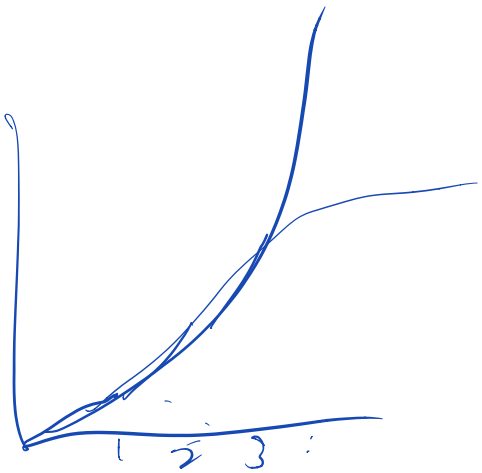
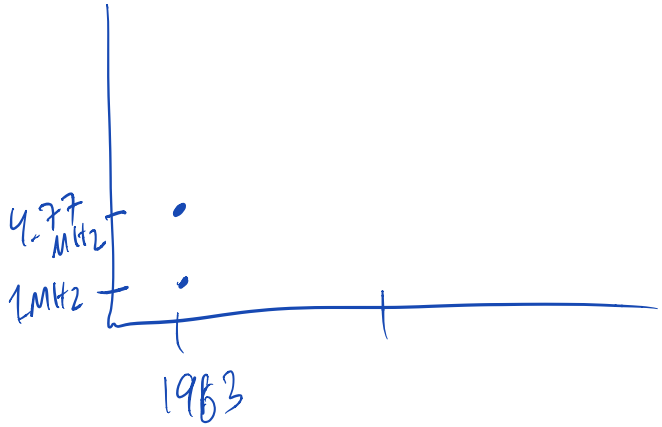
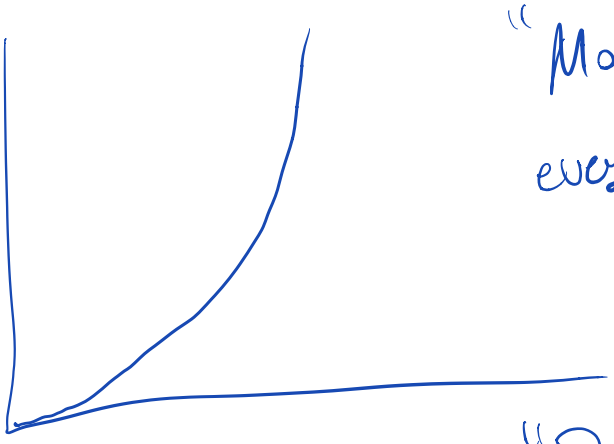
"cores"



20 MHz



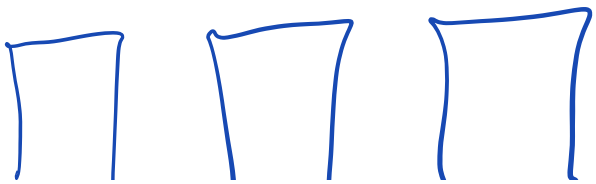


"Moore's Law"

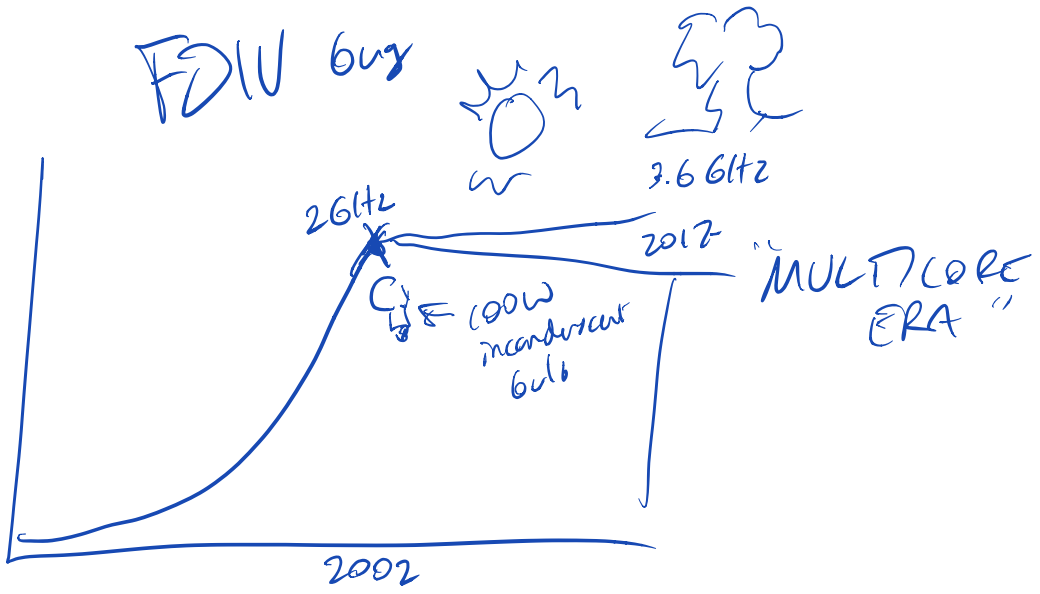
every 18 mos.
# of transistors on chip doubles

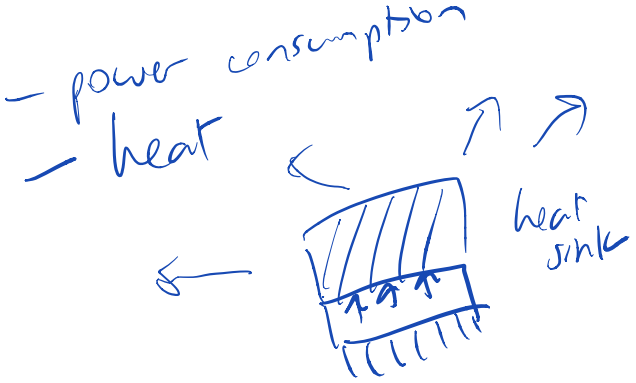
"Dennard scaling"

clock speed scales linearly w/ # of transistors



Synchronous



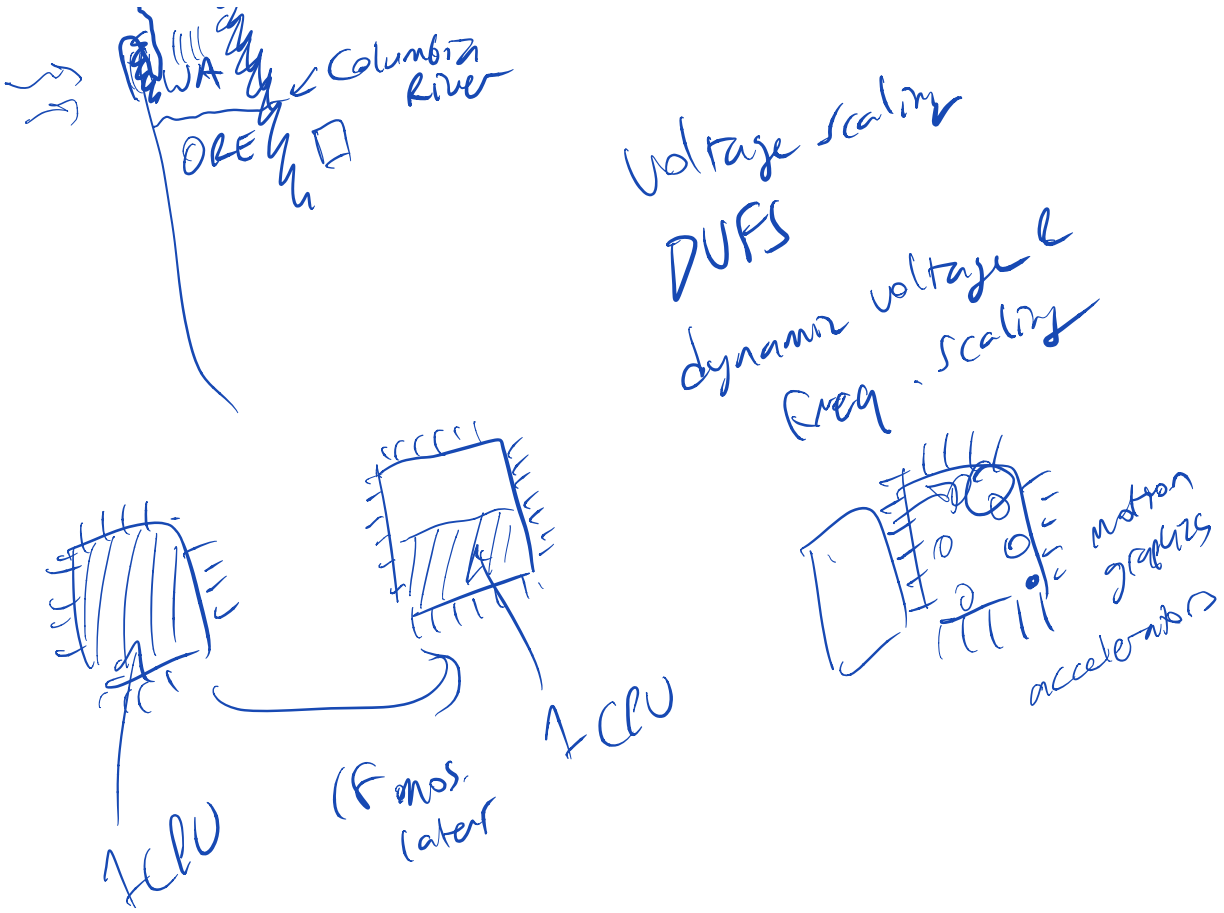


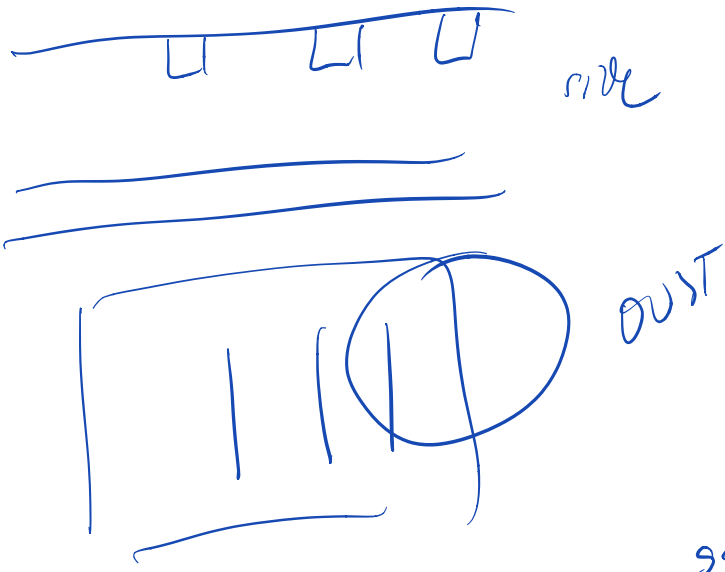
COOLING

FB, Amazon, Google

data center

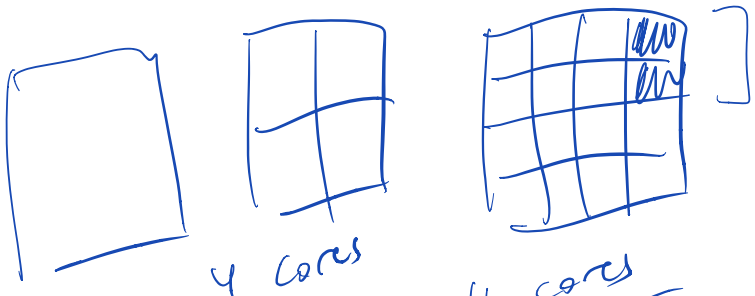
- Cascades





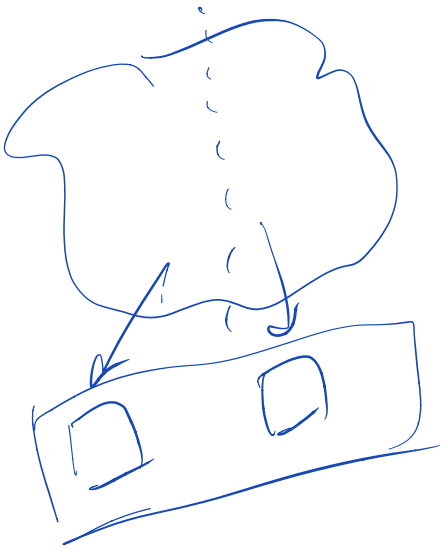
"MULTICORE ERA"

yield $\frac{\text{good}}{\text{total}} \approx 85\%$



2GHz

same heat dissipation & same power



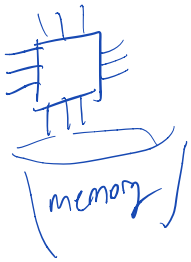
2 cores ⇒

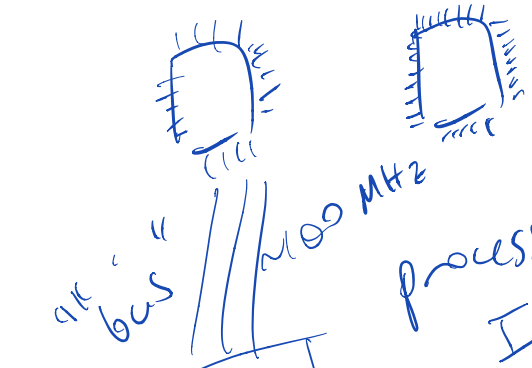
2x speedup

"linear speedup"

one core heat diss. ≈ 2 cores w/ same clock speed

straws ≈ pins





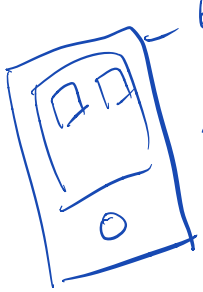
processes:

I/O
reading from or writing memory

CPU compute
addition
multiplication
vector ops
crypto

bus saturation

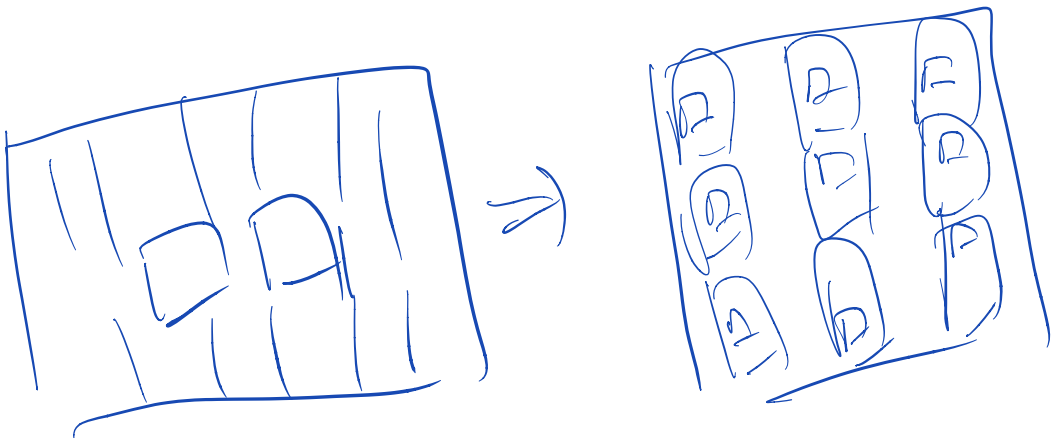
harnessing cores?
- multithreading
- multiprocessing



background apps
→ ~~music~~ video
→ email
→ location
→ GC
→ grammar checking
→ voice recognition
→ OS
→ antivirus
→ UI
→ web server
→ JIT compiler

~~I~~ RAM
intelligent RAM
Patterson
Kozyrakis

Dave Christos



Systems-for-data-science.cs.umass.edu