

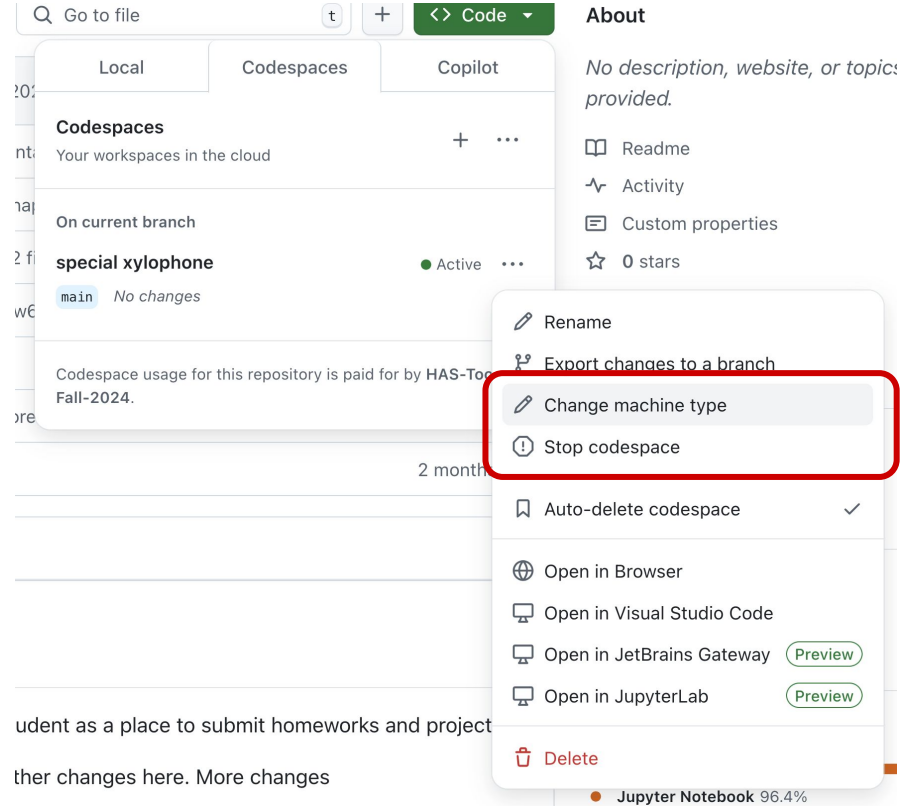
HAS Tools:

Parallel processing concepts

November 15, 2024

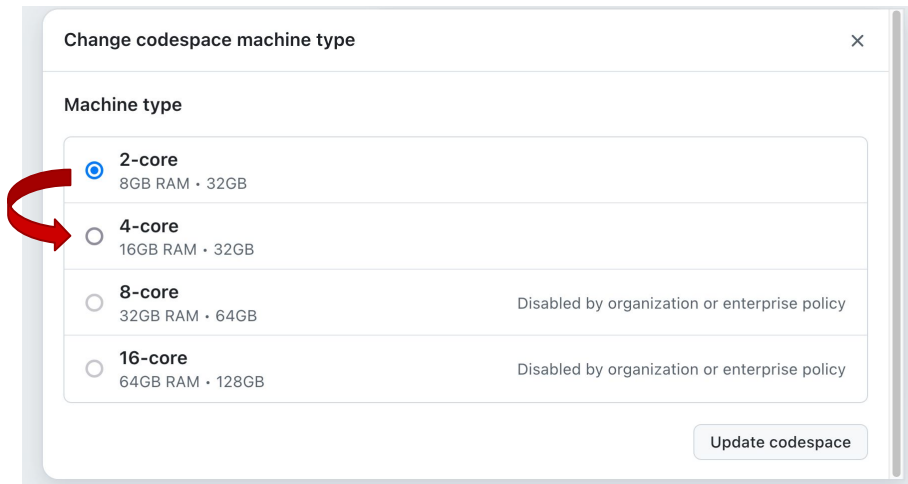
First off, some setup

- If you have your codespace already started please go back to your homework repo and “Stop codespace”



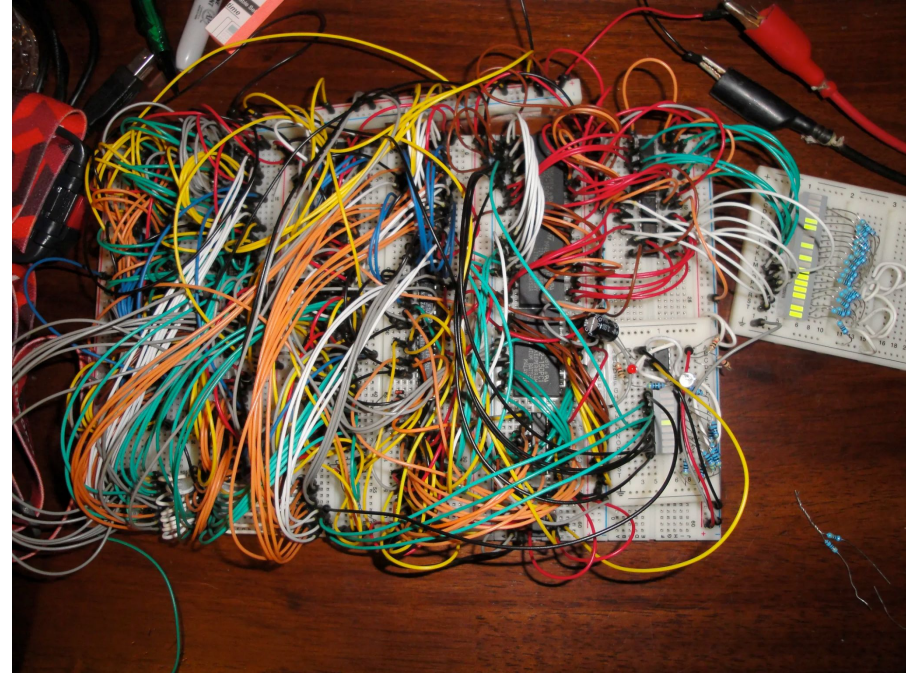
First off, some setup

- If you have your codespace already started please go back to your homework repo and “Stop codespace”
- Then, wait for the codespace to shut down, and go back into the menu and select “Change Machine Type”
- Select “4-core”



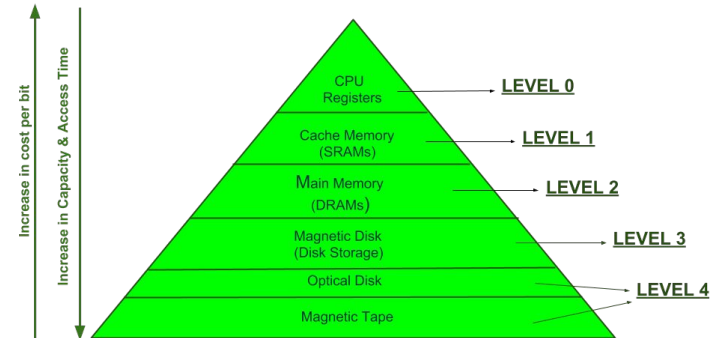
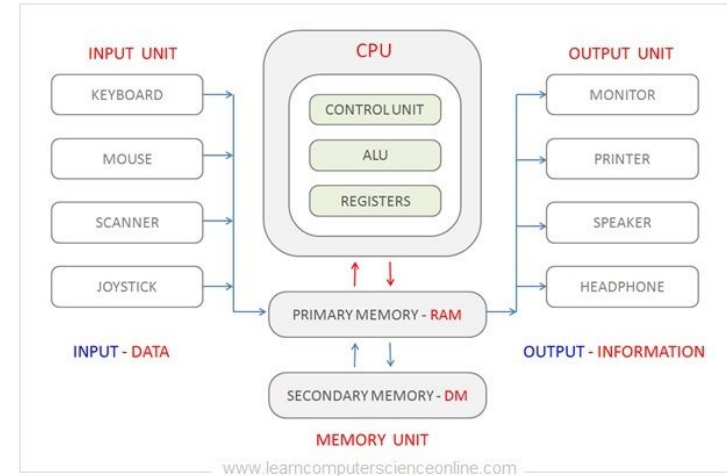
Some light computer architecture

- Quick plug - if you have time and interest, take a computer architecture class
- Computers are essentially magic



Some light computer architecture

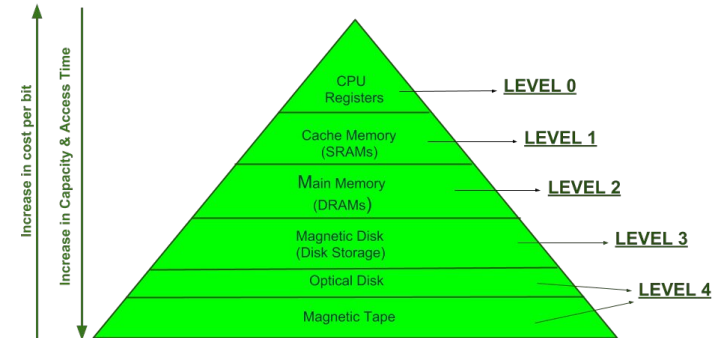
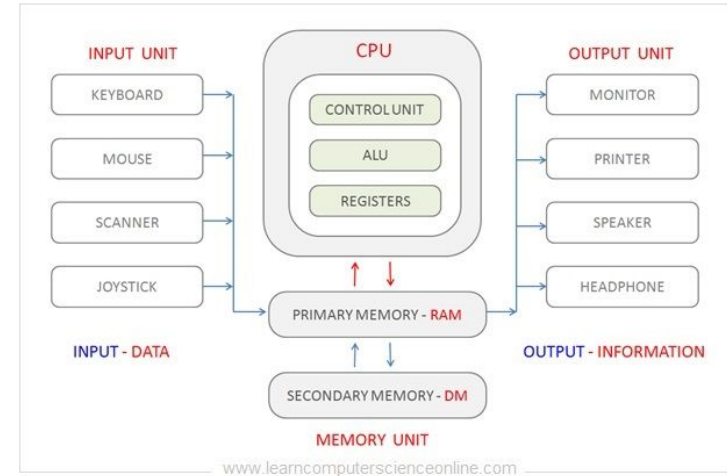
- Quick plug - if you have time and interest, take a computer architecture class
- Computers are essentially magic
- Key components:
 - CPU
 - Primary memory/RAM
 - Storage/Hard Drive
 - Input/output devices
 - GPU (more advanced)



MEMORY HIERARCHY DESIGN

Some light computer architecture

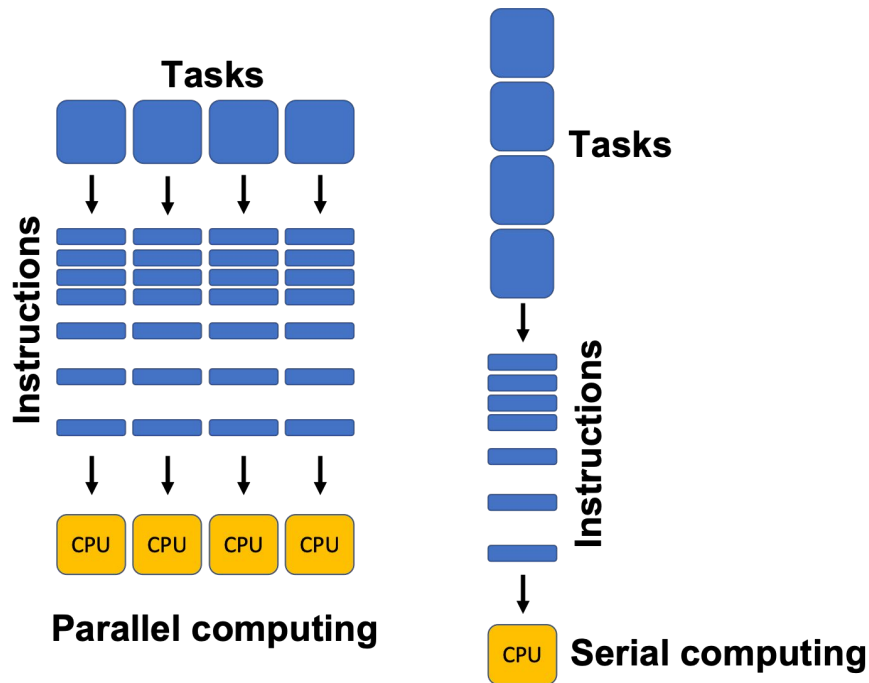
- Quick plug - if you have time and interest, take a computer architecture class
- Computers are essentially magic
- Key components:
 - CPU
 - Primary memory/RAM
 - Storage/Hard Drive
 - Input/output devices
 - GPU (more advanced)
- So, what are those “cores”?



MEMORY HIERARCHY DESIGN

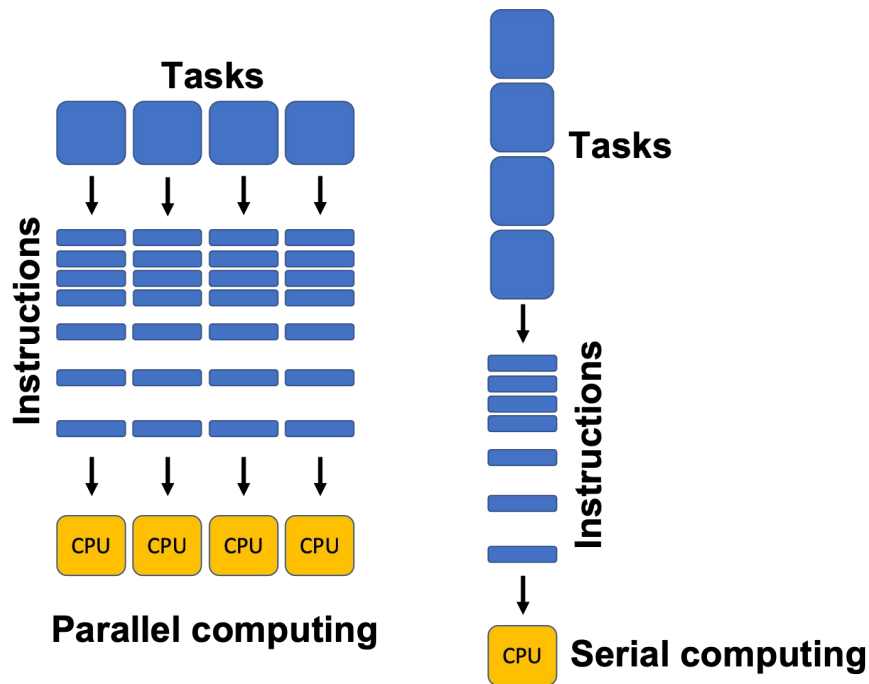
Cores are essentially “copies” of the cpu that can work on problems independently

- Many tasks that we want computers to do don't need to be done in a fixed specific order



Cores are essentially “copies” of the cpu that can work on problems independently

- Many tasks that we want computers to do don't need to be done in a fixed specific order
- So, you can break the tasks up and do them simultaneously, or “in parallel



Historic side note: Parallel computing in history

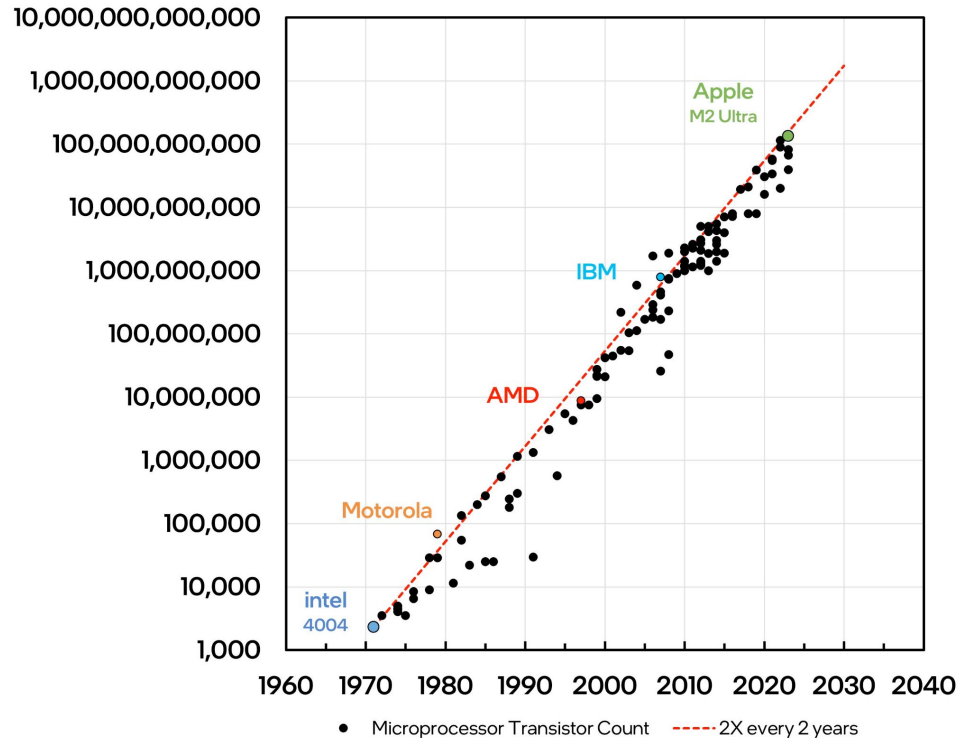
- The term “computer” originally referred to a profession for somebody who “does computation”
- Mostly (uncredited/unrecognized) women who did calculations across a wide range of large projects for NASA, etc
- Often they would work on sub-problems to be combined back into a larger result later - parallel computing in action



Seems like a good idea! Why not do everything in parallel?

There are mainly two reasons why parallel computing was generally niche up until ~2015:

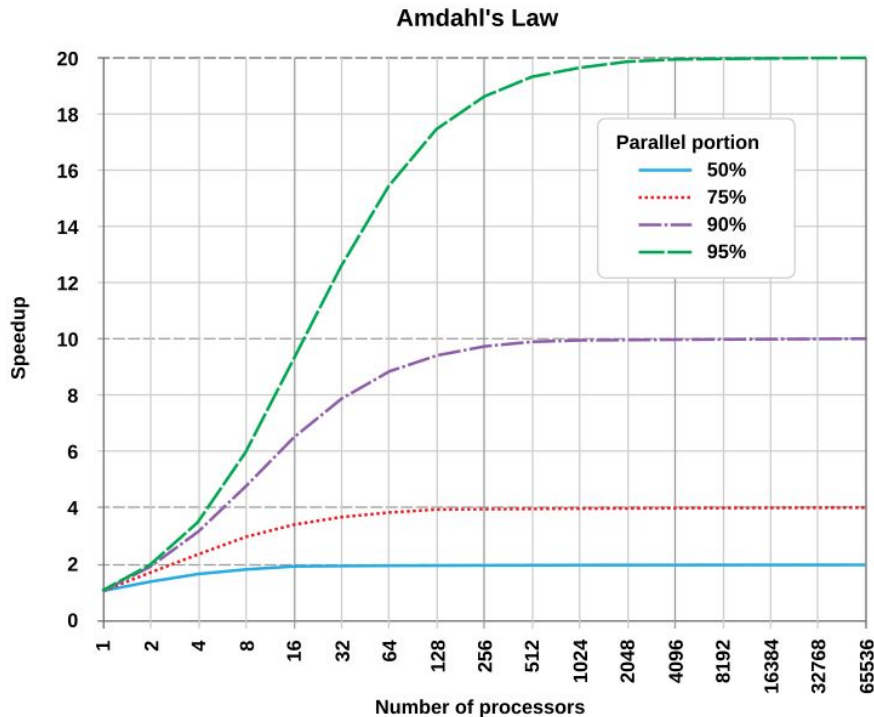
- Parallel computing can be difficult to get right.
- Moore's law: CPUs keep getting faster.



Seems like a good idea! Why not do everything in parallel?

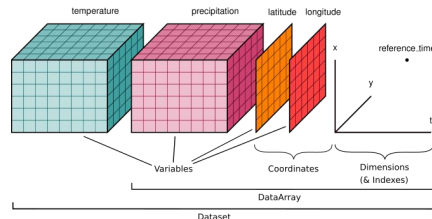
There are mainly two reasons why parallel computing was generally niche up until ~2015:

- Parallel computing can be difficult to get right.
- Moore's law: CPUs keep getting faster.
- Amdahl's law: Diminishing returns due to parts you can't effectively parallelize



Wait, then why are we talking about parallel processing?

- Scaling processor performance is hard, and not always power efficient
- Computers are more general purpose than ever, we often have many simultaneous programs running
- As parallel processing becomes more common, better software support is available
- In fact, you've been doing it without knowing it



prcp	(time, y, x)	float32	dask.a...	
	Array	Chunk		
Bytes	9.25 GiB	230.93 MiB		
Shape	(14965, 584, 284)	(365, 584, 284)		
Dask graph	41 chunks in 2 graph layers			
Data type	float32 numpy.ndarray			

A 3D visualization of a data chunk, showing a rectangular prism with dimensions 14965, 584, and 284.

Jump to new (4 core) codespaces for demo time