Lecture 18: Memory Hierarchy

Wednesday, February 20, 2019 10:59 AM

Outline

- more memory tech
- hierarchy
- locality
- Data movement

nop? → assembler pseudo op

in RISC-V → addi x0, zero, zero, 0 (0x00000013)

```
0:  addi zero, zero, 0      0:  nop     1
4:  addi                    4:  nop     2
8:  addi                    8:  nop     3
12: addi                    12: nop     4
                         → 16: auipc x5, 0
```

"auipc 0"
↳ 0: auipc x5, 0

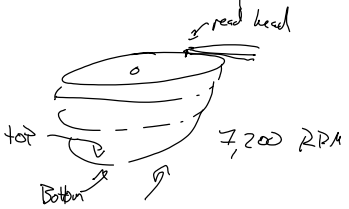
"auipc 1"
↳ 0: nop
4: auipc x5, 0

	SRAM	DRAM	Magnetic disks	flash	new tech
Latency	low (1 cycle) ~1ns	medium 15-45ns	~10ms high latency	~10µs high	~10ns for reads ~1µs for writes
Density (mem per bit ↳cost)	low 10's MB	high 100's GB	very high 10s TBs	very high 10s TBs	very high 1-2TBs
Volatility	Volatile	Volatile	non volatile (storage)	non volatile	non volatile

Magnetic disks

Storage technology
↳ Very high density
↳ Store information @ rear atomic level
↳ many platters

Latency
↳ high (very) → moving parts



10,000 RPM
↳ 167 RPS
1/167 = 0.005 s
5 ms

Solid State "disks"

Flash → floating gate transistor
↳ 1 transistor store data → store multiple bits per transistor

high density
↳ per volume → higher than magnetic disk
↳ cost per bit higher than magnetic disks

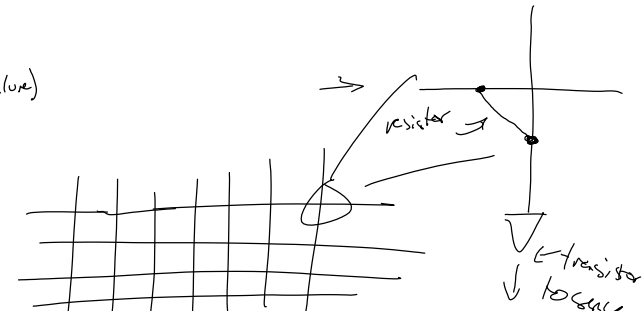
Latency
↳ lower than spinning disk
↳ higher than DRAM/SRAM (Read)
↳ write latency is higher than read

write endurance is low (10s - 100s of writes before failure)

Other persistent memory technologies

"3D X point" → optane

PCM → phase change memory →

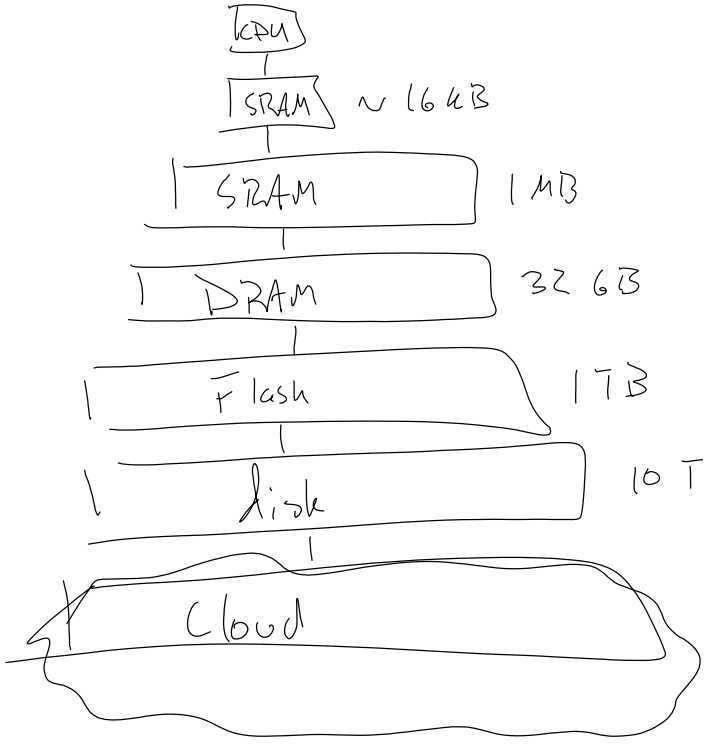


very high density
Latencies ↳ ~ flash

↳ write → high latency (~1μs) ⟷ asymmetric read/write latency
read → faster than DRAM

DINO cpu assumes <1 cycle memory
↳ SRAM → ~100 kB

What can we do to have TBs of data but access it in us
Build system w/ combination of memory technologies



hierarchy of diff. tech