External Memory

Types of External Memory

- Magnetic Disk
  - Fixed disc
  - Removable disc
  - RAID
- Optical
  - CD-ROM
  - CD-Recordable (CD-R)
  - CD-R/W Compact Disc-ReWritable
  - DVD („Digital Versatile Disc„ or „Digital Video Disc")
- Magnetic Tape

DVD

- DVD (an abbreviation of "digital versatile disc„ or "digital video disc") is a digital optical disc storage format invented and developed by Philips, Sony, Toshiba, and Panasonic in 1995. The medium can store any kind of digital data and is widely used for <u>software and other computer files as well as video programs</u> watched using DVD players.

DVD Types

- One of the most common DVDs is the single-sided, single-layer disc, capable of holding 4.7 GB.
- The single-sided, double-layer disc is capable of holding between 8.5-8.7 GB.
- The double-sided, single-layer disc is capable of holding 9.4 GB.
- Although rare, the double-sided, double-layer disc is capable of holding up to 17.08 GB.
- At present: 25 GB / 50 GB.

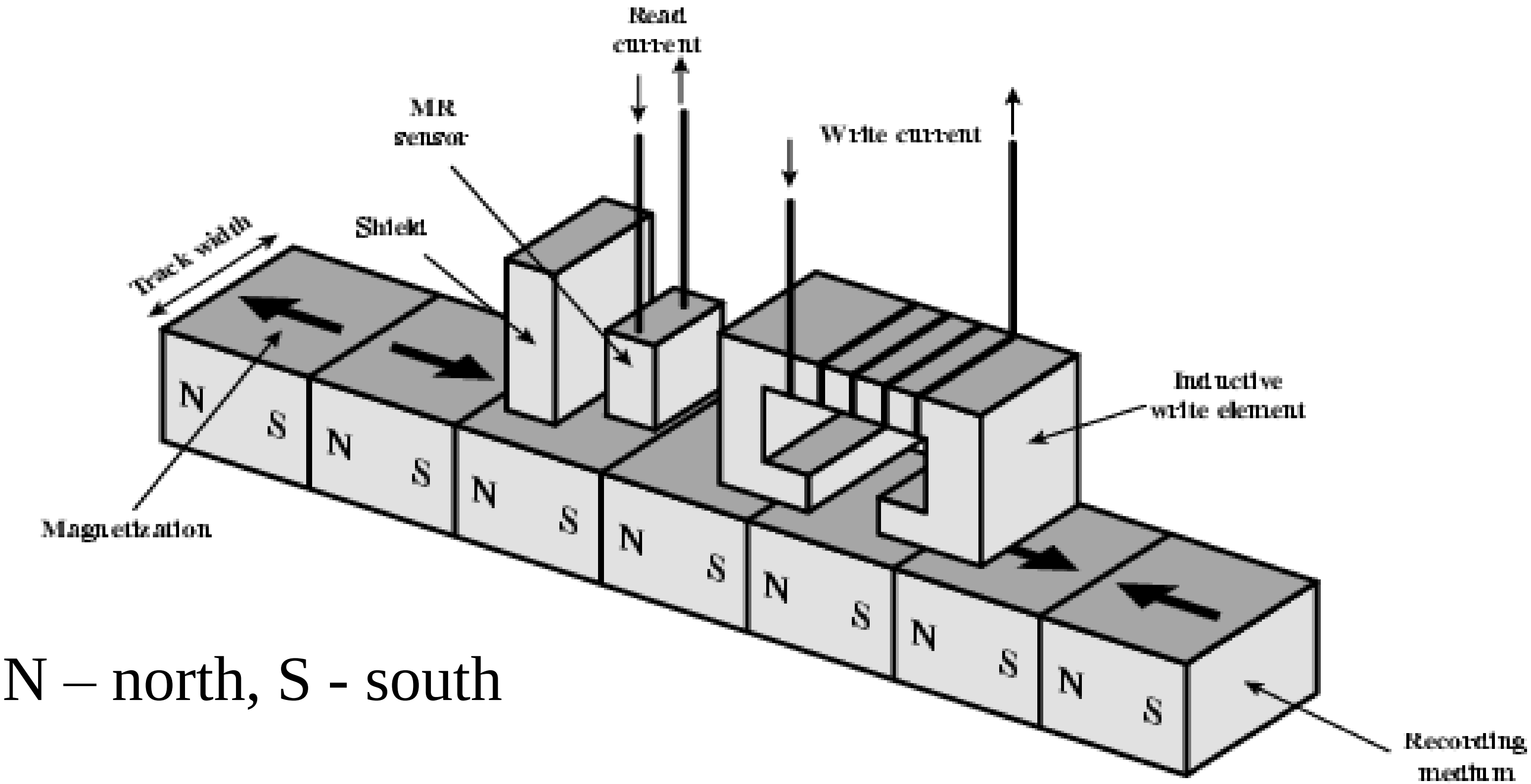
Magnetic Disk

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

Read and Write Mechanisms

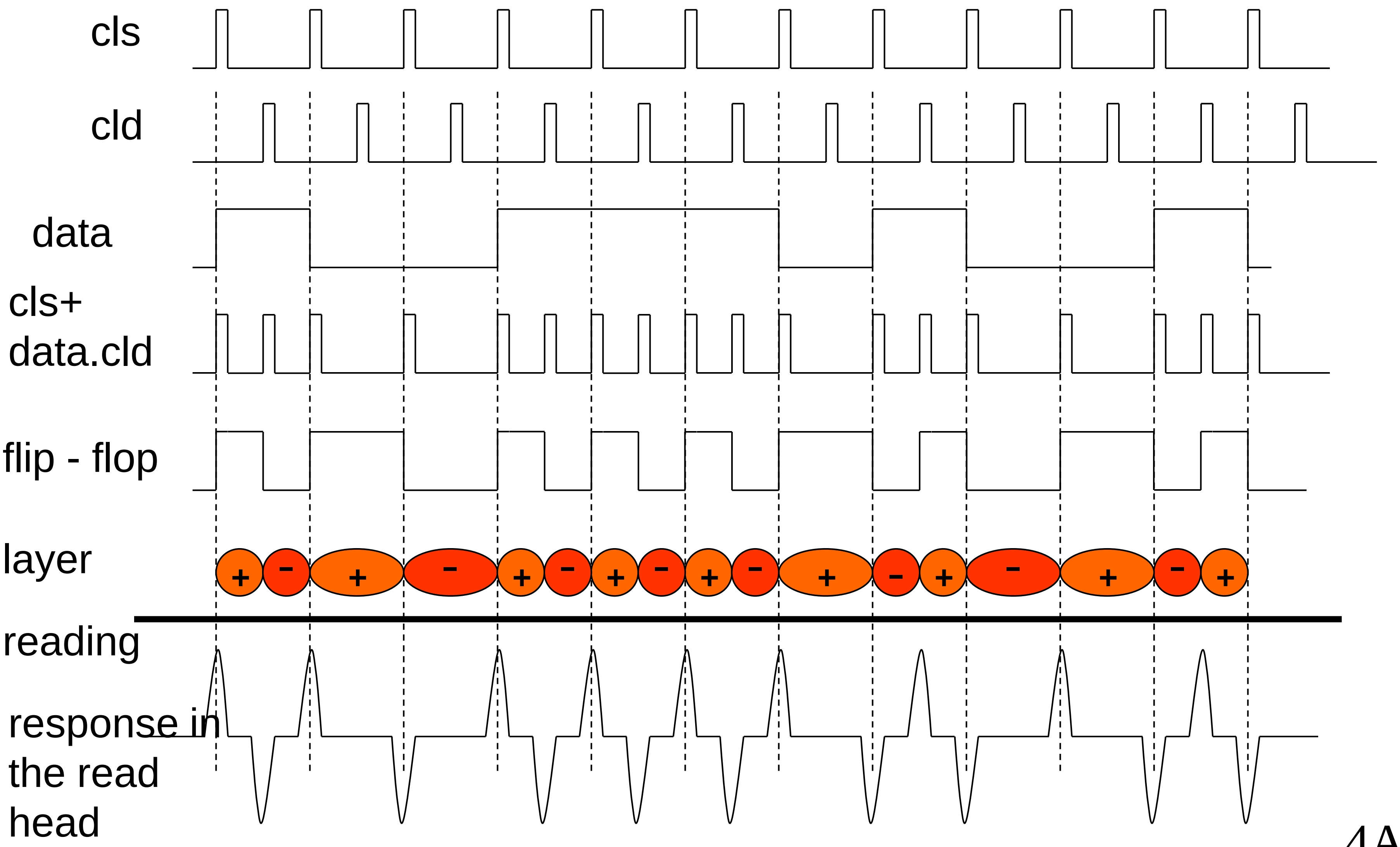
- Recording & retrieval via conductive coil called a head
- May be single read/write head or separate ones
- During read/write, head is stationary, platter rotates
- Write
  - Pulses sent to head
  - Current through coil produces magnetic field
  - Magnetic pattern is recorded on surface below head
- Read (traditional)
  - Magnetic field moving relative to coil induces voltage in head
  - Coil is the same for read and write
- Read (contemporary)
  - Separate read head, close to write head
  - Partially shielded magneto resistive (MR) sensor
  - <u>Electrical resistance depends on the direction of magnetic field</u>
  - High frequency operation
    - Higher storage density and speed

Inductive Write MR Read



N – north, S - south

FM recording (Frequency Modulation)

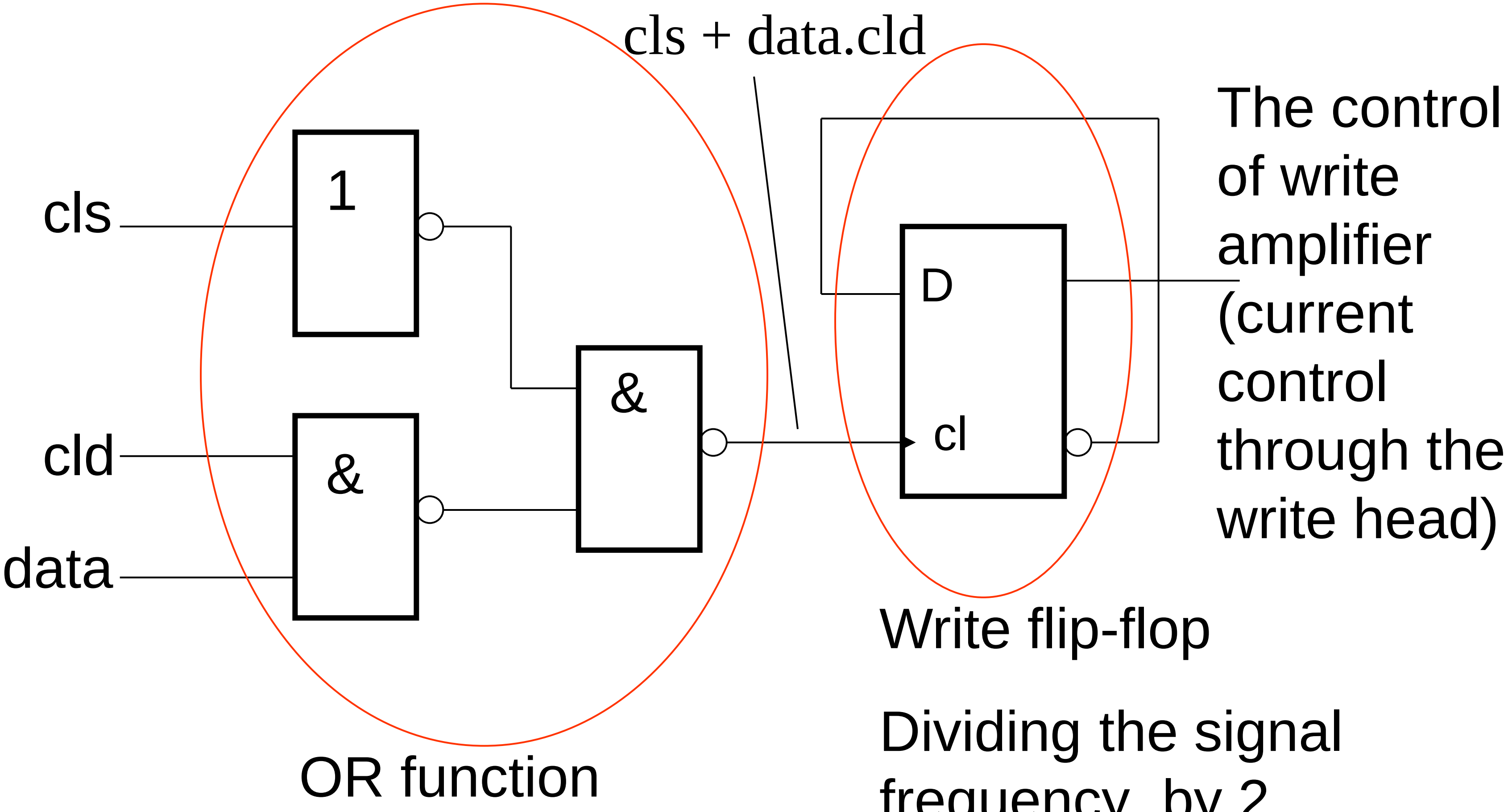


The principles of recording data in FM methodology

Recording „0“ – just cls is included into the resulting signal

Recording „1“ – cld is also included into the resulting signal, resulting in higher frequency of recorded signal (2x) – the worst sample of data.

Encoding data for write procedure

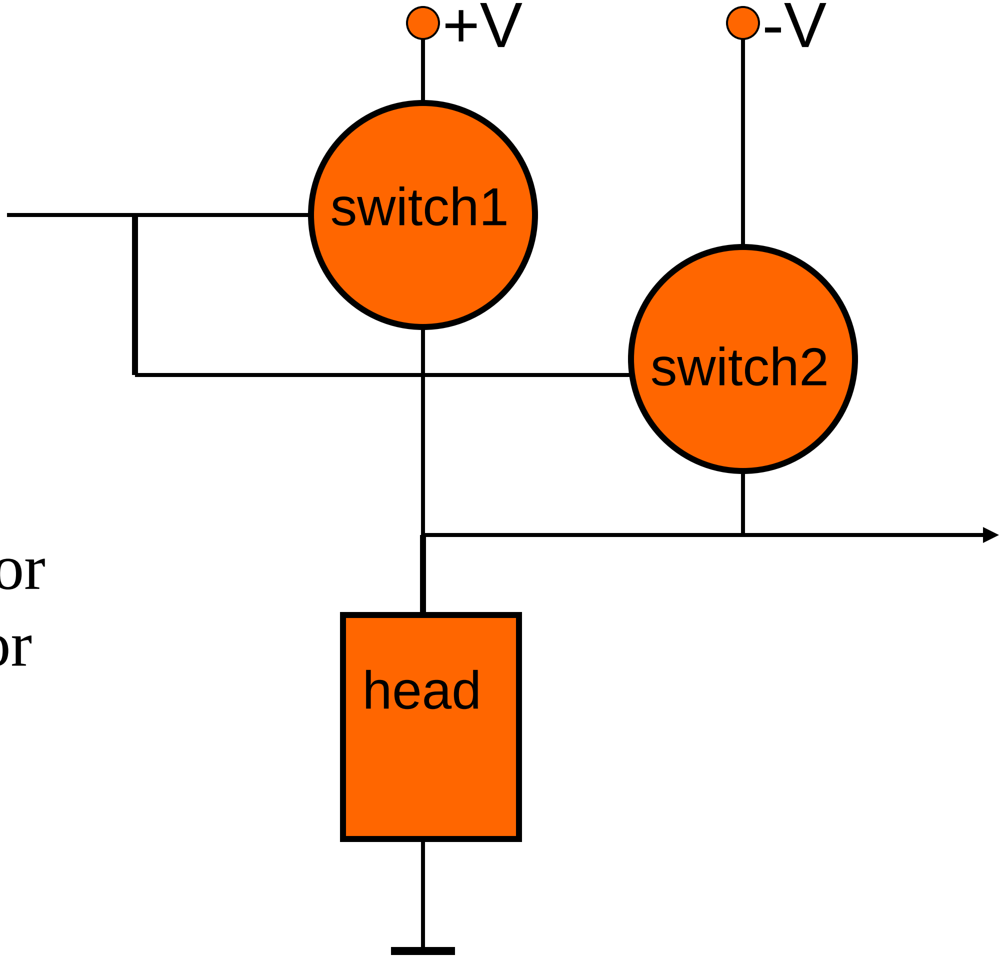


The control of write amplifier (current control through the write head)

Dividing the signal frequency by 2

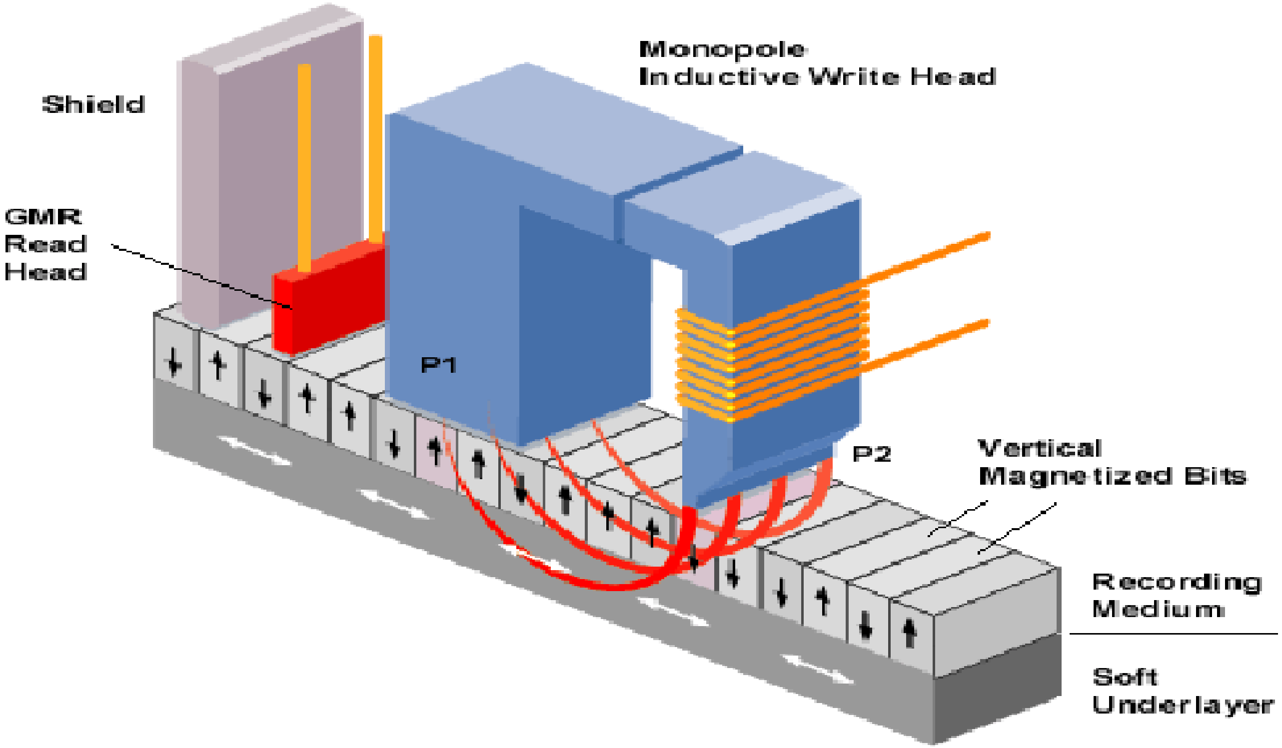
Control of the current through the head during write operation

The output of the write flip-flop



switch 1 – NPN transistor
switch 2 – PNP transistor

Perpendicular Magnetic Recording - PMR

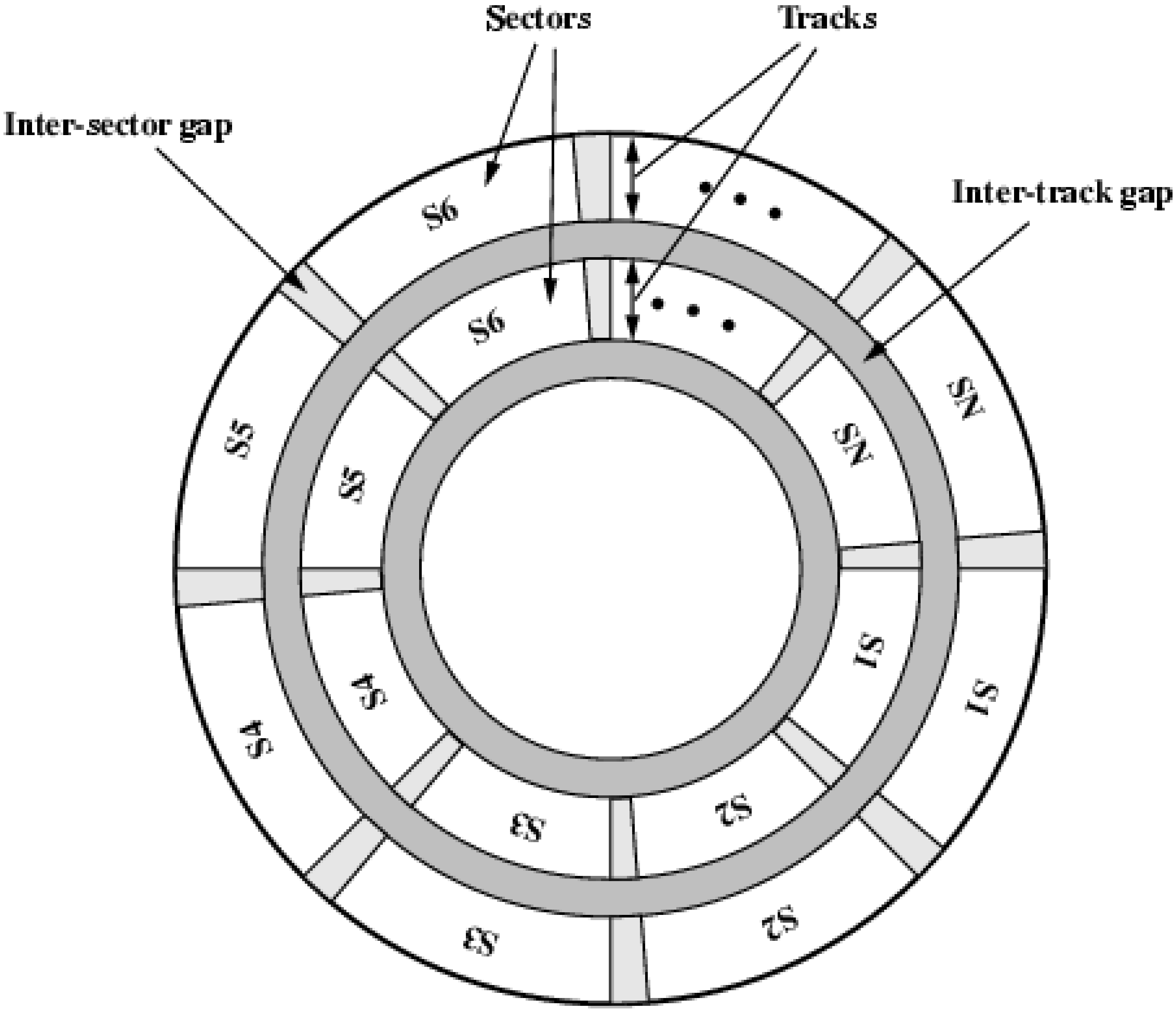


By means of PMR, significantly higher capacity can be reached.

Data Organization and Formatting

- Concentric rings or tracks
  - —Gaps between tracks (track density)
  - —Track density: 30 000 tracks per inch (TPI)
  - —Reduced gap to increase capacity
  - —Same number of bits per track (variable packing density) – not true now
  - —Constant angular velocity
- Tracks divided into sectors.
- One sector is the minimum block size.

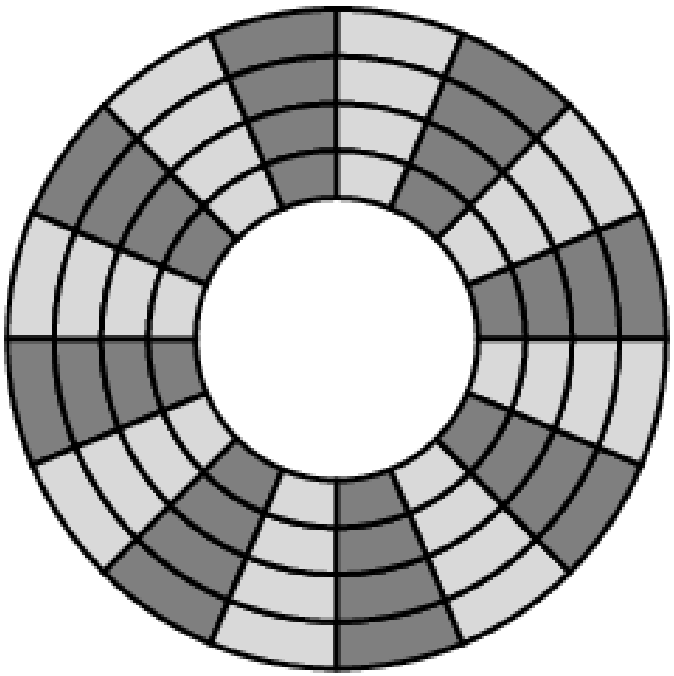
Disk Data Layout



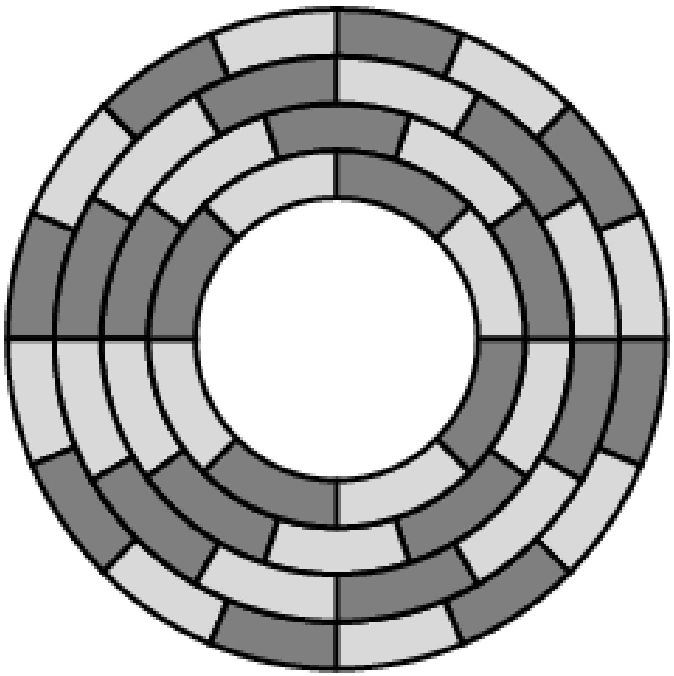
Disk Velocity

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

Disk Layout Methods Diagram



(a) Constant angular velocity

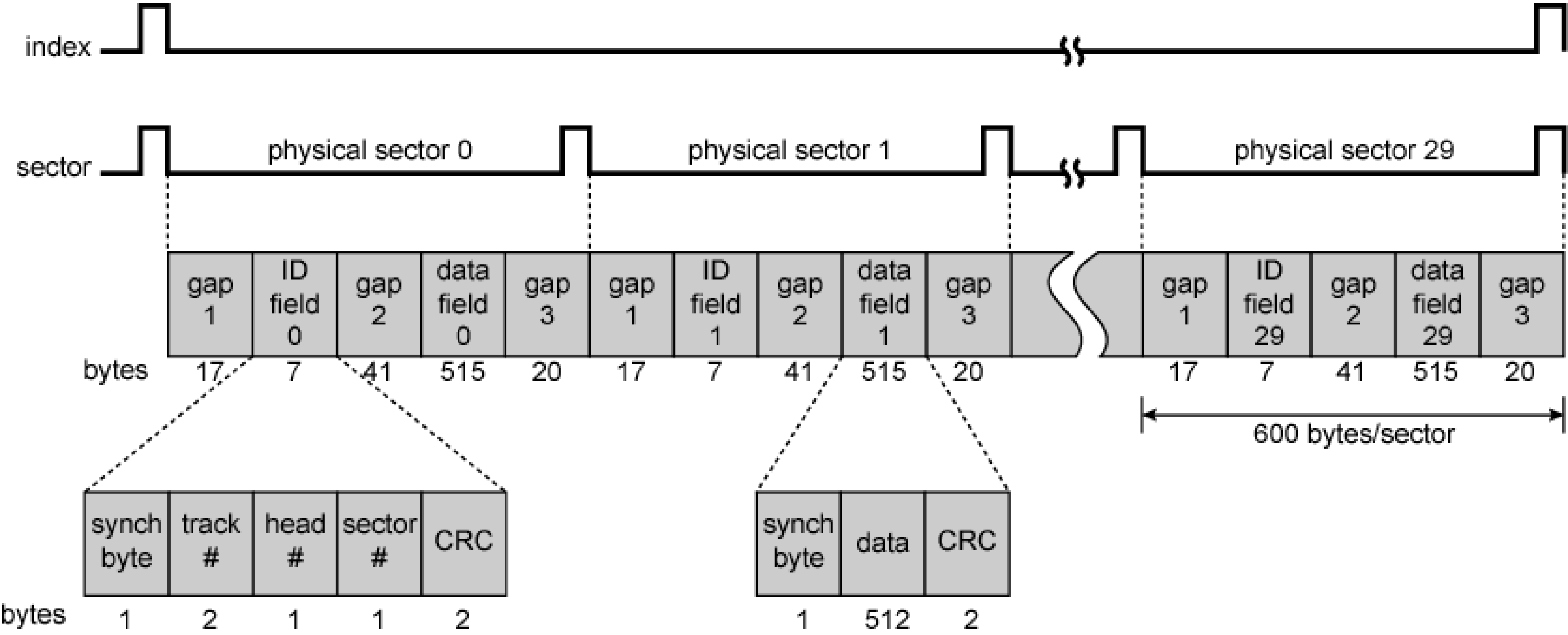


(b) Multiple zoned recording

Finding Sectors

- Must be able to identify start of track and sector.
- Format disk
  - Additional information not available to user
  - Marks tracks and sectors

Winchester Disk Format
Seagate ST506



Characteristics

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

Fixed/Movable Head Disk

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

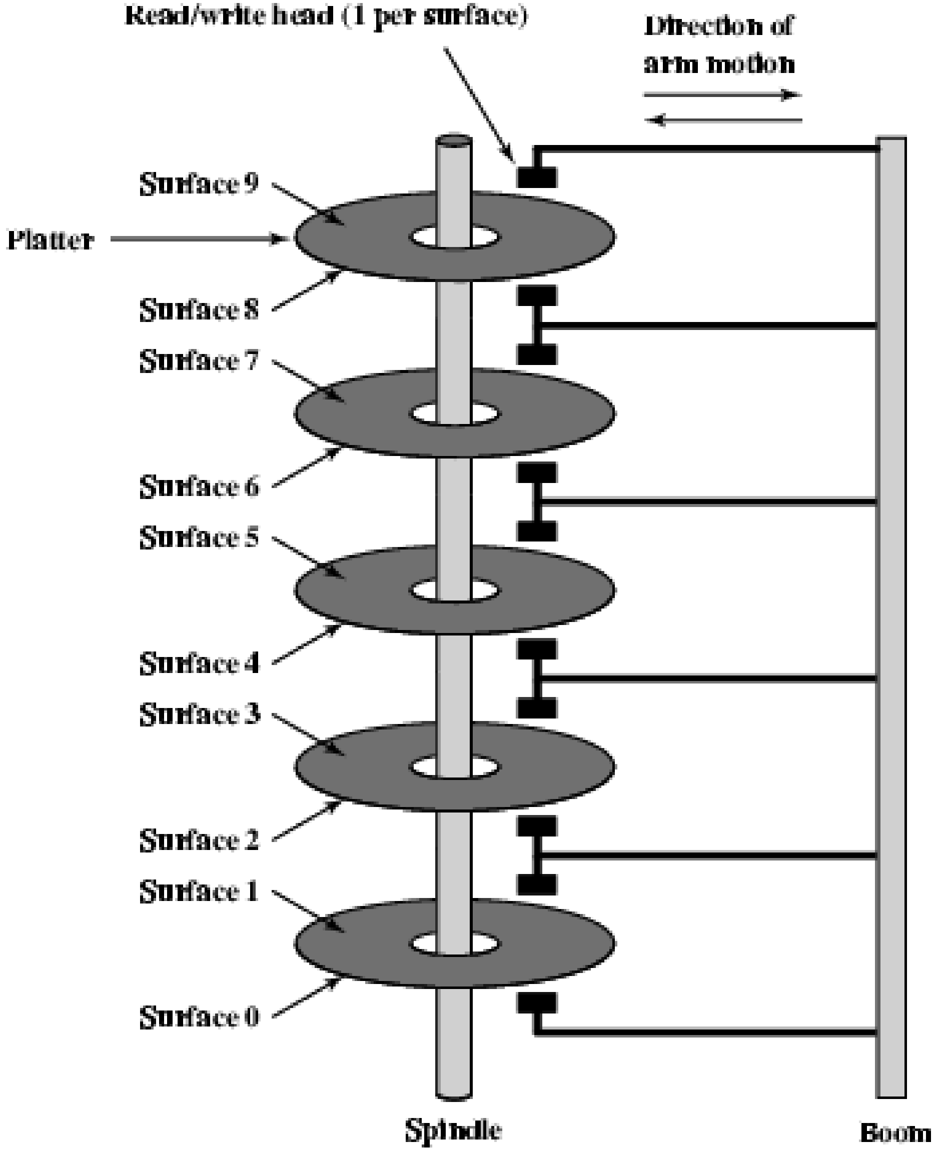
Removable or Not

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

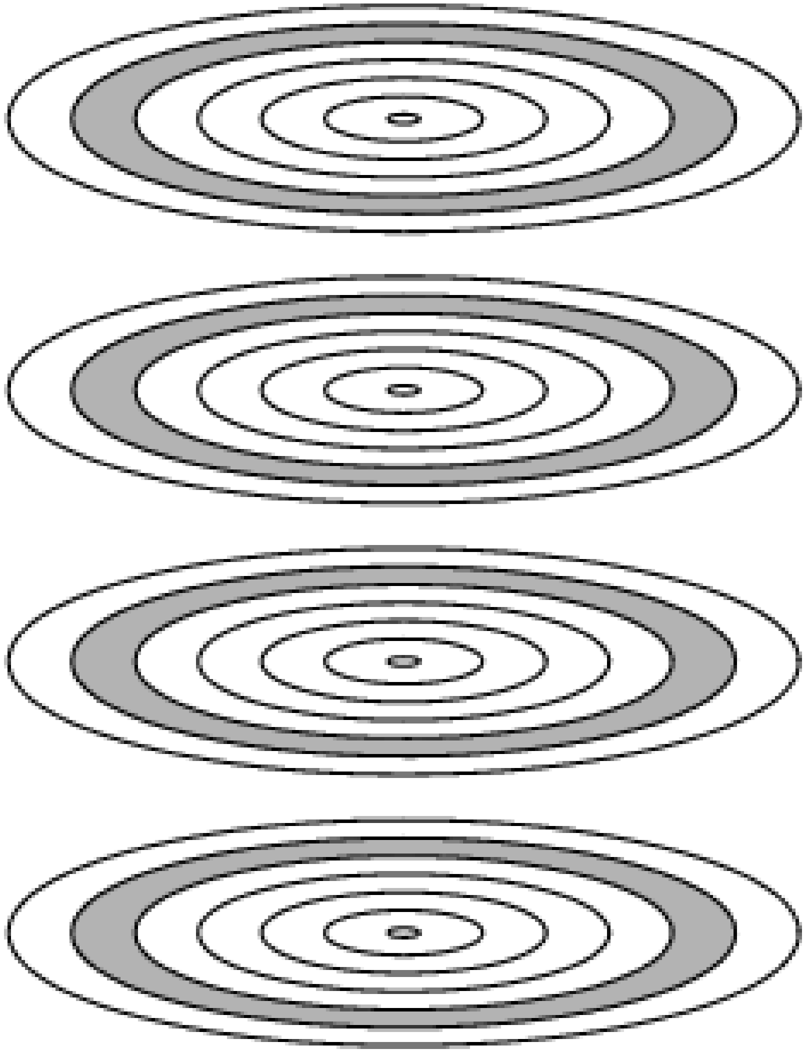
Multiple Platter

- One head per side
- Heads are joined and aligned together
- Aligned tracks on each platter form cylinders
- Data is organized into cylinders
  - it reduces head movement
  - it increases speed (transfer rate)

Multiple Platters



Tracks and Cylinders



Floppy Disk

- 8”, 5.25”, 3.5”
- Small capacity
  - Up to 1.44Mbyte (2.88M never popular)
- Slow
- Universal
- Cheap
- Obsolete, not used now.

Winchester Hard Disk (1)

- Developed by IBM in Winchester (USA).
- Sealed unit (disc is invisible).
- One or more platters (disks).
- Heads fly on boundary layer of air as disk spins.
- Very small head to disk gap.

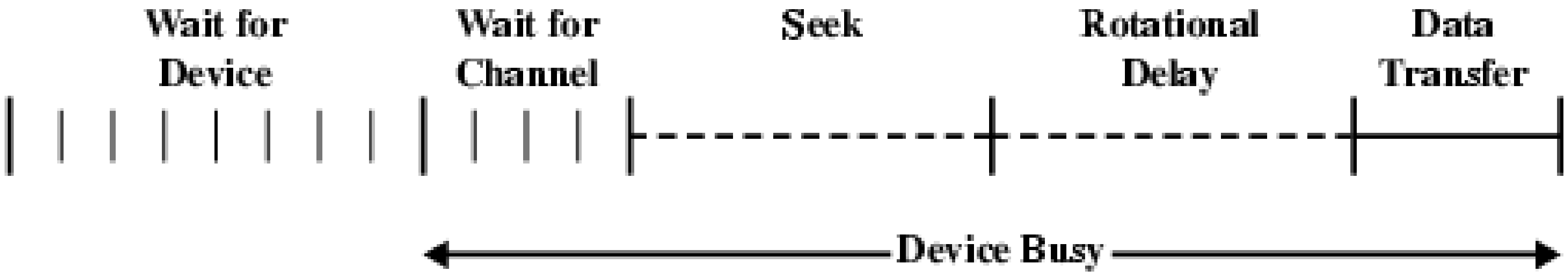
Winchester Hard Disk (2)

- Universal
- Cheap
- Fastest external storage
- Getting larger all the time
  - Terabytes now easily available (up to 10 TB)

Speed

- Seek time
  - Moving head to correct track
- (Rotational) latency
  - Waiting for data to rotate under head
- Access time = Seek time + Rotational delay
- Transfer rate

Timing of Disk I/O Transfer



RAID Architectures

- Redundant Array of Independent Disks (correct).
- Redundant Array of Inexpensive Disks.
- 6 levels in common use.
- Not a hierarchy.
- The set of physical disks viewed by Operating System as single logical drive.
- Data distributed across physical drives.
- The architecture uses redundant capacity to store parity information.
- Two goals:

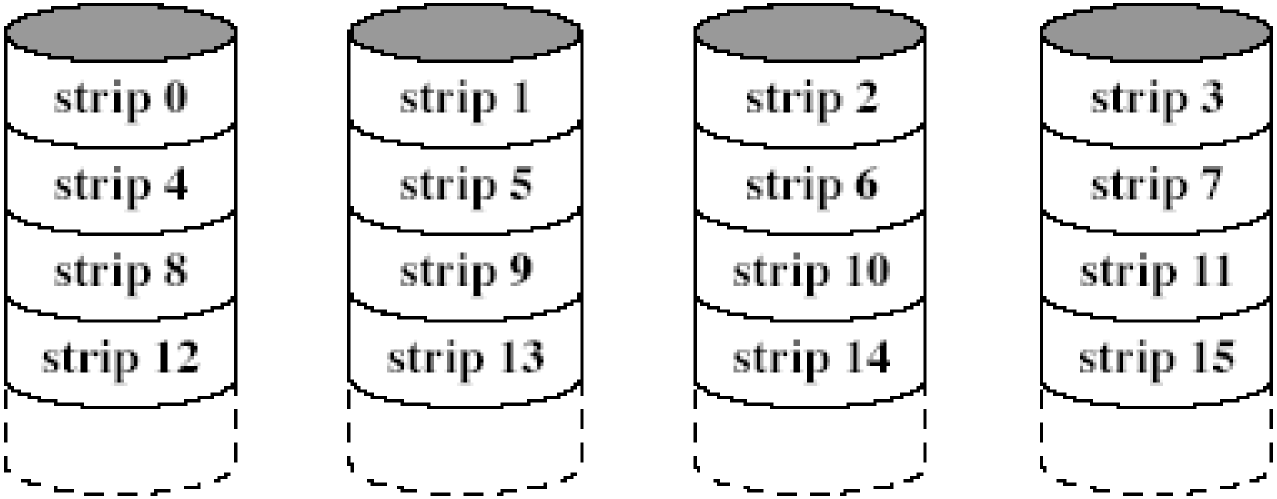
  to reduce the access time

  to increase reliability

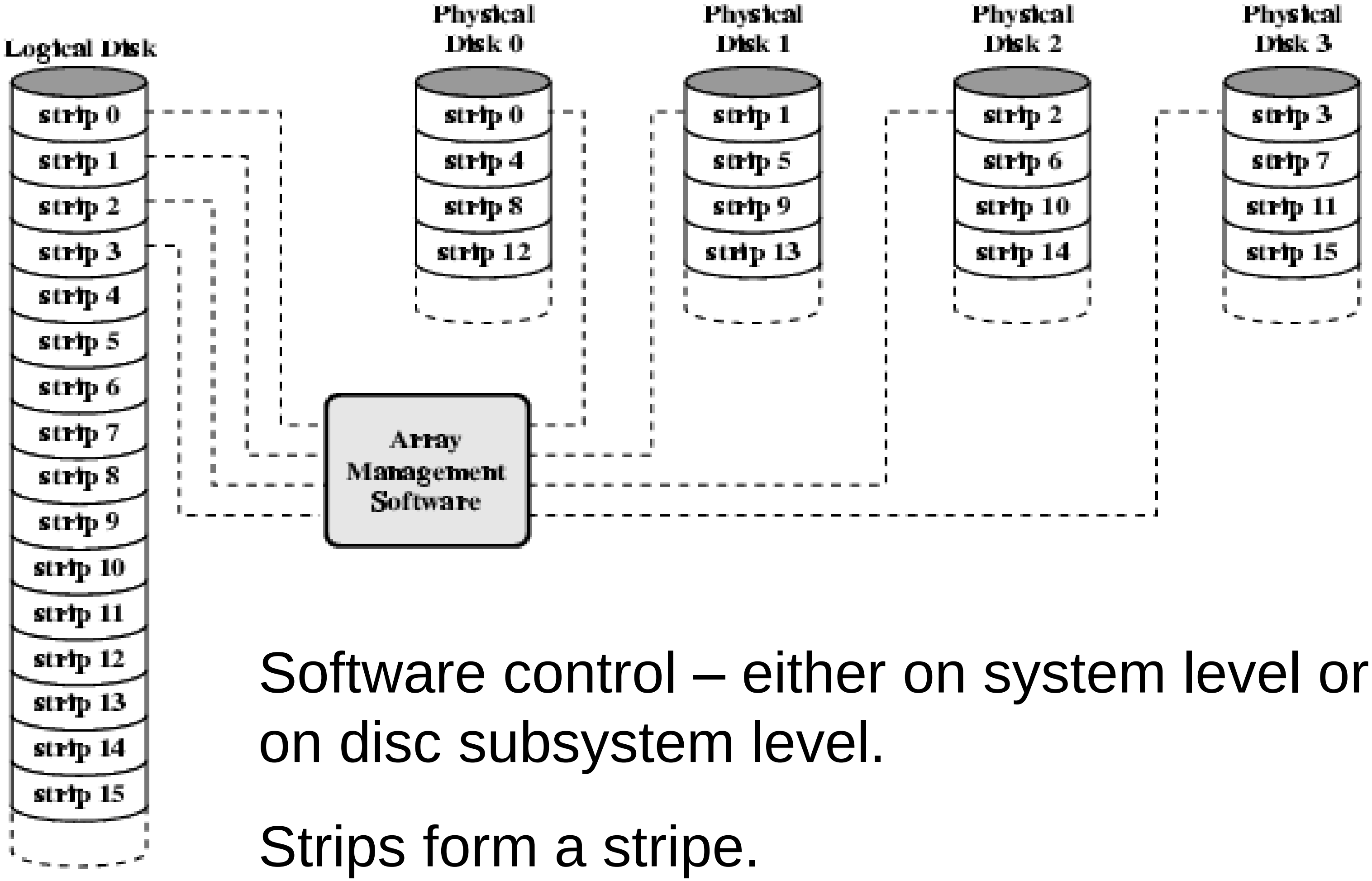
RAID 0

- No redundancy.
- Data striped across all disks.
- The increase of speed:
  - Multiple data requests not on the same disk.
  - Disks seek in parallel.
  - A set of data is striped across multiple disks.

RAID 0


strip 0
strip 1
strip 2
strip 3
strip 4
strip 5
strip 6
strip 7
strip 8
strip 9
strip 10
strip 11
strip 12
strip 13
strip 14
strip 15

Data Mapping for RAID 0



Software control – either on system level or on disc subsystem level.

Strips form a stripe.

RAID 0

- How the operating system cooperates with RAID disks:

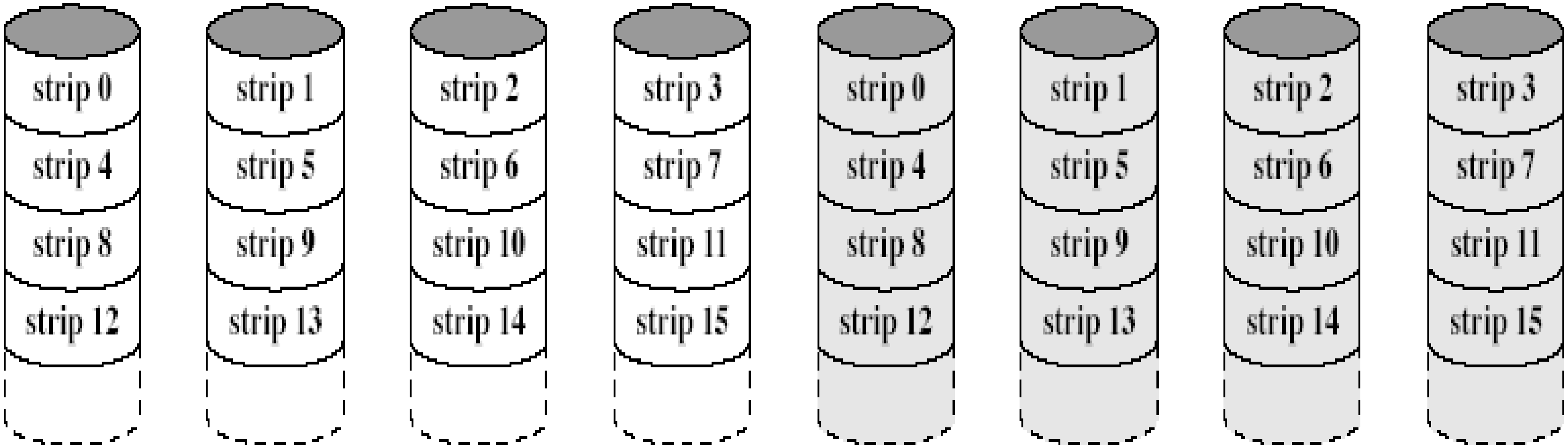
  - the operating system generates „read“ command (data are organized in sequence on strips),

  - RAID controller divides the command to 4 separate commands and sends them to disc controllers,

  - the disc controllers <u>perform the commands in parallel</u>,

  - the best organization of data – data between which a link exists are stored in adjacent strips, i. e. in one stripe (e.g. a file) – data are transferred in parallel.

RAID 1


strip 0
strip 4
strip 8
strip 12
strip 1
strip 5
strip 9
strip 13
strip 2
strip 6
strip 10
strip 14
strip 3
strip 7
strip 11
strip 15
strip 0
strip 4
strip 8
strip 12
strip 1
strip 5
strip 9
strip 13
strip 2
strip 6
strip 10
strip 14
strip 3
strip 7
strip 11
strip 15

RAID 1

- Mirrored disks and backed up.
- 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 (install new disk) & re-mirror
  - No down time (one disk works permanently)
- Expensive.

RAID 1

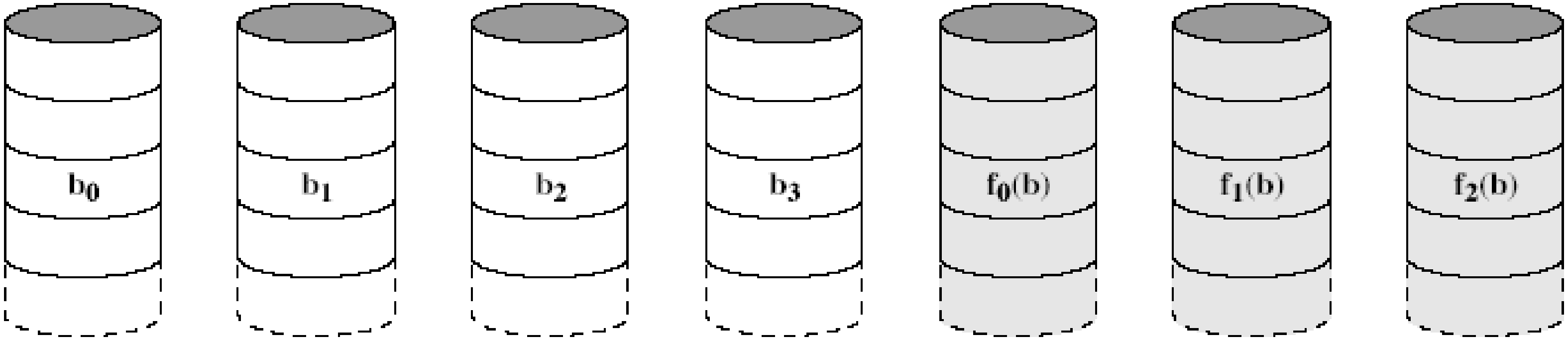
Read data operation – just from one disc (that one which will be faster in search for data (access time = seek time + rotational delay).

Write data operation – performed in parallel to both discs → the performance during write operation will be strongly influenced by the slower disc.

Recovery after fault – data is read from the disc which operates correctly.

Disadvantage – high costs.

RAID 2

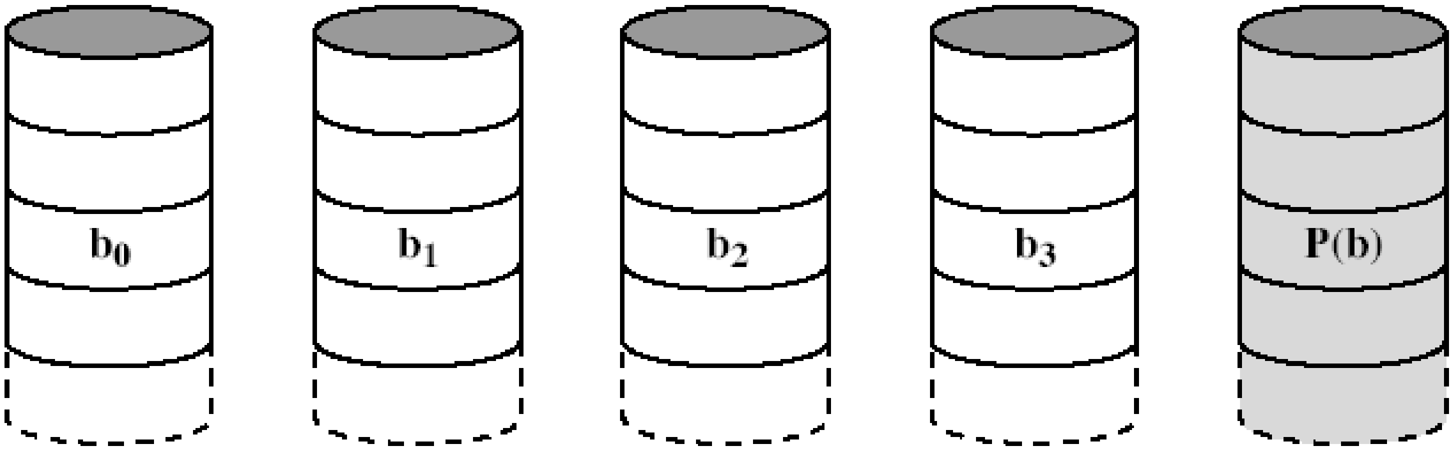


4 discs – data, 3 discs – information allowing to correct data

RAID 2

- The disks are synchronized – the heads are in the same position on all disks (cylinder, rotation).
- Parallel access – all disks are involved in processing I/O commands (they operate in parallel).
- Error correction calculated across corresponding bits on disks.
- Multiple parity disks store Hamming code error correction in corresponding positions.
- Lots of redundancy
  - Expensive
  - Not widely used

RAID 3



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.**

RAID 3

- If a fault appears on a disk, then parity bit is read and data are reconstructed from the remaining bits, the parity bit is used for the reconstruction with other bits.
- After the faulty disk is removed and non-faulty (i. e. new) disc installed, then data can be reconstructed from the remaining bits.
- Data reconstruction:

  - Let RAID system consists of 5 disks:

    X0 – X3 - data
- X4 – parity bit
- How the parity bit is calculated:

  X4(i) = X3(i) xor X2(i) xor X1(i) xor X0(i)

RAID 3

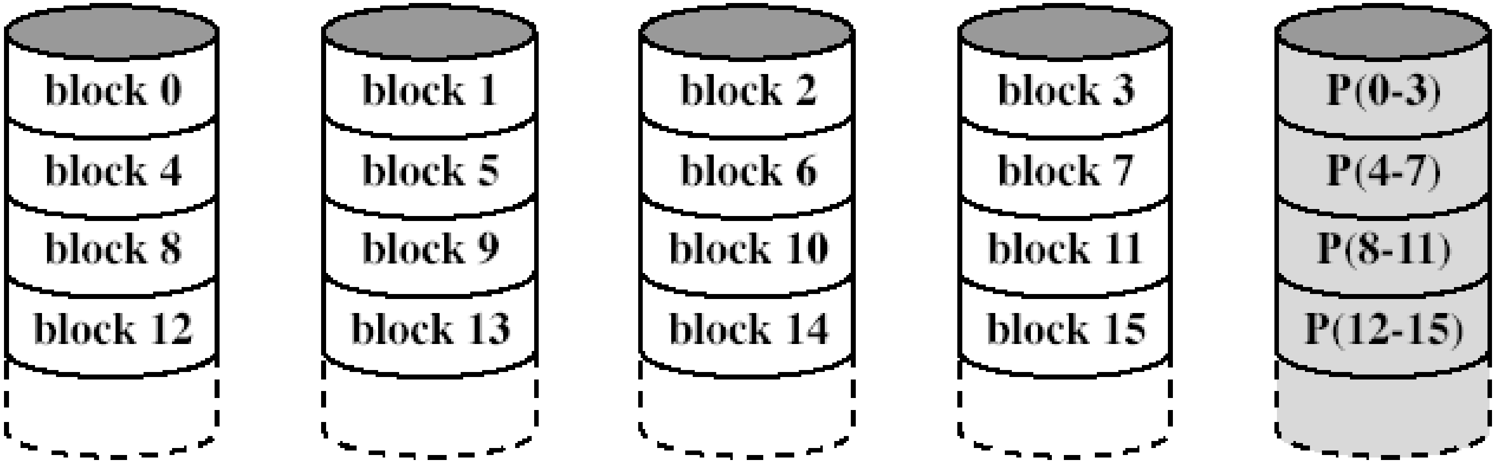
- An example: a failure appeared on disk X1, data stored are not available.
- The parity :

  X4(i) = X3(i) xor X2(i) xor X1(i) xor X0(i)
- To both sides of the equation X4(i) xor X1(i) is xored:

  We receive:

  X4(i) xor X4(i) xor X1(i) = X3(i) xor X2(i) xor X1(i) xor X0(i) xor X4(i) xor X1(i),
- X4(i) xor X4(i) is always equal to 0, the same holds for X1(i) xor X1(i)
- Then: X1(i) = X4(i) xor X3(i) xor X2(i) xor X0(i)
- Thus, the value of the bit from the disk with failure is calculated from remaining bits.
- **The conclusion:** <u>the value of the bit from the disc with a failure is derived from the remaining bits. This principle is used for RAID3 - RAID6 architectures</u>.

RAID 4


block 0
block 1
block 2
block 3
P(0-3)
block 4
block 5
block 6
block 7
P(4-7)
block 8
block 9
block 10
block 11
P(8-11)
block 12
block 13
block 14
block 15
P(12-15)

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

The use of redundancy for data recovery

- The situation: we need to write data to one disk only – how the parity written to disk is derived?
- The field consists of 5 disks: X0 – X3 – data, X4 – parity evaluated in the following way:

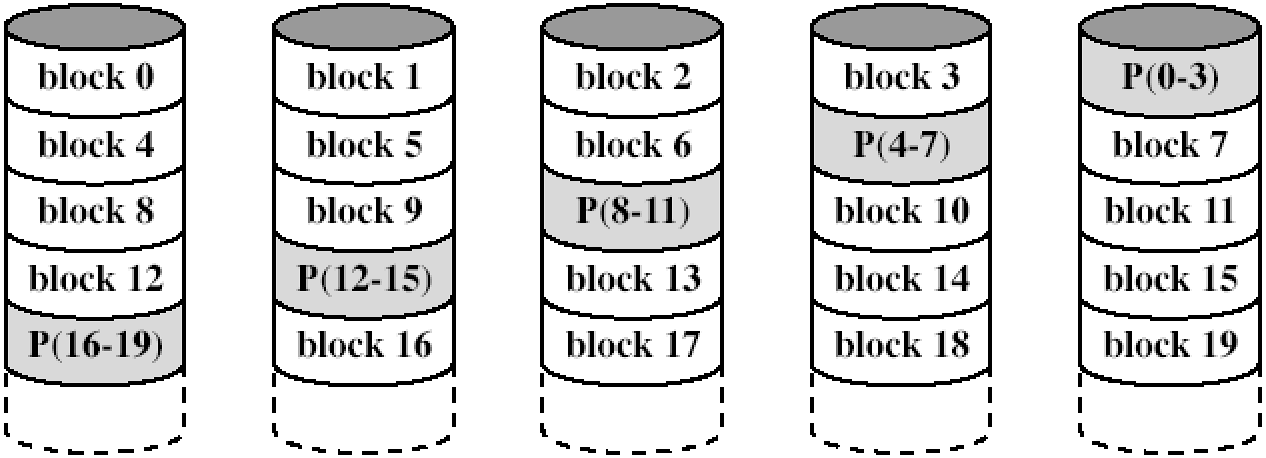
  $X4(i) = X3(i) \text{ xor } X2(i) \text{ xor } X1(i) \text{ xor } X0(i)$
- How the new parity is evaluated when the value of bit X1 is changed:

  $X4'(i) = X3(i) \text{ xor } X2(i) \text{ xor } X1'(i) \text{ xor } X0(i) = X3(i) \text{ xor } X2(i) \text{ xor } X1'(i) \text{ xor } X0(i) \text{ xor } X1(i) \text{ xor } X1(i) =$

  $= X3(i) \text{ xor } X2(i) \text{ xor } X1(i) \text{ xor } X0(i) \text{ xor } X1(i) \text{ xor } X1'(i) =$

  $= X4(i) \text{ xor } X1(i) \text{ xor } X1'(i)$
- The conclusion: to evaluate the new parity, then old parity value, original value and new value of the bit are used.
- To the disk data bit and parity are recorded.

RAID 5

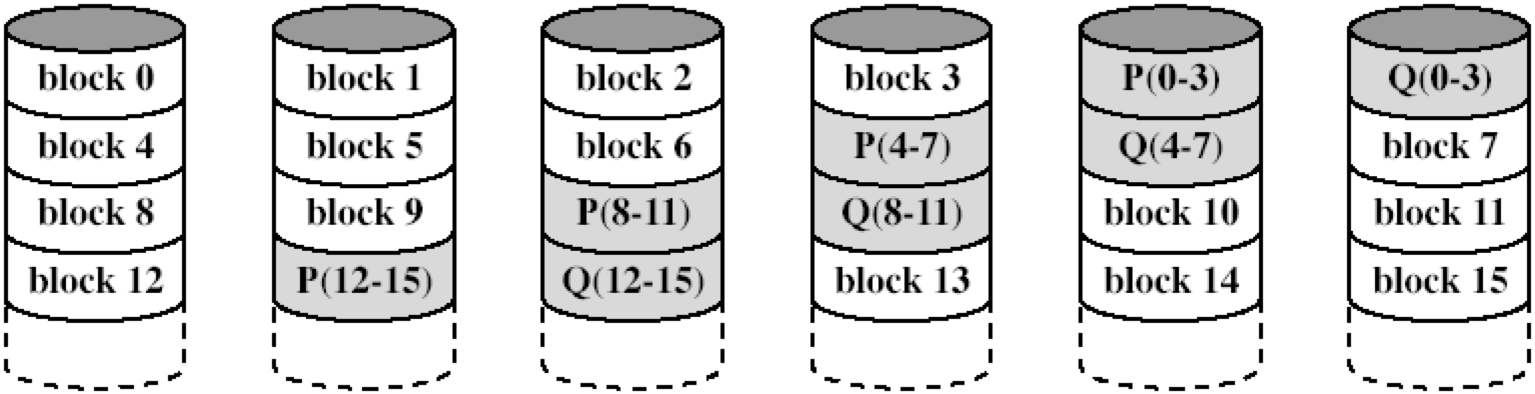


„block-level distributed“ parity type

RAID 5

- Like RAID 4.
- Parity striped across all disks.
- Avoids RAID 4 bottleneck at parity disk.
- Commonly used in network servers.

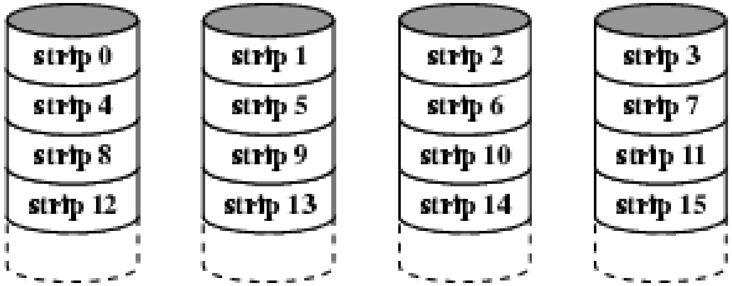
RAID 6


block 0
block 4
block 8
block 12
block 1
block 5
block 9
P(12-15)
block 2
block 6
P(8-11)
Q(12-15)
block 3
P(4-7)
Q(8-11)
block 13
P(0-3)
Q(4-7)
block 10
block 14
Q(0-3)
block 7
block 11
block 15

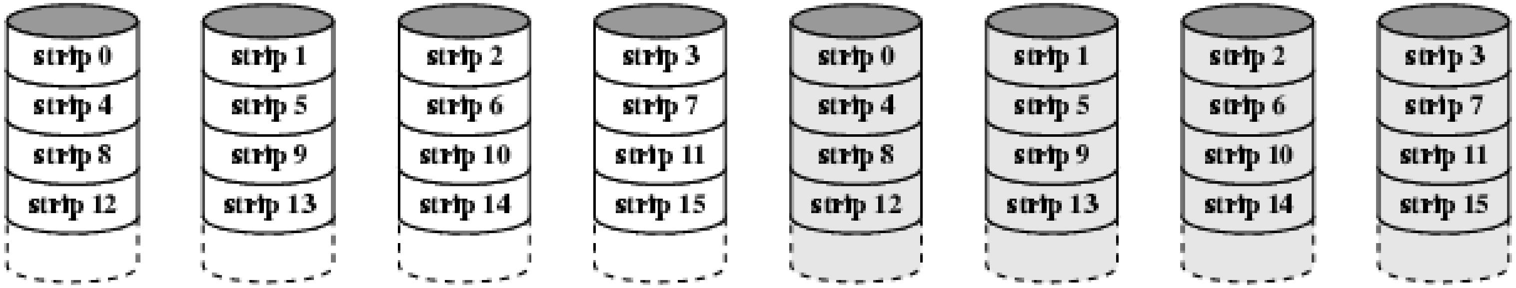
RAID 6

- 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

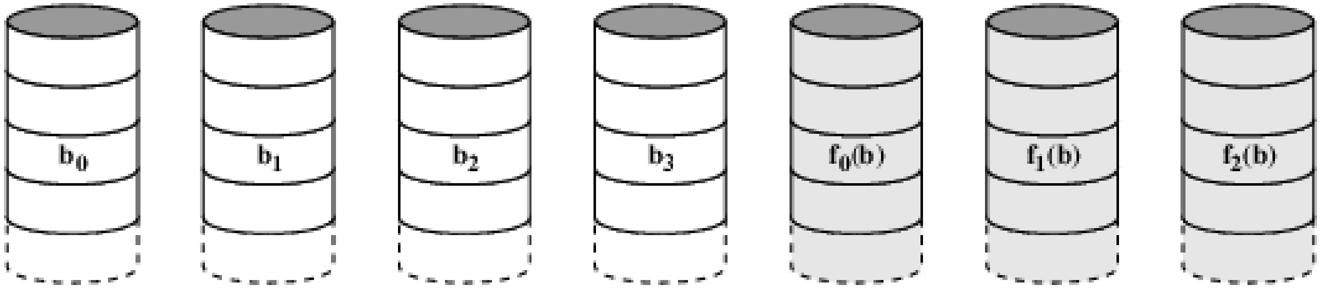
RAID 0, 1, 2



(a) RAID 0 (non-redundant)

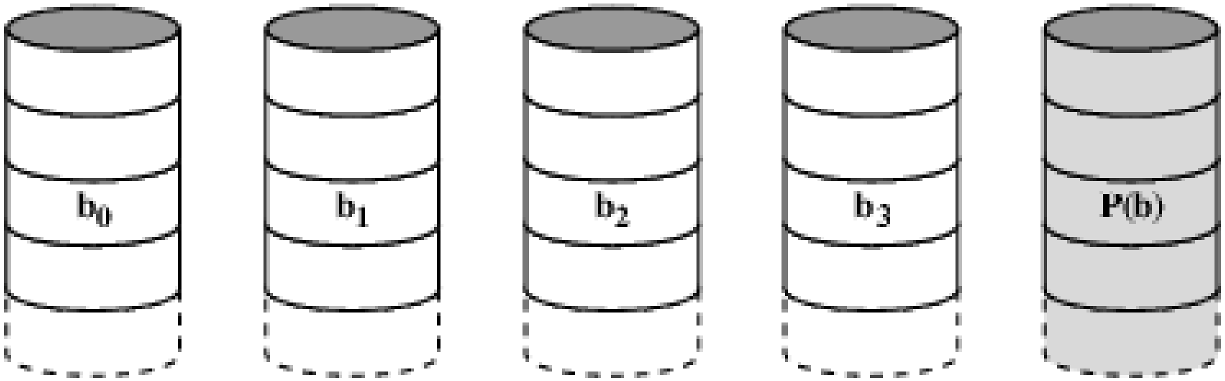


(b) RAID 1 (mirrored)

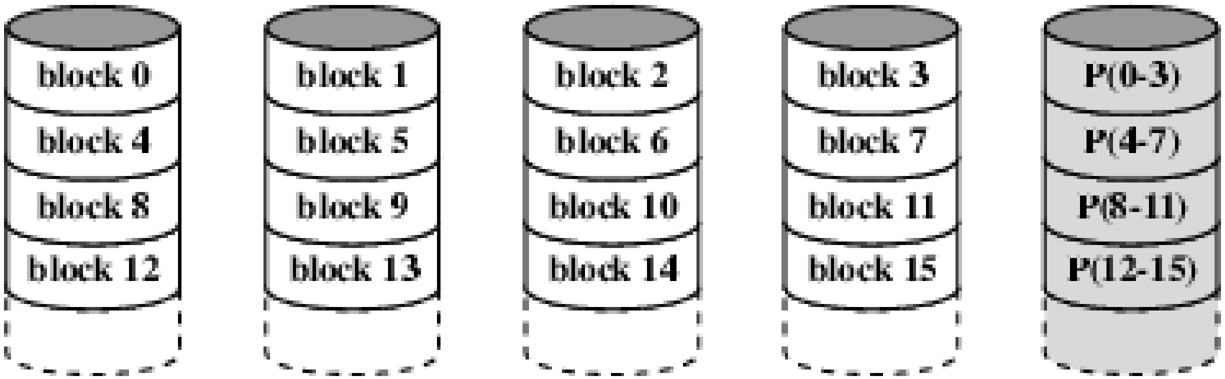


(c) RAID 2 (redundancy through Hamming code)

RAID 3 & 4

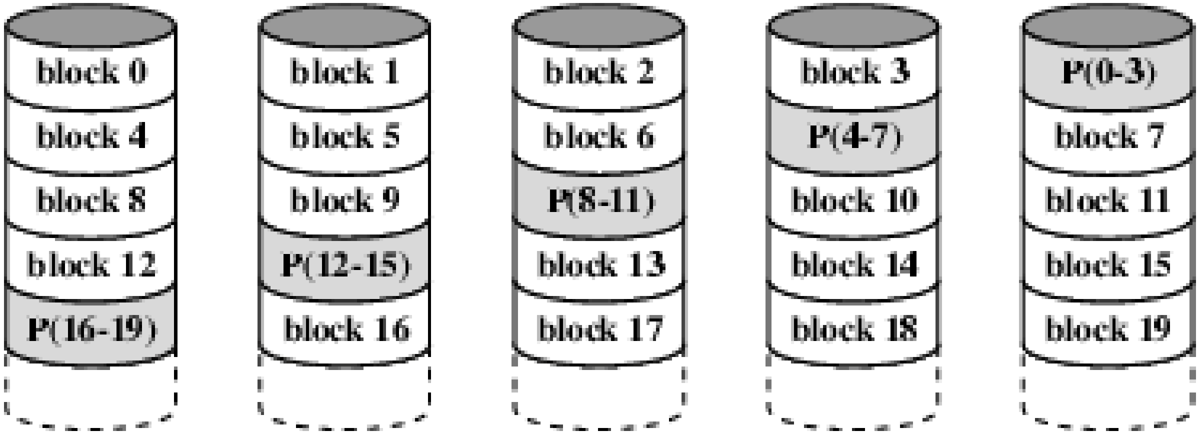


(d) RAID 3 (bit-interleaved parity)

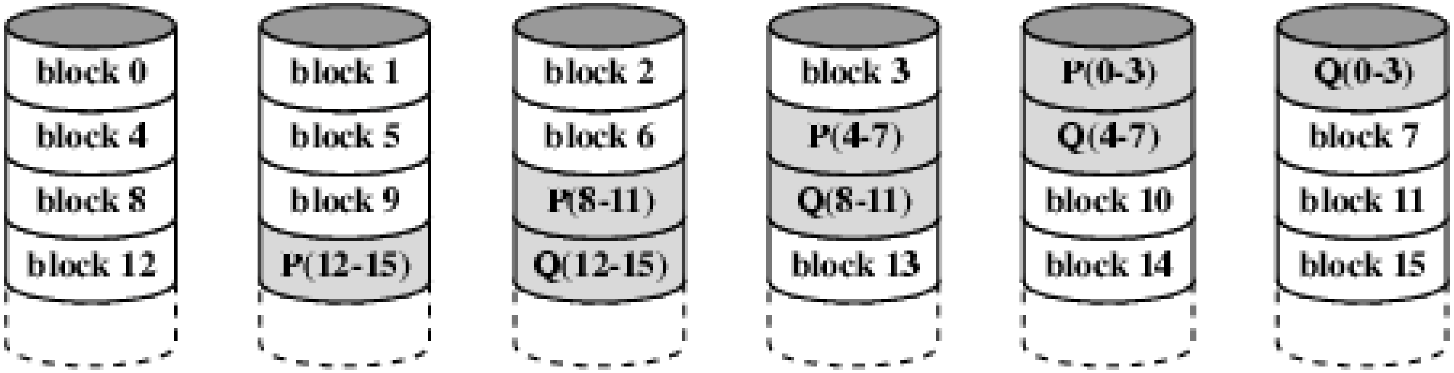


(e) RAID 4 (block-level parity)

RAID 5 & 6



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



(g) RAID 6 (dual redundancy)