Operating Systems

RAID –
Redundant Array of Independent Disks

Submitted by
Ankur Niyogi
2003EE20367

YOUR DATA IS LOST@#!!

- Do we have backups of all our data????
  - The stuff we cannot afford to lose??

- How often do we do backups???
  - Daily, Weekly or Monthly??

- Are they magnetic, optical or physical??

- How long would it take to totally recover from the disaster???

FLOW OF PRESENTATION

- Secondary storage – advantages and limitations
- Increasing Reliability via Redundancy
- RAID
- Mirroring and Data-Striping
- RAID Levels

Secondary Storage Devices

- Significant role in storing large amount of data as memory is expensive
- Plays a vital role when disk is used as virtual memory
- Magnetic in nature
- Characteristically uses a “moving head disk” mechanism to read and write data

RAID : Redundant Array of Inexpensive Disks

Performance limitation of Disks:

- Performance of a single disk is very limited

- Throughput : 125 reqs/s
- Bandwidth : 20-200MB/s (max) 15-30MB/s (sustained)
- Very difficult to significantly improve the performance of disk drives
  - Disks are electromechanical devices
- Speed gap between disks and CPU/Memory is widening
  - CPU speed increases @ 60% / year
  - Disks speed increase @ 10-15% / year
- Improvement in disk technologies is still very impressive BUT only in the capacity / cost area.

What does RAID stand for?

In 1987, Patterson, Gibson and Katz at the University of California Berkeley, published a paper entitled “ A Case for Redundant Array of Inexpensive Disks(RAID)”.

Described the various types of Disk Arrays, referred to as the acronym RAID.

The basic idea of RAID was to combine multiple, small inexpensive disks drive into an array of disk drives which yields performance exceeding that of a Single, Large Expensive Drive(SLED).

Additionally this array of drives appear to the computer as a single logical storage unit or drive.

Improvement of Reliability via Redundancy

•In a SLED Reliabity 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.

Reliabity of a disk in an array of N disks = Reliability of 1 disk / N

100000hrs / 100 = 1000 hrs = 41.66 days !!

➢ Solution ?

Redundancy

Redundancy

- Mirroring
- Data Striping

Mirroring

- Duplicate every disk
- Logical disk consists of two physical disks.
- Every write is carried out on both disks.
- If one of the disk fails, data read from the other
- Data permanently lost only if the second disk fails before the first failed disk is replaced.

Reliability in Mirroring

Suppose mean time to repair is 10 hrs ,
the mean time to data loss of a mirrored disk system is
100,000 ^ 2 / (2 * 10) hrs ~ 57,000 years !

Main disadvantage :
Most expensive approach .

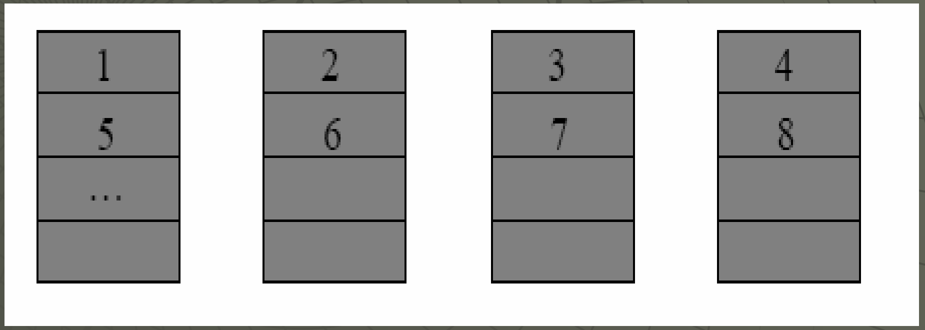
Parallel Disk Systems

• We cannot improve the disk performance significantly as a single drive

- But many applications require high performance storage systems ?

• Solutions :

- Parallel Disk Systems
- Higher Reliability and Higher data-transfer rate.

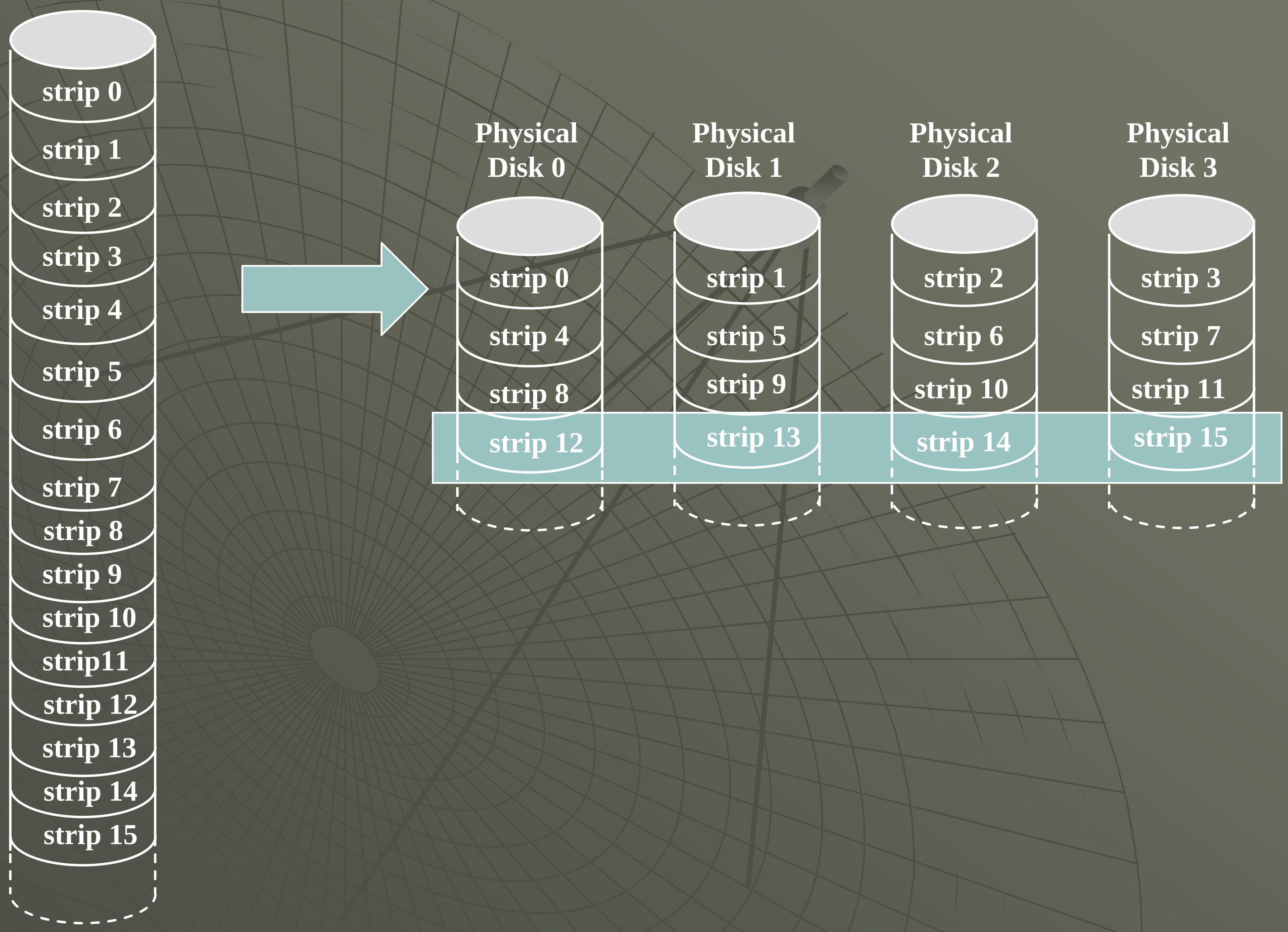


DATA STRIPING

- Fundamental to RAID
- A method of concatenating multiple drives into one logical storage unit.
- Splitting the bits of each byte across multiple disks : *bit – level striping*

  e.g. an array of eight disks, write bit i of each byte to disk I
- Sectors are eight times the normal size
- Eight times the access rate
- Similarly for blocks of file, *block-level striping*

Logical to Physical Data mapping for striping

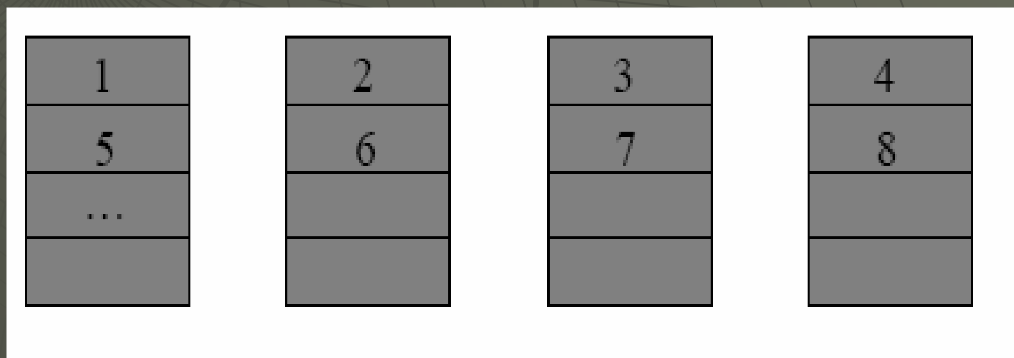


RAID LEVELS

- Data are distributed across the array of disk drives
- Redundant disk capacity is used to store parity information, which guarantees data recoverability in case of a disk failure
- Levels decided according to schemes to provide redundancy at lower cost by using striping and “parity” bits
- Different cost-performance trade-offs

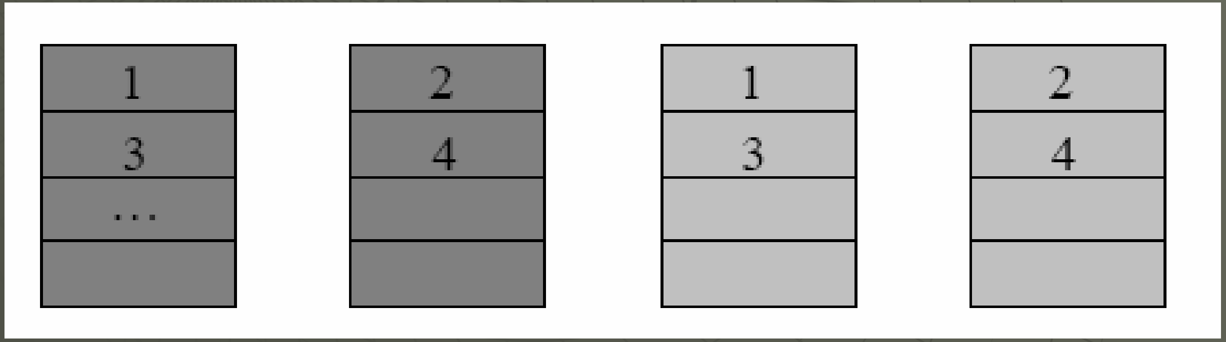
RAID 0

- Striping at the level of blocks
- Data split across in drives resulting in higher data throughput
- Performance is very good but the failure of any disk in the array results in data loss
- RAID 0 commonly referred to as striping
- *Reliability Problems* : No mirroring or parity bits

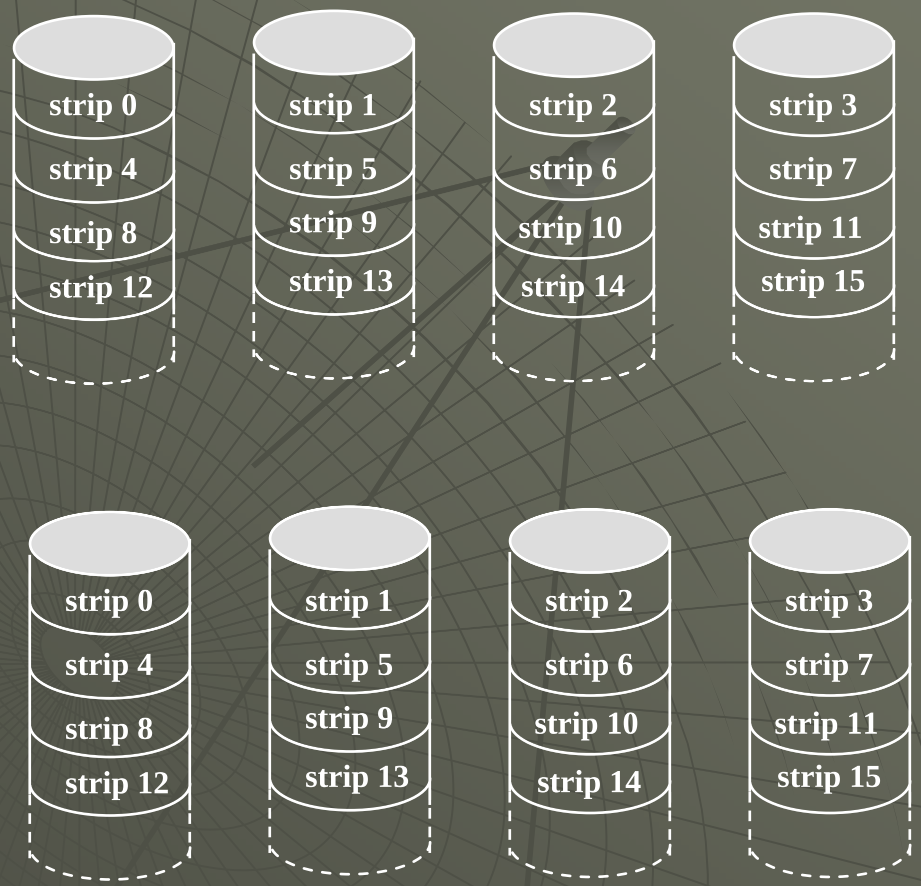


RAID 1

• Introduce redundancy through mirroring
• Expensive
• Performance Issues
-- No data loss if either drive fails
– Good read performance
– Reasonable write performance
• Cost / MB is high
• Commonly referred to as “mirroring”

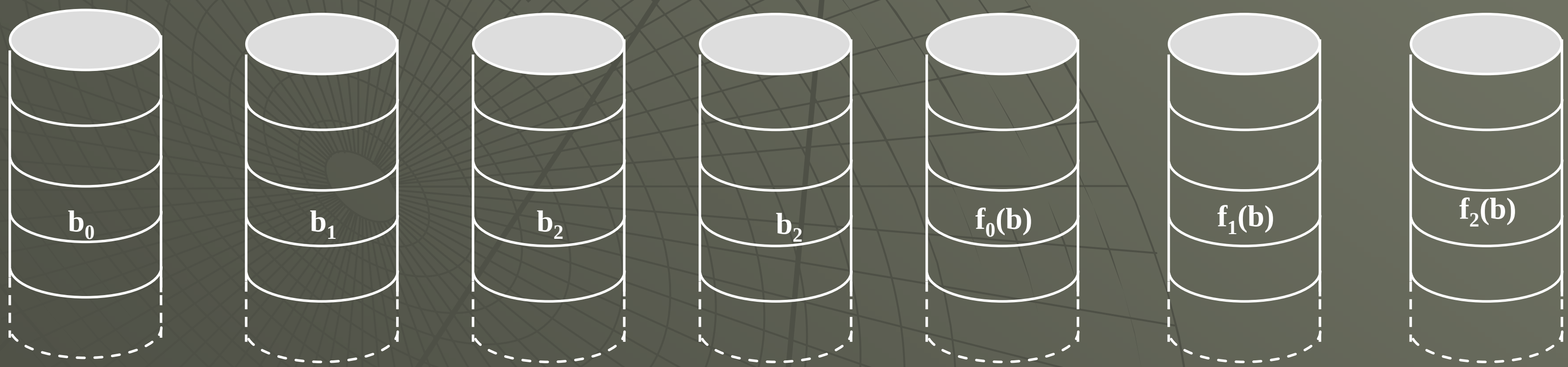


RAID 1(Mirrored)



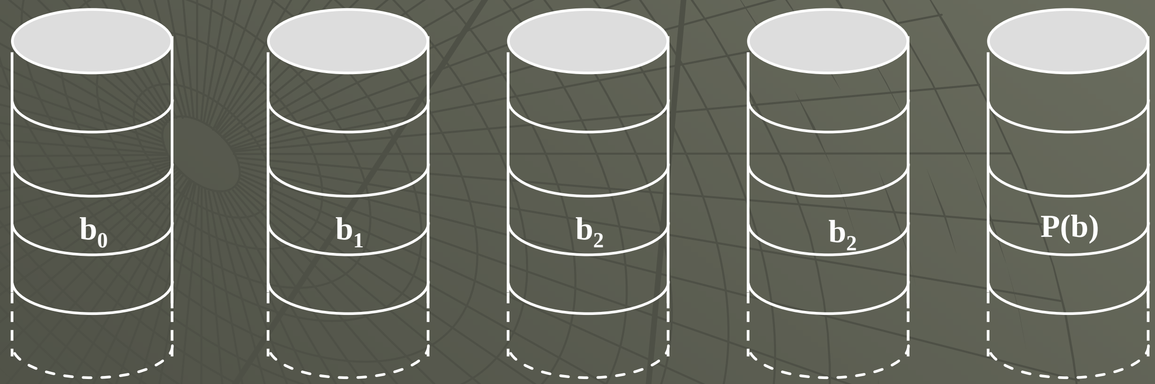
RAID 2

- Uses Hamming (or any other) error-correcting code (ECC)
- Intended for use in drives which do not have built-in error detection
- Central idea is if one of the disks fail the remaining bits of the byte and the associated ECC bits can be used to reconstruct the data
- Not very popular



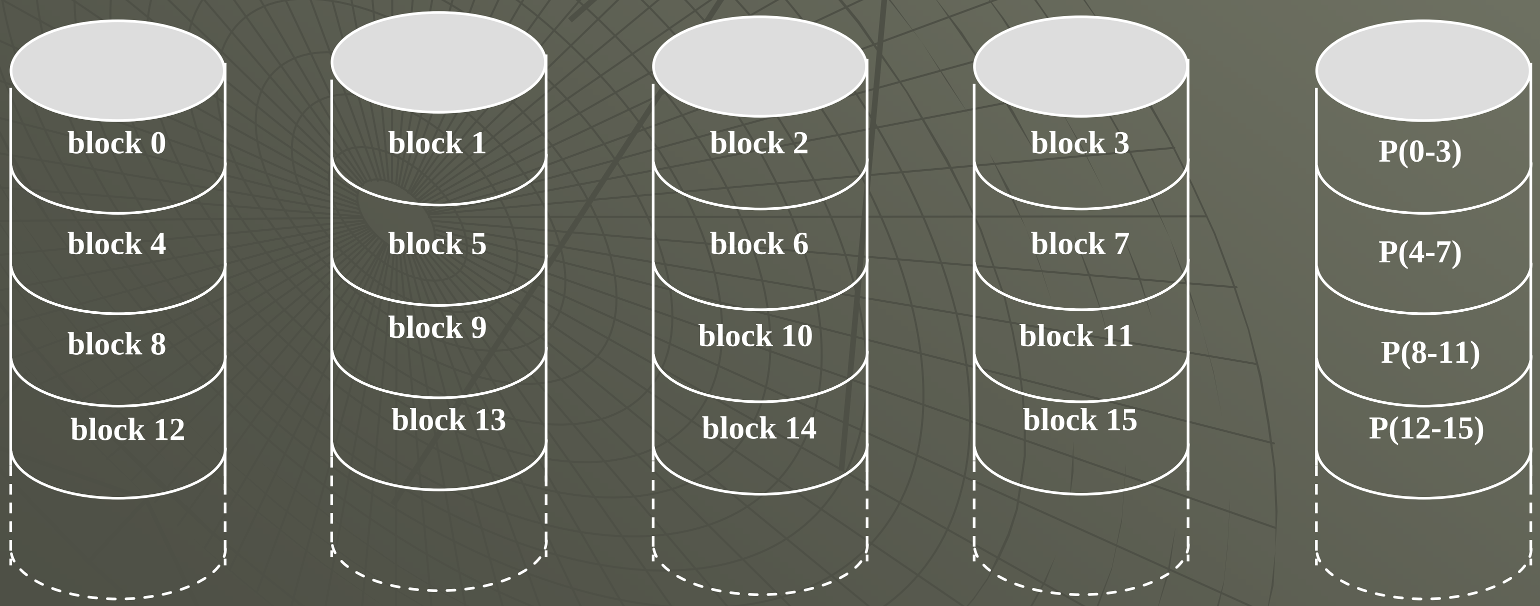
RAID 3

- Improves upon RAID 2, known as Bit-Interleaved Parity
- Disk Controllers can detect whether a sector has been read correctly.
- Storage overhead is reduced – only 1 parity disk
- Expense of computing and writing parity
- Need to include a dedicated parity hardware



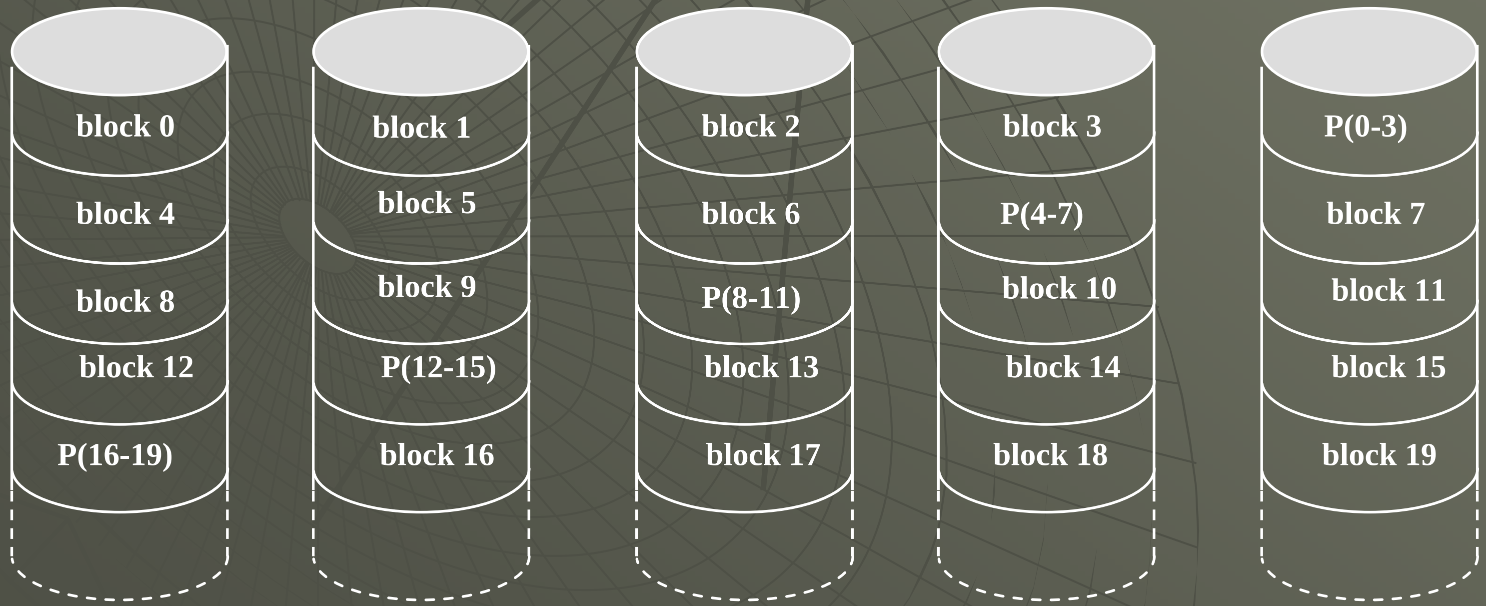
RAID 4

• Stripes data at a block level across several drives, with parity stored on one drive - block-interleaved parity
• Allows recovery from the failure of any of the disks
• Performance is very good for reads
• Writes require that parity data be updated each time. Slows small random writes but large writes are fairly fast



RAID 5

- Block-interleaved Distributed parity
- Spreads data and parity among all N+1 disks, rather than storing data in N disks and parity in 1 disk
- Avoids potential overuse of a single parity disk – improvement over RAID 4
- Most common parity RAID system



RAID 6

- P+Q Redundancy

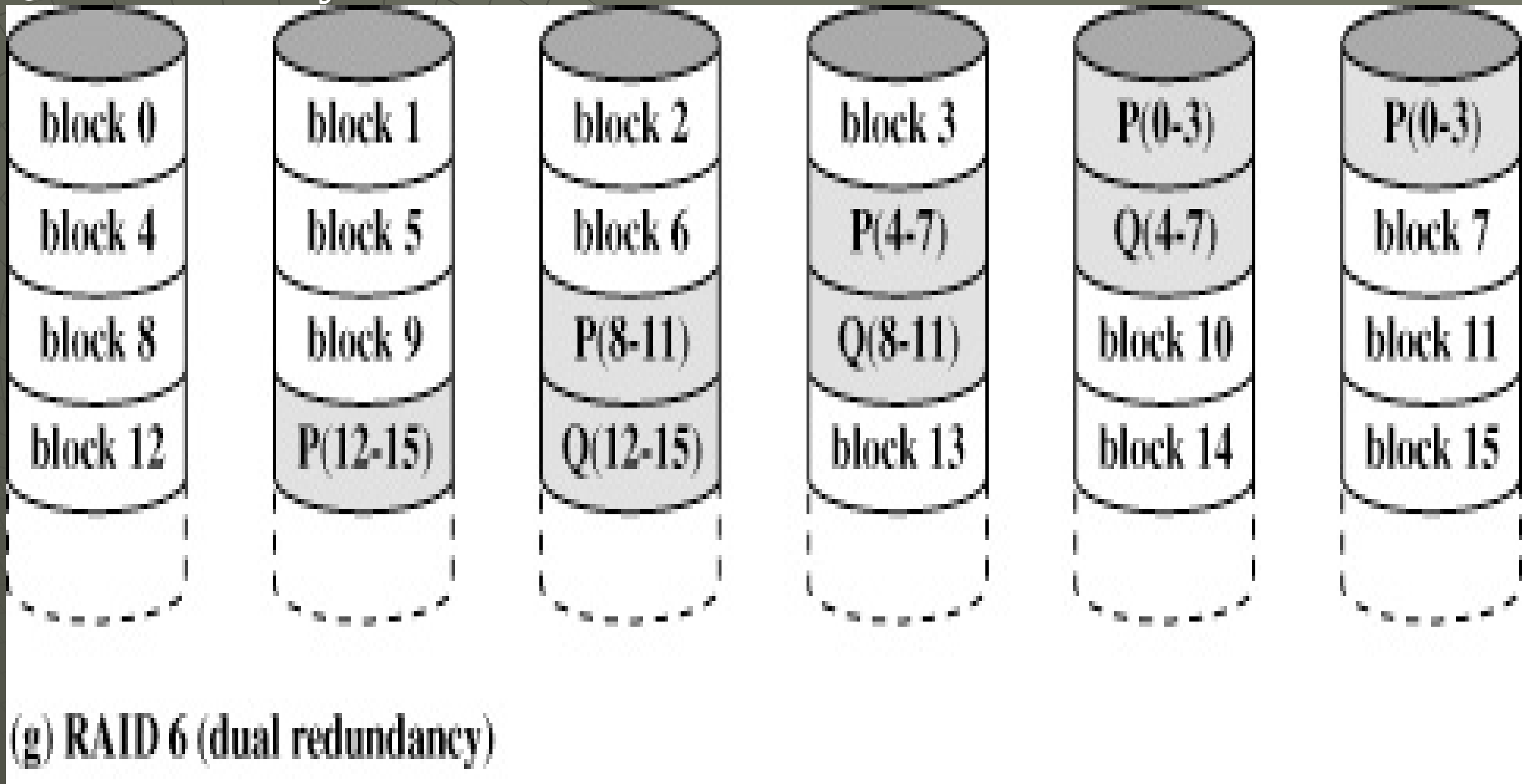
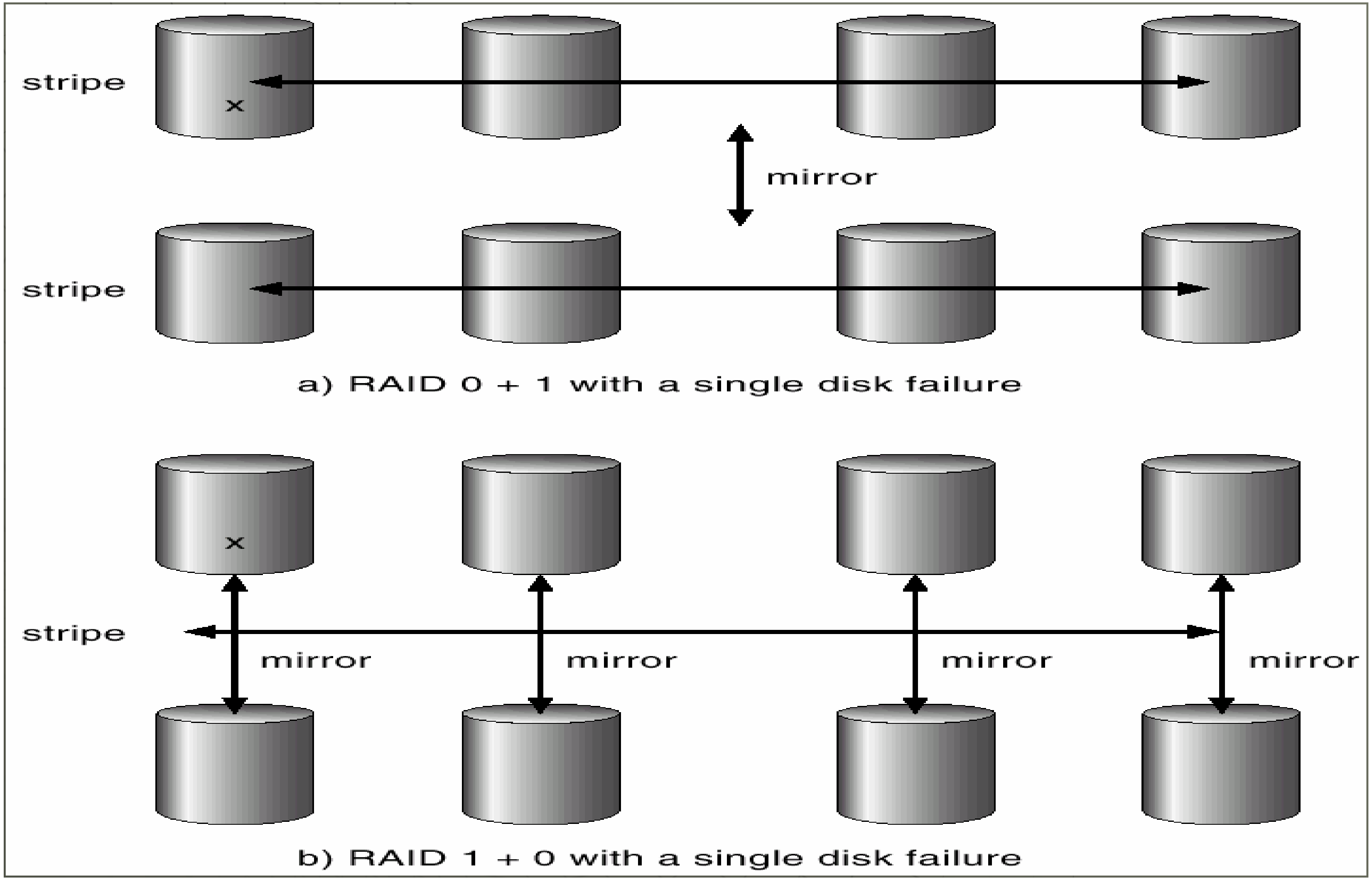


Figure 11.9 RAID Levels (page 2 of 2)

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



a) RAID 0 + 1 with a single disk failure

b) RAID 1 + 0 with a single disk failure

RAID 10

• Advantages
– Highly fault tolerant
– High data availability
– Very good read / write performance

• Disadvantages
– Very expensive

• Applications
– Where high performance and redundancy are critical

Selecting a RAID Level

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

References

1.www.bridgeport.edu/sed/fcourses/cpe473/Lectures/RAID.ppt

2. r61505.csie.nctu.edu.tw/OG/project/extra6-Ch8-RAID.ppt

3. A. Silberschatz, P. B. Galvin, and G. Gagne, Operating System Concepts, 7th Edition, John Wiley & Sons, 2005