calculations
    - Better use of resources
    - Smaller carbon footprint

- But there is more to the story:

Where are the resources to build these machines coming from?

- All hardware components must be built from something:
  - Rare metals?
  - Silicone?
  - Cooling with water --- Where does the water come from?

Is That All?

Thinking about Things Not Seen

- Datacenters and supercomputing centers are becoming very energy efficient

- Emissions are driven by one-time infrastructure and hardware
  - Facility construction
  - Chip manufacturing

- At Facebook, these infrastructure and hardware-related activities accounted for 23 times more carbon emissions than operational use!

- Need to update hardware, software, programming languages, and compilers

- Don’t forget about water consumption and mining raw materials

Chasing Carbon: The Elusive Environmental Footprint of Computing

Discussion

Thoughts on the Future?

- Think about the different components of computing. Which ones do you think contribute to climate change? Why?

- What are the drivers for the development of new architectures? Which new approaches are you most excited about? Why?

- Discuss and expand upon the ecological implications of high performance computing.

Parting Thoughts

Where Do We Go From Here?

- We want to affect positive change in the world
- This requires social and environmental responsibility

It's hard to save the world if we cut the rainforests down to do so.

Wrap-up

CS205: Aim and Objectives

Learn Parallel Computational Thinking and Tools

Practical overview of:

- Foundations of “parallel thinking”
- Aspects to consider when designing large-scale applications
- Parallel programming models for compute- and data-intensive applications, and
- Existing platforms, open-source tools and cloud services to support their execution

After the course, you will be in a great position to:

- Make effective use of the diverse, and rapidly changing, landscape of programming models, platforms and computing architectures for high performance computing and big data
- Decide which kind of programming model and platform is appropriate to meet your scalability and performance
- Apply the enduring principles behind these rapid changes in technology that remain true, no matter which version of a particular platform you are using

CS205: Contents

A Practical View: From Design to Implementation

INTRO: LARGE-SCALE COMPUTATIONAL AND DATA SCIENCE

A. PARALLEL PROCESSING FUNDAMENTALS
- A.1. Parallel Processing Architectures
- A.2. Large-scale Processing on the Cloud
- A.3. Practical Aspects of Cloud Computing
- A.4. Application Parallelism
- A.5. Designing Parallel Programs

B. PARALLEL COMPUTING
- B.1. Foundations of Parallel Computing
- B.2. Performance Optimization
- B.3. Accelerated Computing
- B.4. Shared-memory Parallel Processing
- B.5. Distributed-memory Parallel Processing

C. PARALLEL DATA PROCESSING
- C.1. Foundations of Data Processing
- C.2. Batch Data Processing
- C.3. Dataflow Processing
- C4. Stream Data Processing

WRAP-UP: ADVANCED TOPICS

CS205: Contents

Programming Models, Platforms and Infrastructures

Application Software

OpenACC

Spark

OpenMP

Map-Reduce

DRMAA

MPI

Programming Model

Platform

Architecture

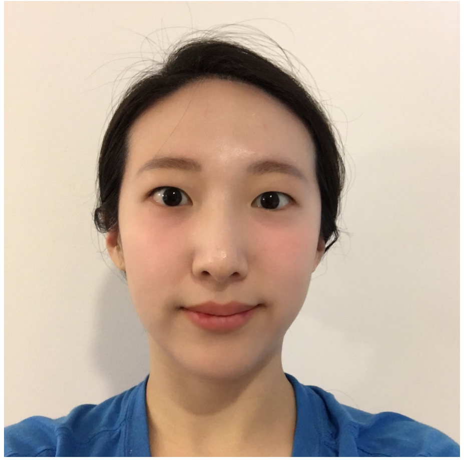
amazon web services

CLOUD

ON-PREMISES

CS205: Staff

Teaching Fellows



Hayoun Oh

Simon Warchol

Oluwatosin (Tosin) Alliyu

Haipeng Lin

George Touloumes

Final Reminders

Course Wrap-up

- Final presentations
  - Monday, May 10th
  - Select a presentation block and presentation slot
  - Sign-up link is on Piazza
  - Read instructions when you sign up!
  - Please submit a pre-recorded version.

- Think about TF-ing in the fall!

- The Q is now open