Chapter 8
Adding a Disk

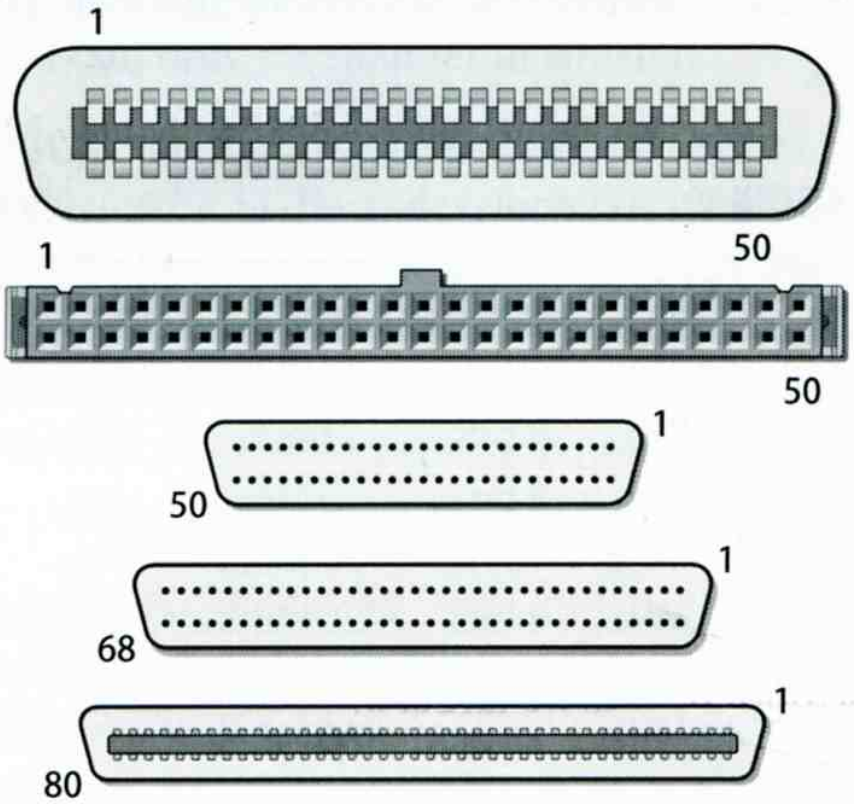
Disk Interface

- SCSI
  - Small Computer Systems Interface
  - High performance and reliability
- IDE (or ATA)
  - Integrated Device Electronics (or AT Attachment)
  - Low cost
  - Become acceptable for enterprise with the help of RAID technology
- SATA
  - Serial ATA
- SAS
  - Serial Attached SCSI
- USB
  - Universal Serial Bus
  - Convenient to use

Disk Interface – SCSI Interface Evolution

Version	Freq.	Width	Speed	Length	Diff.
SCSI-1	5MHz	8 bits	5MB/s	6m	25m
SCSI-2	5MHz	8 bits	5MB/s	6m	25m
SCSI-2 Fast	10MHz	8 bits	10MB/s	3m	25m
SCSI-2 Fast Wide	10MHz	16 bits	20MB/s	3m	25m
Ultra SCSI	20MHz	8 bits	20MB/s	1.5m	25m
Ultra Wide SCSI	20MHz	16 bits	40MB/s	1.5m	25m
Ultra2 SCSI	40MHz	16 bits	80MB/s	-	12m
Ultra160 SCSI	80MHz	16 bits	160MB/s	-	12m
Ultra320 SCSI	160MHz	16 bits	320MB/s	-	12m

Disk Interface – SCSI Interface Connector



Centronics
50 pins, SCSI-1/2, external

Ribbon connector (female)
50 pins, SCSI-1/2, internal

Mini-micro, aka HD50
50 pins, SCSI-2, external

Wide mini-micro, aka HD68
68 pins, SCSI-2/3, int/ext

SCA-2
80 pins, SCSI-3, internal

Disk Interface – SCSI Interface

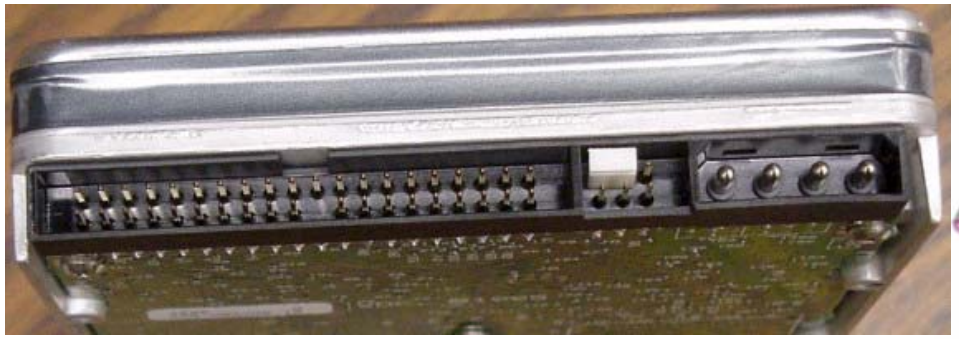
- Daisy chain on SCSI bus
  - Most external devices have two SCSI ports
  - Terminator
- Each SCSI device has a SCSI ID

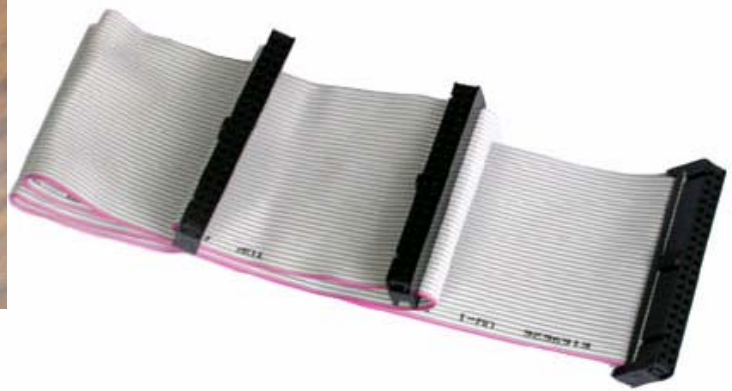
Disk Interface – ATA & SATA

- ATA (AT Attachment)
  - ATA2
    - PIO, DMA
    - LBA (Logical Block Addressing)
  - ATA3, Ultra DMA/33/66/100/133
  - ATAPI (ATA Packet Interface)
    - CDROM, TAP
  - Only one device can be active at a time
    - **SCSI support overlapping commands, command queuing, scatter-gather I/O**
  - Master-Slave
  - 40-pin ribbon cable
- SATA
  - Serial ATA

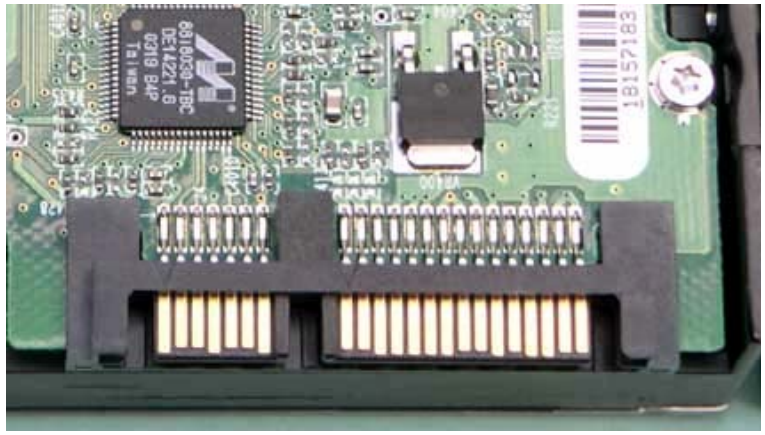
Disk Interface – ATA & SATA Interface

- ATA interface and it's cable





- SATA interface and it's cable



Disk Interface – SAS

- **SAS – Serial Attached SCSI**
- **SAS vs parallel SCSI**
  - SAS uses Serial transfer protocol to interface multiple devices hence lesser signaling overhead than parallel SCSI, resulting in higher speed.
  - No bus contention as SAS bus is point-to-point while SCSI bus is multidrop. Each device is connected by a dedicated bus to the initiator. Connection through expanders may appear to cause some contention, but this is transparent to the initiator.
  - SAS has no termination issues and does not require terminator packs like parallel SCSI.
  - SAS eliminates skew.
  - SAS supports higher number of devices (> 16384) while Parallel SCSI limits it to 16 or 32.
  - SAS supports higher transfer speed (1.5, 3.0 or 6.0 Gbps). The speed is realized on each initiator-target connection, hence higher throughput whereas in parallel SCSI the speed is shared across the entire multidrop bus.
  - SAS supports SATA devices.
  - SAS uses SCSI commands to interface with SAS End devices.

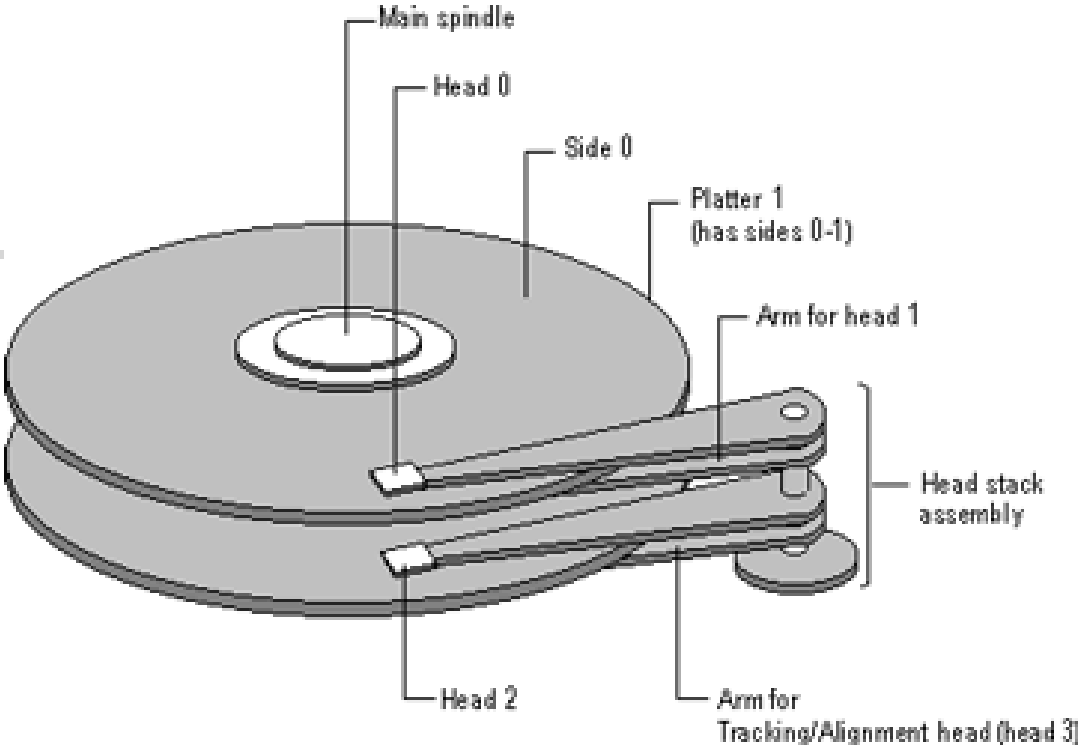
Disk Interface – USB

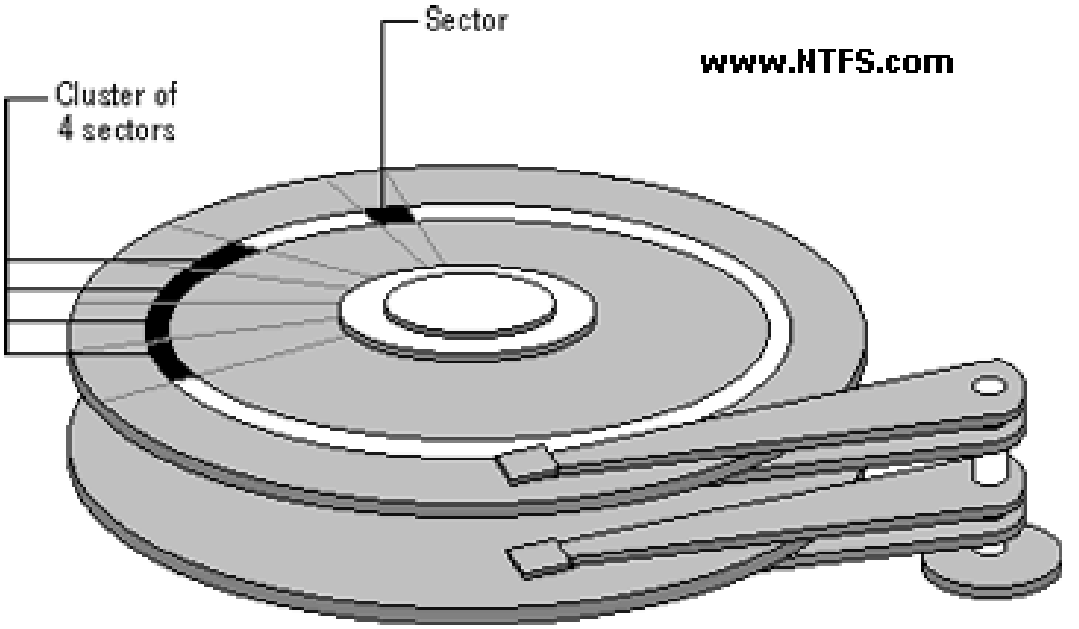
- USB to IDE/SATA Converter

Disk Geometry (1)

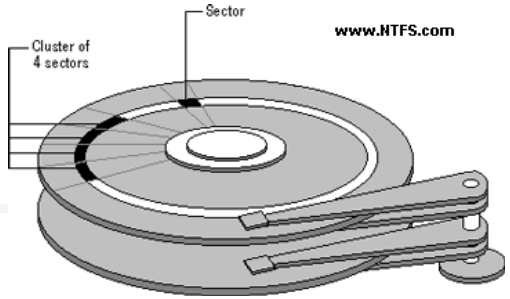
- sector
  - Individual data block
- track
  - circle
- cylinder
  - circle on all platters

- Position
  - CHS
  - Cylinder, Head, Sector





Disk Geometry (2)



- 40G HD
  - 4866 cylinders, 255 heads
  - 63 sectors per track, 512 bytes per sector
  - 512 * 63 * 4866 * 255 = 40,024,212,480 bytes
  - 1KB = 1024 bytes
  - 1MB = 1024 KB = 1,048,576 bytes
  - 1GB = 1024 MB = 1,073,741,824 bytes
  - 42,278,584,320 / 1,073,741,824 ≒ 39.375 GB

Disk Installation Procedure (1)

- The procedure involves the following steps:
  - **Connecting the disk to the computer**
    - **IDE: master/slave**
    - **SCSI: ID, terminator**
    - **power**
  - **Creating device files**
    - **/dev**
    - **Now auto created by devfs (man defvs)**
  - **Formatting the disk**
    - **Low-level format**
      - Address information and timing marks on platters
      - bad sectors
    - **Manufacturer diagnostic utility**

Disk Installation Procedure (2)

- **Partitioning and Labeling the disk**
  - **Allow the disk to be treated as a group of independent data area**
  - **root, home, swap partitions**
  - **Suggestion:**
    - **/var, /tmp ➔ separate partition**
    - Make a copy of root filesystem for emergency
- **Establishing logical volumes**
  - **Combine multiple partitions into a logical volume**
  - **Software RAID technology**
    - **FreeBSD (gvinum)**
    - **Linux (Linux LVM)**
    - **Sun (Solstice Disk Suite)**

Disk Installation Procedure (3)

- **Creating UNIX filesystems within disk partitions**
  - **Use "newfs" to install a filesystem for a partition**
  - **Filesystem components**
    - **A set of inode storage cells**
    - **A set of data blocks**
    - **A set of superblocks**
    - **A map of the disk blocks in the filesystem**
    - **A block usage summary**

Disk Installation Procedure (4)

- Superblock contents
  - The length of a disk block
  - Inode table's size and location
  - Disk block map
  - Usage information
  - Other filesystem's parameters

※ sync system call

Flush the cashed superblocks in-memory copy to the permanent place in disk

Disk Installation Procedure (5)

- **mount**
  - **Bring the new partition to the filesystem tree**
  - **mount point can be any directory**
  - **% mount /dev/ad1s1e /home2**
- **Setting up automatic mounting**
  - **Automount at boot time**
    - **/etc/fstab**
    - **% mount –t ufs /dev/ad2s1a /backup**
    - **% mount –t cd9600 –o ro,noauto /dev/acd0c /cdrom**

```
chwong@sabsd:/etc> less fstab
# Device            Mountpoint    FStype    Options       Dump    Pass#
/dev/ad0s1b          none          swap      sw            0       0
/dev/ad2s1b          none          swap      sw            0       0
/dev/ad0s1a          /             ufs       rw            1       1
/dev/acd0c           /cdrom        cd9660    ro,noauto     0       0
proc                 /proc         procfs    rw            0       0
/dev/ad2s1a          /backup       ufs       rw,noauto     1       1
ccduty:/bsdhome      /bsdhome      nfs       rw,noauto     0       0
```

Disk Installation Procedure (6)

- **Setting up swapping on swap partitions**
  - **swapon command**

fsck – check and repair filesystem (1)

- System crash will cause
  - Inconsistency between memory image and disk contents
- fsck –p
  - Examine all local filesystem listed in /etc/fstab at boot time
  - Automatically correct the following damages:
    - Unreferenced inodes
    - Inexplicably large link counts
    - Unused data blocks not recorded in block maps
    - Data blocks listed as free but used in file
    - Incorrect summary information in the superblock

fsck – check and repair filesystem (2)

- Run fsck in manual to fix serious damages
  - Blocks claimed by more than one file
  - Blocks claimed outside the range of the filesystem
  - Link counts that are too small
  - Blocks that are not accounted for
  - Directories that refer to unallocated inodes
  - Other errors
- fsck will suggest you the action to perform
  - Delete, repair, …

Adding a disk to FreeBSD (1)

1. Check disk connection
   - > Look system boot message

     ad3: 16383MB <Virtual HD> [33288/16/63] at ata1-slave WDMA2

2. Use /stand/sysinstall to install the new HD
   - > Configure → Fdisk → Label
   - > Don't forget to "W" the actions
3. Make mount point and mount it
   - > % mkdir /home2
   - > % mount –t ufs /dev/ad3s1e /home2
   - > % df
4. Edit /etc/fstab

Adding a disk to FreeBSD (2)

- If you forget to enable soft-update when you add the disk
  - % umount /home2
  - % tunefs –n enable /dev/ad3s1e
  - % mount –t ufs /dev/ad3s1e /home2
  - % mount

```
/dev/ad0s1a on / (ufs, local, soft-updates)
/dev/ad1s1e on /home (ufs, local, soft-updates)
procfs on /proc (procfs, local)
/dev/ad3s1e on /home2 (ufs, local, soft-updates)
```

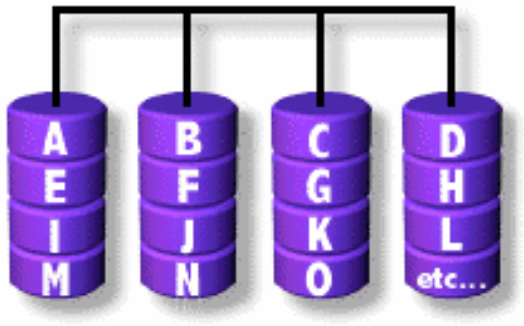
RAID (1/2)

- Redundant Array of Inexpensive Disks
  - A method to combine several physical hard drives into one logical unit
- Depending on the type of RAID, it has the following benefits:
  - Fault tolerance
  - Higher throughput
  - Real-time data recovery
- RAID Level
  - RAID 0, 1, 0+1, 2, 3, 4, 5, 6
  - Hierarchical RAID

RAID (2/2)

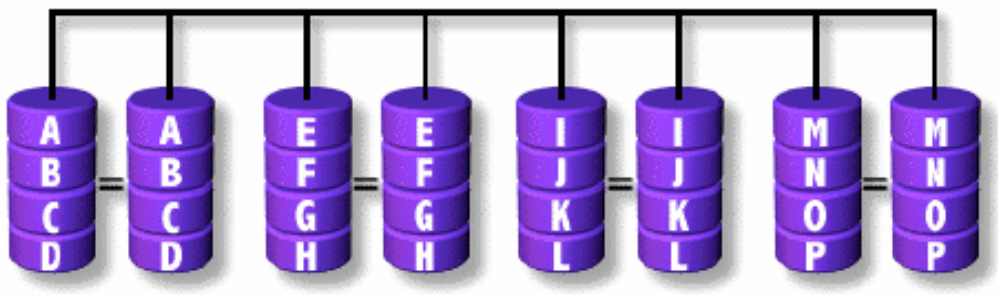
- Hardware RAID
  - There is a dedicate controller to take over the whole business
  - RAID Configuration Utility after BIOS
    - Create RAID array, build Array
- Software RAID
    - **FreeBSD (gvinum)**
    - **Linux (Linux LVM)**
    - **Sun (Solstice Disk Suite)**

RAID 0



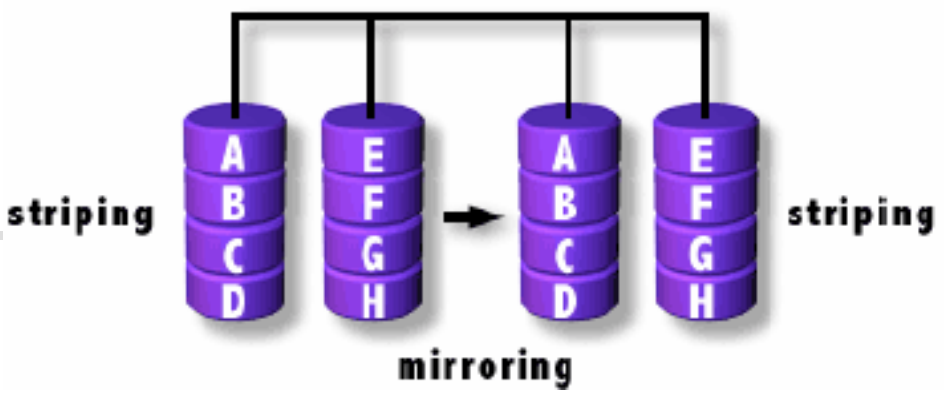
- Stripped data intro several disks
- Minimum number of drives: 2
- Advantage
  - Performance increase in proportional to n theoretically
  - Simple to implement
- Disadvantage
  - No fault tolerance
- Recommended applications
  - Non-critical data storage
  - Application requiring high bandwidth (such as video editing)

RAID 1



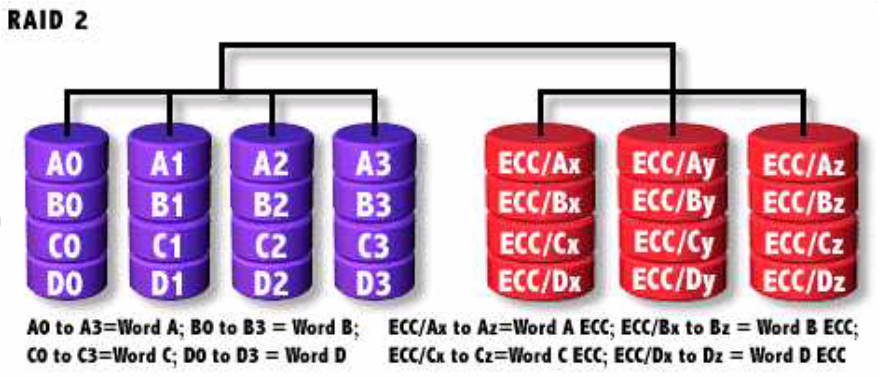
- Mirror data into several disks
- Minimum number of drives: 2
- Advantage
  - 100% redundancy of data
- Disadvantage
  - 100% storage overage
  - Moderately slower write performance
- Recommended application
  - Application requiring very high availability (such as home)

RAID 0+1



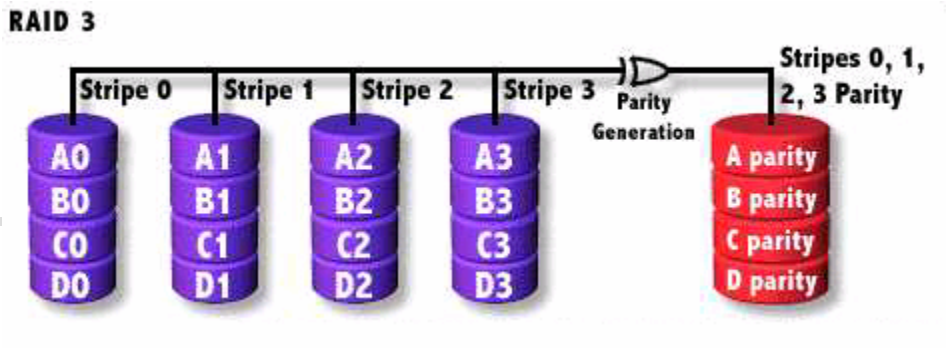
- Combine RAID 0 and RAID 1
- Minimum number of drives: 4

RAID 2



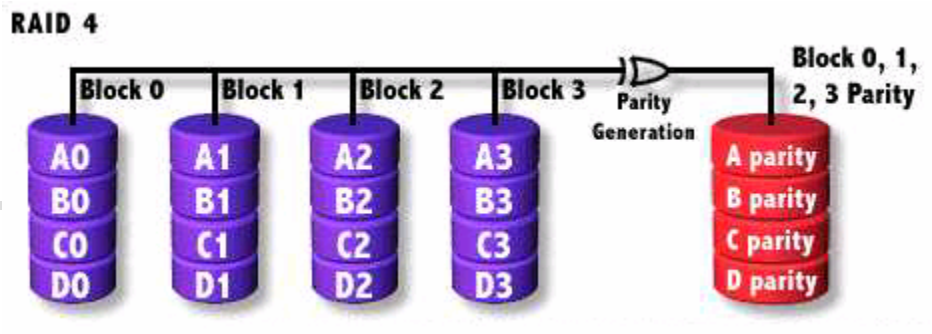
- Hamming Code ECC Each bit of data word
- Advantages:
  - "On the fly" data error correction
- Disadvantages:
  - Inefficient
  - Very high ratio of ECC disks to data disks
- Recommended Application
  - No commercial implementations exist / not commercially viable

RAID 3



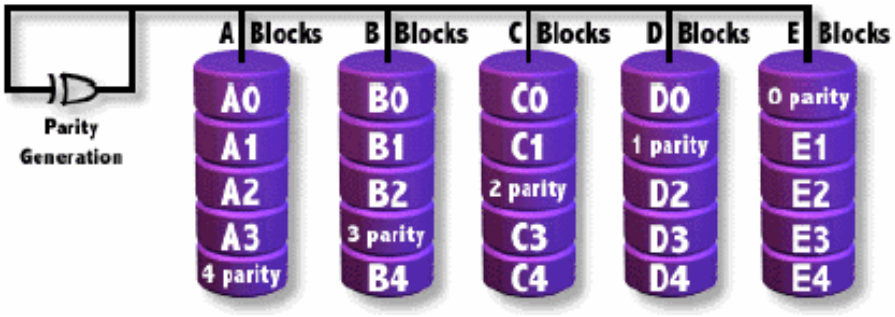
- Parallel transfer with Parity
- Minimum number of drives: 3
- Advantages:
  - Very high data transfer rate
- Disadvantages:
  - Transaction rate equal to that of a single disk drive at best
- Recommended Application
  - Any application requiring high throughput

RAID 4



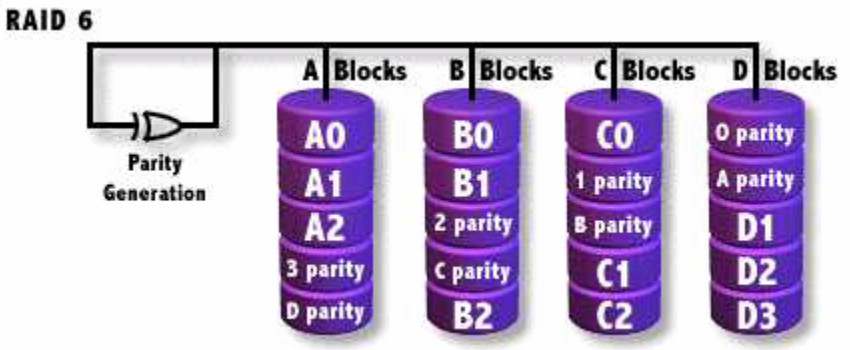
- Similar to RAID3
- RAID 3 V.S RAID 4
  - Byte Level V.S Block Level

RAID 5



- Independent Disk with distributed parity blocks
- Minimum number of drives: 3
- Advantage
  - Highest read data rate
  - Medium write data rate
- Disadvantage
  - Disk failure has a medium impact on throughput
  - Complex controller design
  - When one disk failed, you have to rebuild the RAID array

RAID 6



- Similar to RAID5
- Minimum number of drives: 4
- 2 parity checks, 2 disk failures tolerable.