# CS152
# Computer Architecture and Engineering
# Lecture 19: I/O Systems

April 5, 1995

Dave Patterson (patterson@cs) and
Shing Kong (shing.kong@eng.sun.com)

Slides available on http://http.cs.berkeley.edu/~patterson




## Recap: Virtual Memory

- Virtual Memory invented as another level of the hierarchy
- Controversial at the time: can SW automatically manage 64KB across many programs?
- DRAM growth removed the controversy
- Today VM allows many processes to share single memory without having to swap all processes to disk, protection more important
- (Multi-level) page tables to map virtual address to physical address
- TLBs are important for fast translation
- TLB misses are significant in performance

## Outline of Today's Lecture

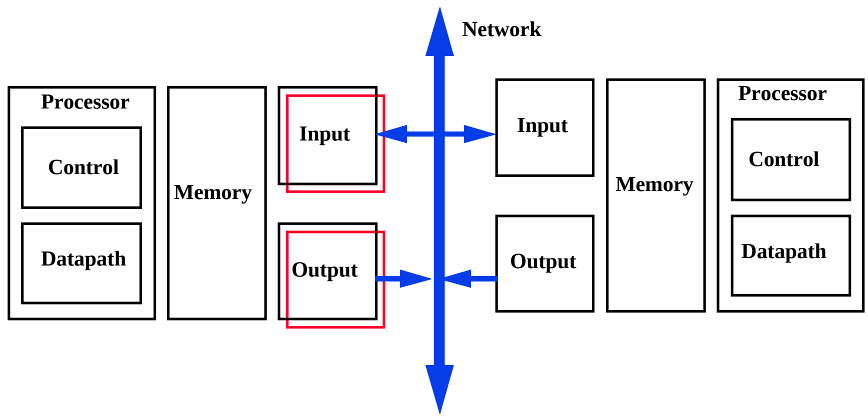
- Recap and Introduction (5 minutes)
- I/O Performance Measures (15 minutes)
- Questions and Administrative Matters (5 minutes)
- Types and Characteristics of I/O Devices (10 minutes)
- Magnetic Disks (15 minutes)
- Break (5 minutes)
- Graphic Displays and Video RAM (20 minutes)
- Summary (5 minutes)





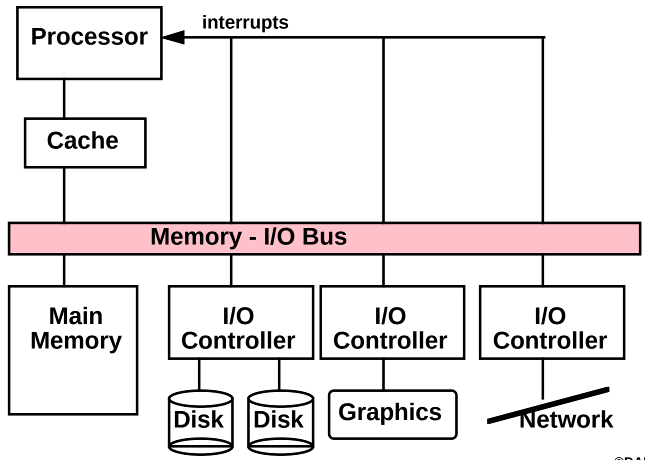
## The Big Picture: Where are We Now?

- Today's Topic: I/O Systems



## I/O System Design Issues

- Performance
- Expandability
- Resilience in the face of failure





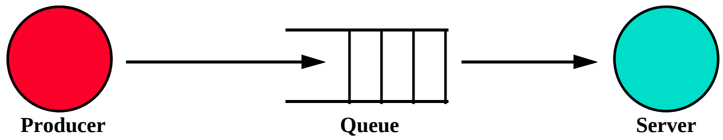


## I/O System Performance

° I/O System performance depends on many aspects of the system:
- The CPU
- The memory system:
  - Internal and external caches
  - Main Memory
- The underlying interconnection (buses)
- The I/O controller
- The I/O device
- The speed of the I/O software
- The efficiency of the software's use of the I/O devices

° Two common performance metrics:
- Throughput: I/O bandwidth
- Response time: Latency

## Producer-Server Model

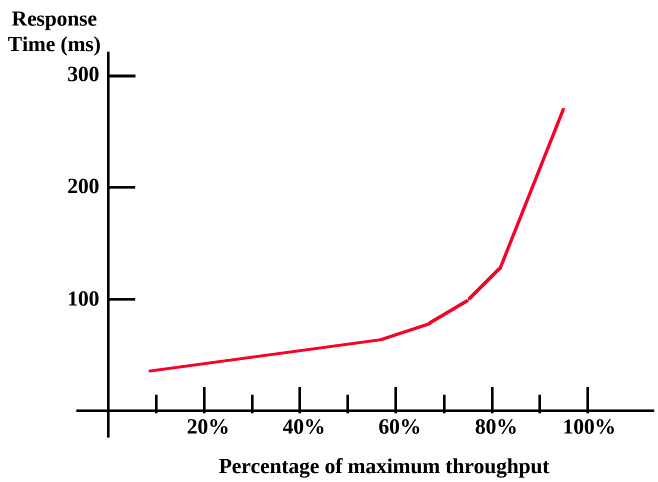



- Throughput:
  - The number of tasks completed by the server in unit time
  - In order to get the highest possible throughput:
    - The server should never be idle
    - The queue should never be empty
- Response time:
  - Begins when a task is placed in the queue
  - Ends when it is completed by the server
  - In order to minimize the response time:
    - The queue should be empty
    - The server will be idle

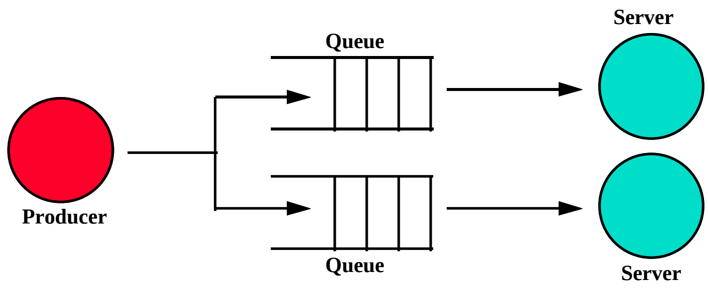




## Throughput versus Respond Time



## Throughput Enhancement




- In general throughput can be improved by:
  - Throwing more hardware at the problem
- Response time is much harder to reduce:
  - Ultimately it is limited by the speed of light
  - You cannot bribe God!





## I/O Benchmarks for Magnetic Disks

- Supercomputer application:
  - Large-scale scientific problems
- Transaction processing:
  - Examples: Airline reservations systems and banks
- File system:
  - Example: UNIX file system

## Supercomputer I/O

- Supercomputer I/O is dominated by:
  - Access to large files on magnetic disks
- Supercomputer I/O consists of one large read (read in the data)
  - Many writes to snapshot the state of the computation
- Supercomputer I/O consists of more output than input
- The overriding supercomputer I/O measures is data throughput:
  - Bytes/second that can be transferred between disk and memory




## Transaction Processing I/O

- Transaction processing:
  - Examples: airline reservations systems, bank ATMs
  - A lot of small changes to a large body of shared data
- Transaction processing requirements:
  - Throughput and response time are important
  - Must be gracefully handling certain types of failure
- Transaction processing is chiefly concerned with I/O rate:
  - The number of disk accesses per second
- Each transaction in typical transaction processing system takes:
  - Between 2 and 10 disk I/Os
  - Between 5,000 and 20,000 CPU instructions per disk I/O

## File System I/O

- Measurements of UNIX file systems in an engineering environment:
  - 80% of accesses are to files less than 10 KB
  - 90% of all file accesses are to data with sequential addresses on the disk
  - 67% of the accesses are reads
  - 27% of the accesses are writes
  - 6% of the accesses are read-write accesses





## Questions and Administrative Matters (5 Minutes)

## Types and Characteristics of I/O Devices

- **Behavior: how does an I/O device behave?**
  - **Input: read only**
  - **Output: write only, cannot read**
  - **Storage: can be reread and usually rewritten**
- **Partner:**
  - **Either a human or a machine is at the other end of the I/O device**
  - **Either feeding data on input or reading data on output**
- **Data rate:**
  - **The peak rate at which data can be transferred:**
    - **Between the I/O device and the main memory**
    - **Or between the I/O device and the CPU**

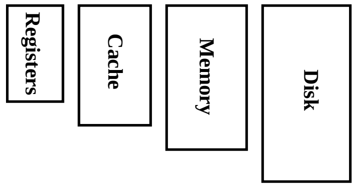




## I/O Device Examples

| Device | Behavior | Partner | Data Rate (KB/sec) |
|---|---|---|---|
| Keyboard | Input | Human | 0.01 |
| Mouse | Input | Human | 0.02 |
| Line Printer | Output | Human | 1.00 |
| Laser Printer | Output | Human | 100.00 |
| Graphics Display | Output | Human | 30,000.00 |
| Network-LAN | Input or Output | Machine | 200.00 |
| Floppy disk | Storage | Machine | 50.00 |
| Optical Disk | Storage | Machine | 500.00 |
| Magnetic Disk | Storage | Machine | 2,000.00 |

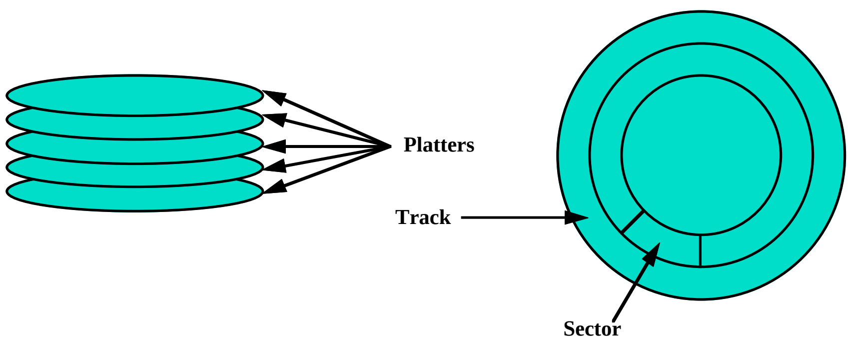
## Magnetic Disk




- Purpose:
  - Long term, nonvolatile storage
  - Large, inexpensive, and slow
  - Lowest level in the memory hierarchy
- Two major types:
  - Floppy disk
  - Hard disk
- Both types of disks:
  - Rely on a rotating platter coated with a magnetic surface
  - Use a moveable read/write head to access the disk
- Advantages of hard disks over floppy disks:
  - Platters are more rigid ( metal or glass) so they can be larger
  - Higher density because it can be controlled more precisely
  - Higher data rate because it spins faster
  - Can incorporate more than one platter

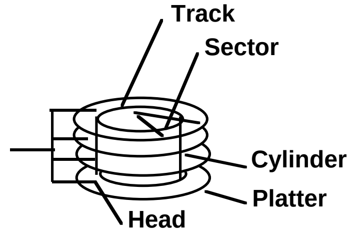



## Organization of a Hard Magnetic Disk




- Typical numbers (depending on the disk size):
  - 500 to 2,000 tracks per surface
  - 32 to 128 sectors per track
    - A sector is the smallest unit that can be read or written
- Traditionally all tracks have the same number of sectors:
  - Constant bit density: record more sectors on the outer tracks

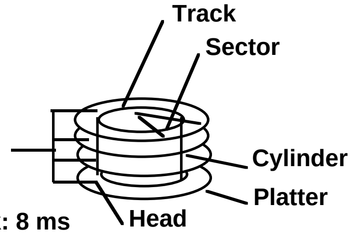
## Magnetic Disk Characteristic




- Cylinder: all the tacks under the head at a given point on all surface
- Read/write data is a three-stage process:
  - Seek time: position the arm over the proper track
  - Rotational latency: wait for the desired sector to rotate under the read/write head
  - Transfer time: transfer a block of bits (sector) under the read-write head
- Average seek time as reported by the industry:
  - Typically in the range of 12 ms to 20 ms
  - (Sum of the time for all possible seek) / (total # of possible seeks)
- Due to locality of disk reference, actual average seek time may:
  - Only be 25% to 33% of the advertised number

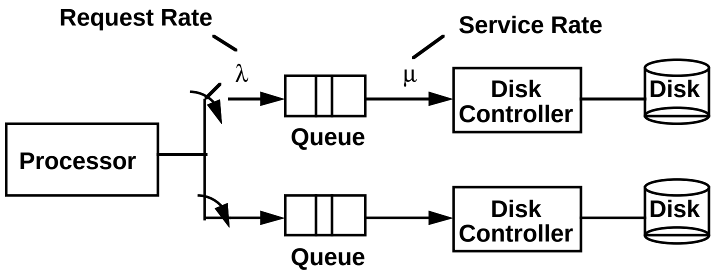



## Typical Numbers of a Magnetic Disk




- Rotational Latency:
  - Most disks rotate at 3,600 RPM
  - Approximately 16 ms per revolution
  - An average latency to the desired information is halfway around the disk: 8 ms
- Transfer Time is a function of :
  - Transfer size (usually a sector): 1 KB / sector
  - Rotation speed: 3600 RPM to 5400 RPM
  - Recording density: typical diameter ranges from 2 to 14 in
  - Typical values: 2 to 4 MB per second

## Disk I/O Performance




- Disk Access Time = Seek time + Rotational Latency + Transfer time
  + Controller Time + Queueing Delay

- **Estimating Queue Length:**
  - Utilization = U = Request Rate / Service Rate
  - Mean Queue Length = U / (1 - U)
  - As Request Rate -> Service Rate
    - Mean Queue Length -> Infinity




## Magnetic Disk Examples

| Characteristics | IBM 3090 | IBM0663 | Integral 1820 |
|---|---|---|---|
| Disk diameter (inches) | 10.88 | 3.50 | 1.80 |
| Formatted data capacity (MB) | 22,700 | 1,000 | 21 |
| MTTF (hours) | 50,000 | 400,000 | 100,000 |
| Number of arms/box | 12 | 1 | 1 |
| Rotation speed (RPM) | 3,600 | 4,318 | 3,800 |
| Transfer rate (MB/sec) | 4.2 | 4 | 1.9 |
| Power/box (watts) | 2,900 | 12 | 2 |
| MB/watt | 8 | 102 | 10.5 |
| Volume (cubic feet) | 97 | 0.13 | 0.02 |
| MB/cubic feet | 234 | 7692 | 1050 |

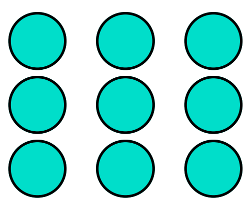
## Reliability and Availability

- Two terms that are often confused:
  - Reliability: Is anything broken?
  - Availability: Is the system still available to the user?
- Availability can be improved by adding hardware:
  - Example: adding ECC on memory
- Reliability can only be improved by:
  - Bettering environmental conditions
  - Building more reliable components
  - Building with fewer components
    - Improve availability may come at the cost of lower reliability





## Disk Arrays



- A new organization of disk storage:
  - Arrays of small and inexpensive disks
  - Increase potential throughput by having many disk drives:
    - Data is spread over multiple disk
    - Multiple accesses are made to several disks
- Reliability is lower than a single disk:
  - But availability can be improved by adding redundant disks:
    Lost information can be reconstructed from redundant information
  - MTTR: mean time to repair is in the order of hours
  - MTTF: mean time to failure of disks is three to five years

## Secondary Storage using DRAMs

- There are two ways DRAMs can be used for secondary storage:
  - Sold state disk
  - Expanded storage
- Solid state disks:
  - Function just like magnetic disk but:
    - Much faster
    - Much more expensive
  - Battery is used to make the system nonvolatile
- Expanded storage:
  - A large memory that allows only block transfers to or from main memory





## Optical Compact Disks

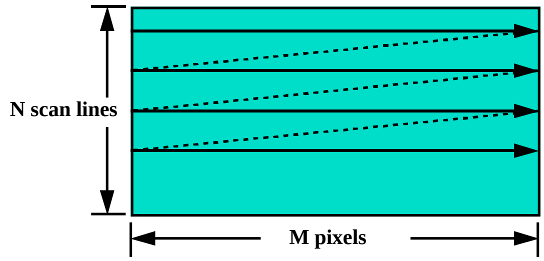
- Disadvantage:
  - It is a read-only media
- Advantages of Optical Compact Disk:
  - It is removable
  - It is inexpensive to manufacture
  - Have the potential to compete with new tape technologies for archival storage

## Break (5 Minutes)





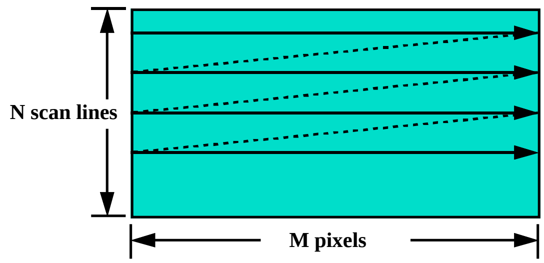
## Graphics Displays




° **Raster Cathode Ray Tube (CRT) Displays:**
  - **Resolution: (M pixels) x (N horizontal scan lines)**

° **Typical Sizes:**
  - **Studio Quality: 720 x 480**
  - **High Resolution Workstation Monitor: 1280 x 1024**
  - **High Definition TV (proposed standards):**
    - **U.S.A.: 1440 x 960**
    - **Others: 1920 x 1080**

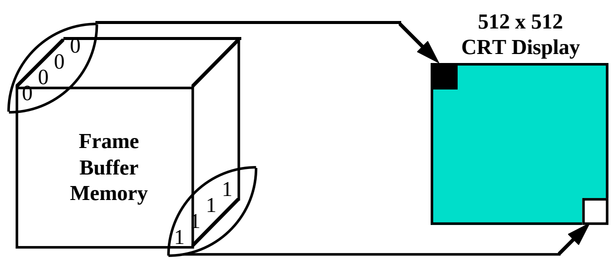
## Graphics Displays and Bit Map




- The M x N image is represented by a bit map in memory:
  - Black & White: 1bit per pixel
    - The pixel is either ON or OFF.  No longer widely used.
  - Gray scale: 8 bits per pixel
    - Each pixel can have 256 shades of black white
  - Color: 24 bits per pixel
    - 8 bits for each of the primary color: Red Green Blue




## Frame Buffer for a 512 x 512 4-bit Gray-Scale Display




- Examples:
  - Bit map value for location (0, 0)  = 0000  => Location (0, 0) is black
  - Bit map value for location (511, 511) is 1111 => L(511, 511) is white
- Total Number of Bits: 512 x 512 x 4 = 2**20 = 1 Mb = 128 KB

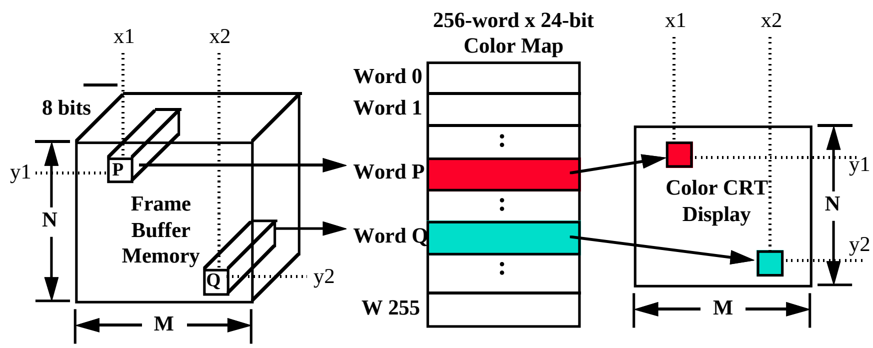
## Cost of Computer Graphics

- Size of Frame Buffer = M x N x P
  - M: Number of pixels per scan line
  - N: Number of horizontal scan lines
  - P: Number of bits per pixel
- Color Frame Buffer for High Resolution Workstation:
  - 1280 x 1024 x 24-bit = 3840 KB ~= 4 MB
- The size of the color frame buffer can be reduced:
  - Most pictures do not need the full palette (2**24) of possible colors
  - Use a two-level representation



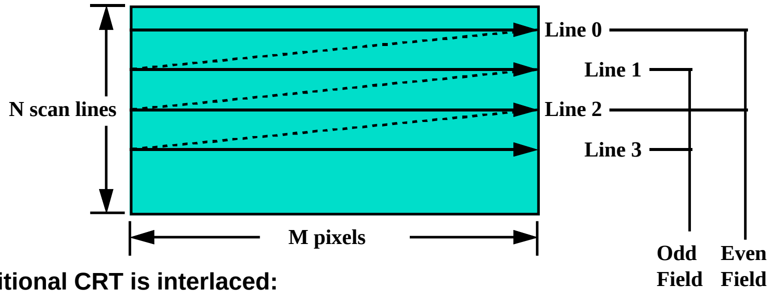


## Color Map: Reduce the Cost of Color Frame Buffer




- Color Map Size is 256-word x 24-bit:
  - Only 256 different colors can appear on the screen at a time
- The color map is loaded by the application program:
  - Each picture can have its own palette of color to chose from

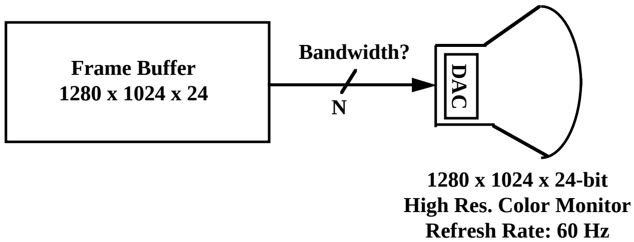
## Dynamic Feature: Frames and Fields




- ° Traditional CRT is interlaced:
  - Each Frame consist of 2 Fields:
    - Even Field: Lines 0, 2, 4, ... etc. Odd Field: Lines 1, 3, 5, ... etc.
  - The screen is updated 30 times per second:
    - 30 Frames per second 60 Fields per second
- ° High resolution (1280 x 1024) workstation monitor:
  - The entire screen is refreshed at 60 to 72 times per second

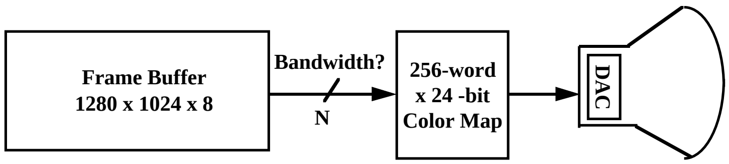



## Bandwidth Requirement for the Frame Buffer




- ° Assume no color map:
  - Frame Buffer Size: 1280 x 1024 x 24 = 3840 KB ~= 4 MB
  - Bandwidth Requirement: 4 MB x 60 Hz = 240 MB / sec
- ° Assume N = 32-bit = 4-byte:
  - The frame buffer needs to response at: 240/4 = 60 MHz => 16.6 ns
  - This is faster than most of the low cost SRAM
  - Another reason to use a color map!

## Bandwidth Requirement for the Frame Buffer (Cont.)

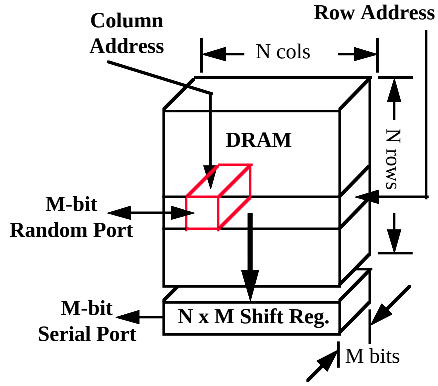



- With a 256-word by 24-bit color map:
  - Frame Buffer Size: 1280 x 1024 x 8 = 1280 KB
  - Bandwidth Requirement: 1280 KB x 60 Hz = 75 MB / sec
- Assume N = 32-bit = 4-byte:
  - The frame buffer needs to response at: 75/4 =18.75 MHz => 53.3 ns
  - DRAM is still too slow and SRAM is too expensive.
  - Solution: Video RAM (VRAM)

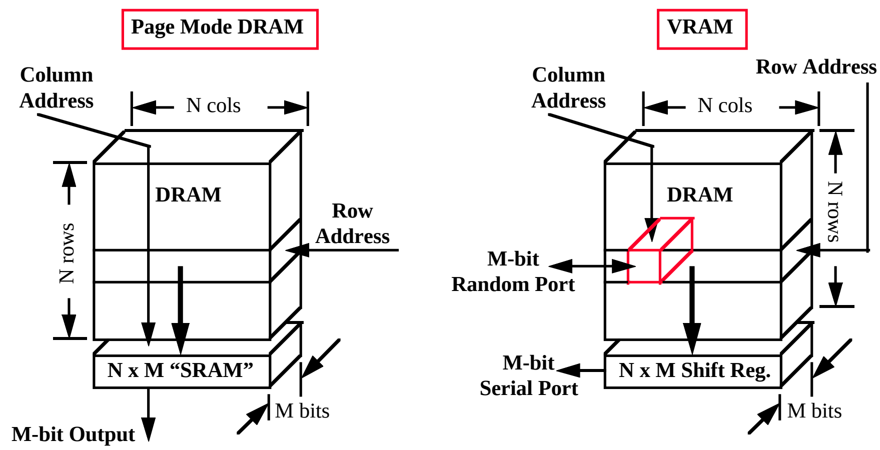



## Video Random Access Memory (VRAM)

- A high-speed shift register
  - Hold one row of DRAM
- Random Port:
  - Access the DRAM array
  - Works like regular DRAM (and just as slow)
- Serial Port:
  - Access the shift register



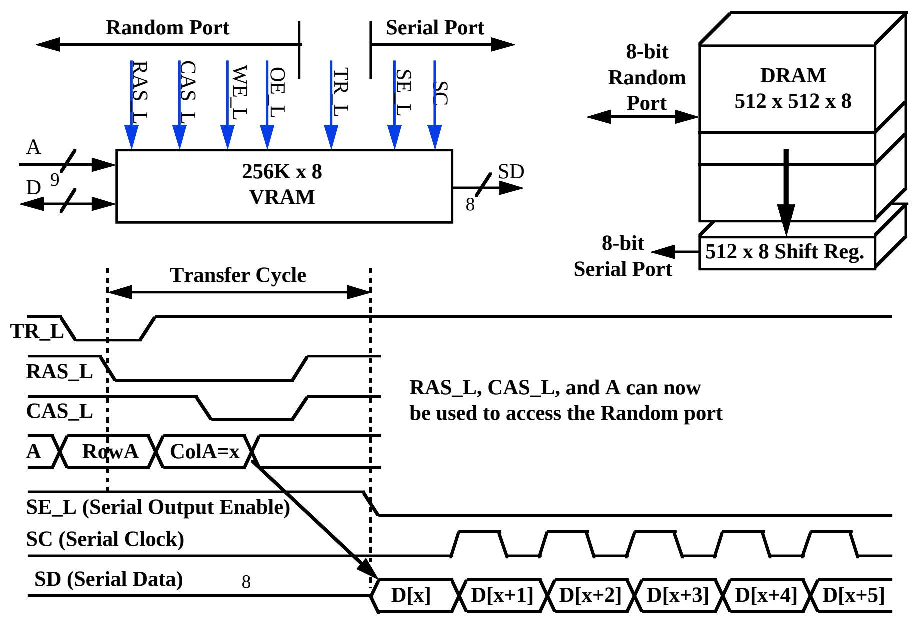
# VRAM versus Fast Page Mode DRAM



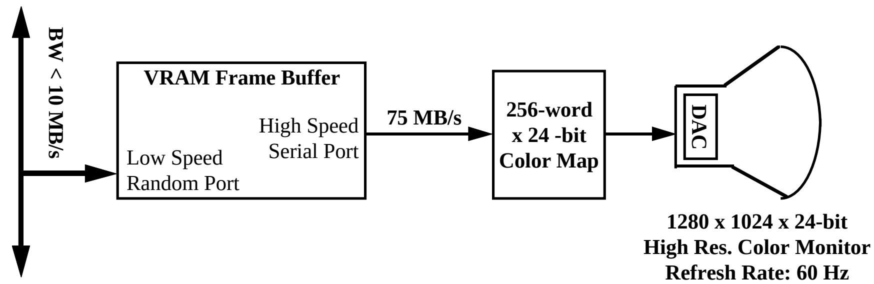

° Serial Port access is similar to Fast Page Mode operation except:
- N x M shift register is much faster than the N x M SRAM
- Fast Page Mode can access any location within the SRAM
- Serial port operation only  allow sequential access (shift out)




# VRAM Logic Diagram and Serial Port Timing



## Dual Port Frame Buffer using VRAM

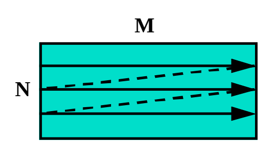



- High Speed Serial Port
  - Connects to the graphic display
  - Supply data to refresh the image on the screen
- Low Speed Random Port
  - Connects to the I/O Bus
  - CPU writes to the Frame Buffer to change the image




## Summary:

- Disk I/O Benchmarks:
  - Supercomputer Application: main concern is data rate
  - Transaction Processing: main concern is I/O rate
  - File System: main concern is file access
- Three Components of Disk Access Time:
  - Seek Time: advertised to be 12 to 20ms. May be lower in real life.
  - Rotational Latency: 5.6 ms at 5400 RPM and 8.3 ms at 3600 RPM
  - Transfer Time: 2 to 4 MB per second




- Graphic Display:
  - Resolution: (M pixels) x (N scan lines)
  - Frame Buffer size and bandwidth requirement can be reduced by placing a Color Map between the Frame Buffer and CRT display
  - VRAM: a DRAM core with a high speed shift register

## Where to get more information?

- To be continued ...