MODULE 6 – PART B

Redundant Array of Independent Disks (RAID)

Outline

- Introduction
- Need for RAID
- Levels of RAID
- RAID Level Selection
- I/O Performance
- References

Redundant Array of Independent Disks

- RAID stands for Redundant Array of Independent Disks, which is a technology to associate multiple secondary storage devices and utilize as a single storage media.
- Single, Large Expensive Drive (SLED) is less performance.
- RAID chain multiple, small inexpensive disks drive into an array of disk drives which enhances performance.
- In addition this array of drives perform to the computer as a single logical storage unit.

Secondary Storage Devices

- Significant role in storing large amount of data as memory is expensive
- Plays a major role when disk is used as virtual memory
- Magnetic structure
- Typically uses a "moving head disk" mechanism in order to read data and write data

Raid and performance

- Disk organization techniques that exploit a collection of disks, i.e., striping, mirroring, parity
- Originally "inexpensive" disks
- Enhance Speed, Improved reliability
- Combination of Speed and reliability improves
- Collection appears as a single disk to the system

Data loss and recovery

- Need to have backups of all the data
- Often take backups Periodically - Daily, Weekly or Monthly
- Ensure the type of device - magnetic, optical or physical
- Time taken to recover from the disaster and how long it takes?
- The usage of multiple disks increases the risk of failure (not data loss):
  - A system with 100 disks, each with MTTF of 100,000 hours (approx. 11 years), will have a system MTTF of 1000 hours (approximately 41 days)

Improvement of Reliability via Redundancy

- In a SLED Reliability becomes a big problem as the data in an entire disk may be lost
- As the number of disks per component increases, the probability of failure also increases
  - Suppose a (reliable) disk fails every 100,000 hrs.
  - Reliability of a disk in an array of N disks = Reliability of 1 disk/ N
  - 100000hrs / 100 = 1000 hrs. = 41.66 days !!
- Solution ? Redundancy

RAID – Mirroring

- Disk mirroring, is the replication of data to two or more disks / Duplicate every disk
- Disk mirroring is used for applications such as transactional applications, email and operating systems which require high performance and availability,
- If one of the disk fails, the lost data read from the other
- The second disk fails before the first failed disk is replaced if the Data is permanently lost.

RAID DATA STRIPING

- RAID striping is ideal for transfer rate-intensive environments
- RAID is used in highly multi-tasking, request rate-intensive environments
- A method of concatenating multiple drives into one logical storage unit.
- Across multiple disks, splitting the bits of each byte
  - Bit – level striping
  - Block-level striping
- When a request raises, in parallel, the request less than the stripe size is directed to one HD. The request greater than the stripe size is distributed into multiple requests and directed to multiple disks

RAID LEVEL 0

- Striping is carried out in the level of blocks
- Data split across drives bring about in increased data throughput.
- Hard disks are striped using the stripe size stated during configuration.
- Typically used in data rate intensive applications
- Performance is reasonable but the failure of any disk in the array results in data loss
- Reliability Issues : No parity bits or mirroring.

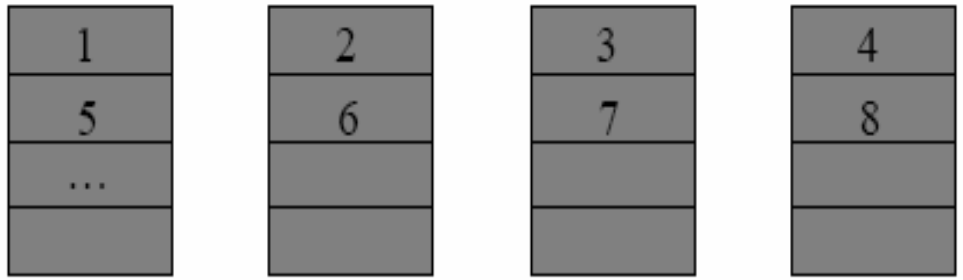
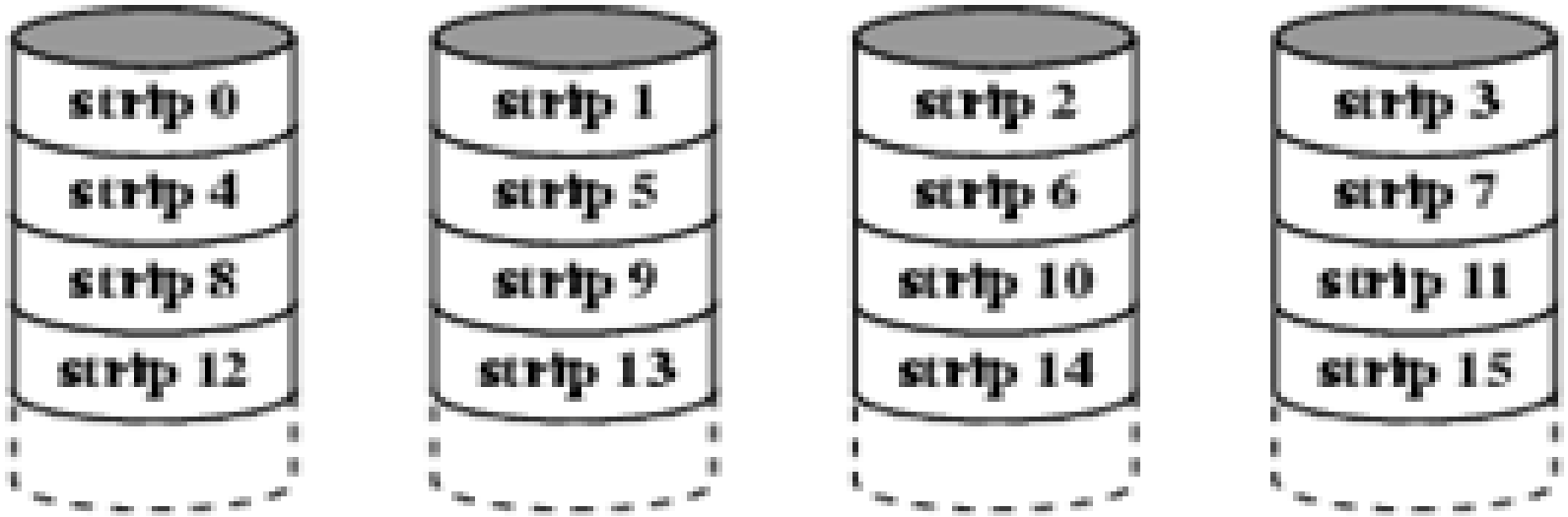


Fig. 1 Raid Level 0

RAID LEVEL 0

- The data are *striped across the available disks. This is* best understood by considering Figure .

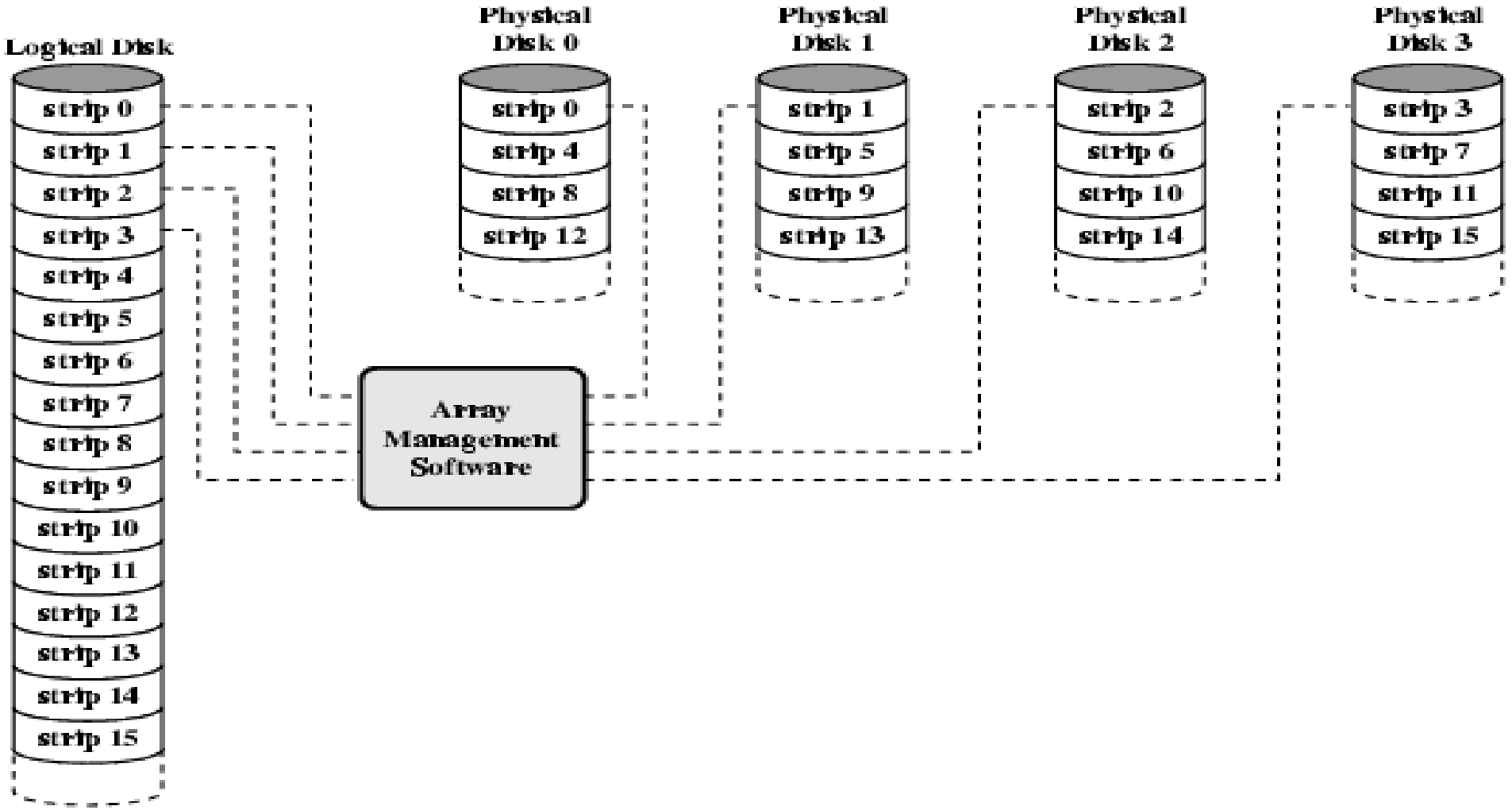


(a) RAID 0 (non-redundant)

RAID LEVEL 0

- All of the user and system data are viewed as being stored on a logical disk. The logical disk is divided into strips.
- these strips may be physical blocks, sectors, or some other unit. The strips are mapped round robin to consecutive physical disks in the RAID array.
- A set of logically consecutive strips that maps exactly one strip to each array member is referred to as a **stripe**

Data Mapping For RAID 0



RAID LEVEL 1

- RAID 1 delivers the highest level of fault tolerance to a single hard drive.
- Usually denoted to as "mirroring"
- Provides redundancy through mirroring concept which is expensive.
- Performance Issues: No data loss if either drive fails.
- Mirroring is suitable for any application where protection from drive failure is vital.
- Delivers instantaneous disaster recovery when a drive fails.
- Good read and write performance.

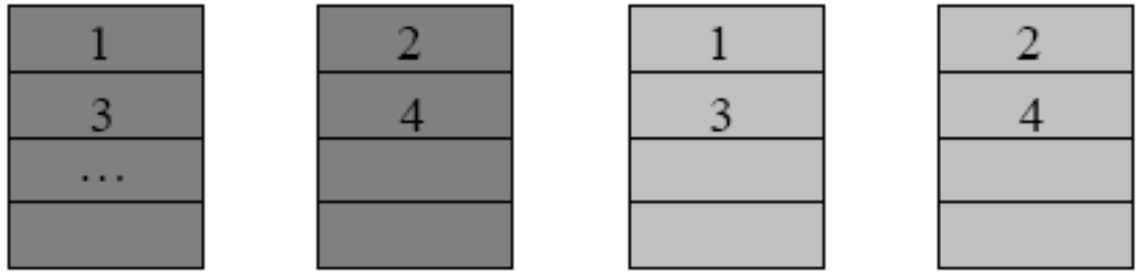


Fig. 2 Mirroring

RAID 1(MIRRORED)

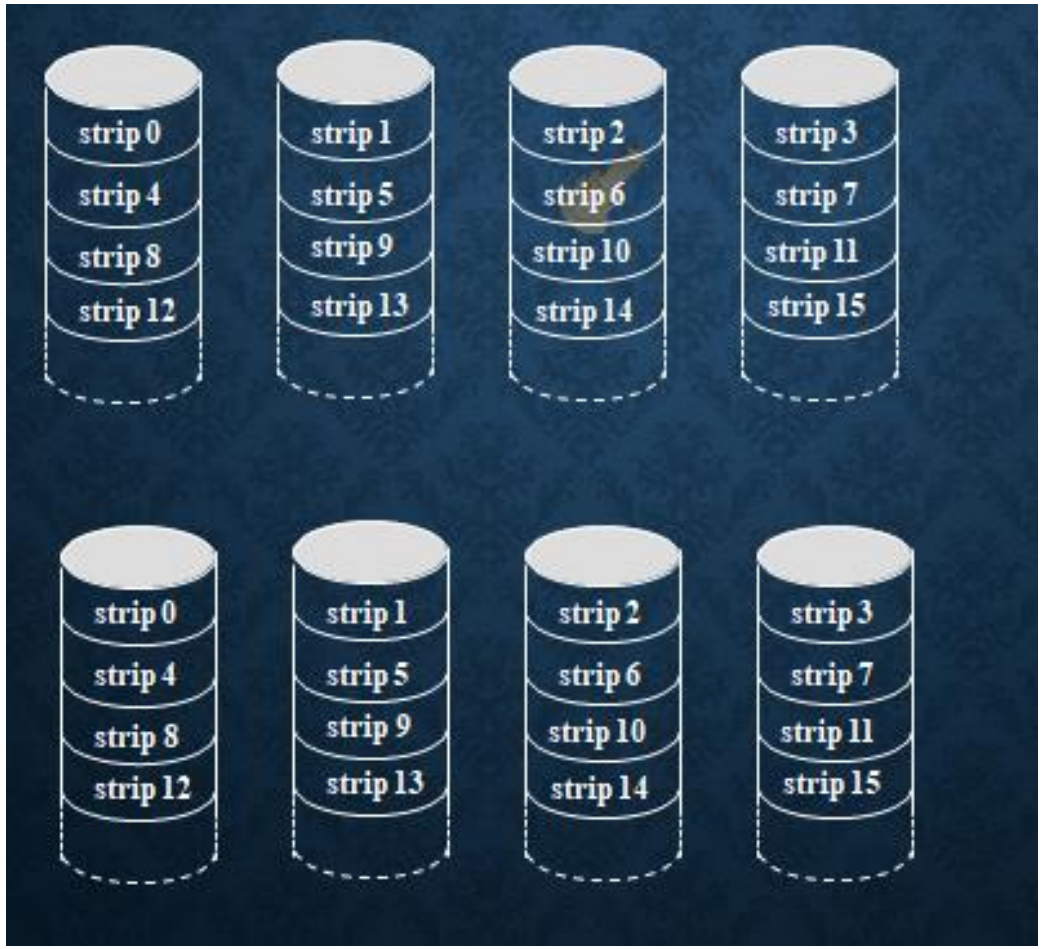


Fig. 3 Mirroring

RAID LEVEL 2

- RAID Level 2 use Hamming code, error-correcting code (ECC)
- Anticipated for utilize in drives which doesn't have built-in error detection
- The basic idea is if disks fail the residual bits of the byte and the associated ECC bits can be used to rebuild the lost data.
- Good solution in extent of data security.
- In case of HDD failure – no matter if it was the disk with data or the Hamming code –the array may be reconstructed by the other disks available.

Fig. 3 Error Correcting Code

RAID LEVEL 3

- Developed upon RAID 2, called as Bit-Interleaved Parity
- It is used to store checksums and it supports a special processor in parity codes calculating "the parity disk".
- Disk Controllers can detect whether a sector has been read correctly.
- Storage overhead is reduced – only 1 parity disk
- Require a dedicated parity hardware
- Expense of Read/Write parity

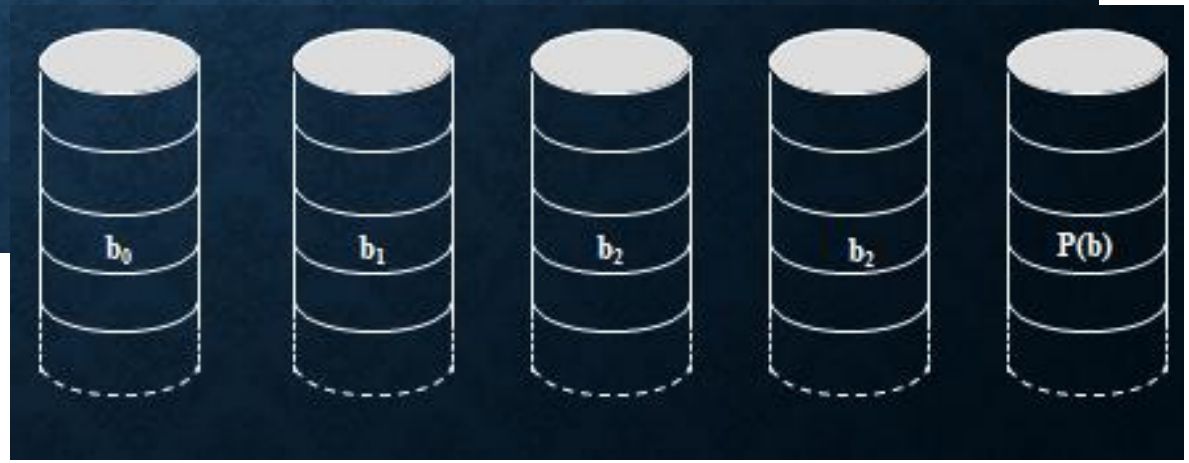


Fig. 5 Bit Interleaved Parity

RAID LEVEL 4

- RAID 4 used block-interleaved parity.
- Stripes data at a block level (16, 32, 64 or 128 kB) across several drives, with parity stored on one drive.
- For each row of written data, any recorded block is written on a parity disk.
- Allows recovery from the failure of any of the disks. Effective Read/Write operations.

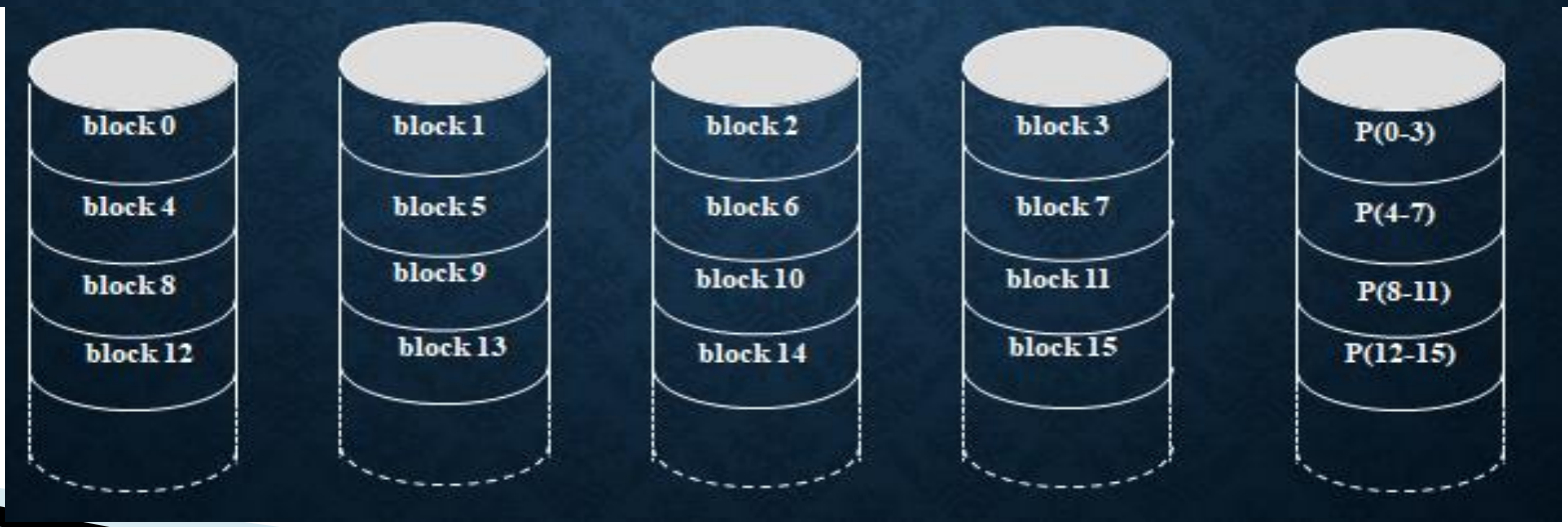


Fig. 6 Block – interleaved parity

RAID LEVEL 5

- Block-interleaved Distributed parity which requires at least 3 drives.
- Rather than storing data in N disks and parity in 1 disk, RAID 5 Spreads parity and data between all N+1 disks.
- Due to the parity calculation, Read data transactions are very fast while write data transactions are slower.

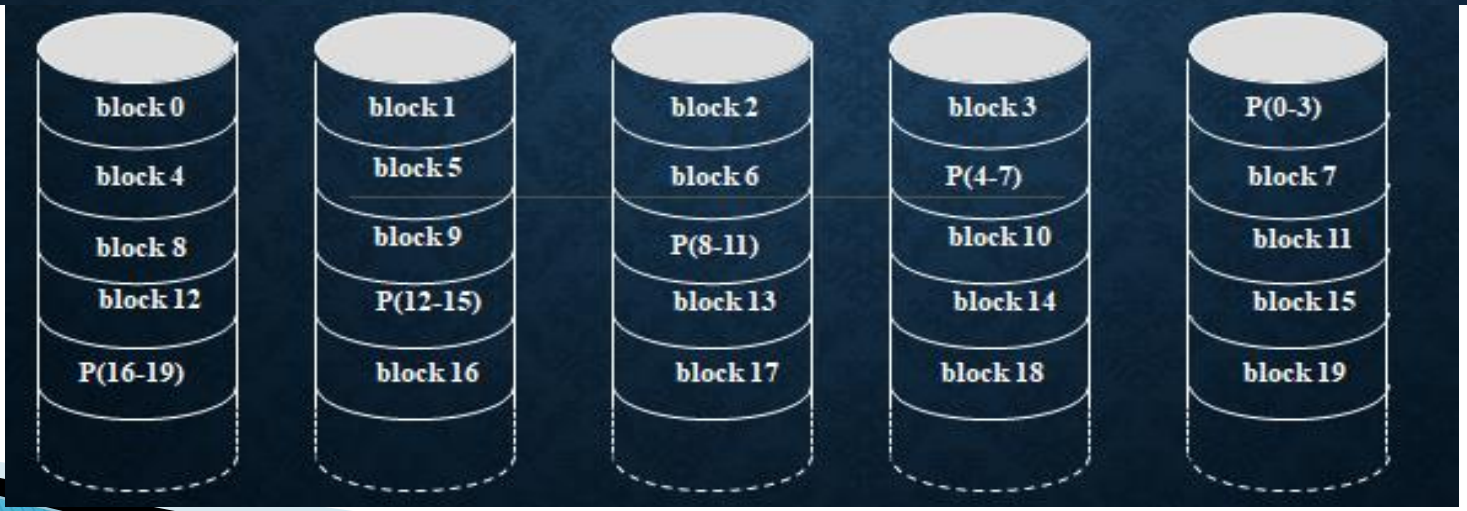
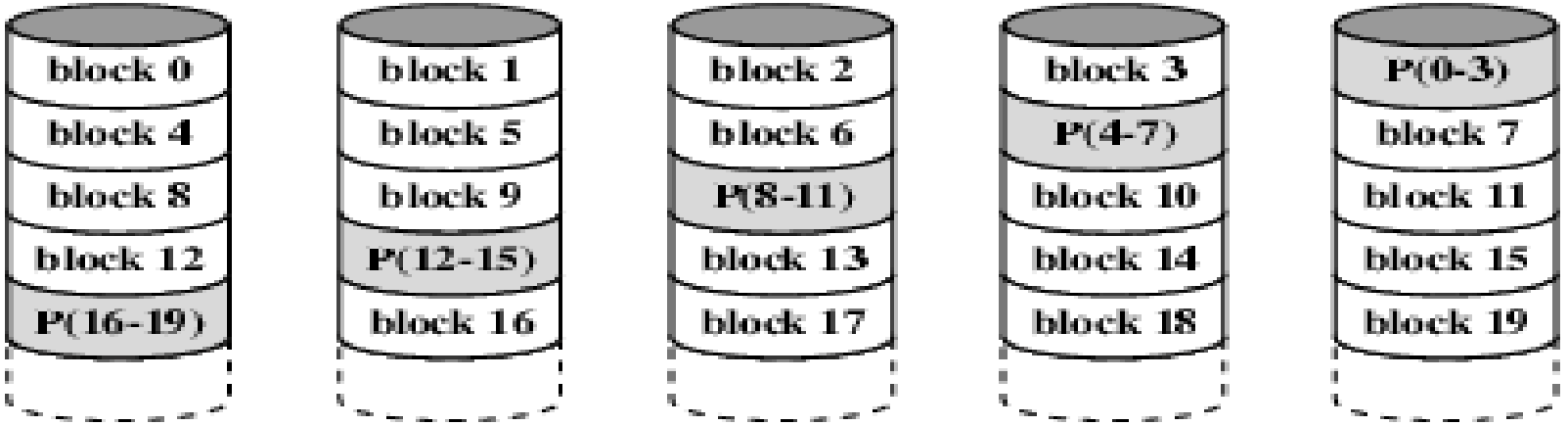


Fig. 7 Distributed Parity

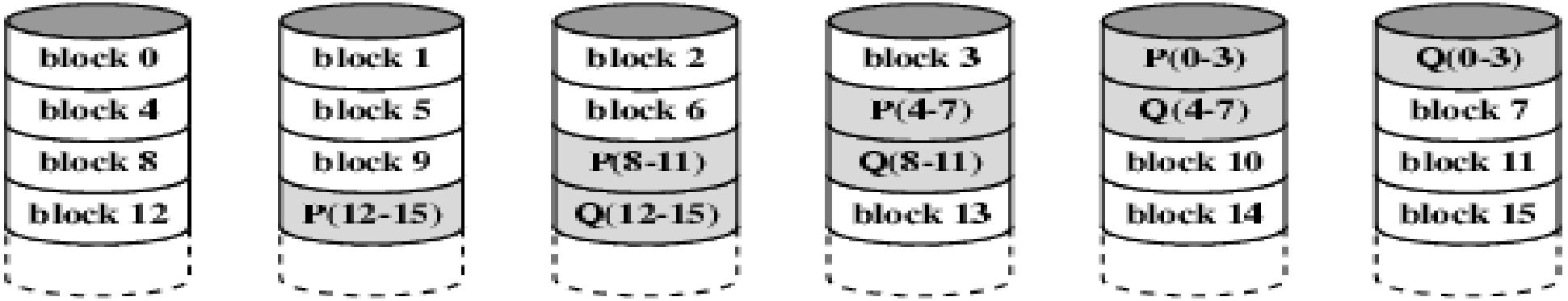
RAID 6

- 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 5 & 6



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



(g) RAID 6 (dual redundancy)

RAID LEVEL 10

RAID(0+1) and RAID(1+0)

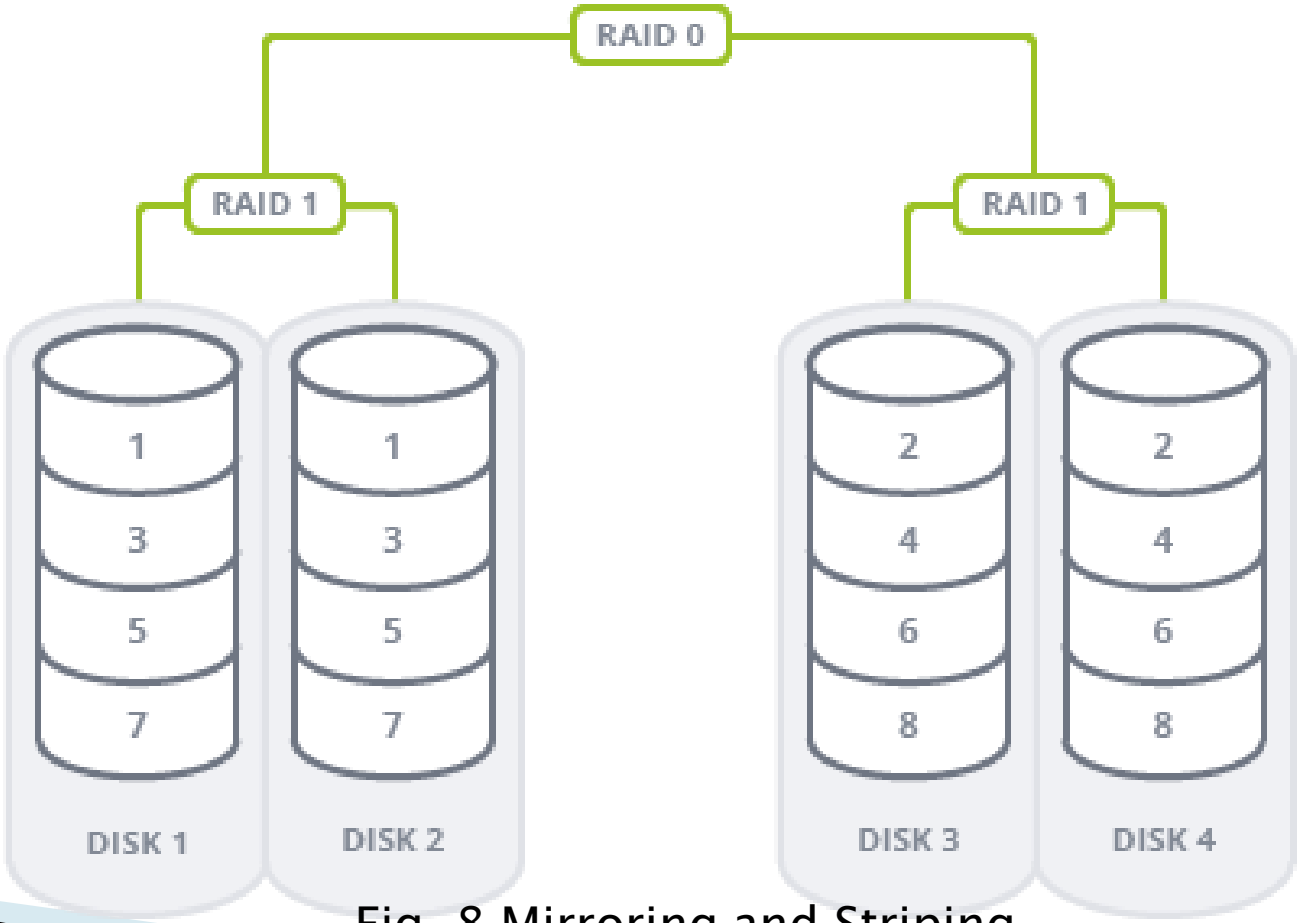


Fig. 8 Mirroring and Striping

RAID LEVEL 10

- **RAID 10**, then, is a combination of **levels** 1 (mirroring) and 0 (striping), which is why it is also sometimes identified as **RAID** 1 + 0.
- Mirroring is writing data to two or more hard drive disks (HDDs) at the same time – if one disk fails, the mirror image preserves the data from the failed disk.
- disks in a RAID 10 rebuild lost data is very fast since copying all the data from the surviving mirror to a new drive.

RAID Level SELECTION

- RAID 0 –High-Performance applications where data loss is not critical
- RAID 1 –High Reliability with fast recovery
- RAID 5 –Preferred for storing large volumes of data
- RAID 6 –Not Supported currently by many RAID implementations
- RAID 10/01 –Both performance and reliability are important, e.g. in small databases

I/O Performance

- **Reliability**: How many disk faults can the system tolerate?
- **Availability**: What fraction of the total session time is a system in uptime mode, i.e. how available is the system for actual use?
- **Performance**: How good is the response time? How high is the throughput (rate of processing work)? Note that performance contains a lot of parameters and not just the two.
- **Capacity**: Given a set of N disks each with B blocks, how much useful capacity is available to the user?

I/O Performance

- Data throughput the quantity of data that can be processed within a specific time interval
- Number of requests (transactions) that can be processed within a specific time interval
- Average response time the average time for the processing of a single request

I/O Performance

- Lost information can be reconstructed from redundant information.
- But availability can be improved by adding redundant disks (RAID).
- MTTR: mean time to repair is in the order of hours
- MTTF: mean time to failure of disks is tens of years

References

1. David A. Patterson and . John L. Hennessy —Computer Organization and Design-The Hardware/Software Interface, 5th edition, Morgan Kaufmann, 2011.
2. Carl Hamacher, Zvonko Vranesic, Safwat Zaky, Computer organization, Mc Graw Hill, Fifth edition ,Reprint 2011.