08-Memory

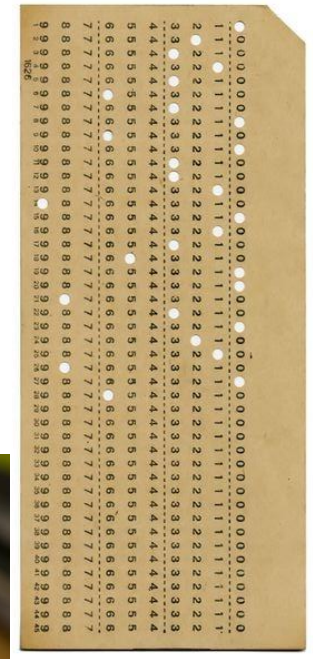
Ahmad Abdullah, PhD
abdullah@uvic.ca
https://web.uvic.ca/~abdullah/csc230

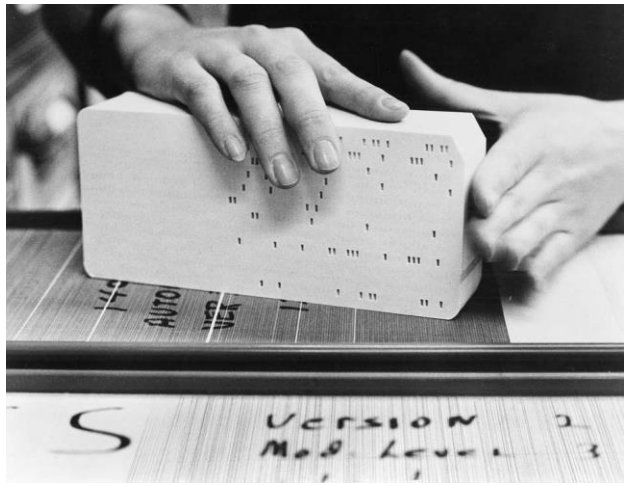
Lectures: MR 10:00 – 11:20 am
Location: ECS 125

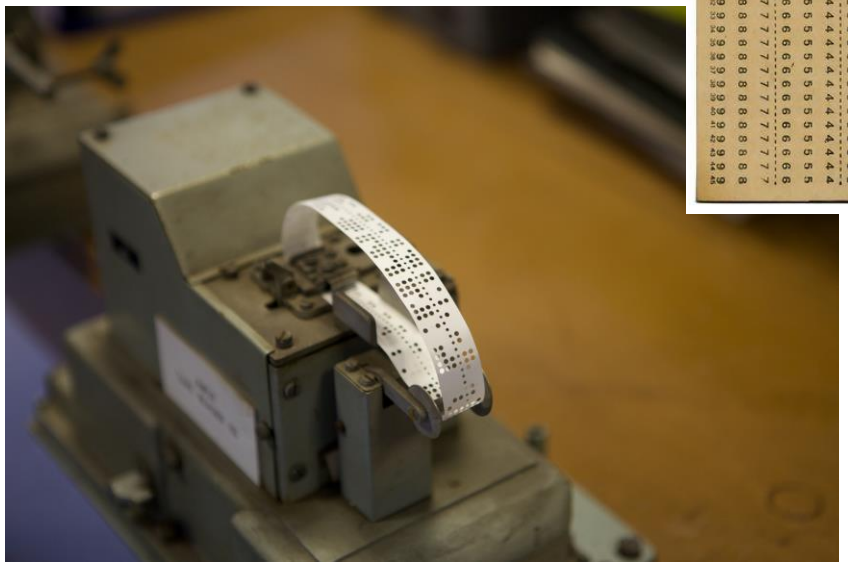
Outline

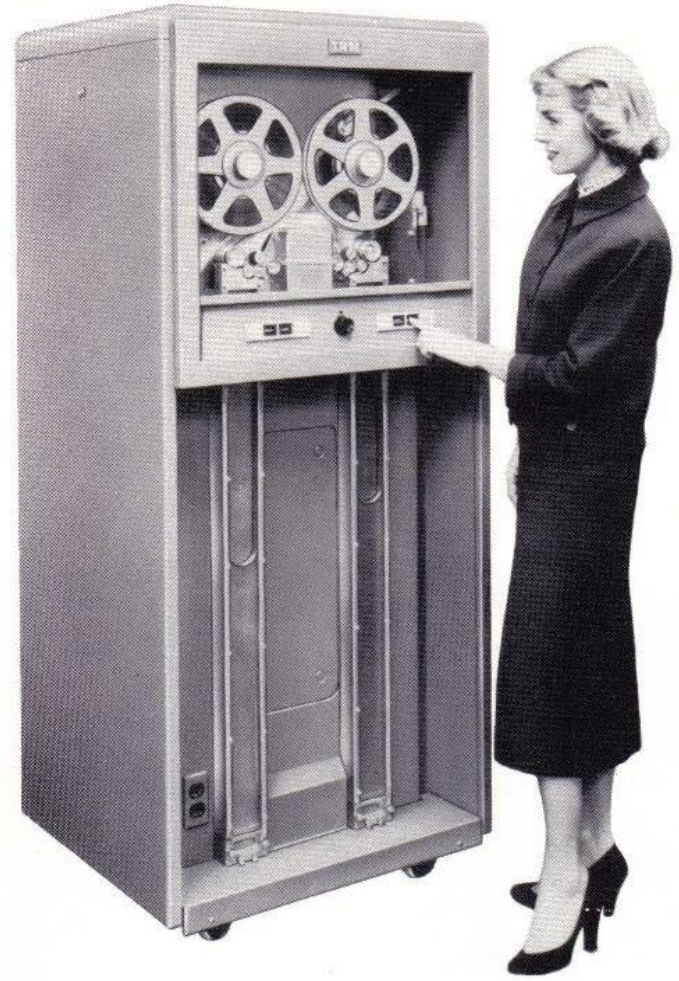
- History
- Memory Types
- Technology
- Memory Hierarchy
- Memory System Characteristics

History







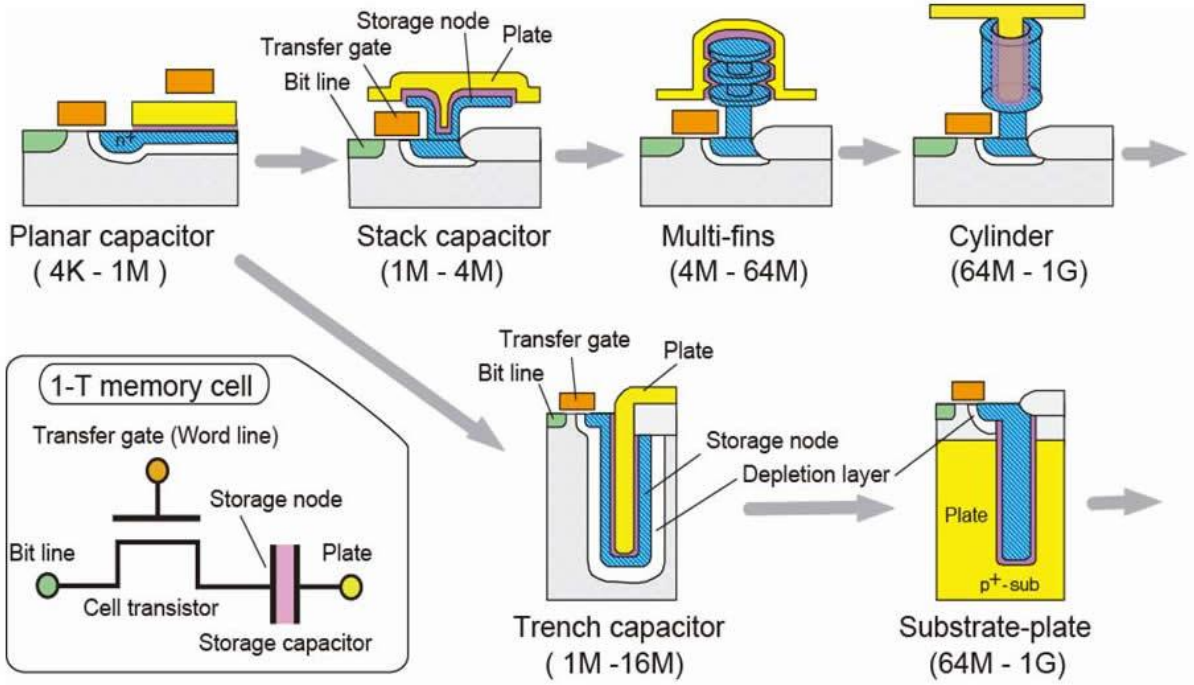


Memory Types

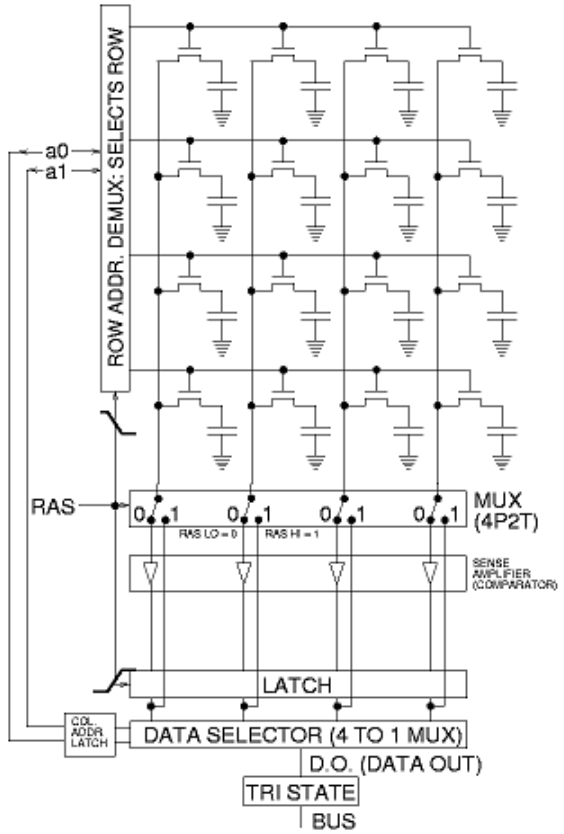
- Volatile:
  - RAM
    - Dynamic RAM
    - Static RAM
- Non-volatile:
  - ROM, Disk, Flash Memory

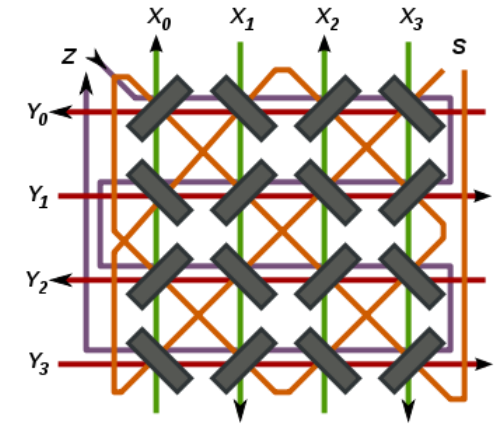
Dynamic Memory

- Using dynamic electrical elements such as inductors or capacitors.
  - Time constant: $\tau = RC = \frac{L}{R}$



https://electronics.stackexchange.com/questions/306002/why-does-a-dram-cell-necessarily-contain-a-capacitor

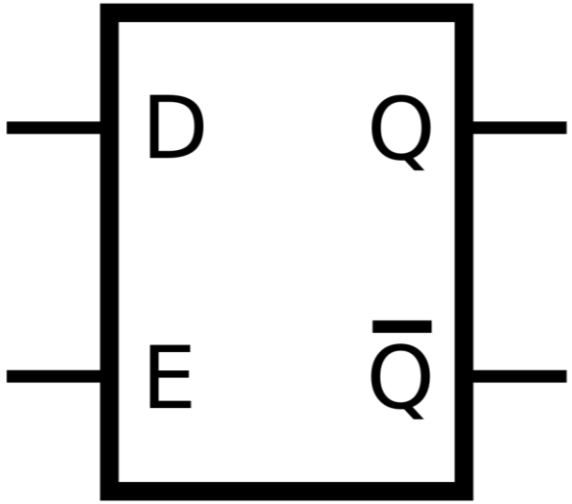




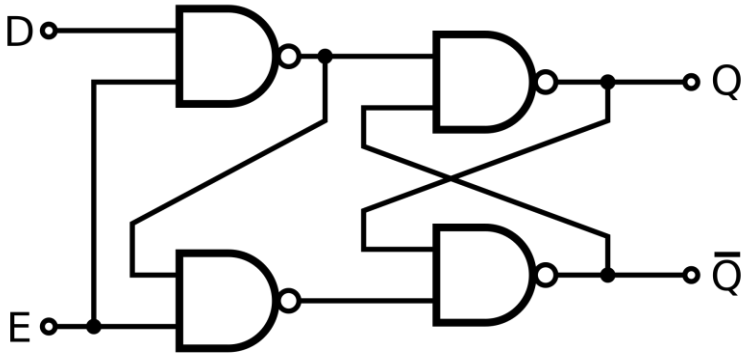
https://en.wikipedia.org/wiki/Magnetic-core_memory

https://commons.wikimedia.org/wiki/File:Magnetic-core_memory.JPG

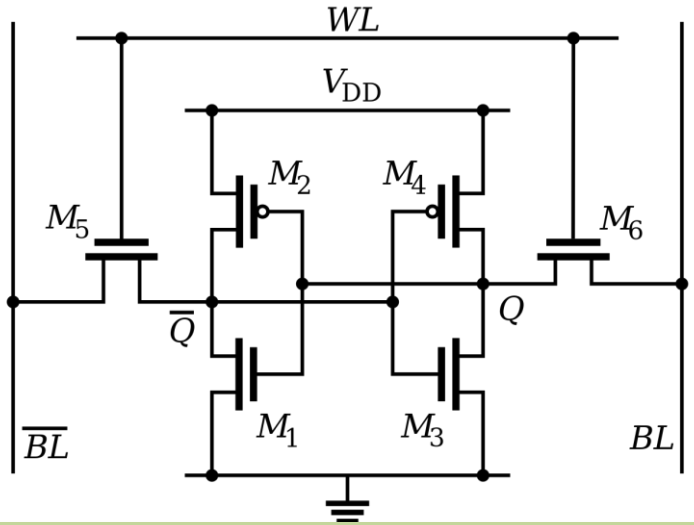
Static Memory



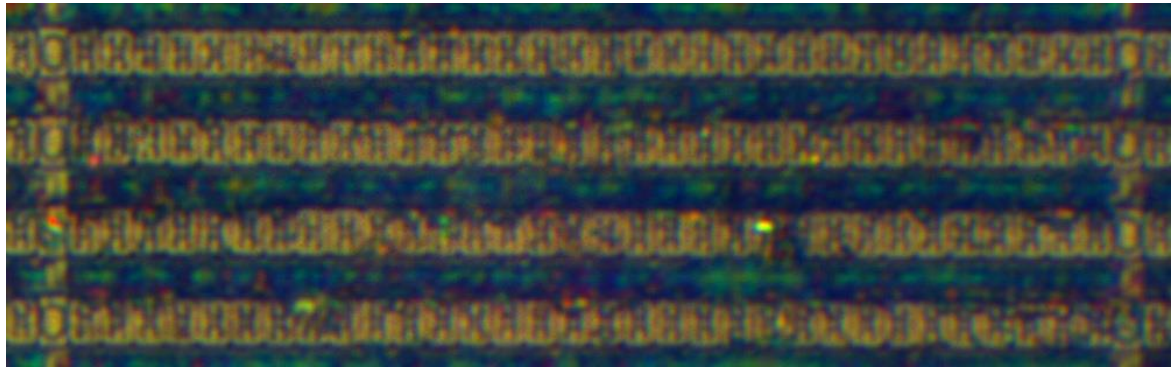
Each bit is stored in one D-FlipFlop



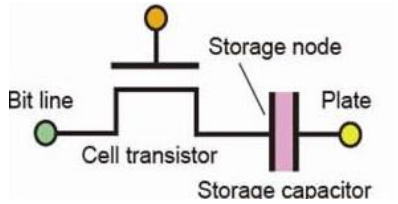
Each D-FlipFlop consists of 4 NAN Gates



Each NAN Gates consists of 6 transistors



SRAM cells on a STM32F103VGT6 microcontroller as seen by an optical microscope



Compare to one dynamic memory cell!

Memory Hierarchy

- The constrains a computer-system designer consider:
  - How much memory of a certain type can be obtained (capacity)?
  - How fast is that memory (access time)?
  - How expensive is it (cost per bit)
- And there are tradeoffs
  - Faster access time → greater cost per bit
  - Greater capacity → smaller cost per bit
  - Greater capacity → slower access time
- This is a dilemma!
  - And the only way out is to not rely on a single memory component or technology
  - Rather, we have a hierarchy of memory components which we exploit

Memory Hierarchy (Cont.)

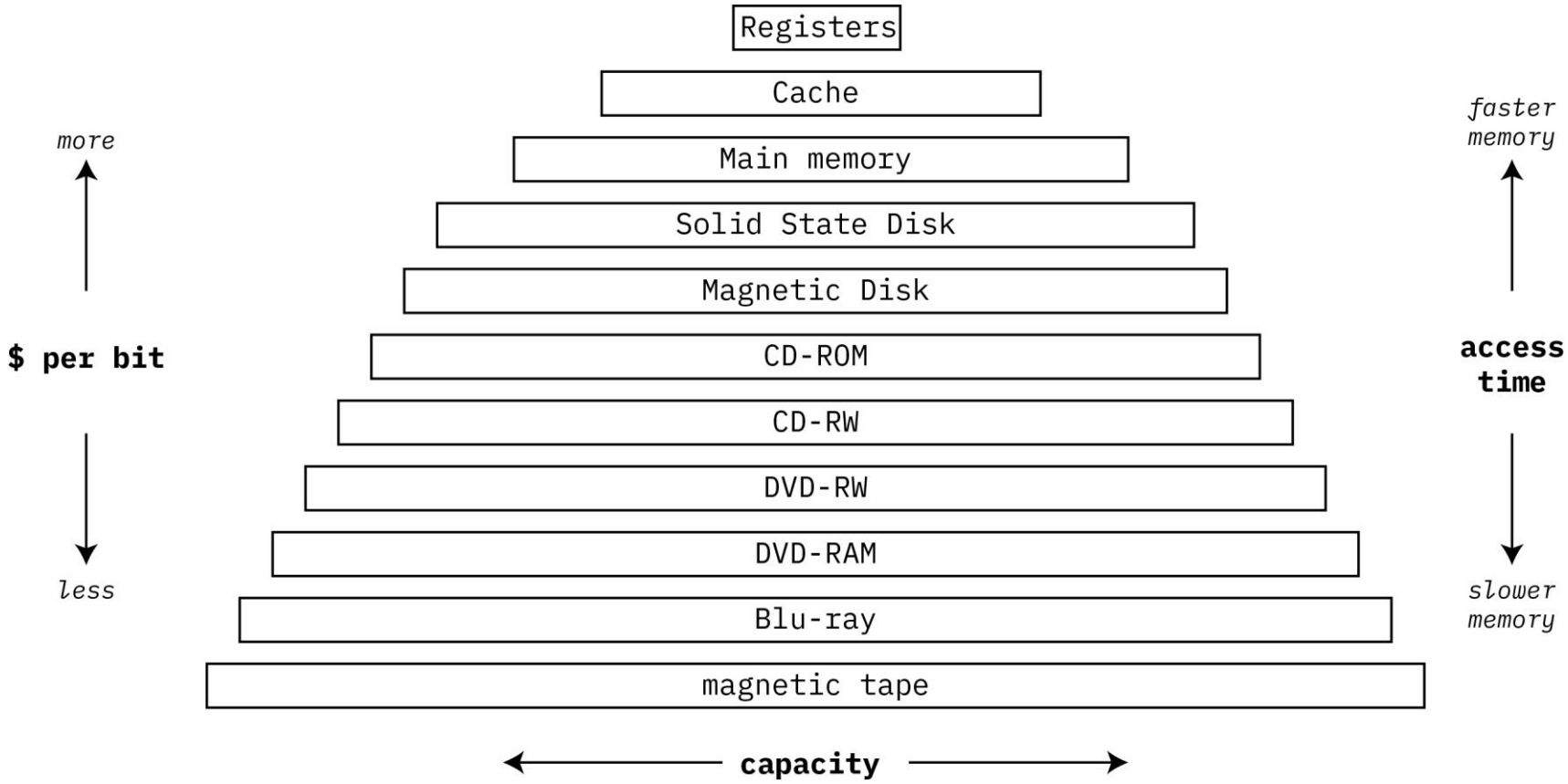


Diagram is definitely not to scale!

Memory Hierarchy (Cont.)

- As we go down the hierarchy, we see that:
  a) cost per bit decreases
  b) capacity increases
  c) access time **increases** (i.e., slower)
  d) decreasing frequency of access of the memory by the processor
- Idea:
  - small, more expensive, faster memories...
  - ... are supplemented by larger, cheaper, slower memories.
- We need to explore, however, the meaning of d) above...

Memory system characteristics

1. location
2. capacity
3. unit of transfer
4. access method
5. performance
6. physical type
7. physical characteristics
8. organization

In the following, we'll organize our coverage by using eight key characteristics to describe memory systems

1. Location

- Is the memory internal to the computer ...
- ... or external?
- Main memory is internal...
  - ... where “internal” refers more to the nature of the address/data buses used to read/write such memory.
  - ... but there may be other forms of internal memory
- External storage:
  - usually peripheral storage (disk, SSD, tape)
  - normally accessed by the CPU via I/O controllers

2. Capacity

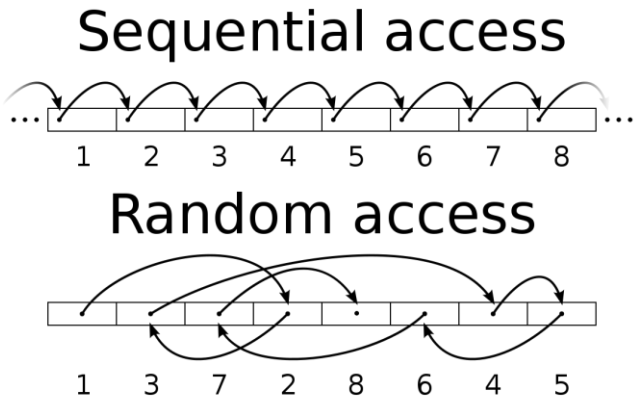
- A simple statement of either:
  - number of words available in the memory device
  - number of bytes
- Note that **word** has different definitions
  - For our course, a AVR words are two bytes (16 bits)
  - In other architectures (i.e., MIPS32 or RISC-V), a word is four bytes (32 bits)
  - Some earlier computers had a variable word size!

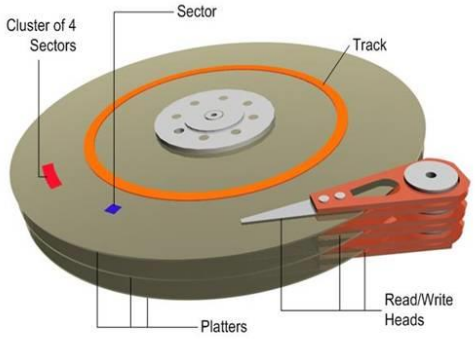
3. Unit of Transfer

- Determined largely by:
  - Number of electrical lines leading into and out of memory module
- Related concept: **addressable units**
  - Unit of memory addressed at level of byte? of word? or something even bigger?
- For internal memory:
  - Number of bits read out of / written into memory at one time
- For external memory:
  - Data transferred in units larger than a word (sometimes called **blocks**)

4. Access method

- Sequential access:
  - Access is made in a specific linear sequence
  - Usually consequence of physical medium
- Random access:
  - Each memory location can be accessed independent of others at no extra cost (time & energy)
- Direct access:
  - Location of memory block determined by location
  - Searching, counting, waiting used to reach final location
- Associative
  - Data value retrieved based on the contents of some key to which the value is mapped.





5. Performance

Three parameters are used: **access time, memory-cycle time, transfer rate**

a) **access time** (latency)
   - for **random access**: duration of time from the **instant** an address is presented to the **instant** the data is available for us
   - for other memory types: time to position read/write mechanism at needed location

b) **memory cycle time**
   - Access time...
   - ... plus additional time for electrical transients to die out on signal lines
   - This is almost exclusively concerned with the system bus

5. Performance (Cont.)

c) transfer rate

- Rate at which memory can be transferred into or out of a memory unit

$$T_n = T_A + \frac{n}{R}$$

- T_n: average time to read or write n bits
- T_A: average access time
- n: number of bits
- R: transfer rate in bits per second (bps)

For a hard drive,

- TA = time to position arm over correct track, rotational latency for sector to arrive under read/write head
- R = number of bytes transferred per second with read/write head in position

For Dynamic RAM

- TA = time from when the appropriate page of memory cells is first “strobed” to when bytes from that page can be read/written
- R = number of bytes transferred per second with needed page active

Note: Usually there exist many banks of DRAM working together in parallel!

6. Physical type

- Most common today:
  - semiconductor memory (SRAM, DRAM, flash, etc.)
  - magnetic surface memory
  - tape, optical and magneto-optical

Linear-Tape Open (LTO) format tape & tape unit.
(LTO-4: 0.8 TB per tape; LTO-8: 12 TB per tape)

7. Physical characteristics

- **Volatile** memory
  - Information that decays “naturally”...
  - ... or is lost when electrical power is removed
- **Non-volatile** memory
  - once recorded, information remains without deterioration...
  - ... although a timescale does need to be indicated (i.e., non-volatile does not necessarily mean **archival**)
- Semiconductor memory: some non-volatile examples
  - Non-erasable version: read-only memory (ROM)
  - Re-writable version: EEPROM (Electrically Erasable Programmable Read-Only Memory)

8. Organization

- How are bits physically arranged to form words?
  - Banks
  - Ranks
  - Pages
  - NAND, NOR memory (for flash memory)?
  - etc. etc.
  - Answer is not always obvious
- Other bits in addition to data bits?
  - error **detection**?
  - error **correction**?

Any Questions?