Genome 540 discussion

January 23st, 2025

Joe Min

Genome Sciences
UNIVERSITY OF WASHINGTON

Agenda

Homework 2 questions?

GPUs

Climate implications of AI and machine learning

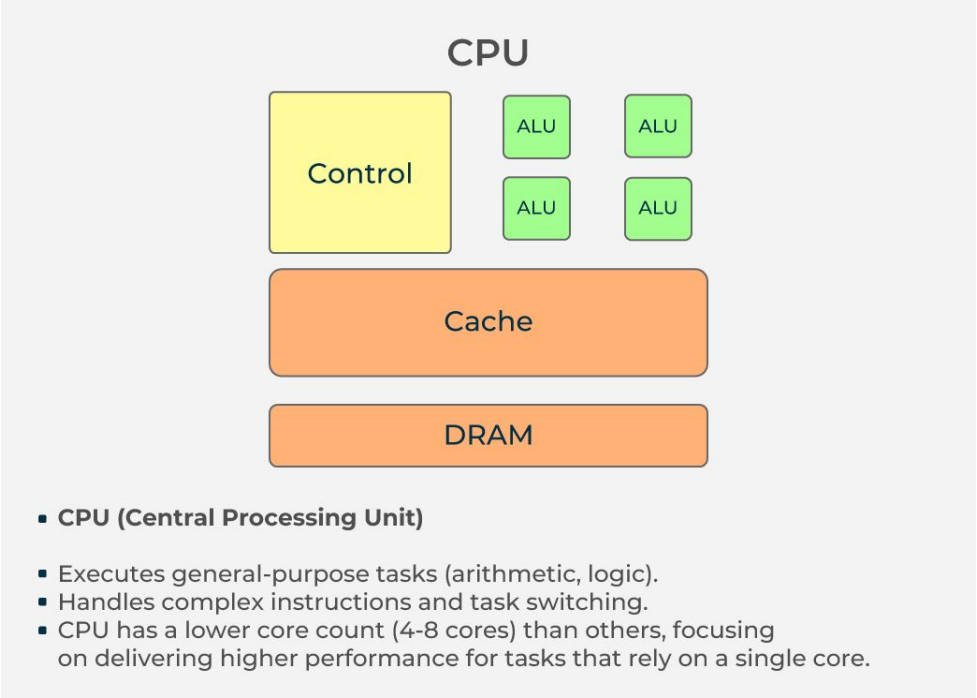
Homework 2 questions?

GPUs

But first, central processing units (CPUs)

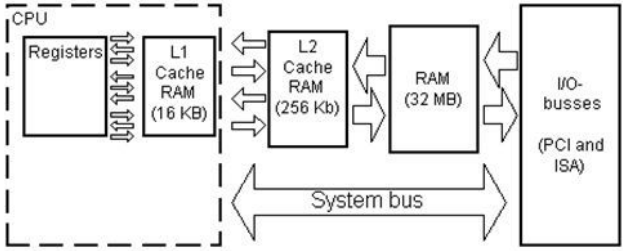
Major units:

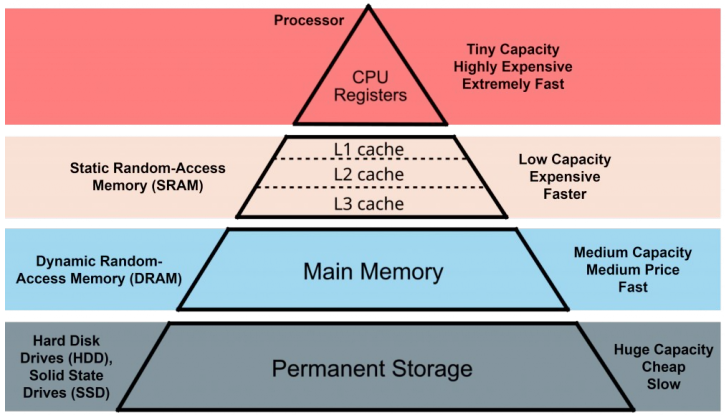
- Storage unit
  - RAM
  - Cache
- Control unit
- Arithmetic Logic Unit (ALU)



https://www.geeksforgeeks.org/difference-between-cpu-and-gpu/

CPUs: storage unit





https://spear-itn.eu/wp-content/uploads/2024/01/pyramid_upgrade-1024x603.png
https://images.slideplayer.com/27/9213216/slides/slide_9.jpg

"The holder"

CPUs operate on their internal memory (fast)

CPUs utilize Random Access Memory (RAM) as well as other caches to reduce trips to disk (slow)

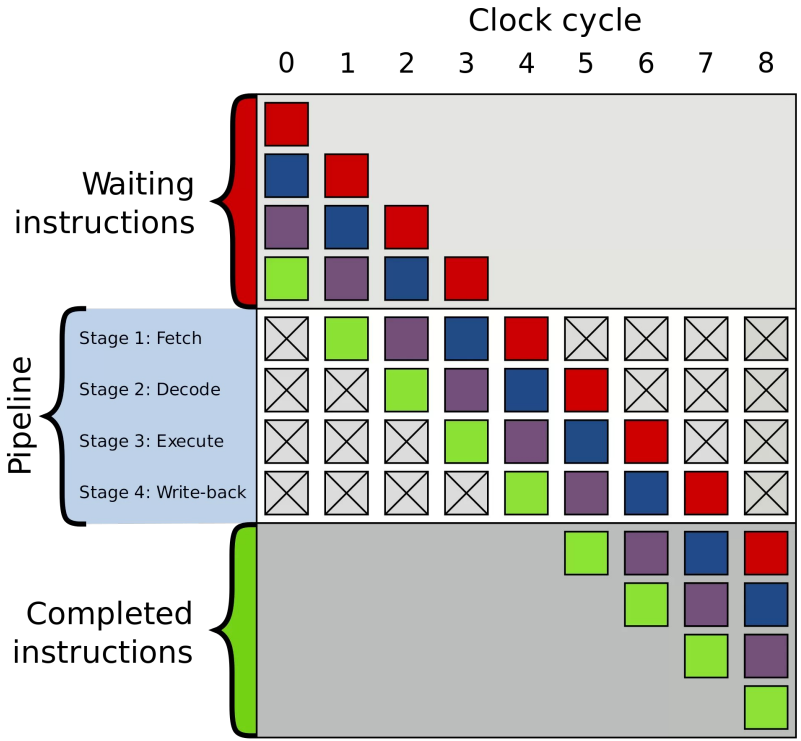
CPUs: control unit

"The manager"

Translates instructions into arithmetic operations

Routes instructions and data to the correct executing units

Sends data back as needed



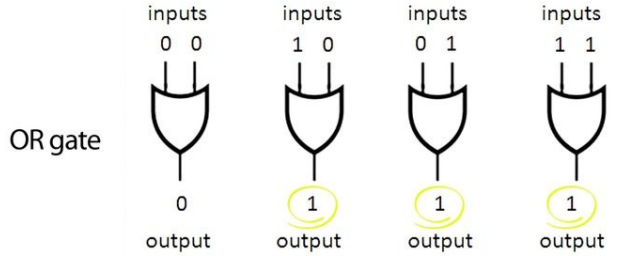
https://www.techspot.com/articles-info/1821/images/2019-04-21-image-4.png

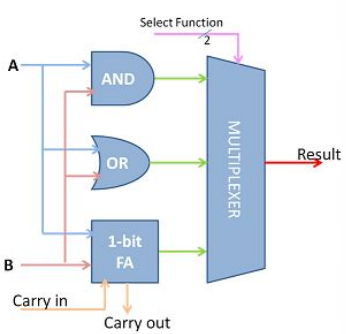
CPUs: ALUs

"The workhorse"

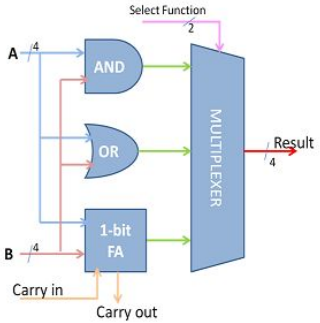
Actually carries out the instructions on the data

Simple binary logic builds into complex programs





1-bit 3-function Combinational ALU



4-bit 3-function Combinational ALU

https://study.com/cimages/videopreview/videopreview-full/arithmetic-logic-unit-alu-definition-design-and-function_111389.jpg
https://res.cloudinary.com/witspry/image/upload/witscad/public/content/courses/computer-architecture/primitive-alu-supporting-and-or-add-function.png

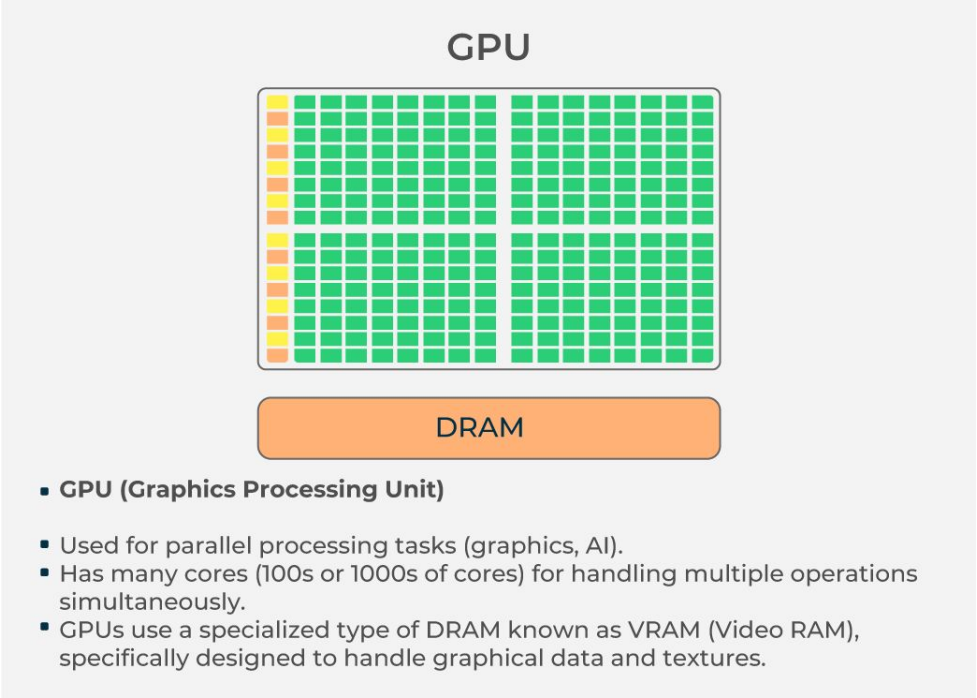
Graphics processing units (GPUs)

Same major units!

But optimized for specific, parallelizable operations

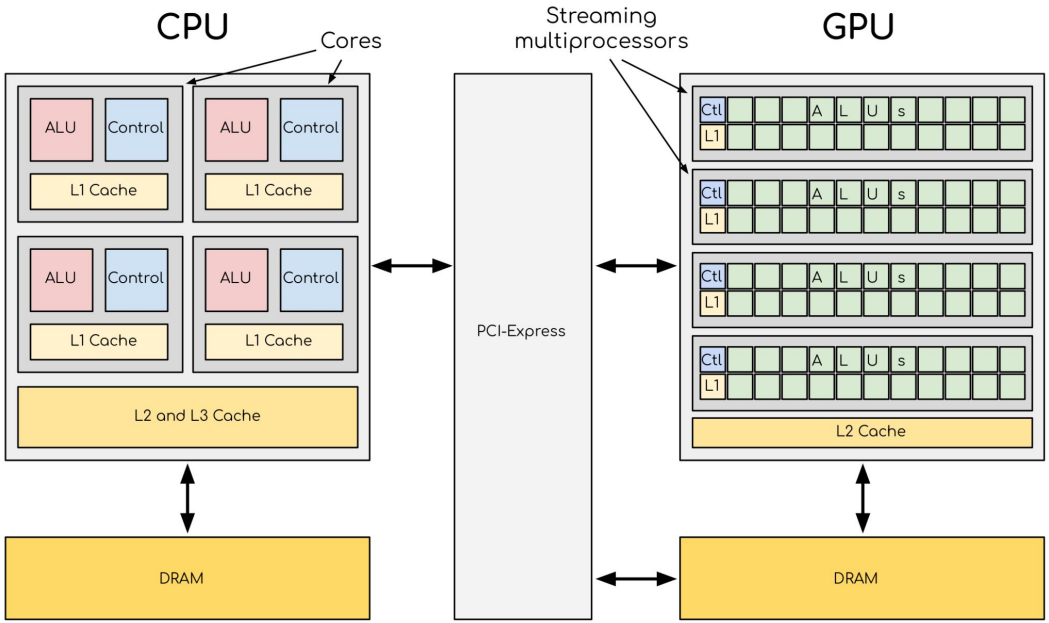
- 100s-1000s of ALUs

Usually larger storage unit



https://www.geeksforgeeks.org/difference-between-cpu-and-gpu/

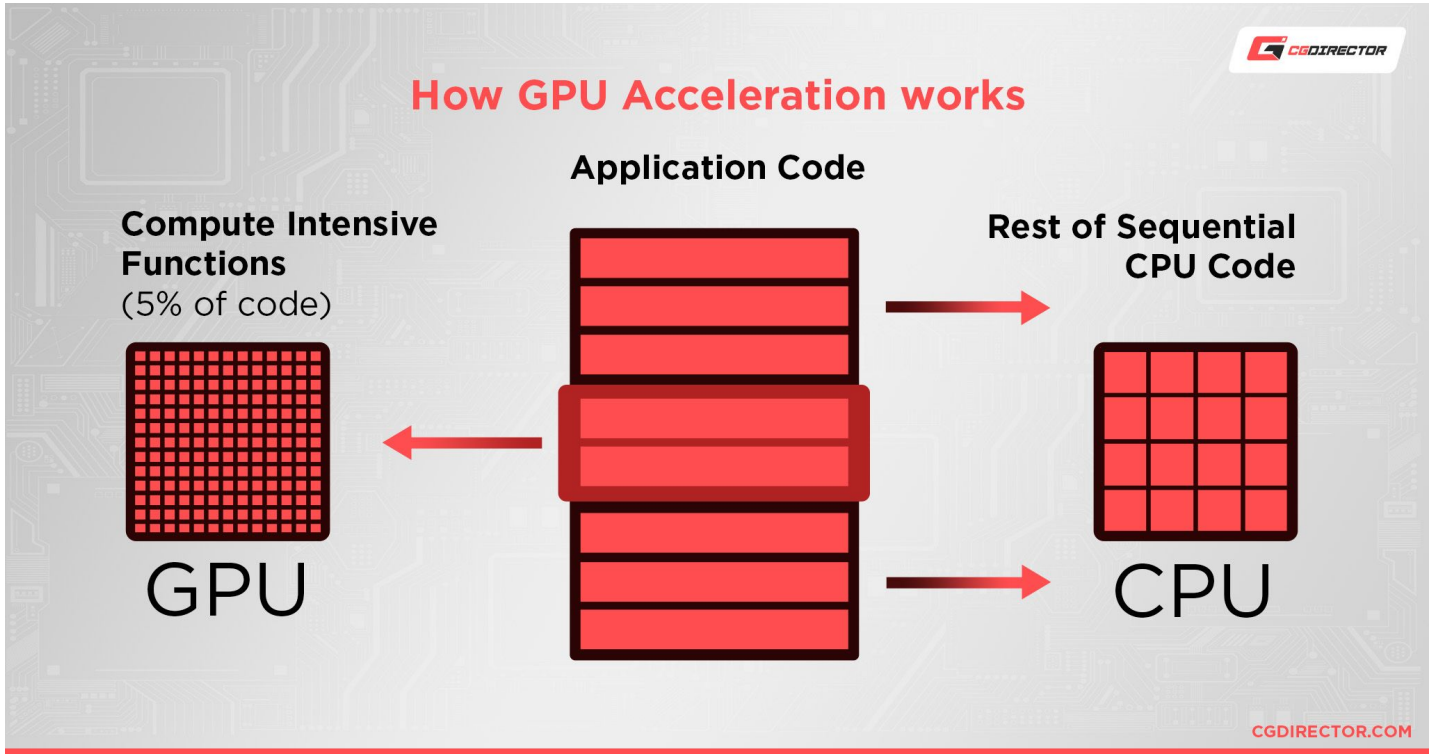
Side-by-side comparison



Parallelization in GPUs allows for incredible performance in matrix math and graphics processing

https://enccs.github.io/gpu-programming/_images/CPUAndGPU.png

Working together



https://www.cgdirector.com/wp-content/uploads/media/2022/06/How-GPU-Acceleration-works.jpg

Climate implications of AI and machine learning

What energy problem?

In 2022, data centers used ~2% of global demand for energy

“By 2026, … data centres’ energy consumption will have increased by between 35% and 128%”

- Equivalent to the energy usage of Sweden or Germany

nature › outlook › article

OUTLOOK | 17 October 2024 | Correction 25 November 2024

Fixing AI’s energy crisis

Hardware that consumes less power will reduce artificial intelligence's appetite for energy. But transparency about its carbon footprint is still needed.

By Katherine Bourzac

https://www.nature.com/articles/d41586-024-03408-z
https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf

AI is exacerbating the issue

A ChatGPT prompt response takes roughly 10x the energy of a Google search

Google estimates its carbon emissions have increased by 48% since 2019

Microsoft, too, at 30% since 2020

https://www.nature.com/articles/d41586-024-03408-z

What uses energy?

“Any time that electrons move through chips, some energy is dissipated as heat.”

Reliance on huge datasets, and moving it between processing units and long term storage, is the biggest factor

- 90% of energy usage is spent accessing memory

https://www.nature.com/articles/d41586-024-03408-z

What uses water?

Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models

Pengfei Li
UC Riverside

Jianyi Yang
UC Riverside

Mohammad A. Islam
UT Arlington

Shaolei Ren[1]
UC Riverside

Paper that originally made the claim that every ChatGPT response uses, on average, 500mL of water

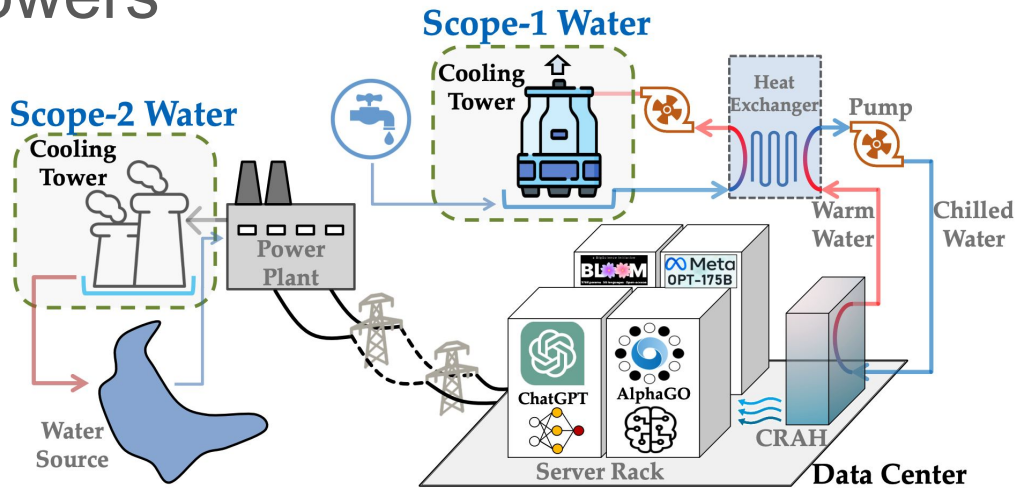
https://arxiv.org/pdf/2304.03271

What uses water?

On-site water for data center cooling (scope 1)

Off-site water for electricity generation (scope 2)

Evaporation from cooling towers is the calculated quantity “wasted water”



Office hours

Good work on homework 1!

Reminder:

Homework 2 is due this Sunday (Jan 26) at 11:59pm!