High Performance Scientific computing
Lecture 4

S. Gopalakrishnan

Von Neumann Architecture

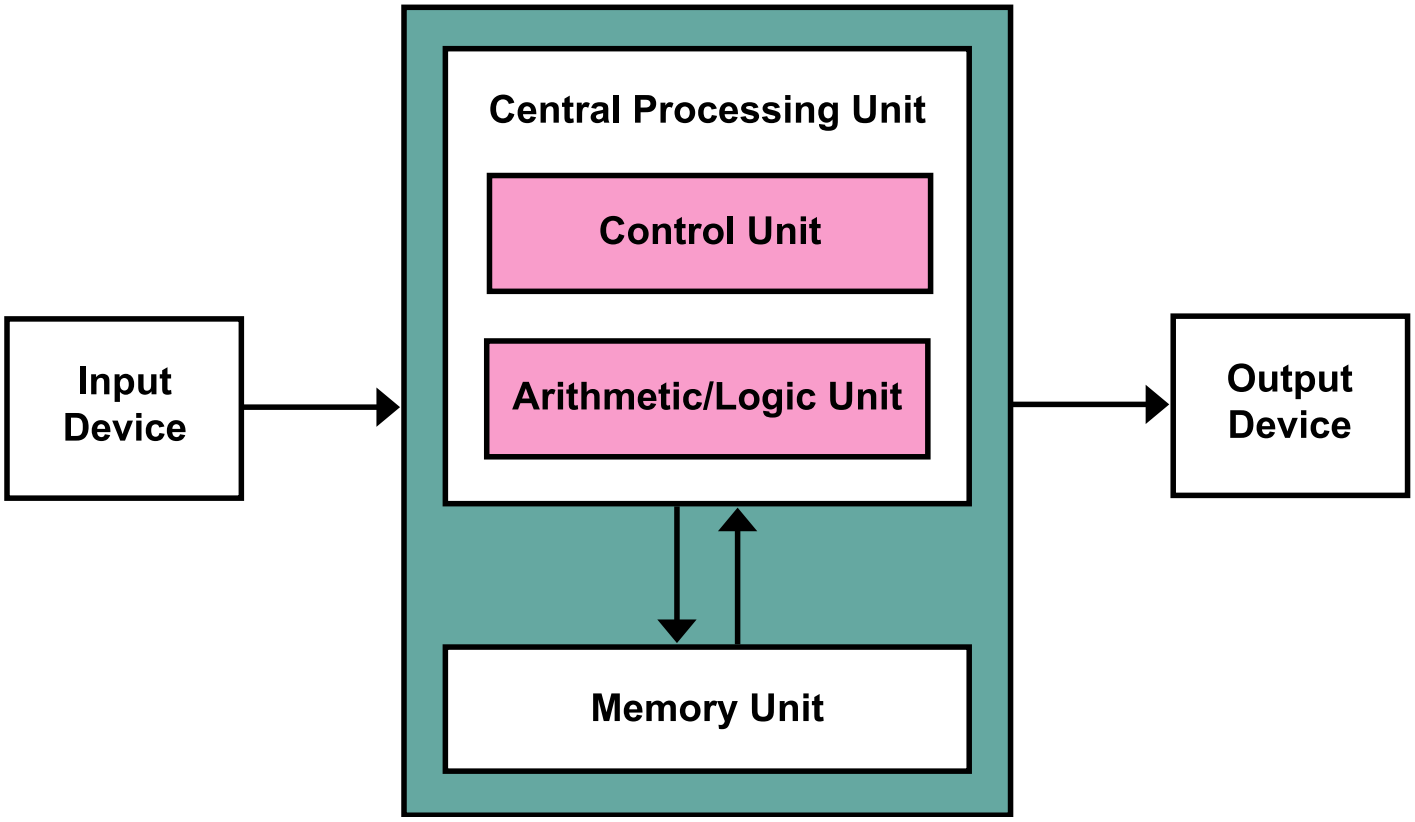
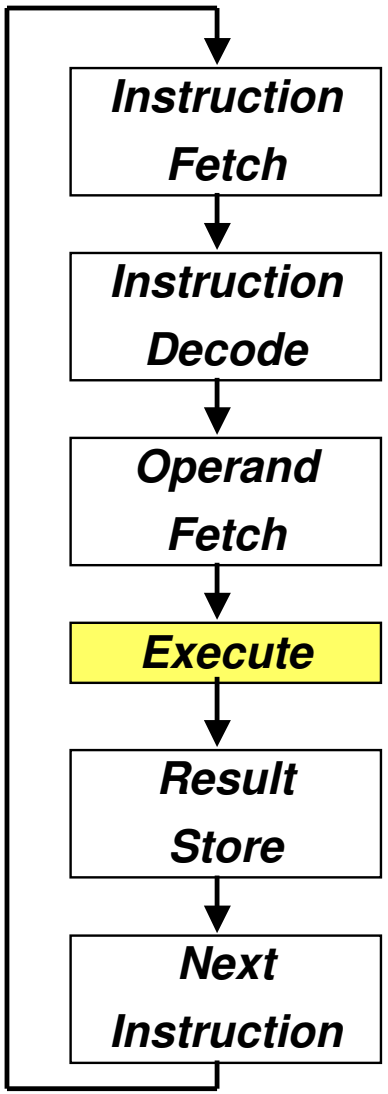


Image: Wikipedia

Execution cycle



Obtain instruction from program storage

Determine required actions and instruction size

Locate and obtain operand data

Compute result value or status

Deposit results in storage for later use

Determine successor instruction

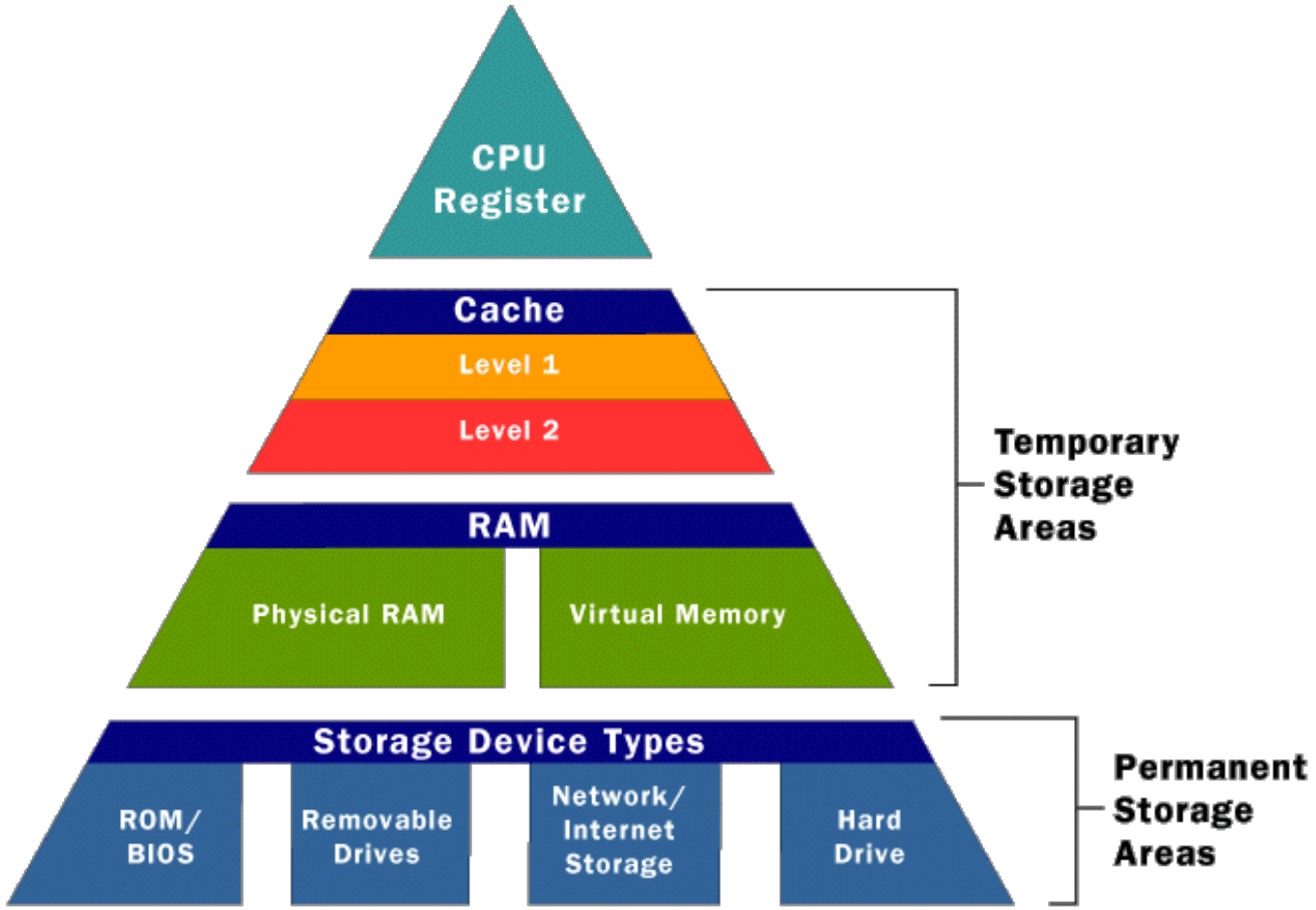
Credit: Based on notes from EC232 UMass – Amherst

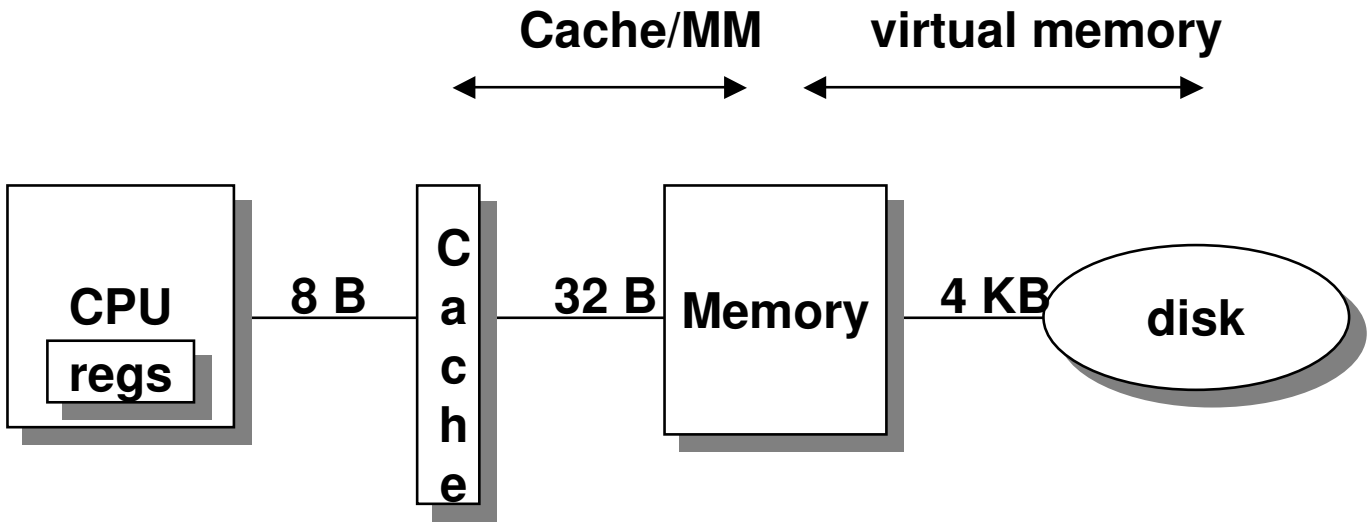
Memory Issues

Memory hierarchy

Faster

Costlier

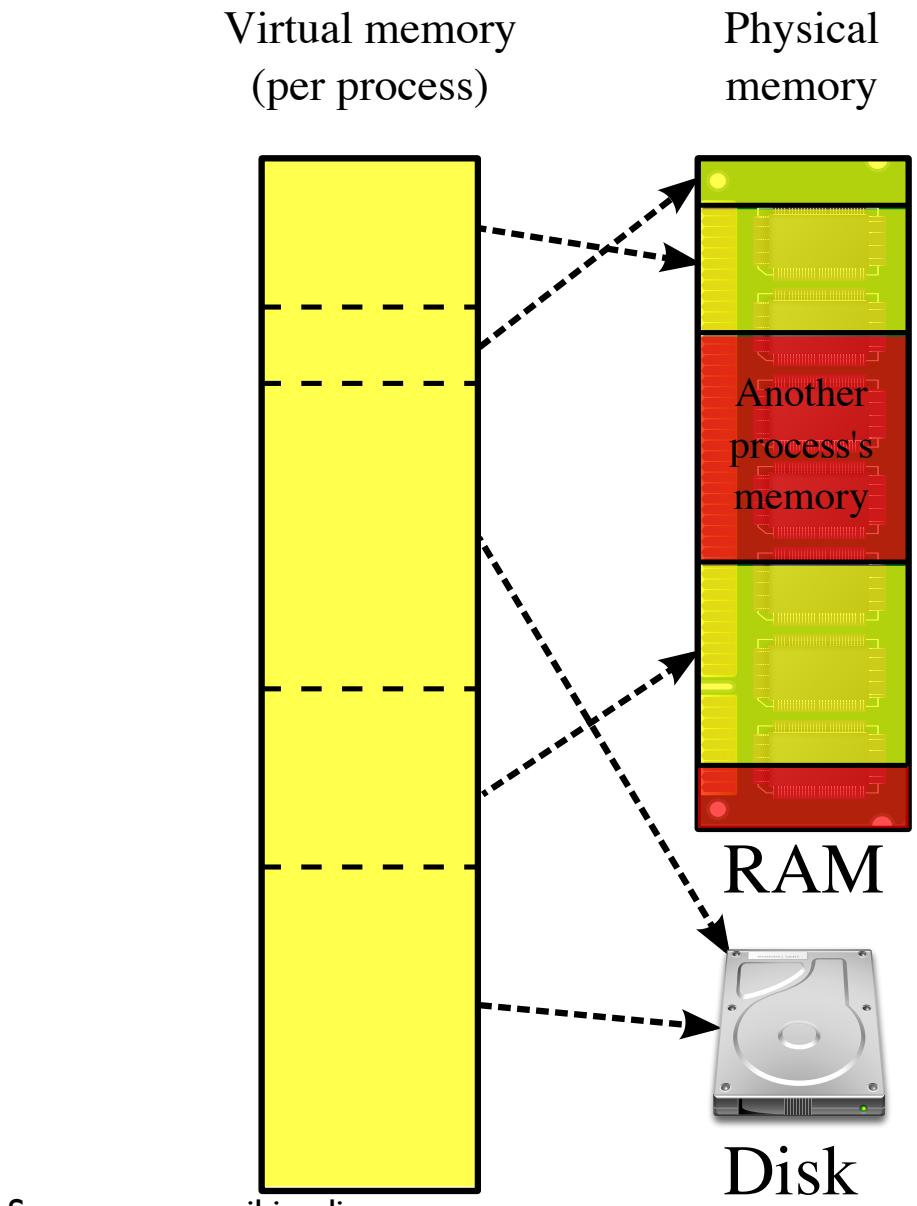




- Notice that the data width is changing
  - Why?
- Bandwidth: Transfer rate between various levels
  - CPU-Cache: 24 GBps
  - Cache-Main: 0.5-6.4GBps
  - Main-Disk: 187MBps (serial ATA/1500)

Source:Ece 232 Umass-Amherst

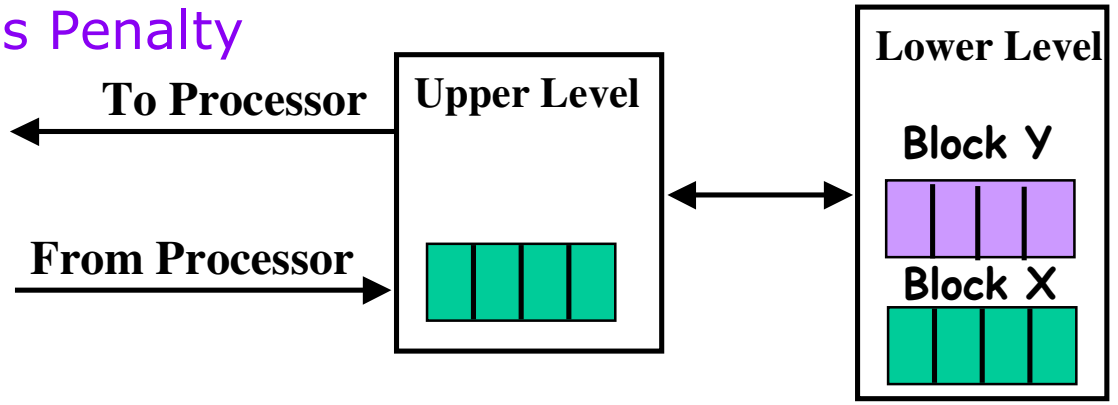
Virtual Memory and Paging



Source: www. wikipedia.com

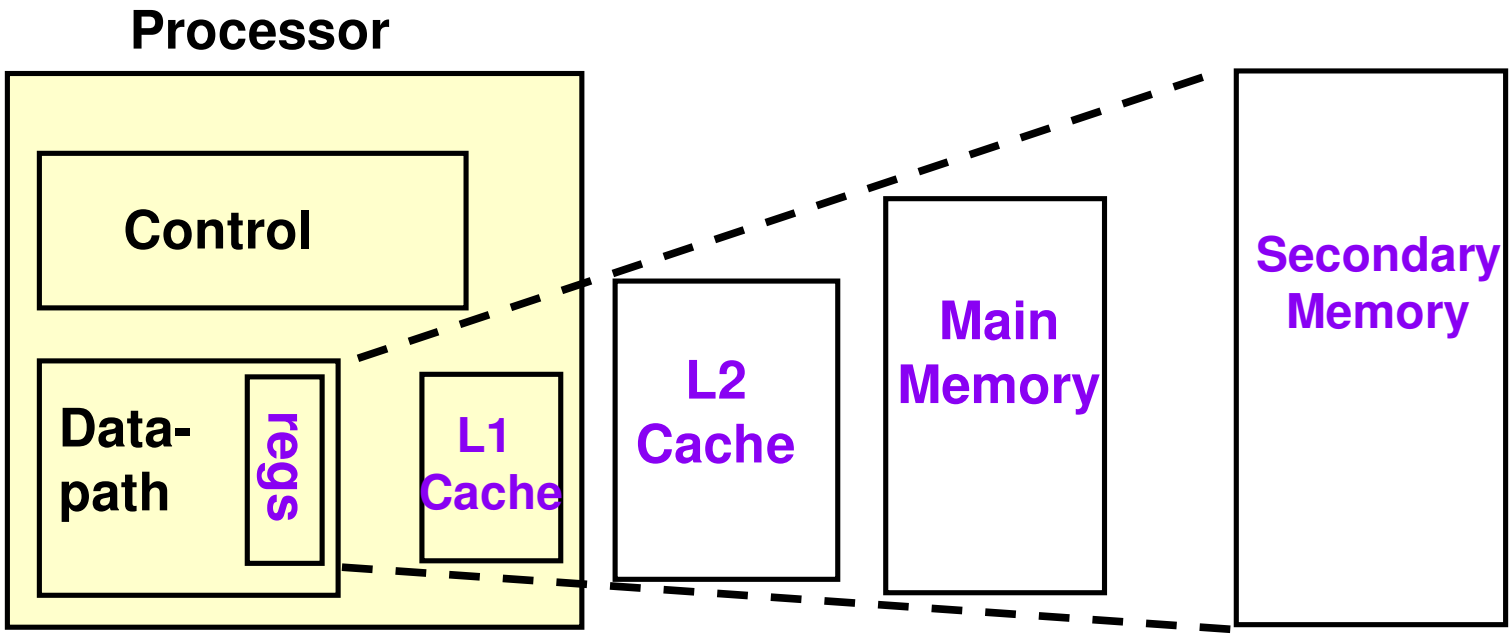
Memory Hierarchy Terminology

- Hit: data appears in upper level in block X
- Hit Rate: the fraction of memory accesses found in the upper level
- Miss: data needs to be retrieved from a block in the lower level (Block Y)
- Miss Rate = 1 - (Hit Rate)
- Hit Time: Time to access the upper level which consists of Time to determine hit/miss + upper level access time
- Miss Penalty: Time to replace a block in the upper level + Time to deliver the block to the processor
- Note: Hit Time << Miss Penalty



Source: ECE 232 Umass-Amherst

Current Memory Hierarchy



Speed(ns):	1ns	2ns	6ns	100ns	10,000,000ns
Size (MB):	0.0005	0.1	1-4	1000-6000	500,000
Cost ($/MB):	--	$10	$3	$0.01	$0.002
Technology:	Regs	SRAM	SRAM	DRAM	Disk

- Cache - Main memory: Speed
- Main memory – Disk (virtual memory): Capacity

Source:Ece 232 Umass-Amherst