Computer Architecture

Prof. Dr. Nizamettin AYDIN

naydin@yildiz.edu.tr
nizamettinaydin@gmail.com

http://www.yildiz.edu.tr/~naydin

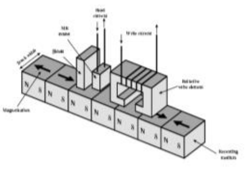
External Memory

Outline

Magnetic Disk

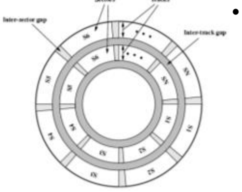
- Disk substrate coated with magnetizable material (iron oxide…rust)
- Substrate used to be aluminium
- Now glass
  - Improved surface uniformity
    - Increases reliability
  - Reduction in surface defects
    - Reduced read/write errors
  - Lower flight heights
  - Better stiffness to reduce disk dynamics
  - Better shock/damage resistance

Write and Read Mechanisms



- During read/write, head is stationary, platter rotates
- Write (inductive)
  - Current through coil produces magnetic field
  - Pulses sent to head
  - Magnetic pattern recorded on surface below
- Read (magneto resistive) (MR)
  - Separate read head, close to write head
  - Partially shielded magneto resistive (MR) sensor
  - Electrical resistance depends on direction of magnetic field
  - High frequency operation
    - Higher storage density and speed

Data Organization and Formatting



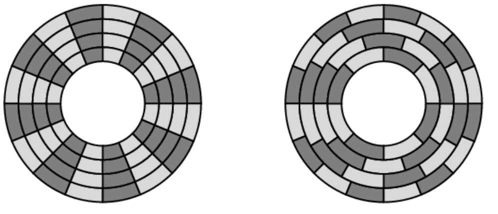
- Concentric rings or tracks
  - Gaps between tracks
  - Reduce gap to increase capacity
  - Same number of bits per track (variable packing density)
  - Constant angular velocity
- Tracks divided into sectors
- Minimum block size is one sector
- May have more than one sector per block

Disk Velocity

- Bit near centre of rotating disk passes fixed point slower than bit on outside of disk
- Increase spacing between bits in different tracks
- Rotate disk at constant angular velocity (CAV)
  - Gives pie shaped sectors and concentric tracks
  - Individual tracks and sectors addressable
  - Move head to given track and wait for given sector
  - Waste of space on outer tracks
    - Lower data density
- Can use zones to increase capacity
  - Each zone has fixed bits per track
  - More complex circuitry

7

Disk Layout Methods Diagram



(a) Constant angular velocity (b) Multiple zoned recording

8

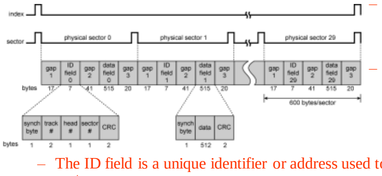
Finding Sectors

- Some means is needed to locate sector positions within a track.
  - There must be some starting point on the track and a way of identifying the start and end of each sector.
- These requirements are handled by means of control data recorded on the disk.
  - Thus, the disk is formatted with some extra data used only by the disk drive and not accessible to the user.

9

Winchester Disk Format-Seagate ST506

- An example of disk formatting:



  - Each track contains 30 fixed-length sectors of 600 bytes each.
  - Each sector holds 512 bytes of data plus control information useful to the disk controller.
  - The ID field is a unique identifier or address used to locate a particular sector.
  - The SYNCH byte is a special bit pattern that delimits the beginning of the field.
  - The track number identifies a track on a surface.
  - The head number identifies a head, because this disk has multiple surfaces.
  - The ID and data fields each contain an error detecting code.

10

Characteristics

- Fixed (rare) or movable head
- Removable or fixed
- Single or double (usually) sided
- Single or multiple platter
- Head mechanism
  - Contact (Floppy)
  - Fixed gap
  - Flying (Winchester)

11

Fixed/Movable Head Disk

- Fixed head
  - One read write head per track
  - Heads mounted on fixed ridged arm
- Movable head
  - One read write head per side
  - Mounted on a movable arm

12

Removable or Not

- Removable disk
  - Can be removed from drive and replaced with another disk
  - Provides unlimited storage capacity
  - Easy data transfer between systems
- Nonremovable disk
  - Permanently mounted in the drive

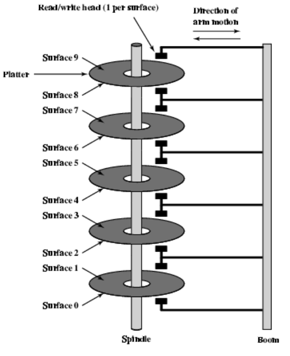
13

Multiple Platter

- One head per side
- Heads are joined and aligned
- Aligned tracks on each platter form cylinders
- Data is striped by cylinder
  - reduces head movement
  - Increases speed (transfer rate)

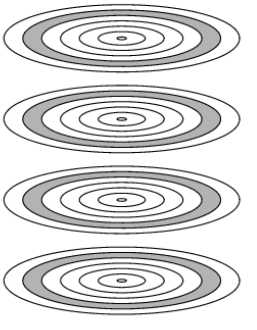
14

Multiple Platters



15

Tracks and Cylinders



16

Floppy Disk

- 8”, 5.25”, 3.5”
- Small capacity
  - Up to 1.44Mbyte (2.88M never popular)
- Slow
- Universal
- Cheap
- Obsolete?

17

Winchester Hard Disk (1)

- Developed by IBM in Winchester (USA)
- Sealed unit
- One or more platters (disks)
- Heads fly on boundary layer of air as disk spins
- Very small head to disk gap
- Getting more robust
- Universal
- Cheap
- Fastest external storage!
- Getting larger all the time
  - Terabyte now easily available

18

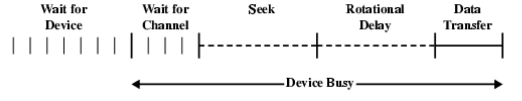
Typical Hard Disk Drive Parameters

- Parameters for typical contemporary high-performance disks

Characteristics	Constellation ES.2	Seagate Barracuda XT	Cheetah NS	Momentus
Application	Enterprise	Desktop	Network attached storage, application servers	Laptop
Capacity	3 TB	3 TB	400 GB	640 GB
Average seek time	8.5 ms read 9.5 ms write	N/A	3.9 ms read 4.2 ms write	13 ms
Spindle speed	7200 rpm	7200 rpm	10, 075 rpm	5400 rpm
Average latency	4.16 ms	4.16 ms	2.98	5.6 ms
Maximum sustained transfer rate	155 MB/s	149 MB/s	97 MB/s	300 MB/s
Bytes per sector	512	512	512	4096
Tracks per cylinder (number of platter surfaces)	8	10	8	4
Cache	64 MB	64 MB	16 MB	8 MB

Timing of Disk I/O Transfer

- The actual details of disk I/O operation depend on the computer system, the operating system, and the nature of the I/O channel and disk controller hardware.
- A general timing diagram of disk I/O transfer:



Disk Performance Parameters-Speed

- Seek time
  - Moving head to correct track
- (Rotational) latency
  - Waiting for data to rotate under head
- Transfer time
  - Depends on the rotation speed of the disk

$$T = \frac{b}{rN}$$

- T: transfer time,
- b: number of bytes to be transferred,
- N: number of bytes on track,
- r: rotation speed, in revolution per sconds

- Access time = Seek + Latency

$$T = T_s + \frac{1}{2r} + \frac{b}{rN}$$

T_s: average seek time

  - Transfer rate

RAID

- Redundant Array of Independent Disks
- Redundant Array of Inexpensive Disks
- 6 levels in common use
- Not a hierarchy
- Set of physical disks viewed as single logical drive by O/S
- Data distributed across physical drives
- Can use redundant capacity to store parity information

RAID 0

- No redundancy
- Data striped across all disks
- Round Robin striping
- Increase speed
  - Multiple data requests probably not on same disk
  - Disks seek in parallel
  - A set of data is likely to be striped across multiple disks

RAID 1

- Mirrored Disks
- Data is striped across disks
- 2 copies of each stripe on separate disks
- Read from either
- Write to both
- Recovery is simple
  - Swap faulty disk & re-mirror
  - No down time
- Expensive

RAID 2

- Disks are synchronized
- Very small stripes
  - Often single byte/word
- Error correction calculated across corresponding bits on disks
- Multiple parity disks store Hamming code error correction in corresponding positions
- Lots of redundancy
  - Expensive
  - Not used

25

RAID 3

- Similar to RAID 2
- Only one redundant disk, no matter how large the array
- Simple parity bit for each set of corresponding bits
- Data on failed drive can be reconstructed from surviving data and parity info
- Very high transfer rates

26

Example: Data reconstruction in RAID3

- Consider an array of five drives (X0,X1,X2,X3 contain data, X4 is parity disk)
- Parity of *i*th bit is calculated as:

  $X4(i)=X3(i)\oplus X2(i)\oplus X1(i)\oplus X0(i)$

- Suppose that drive X1 has failed. The contents of X1 can be regenerated as:

  $X1(i)=X4(i)\oplus X3(i)\oplus X2(i)\oplus X0(i)$

27

RAID 4

- Each disk operates independently
- Good for high I/O request rate
- Large stripes
- Bit by bit parity calculated across stripes on each disk
- Parity stored on parity disk

28

RAID 5

- Like RAID 4
- Parity striped across all disks
- Round robin allocation for parity stripe
- Avoids RAID 4 bottleneck at parity disk
- Commonly used in network servers

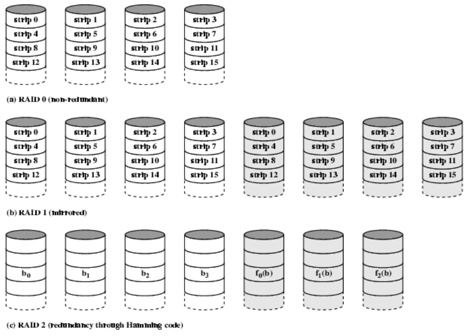
29

RAID 6

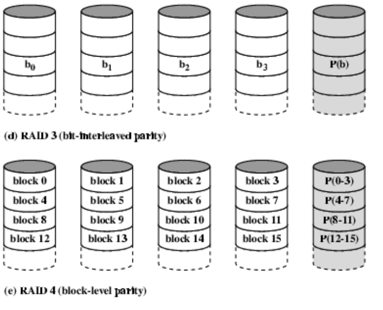
- Two parity calculations
- Stored in separate blocks on different disks
- User requirement of N disks needs N+2
- High data availability
  - Three disks need to fail for data loss
  - Significant write penalty

30

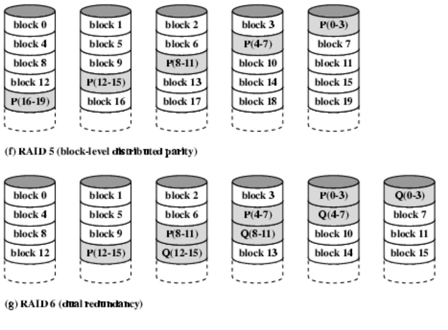
RAID 0, 1, 2



RAID 3 & 4



RAID 5 & 6

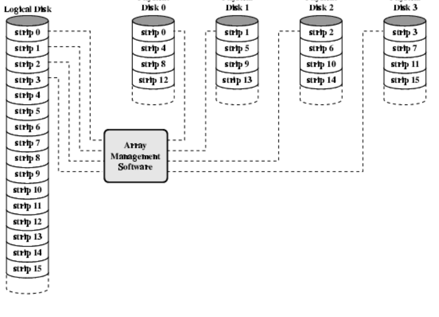


RAID Levels

Category	Level	Description	Disks Required	Data Availability	Large I/O Data Transfer Capacity	Small I/O Request Rate
Striping	0	Nonredundant	N	Lower than single disk	Very high	Very high for both read and write
Mirroring	1	Mirrored	$2N$	Higher than RAID 2, 3, 4, or 5; lower than RAID 6	Higher than single disk for read; similar to single disk for write	Up to twice that of a single disk for read; similar to single disk for write
Parallel access	2	Redundant via Hamming code	$N + m$	Much higher than single disk; comparable to RAID 3, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
	3	Bit-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
Independent access	4	Block-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 5	Similar to RAID 0 for read; significantly lower than single disk for write	Similar to RAID 0 for read; significantly lower than single disk for write
	5	Block-interleaved distributed parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 4	Similar to RAID 0 for read; lower than single disk for write	Similar to RAID 0 for read; generally lower than single disk for write
	6	Block-interleaved dual distributed parity	$N + 2$	Highest of all listed alternatives	Similar to RAID 0 for read; lower than RAID 5 for write	Similar to RAID 0 for read; significantly lower than RAID 5 for write

Note: N = number of data disks; m proportional to log N

Data Mapping For RAID 0



RAID Comparison

Level	Advantages	Disadvantages	Applications
0	I/O performance is greatly improved by spreading the I/O load across many channels and drives No parity calculation overhead is involved Very simple design Easy to implement	The failure of just one drive will result in all data in an array being lost	Video production and editing Image Editing Pre-press applications Any application requiring high bandwidth
1	100% redundancy of data means no rebuild is necessary in case of a disk failure, just a copy to the replacement disk Under certain circumstances, RAID 1 can sustain multiple simultaneous drive failures Simplest RAID storage subsystem design	Highest disk overhead of all RAID types (100%)—inefficient	Accounting Payroll Financial Any application requiring very high availability
2	Extremely high data transfer rates possible The higher the data transfer rate required, the better the ratio of data disks to ECC disks Relatively simple controller design compared to RAID levels 3, 4, & 5	Very high ratio of ECC disks to data disks with smaller word sizes—inefficient Entry level cost very high—requires very high transfer rate requirement to justify	No commercial implementations exist/ not commercially viable

RAID Comparison

3	Very high read data transfer rate Very high write data transfer rate Disk failure has an insignificant impact on throughput Low ratio of ECC (parity) disks to data disks means high efficiency	Transaction rate equal to that of a single disk drive at best (if spindles are synchronized) Controller design is fairly complex	Video production and live streaming Image editing Video editing Prepress applications Any application requiring high throughput
4	Very high Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency	Quite complex controller design Worst write transaction rate and Write aggregate transfer rate Difficult and inefficient data rebuild in the event of disk failure	No commercial implementations exist/ not commercially viable
5	Highest Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency Good aggregate transfer rate	Most complex controller design Difficult to rebuild in the event of a disk failure (as compared to RAID level 1)	File and application servers Database servers Web, e-mail, and news servers Intranet servers Most versatile RAID level
6	Provides for an extremely high data fault tolerance and can sustain multiple simultaneous drive failures	More complex controller design Controller overhead to compute parity addresses is extremely high	Perfect solution for mission critical applications

37

Solid State Drives (SSD)

- A memory device made with solid state components that can be used as a replacement to a hard disk drive.
- In recent years, it is used to complement or even replace hard disk drives (HDDs),
  - both as internal and external secondary memory
- SSDs now on the market and coming on line use a type of semiconductor memory referred to as flash memory

38

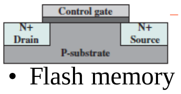
Flash Memory

- a type of semiconductor memory that has been around for a number of years
- is used in many consumer electronic products,
  - smart phones, GPS devices, MP3 players, digital cameras, and USB devices.
- In recent years, the cost and performance of flash memory has evolved to the point where it is feasible to use flash memory drives to replace HDDs.
- Next slide illustrates the basic operation of a flash memory

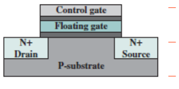
39

Flash Memory

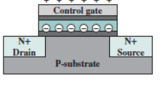
- Transistor structure



  - A small voltage applied to the gate can be used to control the flow of a large current between the source and the drain.
- Flash memory cell in one state



  - A second gate, called a floating gate because it is insulated by a thin oxide layer, is added to the transistor.
  - Initially, the floating gate does not interfere with the operation of the transistor.
  - In this state, the cell is deemed to represent binary 1.
- Flash memory cell in zero state



  - Applying a large voltage across the oxide layer causes electrons to tunnel through it and become trapped on the floating gate, where they remain even if the power is disconnected .
  - In this state, the cell is deemed to represent binary 0.

40

Flash Memory

- Two distinctive types:
  - NOR flash memory
    - the basic unit of access is a bit,
    - the logical organization resembles a NOR logic device
    - provides high-speed random access
    - can read and write data to specific locations,
    - can reference and retrieve a single byte
    - used to store cell phone operating system code and on Windows computers for the BIOS program that runs at startup
  - NAND flash memory
    - the basic unit is 16 or 32 bits,
    - the logical organization resembles NAND devices
    - reads and writes in small blocks
    - used in USB flash drives, memory cards (in digital cameras, MP3 players, etc.), and in SSDs
    - provides higher bit density than NOR and greater write speed
    - does not provide a random-access external address bus so
    - the data must be read on a blockwise basis (also known as page access), where each block holds hundreds to thousands of bits

41

SSD Compared to HDD

- SSDs have the following advantages over HDDs:
  - High-performance input/output operations per second (IOPS):
    - Significantly increases performance I/O subsystems.
  - Durability:
    - Less susceptible to physical shock and vibration.
  - Longer lifespan:
    - SSDs are not susceptible to mechanical wear.
  - Lower power consumption:
    - SSDs use as little as 2.1 watts of power per drive,
      - considerably less than comparable-size HDDs.
  - Quieter and cooler running capabilities:
    - Less floor space required, lower energy costs, and a greener enterprise.
  - Lower access times and latency rates:
    - Over 10 times faster than the spinning disks in an HDD.

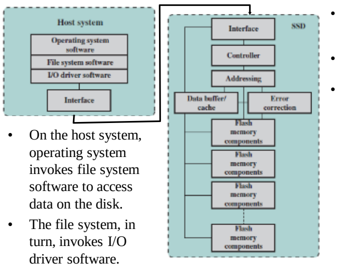
42

SSD Compared to HDD

- Comparison of Solid State Drives and Disk Drives (as of around 2013)

	NAND Flash Drives	Disk Drives
I/O per second (sustained)	Read: 45,000 Write: 15,000	300
Throughput (MB/s)	Read: 200+ Write: 100+	up to 80
Random access time (ms)	0.1	4–10
Storage capacity	up to 256 GB	up to 4 TB

SSD Organization



- On the host system, operating system invokes file system software to access data on the disk.
- The file system, in turn, invokes I/O driver software.
- The I/O driver software provides host access to the particular SSD product.
- If the device is an internal hard drive, a common interface is PCIe.
- For external devices, one common interface is USB.
- In addition to the interface to the host system, the SSD contains the following components:
  - **Controller:**
    - Provides SSD device level interfacing and firmware execution.
  - **Addressing:**
    - Logic that performs the selection function across the flash memory components.
  - **Data buffer/cache:**
    - High speed RAM memory components used for speed matching and to increased data throughput.
  - **Error correction:**
    - Logic for error detection and correction.
  - **Flash memory components:**
    - Individual NAND flash chips.

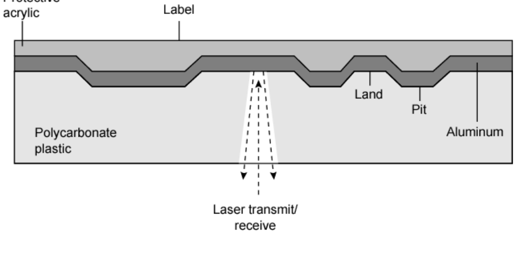
Optical Storage

- CD
  - Compact Disk
- CD-ROM
  - Compact Disk Read-Only Memory
- CD-R
  - CD Recordable
- CD-RW
  - CD Rewritable
- DVD
  - Digital Versatile Disk
- DVD-R
  - DVD Recordable
- DVD-RW
  - DVD Rewritable

Optical Storage CD-ROM

- Originally for audio
- 650Mbytes giving over 70 minutes audio
- Polycarbonate coated with highly reflective coat, usually aluminium
- Data stored as pits
- Read by reflecting laser
- Constant packing density
- Constant linear velocity

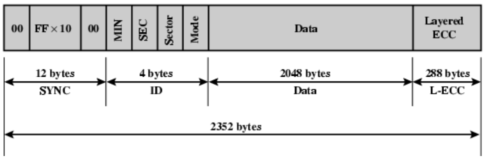
CD Operation



CD-ROM Drive Speeds

- Audio is single speed
  - Constant linier velocity
  - 1.2 ms^{-1}
  - Track (spiral) is 5.27km long
  - Gives 4391 seconds = 73.2 minutes
- Other speeds are quoted as multiples
- e.g. 24x
- Quoted figure is maximum drive can achieve

CD-ROM Format



- Sync: Identifies the beginning of a block. Consists of a byte of all 0s, 10 bytes of all 1s, and a byte of all 0s
- Header: Contains the block address and the mode byte
  - Mode 0=blank data field
  - Mode 1=2048 byte data+error correction
  - Mode 2=2336 byte data
- Data: User data
- Auxiliary: Additional user data in mode 2. In mode 1, this is a 288-byte error correcting code.

49

Random Access on CD-ROM

- Difficult
- Move head to rough position
- Set correct speed
- Read address
- Adjust to required location

50

CD-ROM for & against

- Large capacity (?)
- Easy to mass produce
- Removable
- Robust

- Expensive for small runs
- Slow
- Read only

51

Other Optical Storage

- CD-Recordable (CD-R)
  - WORM
  - Now affordable
  - Compatible with CD-ROM drives
- CD-RW
  - Erasable
  - Getting cheaper
  - Mostly CD-ROM drive compatible
  - Phase change
    - Material has two different reflectivities in different phase states

52

DVD

- Digital Video Disk
  - Used to indicate a player for movies
    - Only plays video disks
- Digital Versatile Disk
  - Used to indicate a computer drive
    - Will read computer disks and play video disks

53

DVD - technology

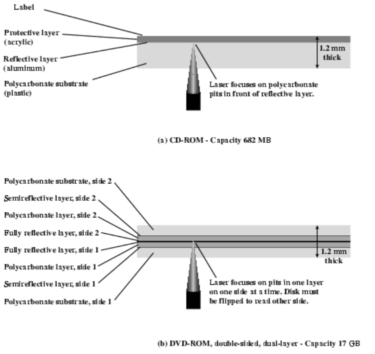
- Multi-layer
- Very high capacity (4.7G per layer)
- Full length movie on single disk
  - Using MPEG compression
- Finally standardized
- Movies carry regional coding
- Players only play correct region films

54

DVD – Writable

- Loads of trouble with standards
- First generation DVD drives may not read first generation DVD-W disks
- First generation DVD drives may not read CD-RW disks

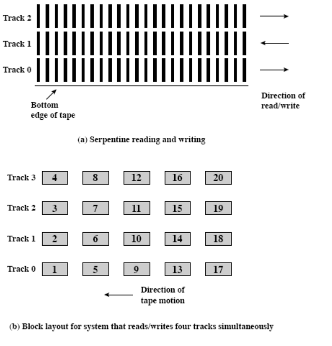
CD and DVD



Magnetic Tape

- Serial access
- Slow
- Very cheap
- Backup and archive
- Capacity: a couple of 100 GB

Typical Magnetic Tape Features



Internet Resources

- Optical Storage Technology Association
  - Good source of information about optical storage technology and vendors
  - Extensive list of relevant links
- DLTtape
  - Good collection of technical information and links to vendors
- Search on RAID