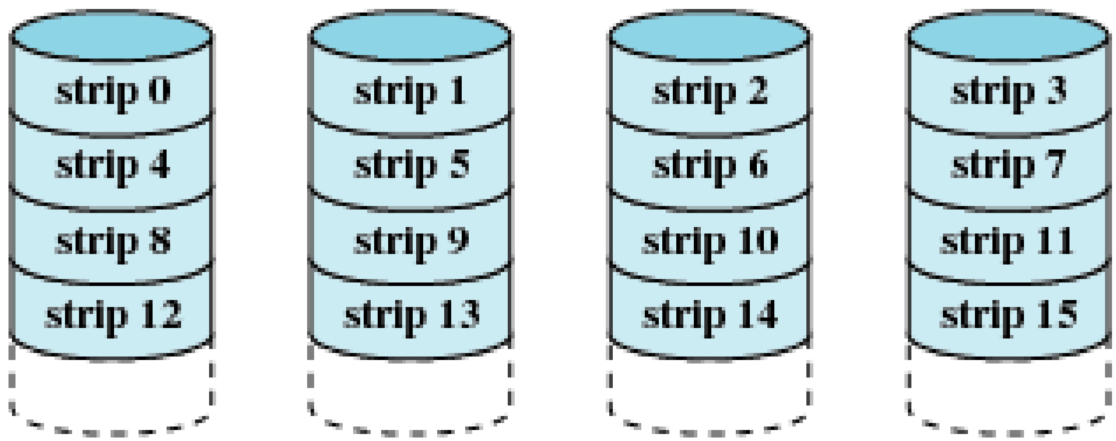
DATA STORAGE TECHNOLOGIES & NETWORKS
(CS C446, CS F446 & IS C446)

LECTURE 19– STORAGE

Disks - Reliability

- Redundancy may be used to improve Reliability
  - Device Level Reliability
    - Improved by redundant disks
      - This of course implies redundant data
  - Data Level Reliability
    - Improved by redundant data
      - This of course implies additional disks
- (RAID) Redundant Array of Inexpensive Disks
  - or *Redundant Array of Independent Disks*
- Different Levels / Modes of Redundancy
  - Referred to as RAID levels

RAID – RAID 0

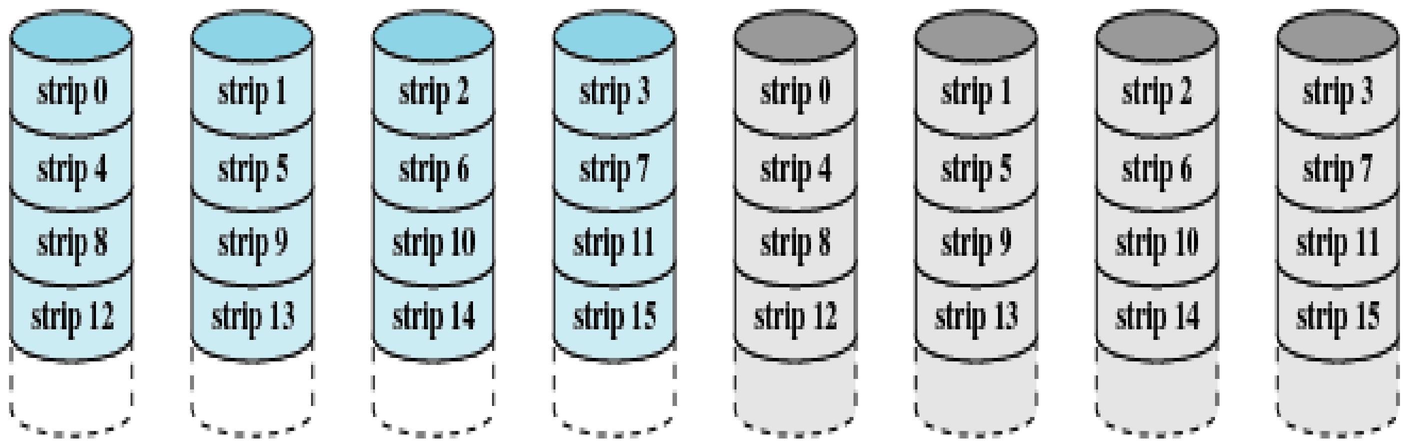


(a) RAID 0 (non-redundant)

RAID – RAID 0

- RAID 0 (Striping)
  - No redundancy
  - Each piece of data (say a logical block) is striped onto D disks in an array
    - A ***stripe*** refers to the entire data block of ***D stripe units***
    - Size of the stripe unit may vary
      - bit-level, byte-level, sector-level, or block-level striping
      - Fine-grained interleaving:
        - increases data transfer rate for all I/O;
        - but only one I/O operation can be performed at a time and all disks must do positioning for every I/O
      - Coarse-grained interleaving:
        - Small I/Os will use fewer disks whereas data transfer rate is improved for large I/Os

RAID 1 (mirrored)

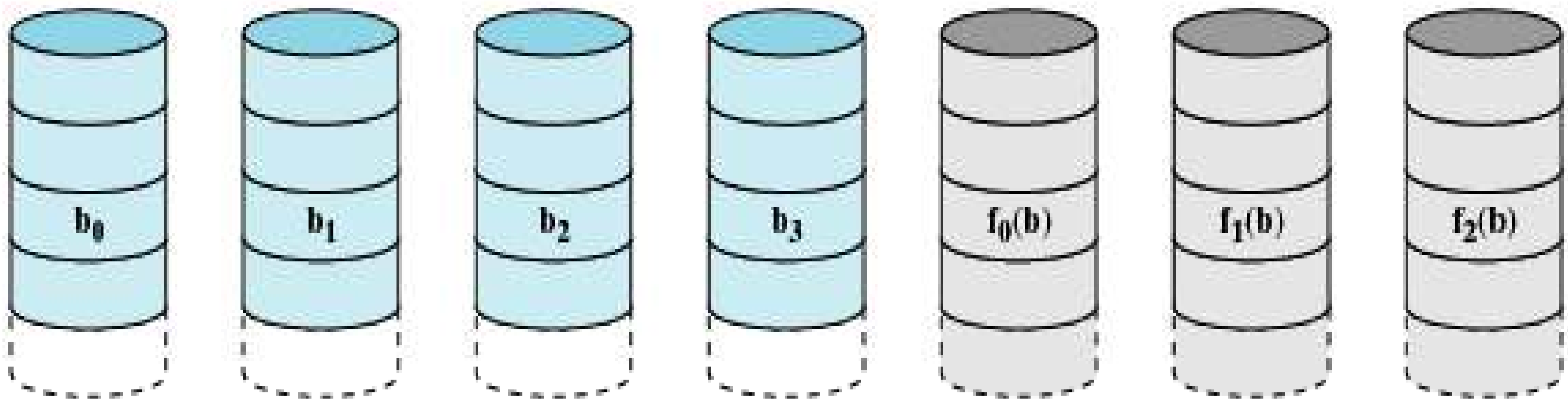


(b) RAID 1 (mirrored)

RAID – RAID 1

- RAID 1 (Mirroring)
  - Each data unit is replicated (i.e. mirrored) in two disks
    - Data disk and Check Disk
  - 2 (independent) reads can be done in parallel
    - I/O rate improves
      - Typically by a factor of 2/s
    - Slowdown factor s
      - 1<= s <=2
      - arising due to synchronization time (s = 1 for synchronized disks)
      - Increases for multiple (more than 2) disks
  - Only 1 write at a time – I/O rate 1/s
  - Can recover from complete disk failure (1 out of 2)

RAID 2 (redundancy through Hamming code)



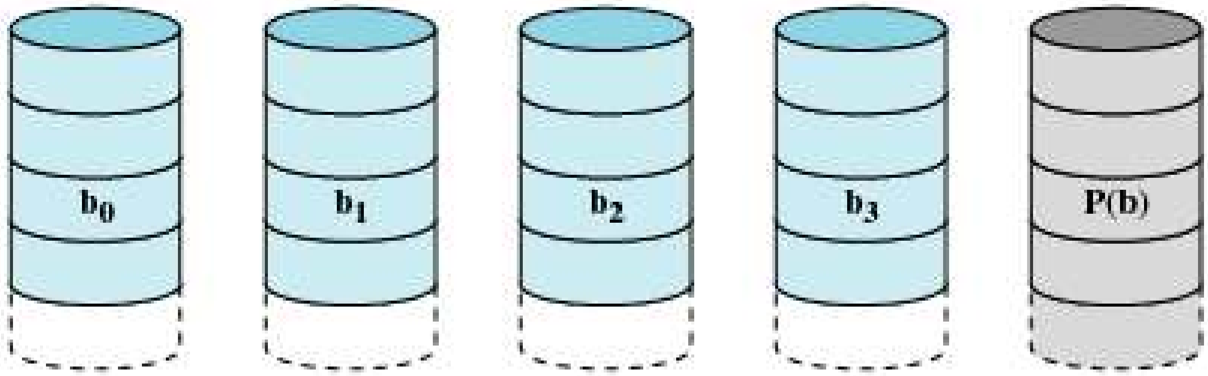
(c) RAID 2 (redundancy through Hamming code)

- Also known as memory style error correction code (ECC) organization
- Striping of bytes across disks (1st bit of each byte in disk 1, 2nd bit in disk2 …. Error correction bits are stored in further disks)

RAID – RAID 2

- RAID 2 (ECC)
  - Error-Correcting Codes have been used in DRAMs for a long time
  - Hamming Code is used typically
  - For Disk Arrays:
    - D bit data and C bit Code added
    - Bit-interleaving:
      - D data disks and C check disks used
    - C=4 when D=10, C=5 when D=25 for Hamming Code (single error correction)
  - Not used commercially
    - Individual disks store ECC along with data

RAID 3 (bit-interleaved parity)



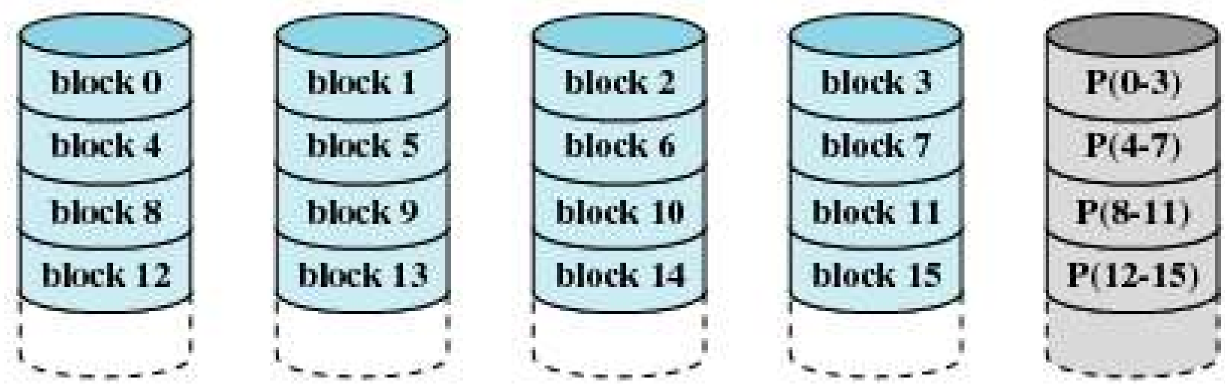
(d) RAID 3 (bit-interleaved parity)

- Striping at the level of bit inter leaved parity organization
- Single bit can be used for error detection and correction
- Advantages
  - Less number of disks, transfer rate is N times faster than RAID 1 as bits are distributed to all disks
- Disadvantages
  - Expense of computing & writing parity, fewer I/O per sec

RAID - RAID 3

- Difference between data level failures and device level failures
  - Disk failures can be detected externally (say by the controller)
- RAID 3
  - 1 bit of parity per D bits of data
    - i.e. 1 check disk per D data disks (bit interleaving)
  - Operations
    - Read (Normal) :
      - All data disks are used
    - Read (1 disk failure):
      - All (other) data disks and parity disk are used
    - Write :
      - All data disks and parity disk are used
      - Read-Modify-Write is not required
  - Always reads and writes complete stripes of data across all disks [drives operate in parallel]
  - No partial writes that update one out of many strips in a stripe
  - Performance:
    - Not used when high I/O rate is required (Why?)

RAID 4 (block-level parity)



(e) RAID 4 (block-level parity)

- Block inter leaved parity organization
  - Data transfer rate is slow, multiple read accesses can proceed in parallel (higher overall I/O rate)
  - Transfer rate of large reads are high (all disks can be read parallel)
  - Transfer rate of large writes are high (data & parity can be written parallel)

RAID – RAID 4

- RAID 3 enables high data transfer rates but
  - allows only one I/O at a time and
  - may suffer from worst case seek and rotational delays unless disks are synchronized.
- RAID 4
  - 1 bit of parity per D bits of data
    - 1 check disk per D data disks but with *block interleaving*
  - Operations:
    - Read (Normal):
      - Small reads:
        - Not all data disks are to be read
        - Independent reads can be on different (data) disks in parallel for reads smaller than stripe unit
      - Large reads:
        - Similar to RAID 3
    - Read (under 1 disk failure): read from all data disks and the check disk

RAID – RAID 4

- RAID 4
  - Operations:
    - Write :
      - Smaller than a stripe unit:
        - New parity = (Old data **XOR** New data) **XOR** old parity
          - So, 2 read operations and 2 write operations (1 data disk and 1 parity disk)
      - Parity disk is a bottleneck