Tutorial 6
External memory

COMP2120B Computer organization

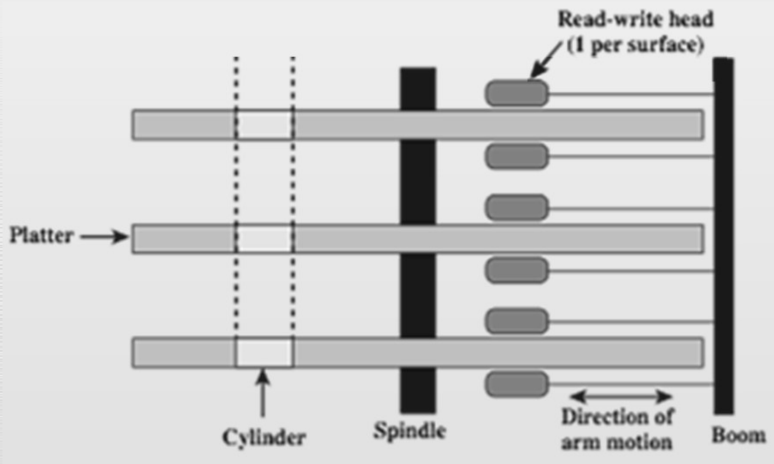
Kevin Lam (yklam2)

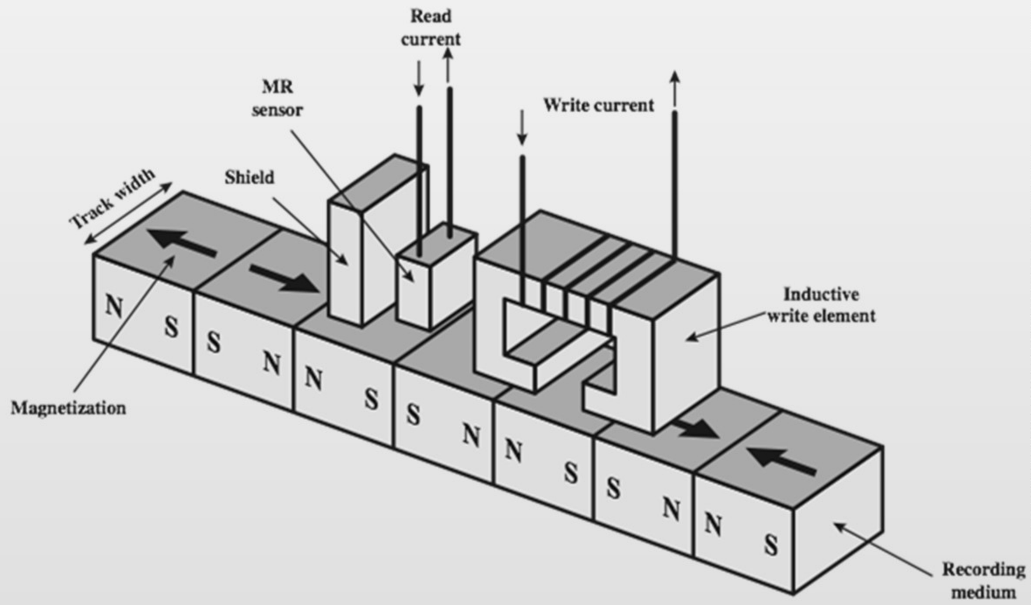
Overview

- Magnetic disk
- Solid state drives
- Other external memory (highlight only)
- RAID

Magnetic Disk

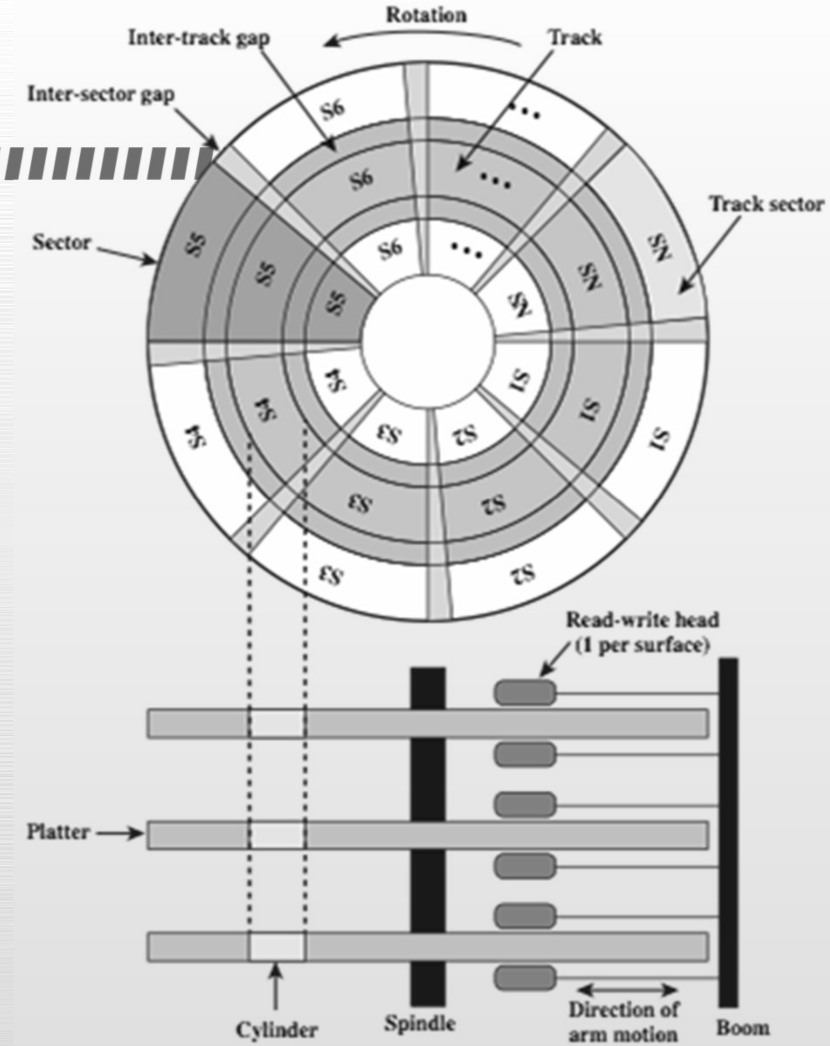
- A magnetic disk is a circular ***platter*** constructed of nonmagnetic material, called the ***substrate***, coated with a magnetizable material





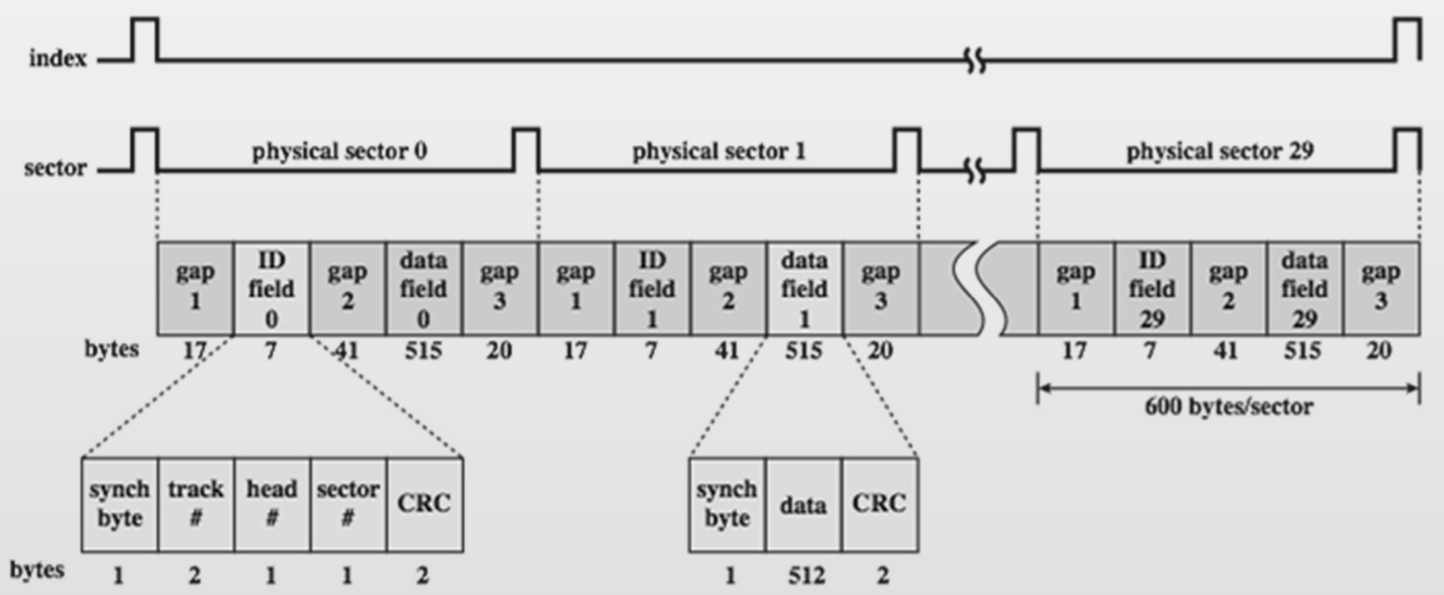
Track, cylinder, sector

- Each surface is divided into circular tracks
- Each track is divided into sectors, usually of 512 bytes
- All the read-write head moves together.
- Tracks of different platter under the head at the same time is called a cylinder.



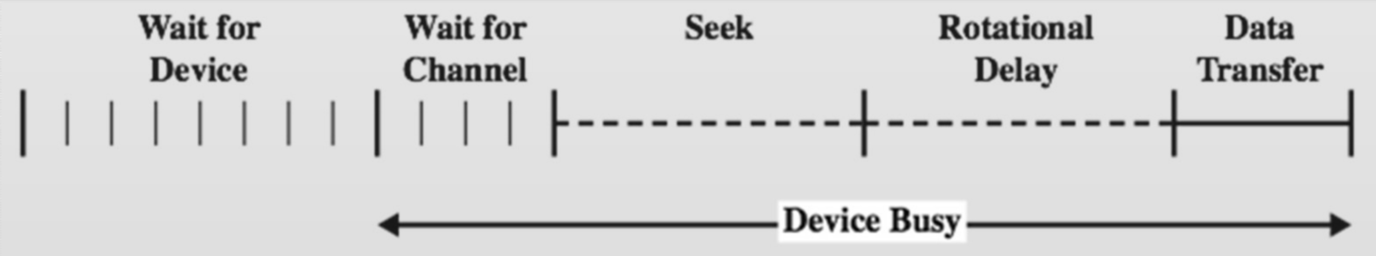
Data stored in magnetic disk

- The data is stored as a stream of preambles, followed by the data, and then an Error Correcting Code (ECC).



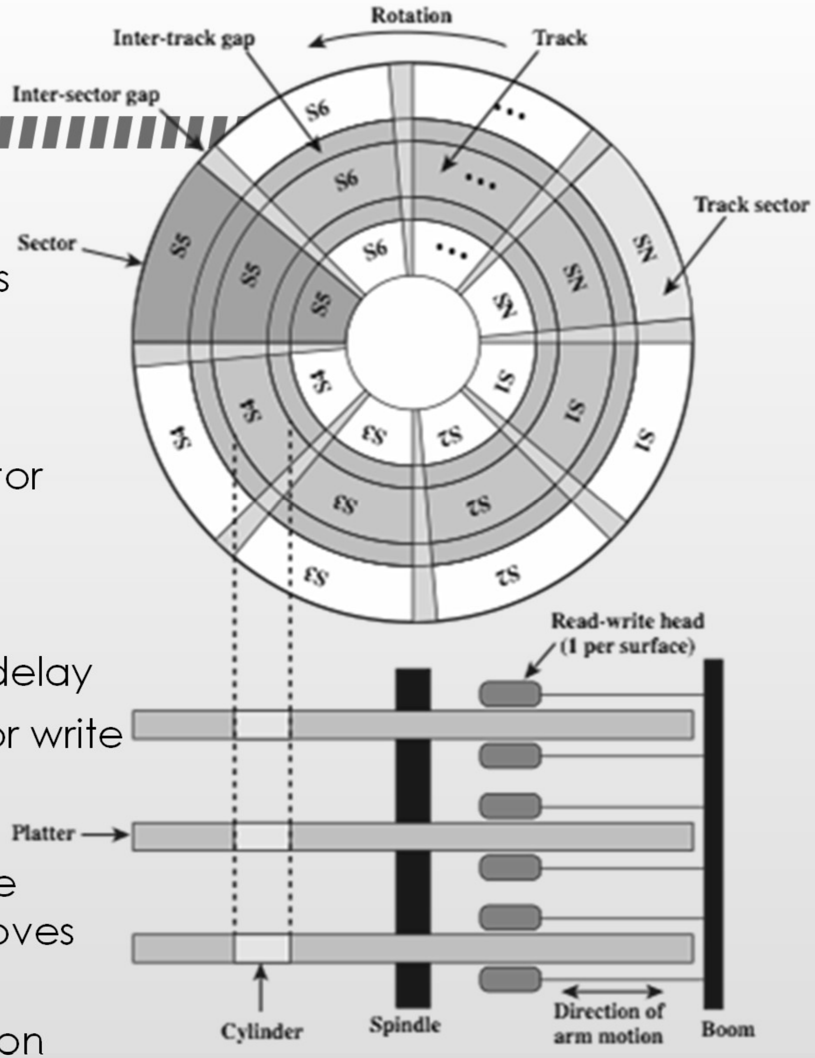
Disk performance parameters

- When the disk drive is operating the disk is rotating at constant speed
- To read or write the head must be positioned at the desired track and at the beginning of the desired sector on the track
  - Track selection involves moving the head in a movable-head system or electronically selecting one head on a fixed-head system
  - Once the track is selected, the disk controller waits until the appropriate sector rotates to line up with the head



Disk performance parameters (cont'd)

- **Seek time**
  - On a movable–head system, the time it takes to position the head at the track
- **Rotational delay** (rotational latency)
  - The time it takes for the beginning of the sector to reach the head
- **Access time**
  - The sum of the seek time and the rotational delay
  - The time it takes to get into position to read or write
- **Transfer time**
  - Once the head is in position, the read or write operation is then performed as the sector moves under the head
  - This is the data transfer portion of the operation



Typical Hard Disk Drive Parameters

Characteristics	Seagate Enterprise	Seagate Barracuda XT	Seagate Cheetah NS	Seagate Laptop HDD
Application	Enterprise	Desktop	Network attached storage, application servers	Laptop
Capacity	6 TB	3 TB	600 GB	2 TB
Average seek time	4.16 ms	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	216 MB/s	149 MB/s	97 MB/s	300 MB/s
Bytes per sector	512/4096	512	512	4096
Tracks per cylinder (number of platter surfaces)	8	10	8	4
Cache	128 MB	64 MB	16 MB	8 MB

Solid State Drives (SSD)

- SSDs have the following advantages over HDDs:
  - High-performance input/output operations per second (IOPS)
  - Durability: Less susceptible to physical shock and vibration
  - Longer lifespan: No mechanical wear
  - Lower power consumption
  - Quieter and cooler running capabilities
  - Lower access times and latency rates

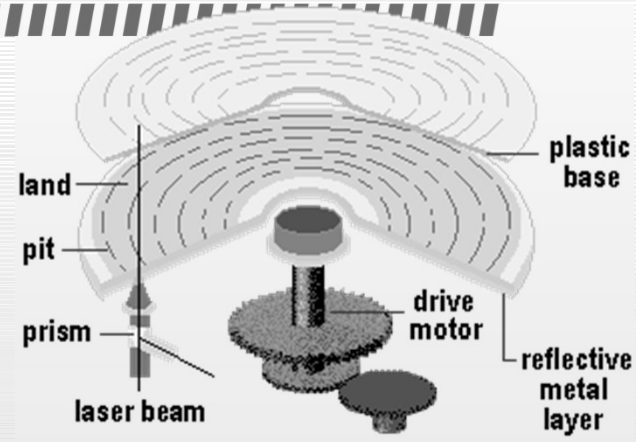
	NAND Flash Drives	Seagate Laptop Internal HDD
File copy/write speed	200—550 Mbps	50—120 Mbps
Power draw/battery life	Less power draw, averages 2–3 watts, resulting in 30+ minute battery boost	More power draw, averages 6–7 watts and therefore uses more battery
Storage capacity	Typically not larger than 512 GB for notebook size drives; 1 TB max for desktops	Typically around 500 GB and 2 TB maximum for notebook size drives; 4 TB max for desktops
Cost	Approx. $0.50 per GB for a 1-TB drive	Approx $0.15 per GB for a 4-TB drive

SSD Practical Issues

- There are two practical issues peculiar to SSDs that are not faced by HDDs:
  - SDD performance has a tendency to slow down as the device is used
    - The entire block must be read from the flash memory and placed in a RAM buffer
    - Before the block can be written back to flash memory, the entire block of flash memory must be erased
    - The entire block from the buffer is now written back to the flash memory
- Flash memory becomes unusable after a certain number of writes
  - Techniques for prolonging life:
    - Front-ending the flash with a cache to delay and group write operations
    - Using wear-leveling algorithms that evenly distribute writes across block of cells
    - Bad-block management techniques
  - Most flash devices estimate their own remaining lifetimes so systems can anticipate failure and take preemptive action

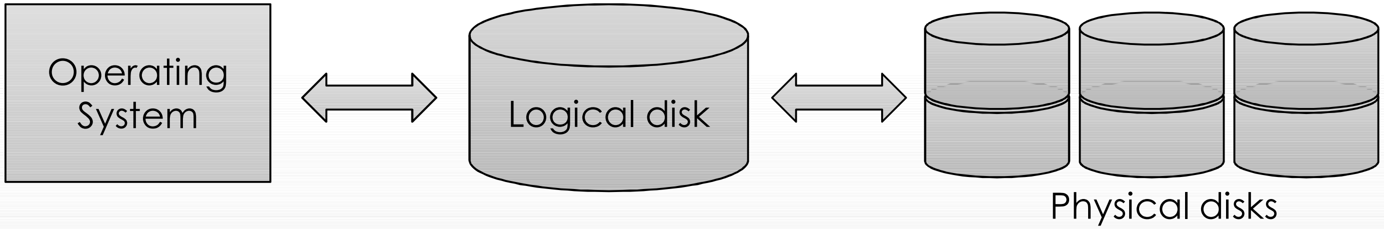
Other External Memory

- Optical Memory
  - CD-R(W) / DVD-R(W)
  - Blu-Ray R(RE)
- Magnetic Tape



Sources:
1. https://www.pctechguide.com/images/32drive.gif
2. http://www.array.cz/htm/qualstar.htm

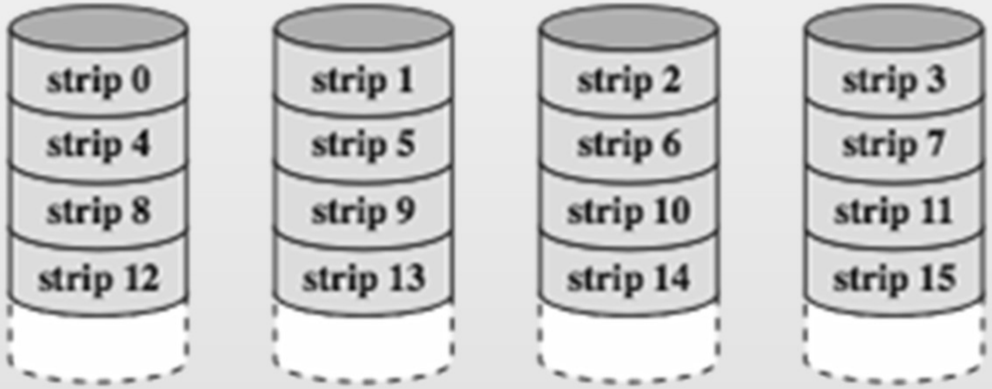
RAID



- RAID stands for Redundant Array of Independent Disks
- Consists of 7 levels
- Levels do not imply a hierarchical relationship but designate different design architectures that share three common characteristics:
  - Set of physical disk drives viewed by the operating system as a single logical drive
  - Data are distributed across the physical drives of an array in a scheme known as striping
  - Redundant disk capacity is used to store parity information, which guarantees data recoverability in case of a disk failure

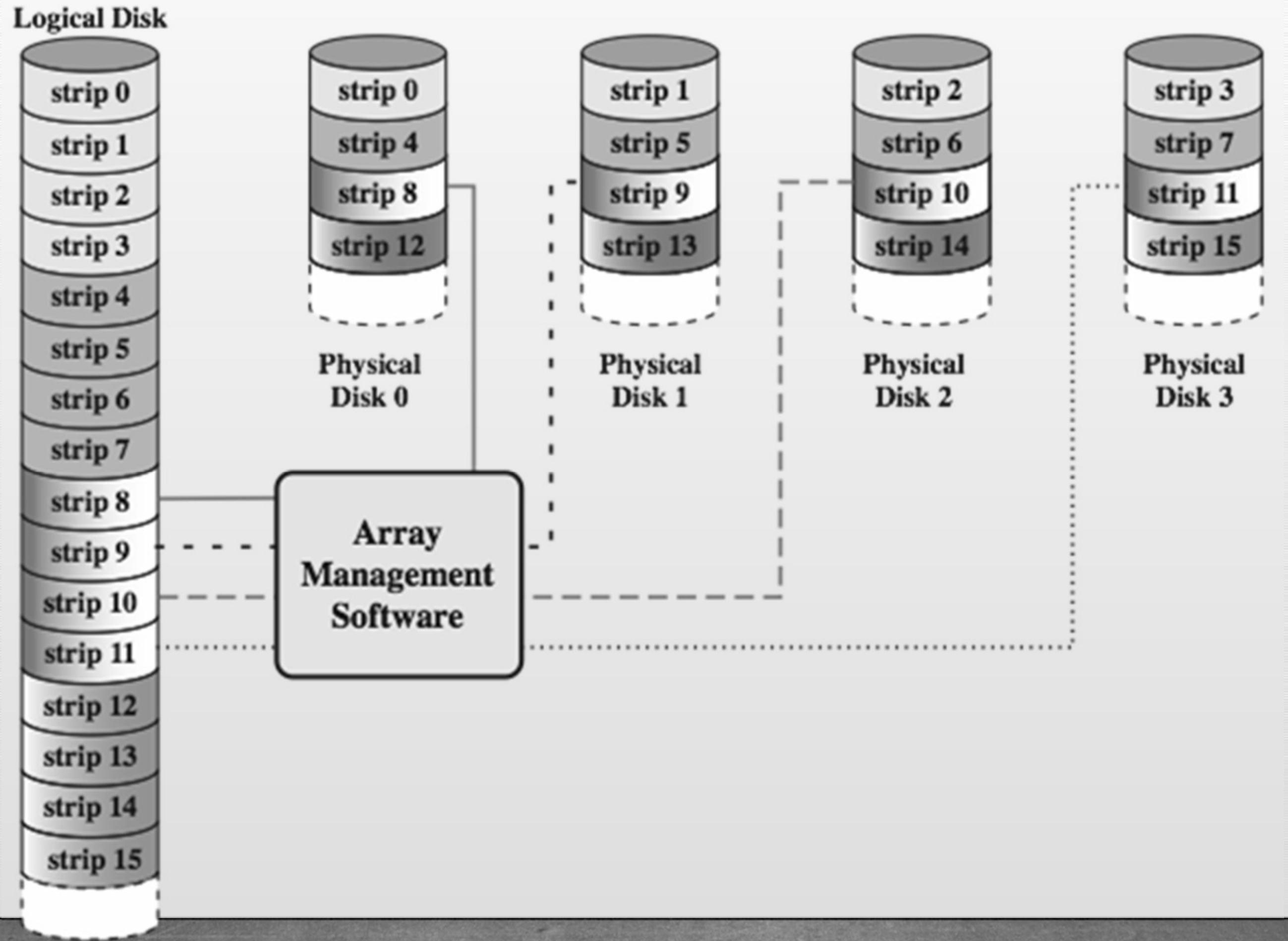
RAID Level 0

- Write consecutive “sectors” over the drives in a round robin fashion
  - Efficient when accessing a block of memory - parallelism
  - Non-redundant



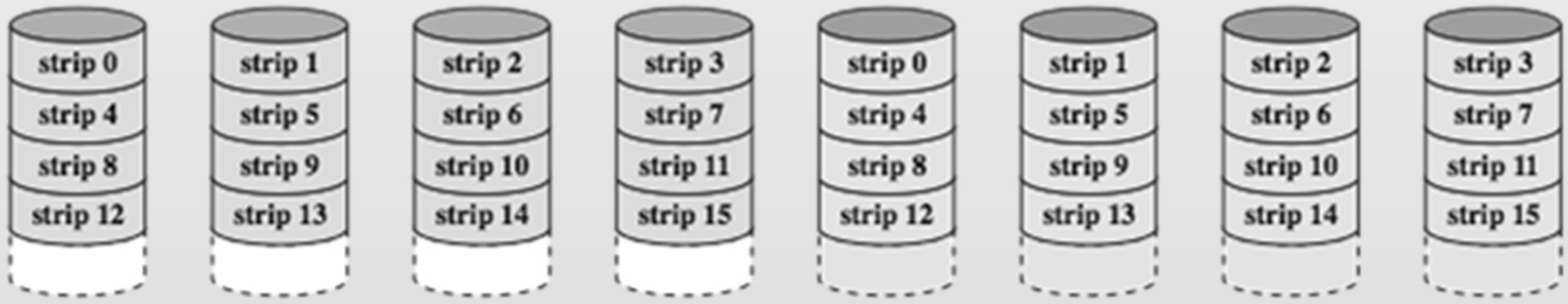
(a) RAID 0 (non-redundant)

RAID Level 0 (cont'd)



Raid Level 1

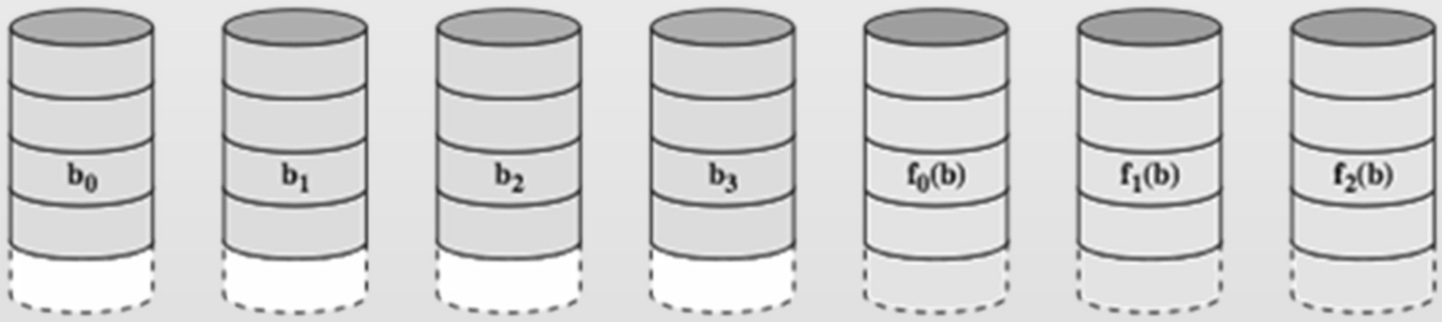
- Duplicate the disk for backup: (mirrored)
  - Fault tolerant
  - During reading, either copy can be used, hence reduced seek time
  - There is no “write penalty”
  - Data recovery is simple
  - Principal disadvantage is the cost



(b) RAID 1 (mirrored)

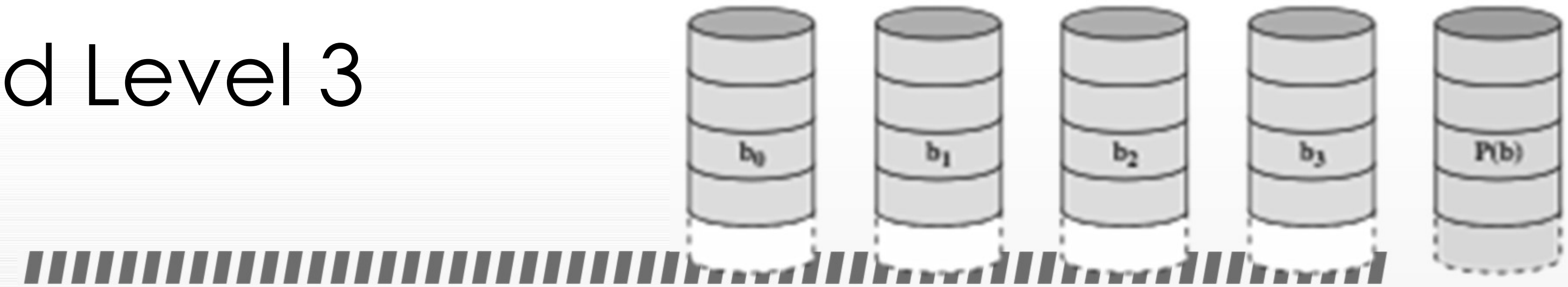
Raid Level 2

- Data strips are very small, often as small as a single byte or word
- Error Correction Code (Hamming code) embedded in extra HDDs
- Expensive, as multiple HDDs required, hence not implemented



(c) RAID 2 (redundancy through Hamming code)

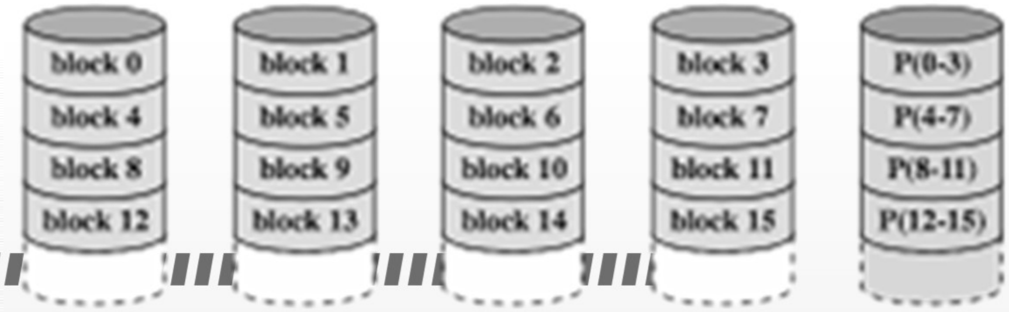
Raid Level 3



(d) RAID 3 (bit-interleaved parity)

- Instead of an error correcting code, a simple parity bit is computed for the set of individual bits in the same position on all of the data disks
- Data distributed in small strips
- The extra HDD contains the parity bit for the corresponding bits of all other HDDs
- If any HDD fails, can easily reconstruct the content for the failed HDD
- Return to full operation requires that the failed disk be replaced and the entire contents of the failed disk be regenerated (rebuild) on the new disk
- In a transaction-oriented environment performance suffers

Raid Level 4

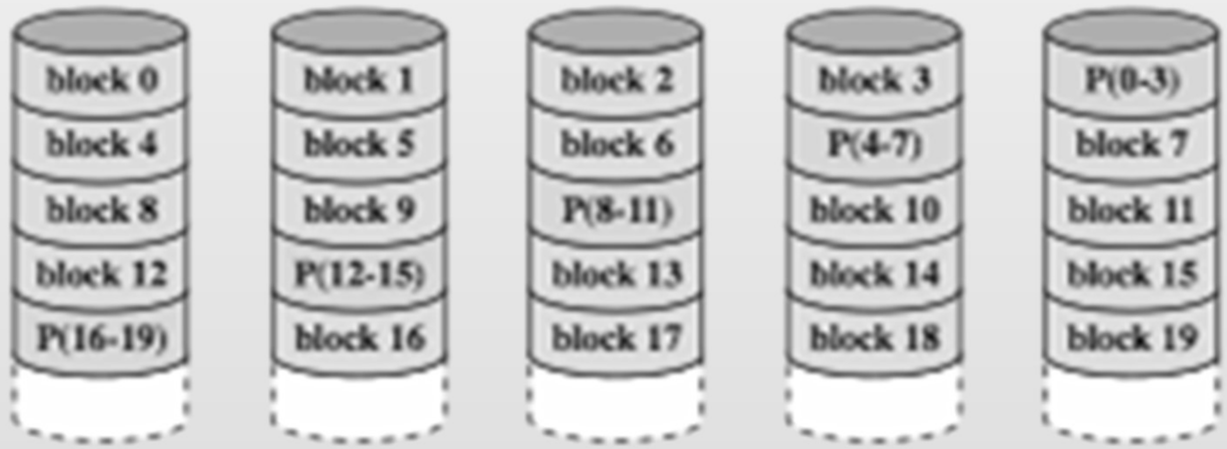


(e) RAID 4 (block-level parity)

- Data striping with relatively large strips
- To calculate the new parity the array management software must read the old user strip and the old parity strip
- Involves a write penalty when an I/O write request of small size is performed
- Each time a write occurs the array management software must update not only the user data but also the corresponding parity bits
- Thus each strip write involves two reads and two writes

Raid Level 5

- Organized in a similar fashion to RAID 4
- Difference is the distribution of the parity strips across all disks
- The distribution of parity strips across all drives avoids the potential I/O bottleneck found in RAID 4



(f) RAID 5 (block-level distributed parity)

Raid Level 6

- Similar to Level 5 RAID, but use two Parity Strips, calculated using different methods
- The two Parity Strips are distributed across different HDDs
- Need two extra HDDs
- Can endure two HDD failure

(g) RAID 6 (dual redundancy)

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
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 ingle 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
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	Provided 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

Summary of RAID Level

Category	Level	Description	Disk Required	Data availability	Large I/O Data Transfer Capacity	Small I/O Request Rate
Striping	0	Non-redundant	N	Lower than single disk	Very high	Very high for both read and write
Mirroring	1	Mirrored	2N	Higher than RAID 3 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	3	Bit-interleaved parity	N+1	Much higher than single disk; comparable to RAID 5	Highest of all listed alternatives	Approximately twice that of a single disk
Independent access	5	Block-interleaved parity	N+1	Much higher than single disk; comparable to RAID 3	Similar to RAID) 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