Memory Hierarchy

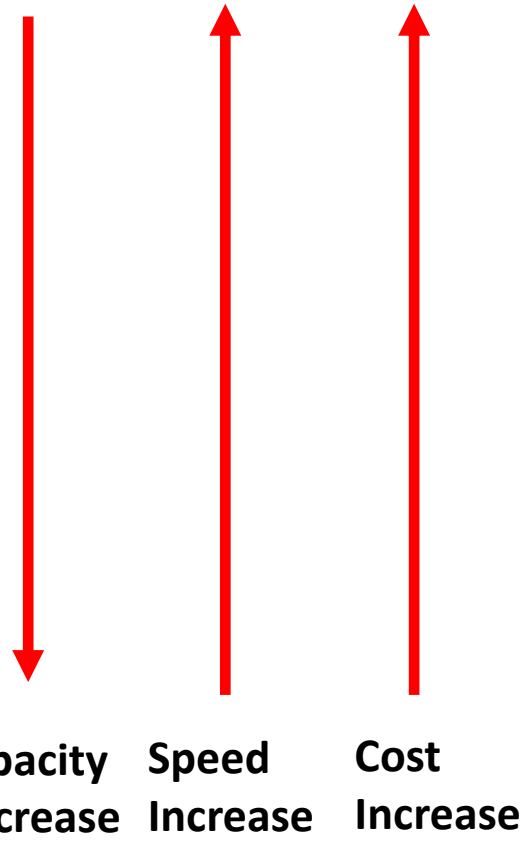
Nahin Ul Sadad
Lecturer
CSE, RUET

Memory Technologies

	Capacity	Latency	Cost/GB
Register	1000s of bits	20 ps	$$$$
SRAM	~10 KB-10 MB	1-10 ns	~$1000
DRAM	~10 GB	80 ns	~$10
Flash (SSD)*	~100 GB	100 us	~$1
Hard disk*	~1 TB	10 ms	~$0.10

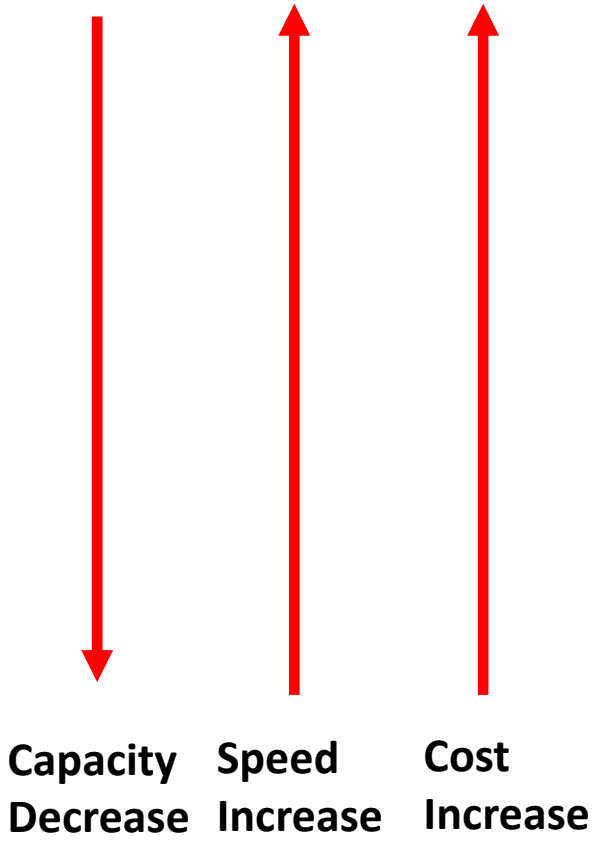
Memory Hierarchy

	Capacity	Latency	Cost/GB
Register	1000s of bits	20 ps	$$$$
SRAM	~10 KB-10 MB	1-10 ns	~$1000
DRAM	~10 GB	80 ns	~$10
Flash (SSD)*	~100 GB	100 us	~$1
Hard disk*	~1 TB	10 ms	~$0.10



Memory Hierarchy

	Capacity	Latency	Cost/GB
Register	1000s of bits	20 ps	$$$$
SRAM	~10 KB-10 MB	1-10 ns	~$1000
DRAM	~10 GB	80 ns	~$10
Flash (SSD)*	~100 GB	100 us	~$1
Hard disk*	~1 TB	10 ms	~$0.10



We want large, fast and cheap memory. But

Large memories are slow!

Fast memories are expensive!

Memory Hierarchy

Idea: Can we use a hierarchal system of memories with different tradeoffs to emulate a large, fast, cheap memory?

Memory Hierarchy

– Programming model: Single memory, single address space

– Machine transparently stores data in fast or slow memory, depending on usage patterns.

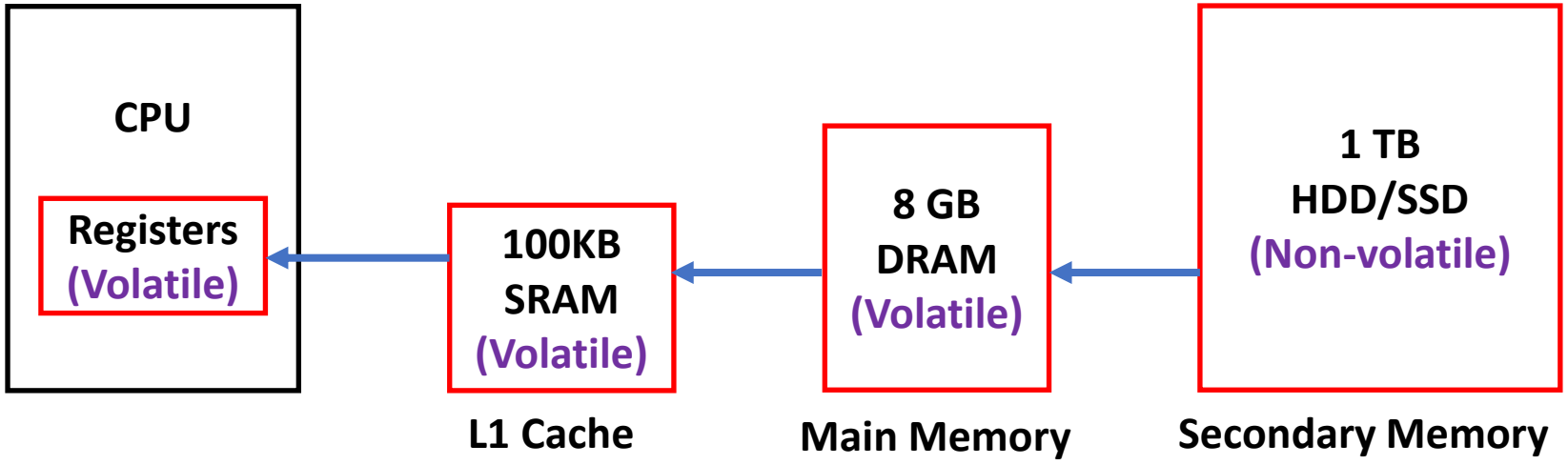
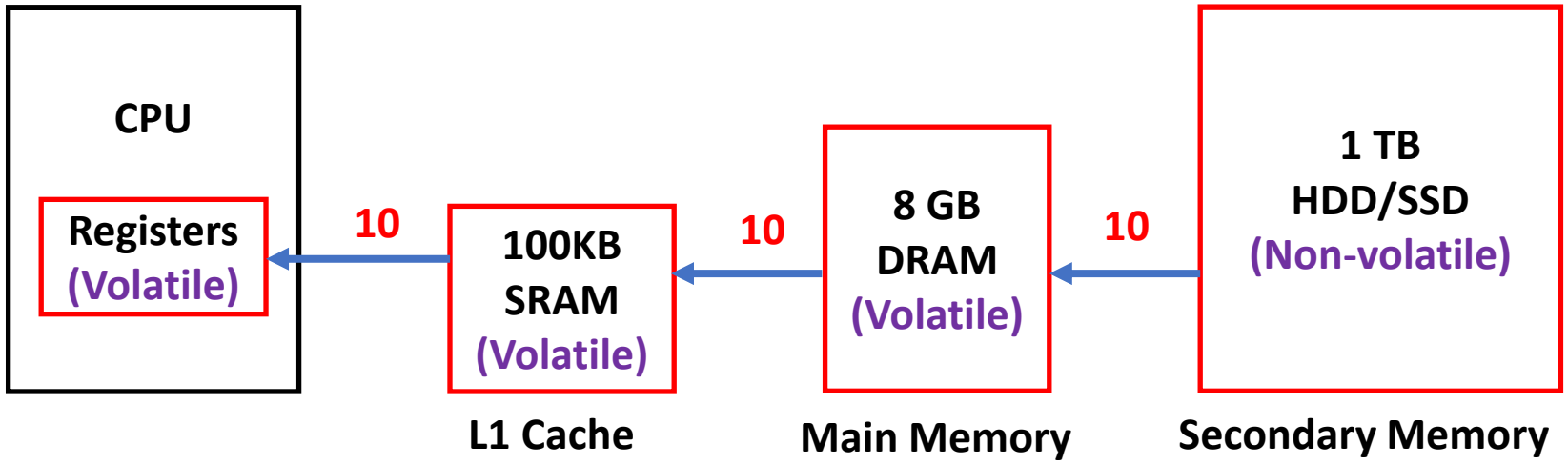


Figure: Memory Hierarchy

Volatile Memory = It only contains data when it is powered ON. It will lose its contents after power off.

Non-volatile Memory = It contains data even after power off.

Memory Hierarchy



Data are stored in Registers before they can be sent to ALU for operation.

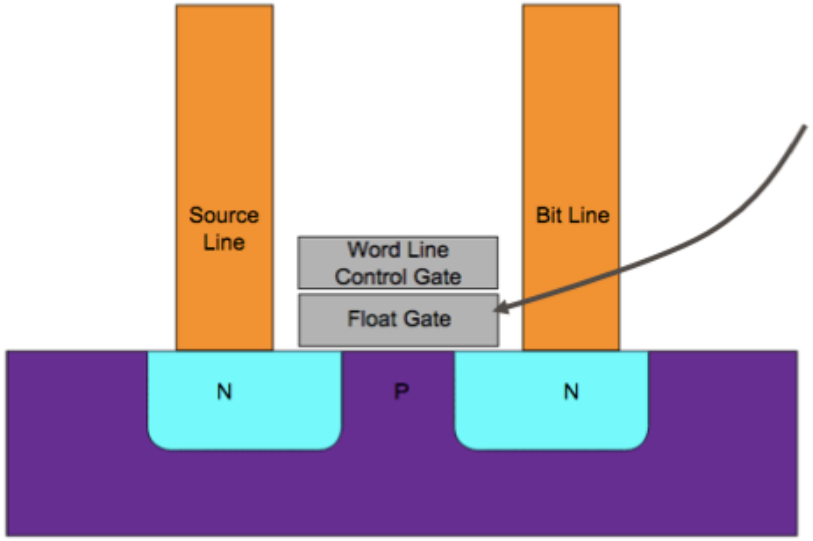
Cache memory stores data that are frequently accessed.

Program's instruction and data are temporarily first loaded into Main Memory before running

OS and other programs are stored in Secondary Memory permanently.

Figure: How Memory Hierarchy works

Non-Volatile Storage: Flash



Electrons here diminish strength of field from control gate ⇒ no inversion ⇒ NFET stays off even when word line is high.

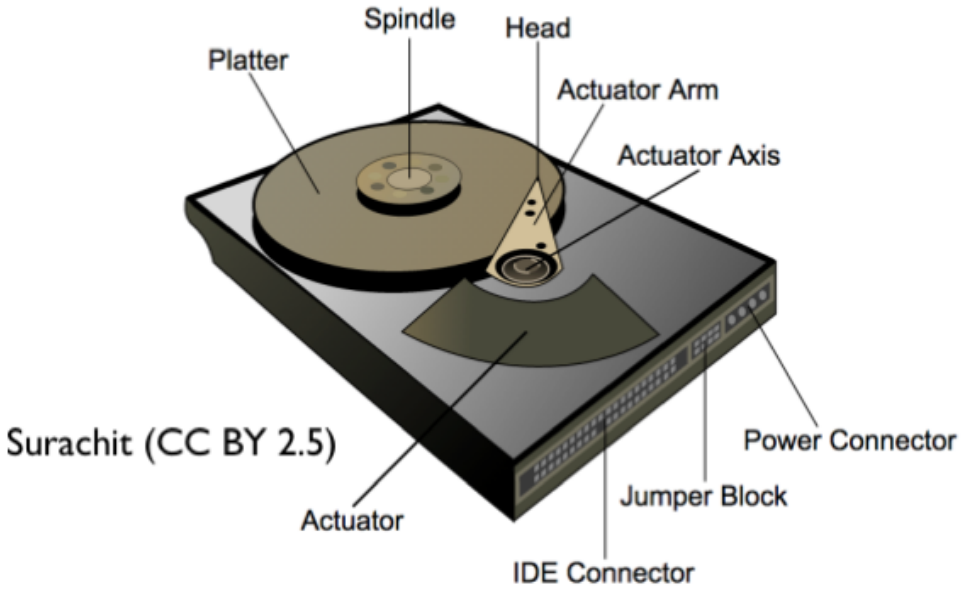
Cyferz (CC BY 2.5)

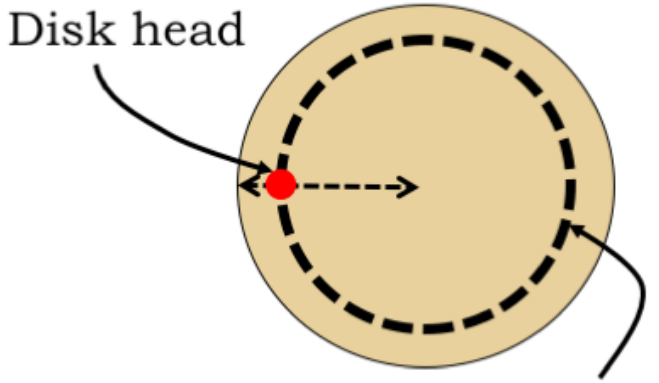
Flash Memory: Use "floating gate" transistors to store charge

- Very dense: Multiple bits/transistor, read and written in blocks
- Slow (especially on writes), 10-100 us
- Limited number of writes: charging/discharging the floating gate (writes) requires large voltages that damage transistor

Examples: Solid State Disk (SSD) + Pen Drive

Non-Volatile Storage: Hard Disk





Circular track divided into sectors

Hard Disk: Rotating magnetic platters + read/write head

- Extremely slow (~10ms): Mechanically move head to position, wait for data to pass underneath head
- ~100MB/s for sequential read/writes
- ~100KB/s for random read/writes
- Cheap

Examples: Hard Drive Disk (HDD)

Thank You ☺