Input/Output

I/O is fundamental to computing -- no computer solves a problem for its own entertainment, so getting results out is a minimum requirement ... and presenting input to be computed on is also useful. Nevertheless, I/O is a peripheral topic.

(c) Copyright Larry Snyder 1998

The Problem with Miss-rate

It doesn't tell the whole story:

Consider increasing direct-mapped cache from 32K to 64K.

Miss Rate drops from 5% to 4%. If the larger cache implies a cycle time of 18ns and the smaller cache implies a cycle time of 15ns, the smaller cache machine has better perfornace.

Postulate: CPI w/o stalls is unchanged
Miss penalty 180ns
Memory references per instruction = 1.5

CPU Time = (CPU execution clock cycles + Memory-stall clock cycles) $\times$ Clock cycle time

(c) Copyright Larry Snyder 1998

Cache Analysis, Continued

$$\text{Memory-stall clock cycles} = \frac{\text{Instructions}}{\text{Program}} \times \frac{\text{Misses}}{\text{Instruction}} \times \text{Miss penalty}$$

$$\text{Misses} = \text{Instruction miss rate} + \text{Data miss rate} \times \frac{\text{Data references}}{\text{Instructions}}$$

Let IC be *instructions per program*

Smaller Cache	Larger Cache
Memory stall clock cycles = $IC \times (0.05 + 0.05 \times 0.5) \times \frac{\text{Absolute miss penalty}}{\text{Clockcycle time}}$	Memory stall clock cycles = $IC \times (0.04 + 0.04 \times 0.5) \times \frac{\text{Absolute miss penalty}}{\text{Clockcycle time}}$
$= IC \times 0.075 \times 180/15 = .9IC$	$= IC \times 0.06 \times 180/18 = .6IC$

(c) Copyright Larry Snyder 1998

Cache Analysis, Continued

Memory-stall clock cycles = 0.9IC (Small) and 0.6IC (Large cache).

Substituting into the CPU time equation, letting CPI w/o stalls be C:

CPU Time = (CPU execution clock cycles + Memory-stalls clock cycles) $\times$ Clock cycle time

Small Cache	Large Cache
CPU time	CPU time
$=((C \times IC) + (0.9 \times IC) \times 15ns$	$=((C \times IC) + (0.6 \times IC) \times 18ns$
$= 15 \times C \times IC + 13.5 \times IC$	$= 18 \times C \times IC + 10.8 \times IC$
$= (15C + 13.5)IC$	$= (18C + 10.8)IC$

For $C \geq 1$ the smaller cache is better

(c) Copyright Larry Snyder 1998

The Scale of Input/Output

Device	Type	Partner	Data Rate (KB/sec)
Keyboard	Input	Human	0.01
Laser Printer	Output	Human	100.00
Graphics Disp	Output	Human	30,000.00
Floppy	Store	Machine	50.00
Tape	Store	Machine	2,000.00

Partners often determine data rate

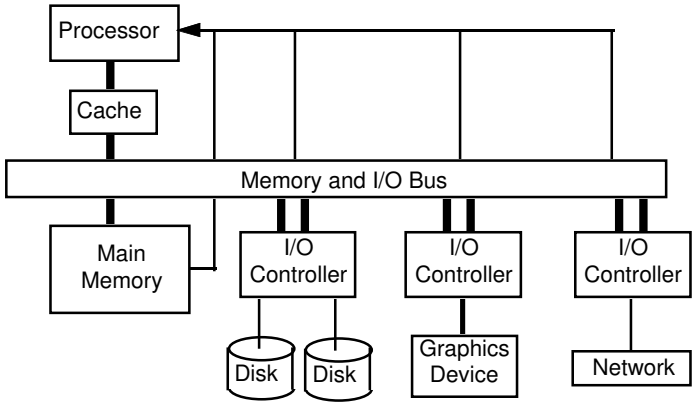
Years	CPU Time	I/O Time	Elapsed Time	%I/O
0	90 Seconds	10 Seconds	100 Seconds	10%
1	90/1.5=60	10 Seconds	70 Seconds	14%
2	60/1.5=40	10 Seconds	50 Seconds	20%
3	40/1.5=27	10 Seconds	37 Seconds	27%
4	27/1.5=18	10 Seconds	28 Seconds	36%
5	18/1.5=12	10 Seconds	22 Seconds	45%

Improving processor performance
at rate of 1.5/year

(c) Copyright Larry Snyder 1998

Typical Architecture

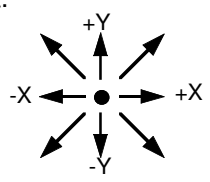
Only interrupts come to the processor directly



(c) Copyright Larry Snyder 1998

Mouse

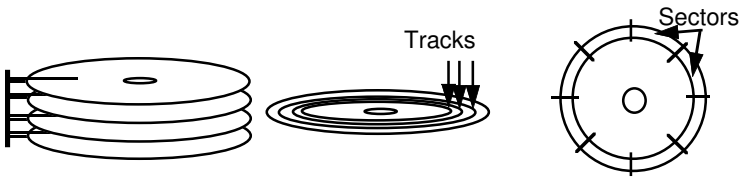
- The physical device has counters that record motion of the ball in the +X, -X, +Y and - Y directions.
- The counter values as well as the button positions are readable by the processor.
- How does the processor know that the mouse has moved? It looks, using a process called polling, or it waits for an interrupt.
    - Polling: A periodic check of status information to detect an unpredictable event.



(c) Copyright Larry Snyder 1998

Disks

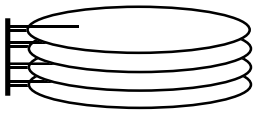
- Floppies are cheap and mobile.
- A hard disk has more favorable properties:
  – It can be larger because it is rigid.
  – It can store more because it can be more precisely controlled.
  – It can transfer data faster because it can spin faster.
  – It can incorporate more than one platter.



(c) Copyright Larry Snyder 1998

Disk Terminology

- Seek time -- time to move the arm so the read head is positioned on a specific track. 12-20 ms average
- Performance measures:
  - Min seek time, Max seek time, Average seek time.
  - Average seek time: Sum of all seeks/number seeks.
  - Observed average is lower than reported average due to consecutive reading and OS efforts at to locality.
- Rotational latency -- delay before the proper sector moves under the disk read head. 0.5 of revolution 3600-5400 RPM
- Transfer rate -- MB/sec rate at which a sector is read from disk and transferred to system bus. 2-4MB/sec



(c) Copyright Larry Snyder 1998

Networks

- RS-232 is a 0.3-19.2Kb/sec. standard for attaching terminals to computers.
- Local Area Network (LAN) or Ethernet is a 10Mb/sec packetized interconnect technology.
- Wide Area Networks typified by ARPANet use 56Kb telephone lines and store-and-forward packetization for long-haul communication.
  - Newer technologies, e.g. fiber optics, can raise transfer rate to 1 Gb/sec.

(c) Copyright Larry Snyder 1998