Lecture 17: Memory Technology

Friday, February 15, 2019 10:57 AM

Outline

- Performance with memory
- Memory technologies
- Hierarchy

50 GB/s memory bandwidth

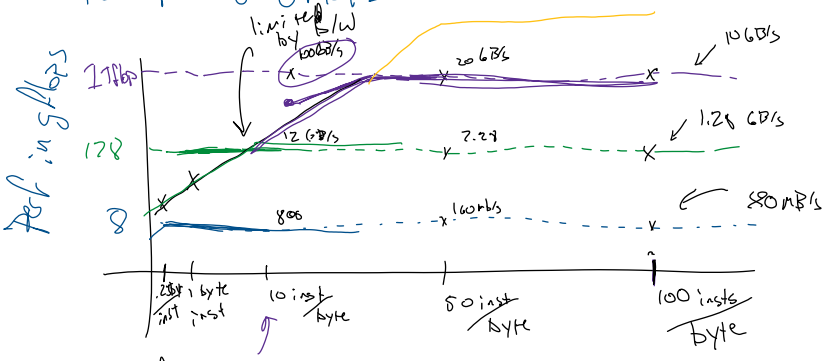
Out of order CPU w/ 2 floating point units @ 4 GHz → 16 core machine

Peak perf: 8 GFlops

128 GFlops
↳ 2 units per core
16 cores
4 billion cycles

8 single precision ops in parallel
256 bit vector instructions
↳ ~1000 gflops
1 true flop

2 units
x 8 flop per unit
x 16 cores
x 4 billion cycles



Roofline model

$\frac{8\ 10^9\ \text{inst/s}}{100\ \text{inst/byte}} = 8 \cdot 10^7\ \text{byte/s}$

80 MB/s

$\frac{1 \cdot 10^{12}}{100} = 1.10$

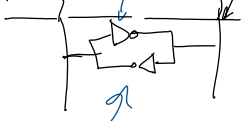
Memory technologies

What characteristics do we care about?

- price ($/bit)
  - bandwidth → packaging
- capacity
  - density area/bit
- latency
- read only vs read write
  - ↳ latency of writes vs reads
- power / energy
- reliability
- destructive reads
- granularity of reads and writes
- volatility

SRAM

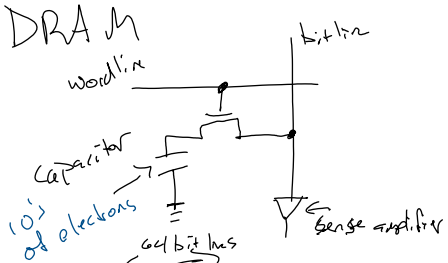
↳ Static RAM → 6-12 transistors



a lot of static power

price $/bit → expensive
density → not dense → big
latency → very fast
1 cycle
power → high
volatile

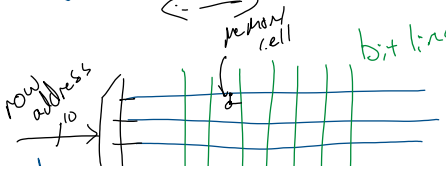
DRAM

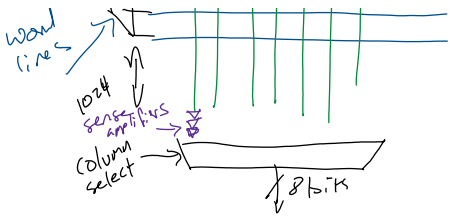


much cheaper than SRAM
denser "

slower than SRAM + higher latency
Volatile → destructive reads
↳ write after reads

64k bit RAM




word lines
1024
sense amplifiers
column select
8 bits