Input/Output and Storage Systems

Objectives

- Understand how I/O systems work, including I/O methods and architectures.
- Become familiar with storage media, and the differences in their respective formats.
- Understand how RAID improves disk performance and reliability.
- Become familiar with the concepts of data compression and applications suitable for each type of compression algorithm.

7.1 Introduction

- Data storage and retrieval is one of the primary functions of computer systems.
  - One could easily make the argument that computers are more useful to us as data storage and retrieval devices than they are as computational machines.
- All computers have I/O devices connected to them, and to achieve good performance I/O should be kept to a minimum!
- In studying I/O, we seek to understand the different types of I/O devices as well as how they work.

7.2 I/O and Performance

- Sluggish I/O throughput can have a ripple effect, dragging down overall system performance.
  - This is especially true when virtual memory is involved.
- The fastest processor in the world is of little use if it spends most of its time waiting for data.
- If we really understand what's happening in a computer system we can make the best possible use of its resources.

7.3 Amdahl's Law

- The overall performance of a system is a result of the interaction of all of its components.
- System performance is most effectively improved when the performance of the most heavily used components is improved.
- This idea is quantified by Amdahl's Law:

$$S = \frac{1}{(1-f) + \frac{f}{k}}$$

where
S is the overall speedup
f is the fraction of work performed by a faster component
k is the speedup of a faster component.

7.3 Amdahl's Law

- Amdahl's Law gives us a handy way to estimate the performance improvement we can expect when we upgrade a system component.
- Example: On a large system,
  - suppose we can upgrade a CPU to make it 50% faster for $10,000 or upgrade its disk drives for $7,000 to make them 2.5 times faster (250% faster).
  - Processes spend 70% of their time running in the CPU and 30% of their time waiting for disk service.
  - An upgrade of which component would offer the greater benefit for the lesser cost?

7.3 Amdahl's Law

- The processor option offers a 130% speedup:

$f = 0.70$, $k = 1.5$ $\quad S = \frac{1}{(1 - 0.7) + 0.7/1.5}$

- And the disk drive option gives a 122% speedup:

$f = 0.30$, $k = 2.5$ $\quad S = \frac{1}{(1 - 0.3) + 0.3/2.5}$

- Each 1% of improvement for the processor costs $333, and for the disk a 1% improvement costs $318.

Should price/performance be your only concern?

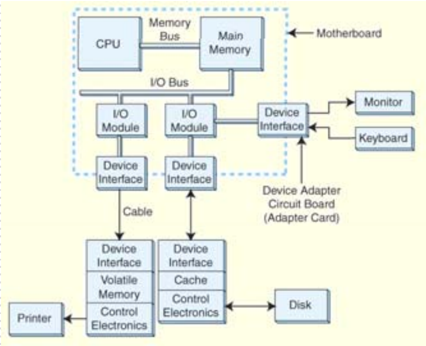
7.3 Amdahl's Law

2. (page 381) Suppose the daytime processing load consists of 60% CPU activity and 40% disk activity. Your customers are complaining that the system is slow. After doing some research, you have learned that you can upgrade your disks for $8,000 to make them 2.5 times as fast as they are currently. You have also learned that you can upgrade your CPU to make it 1.4 as fast for $5,000.
   a. Which would you choose to yield the best performance improvement for the least amount of money?
   b. Which option would you choose if you don't care about the money, but want a faster system?
   c. What is the break-even point for the upgrades? That is, what price would we need to charge for the CPU (or the disk – change only one) so the result was the same cost per 1% increase for both?

7.4 I/O Architectures

- We define input/output as a subsystem of components that moves coded data between external devices and a host system (CPU, main memory).
- I/O subsystems include:
  - Blocks of main memory that are devoted to I/O functions.
  - Buses that move data into and out of the system.
  - Control modules in the host and in peripheral devices
  - Interfaces to external components such as keyboards and disks.
  - Cabling or communications links between the host system and its peripherals.

7.4 Model I/O Configuration



7.4.1 I/O Control Methods

I/O can be controlled in four general ways.

1. *Programmed I/O* reserves a register for each I/O device. Each register is continually polled to detect data arrival.
2. *Interrupt-Driven I/O* allows the CPU to do other things until I/O is requested.
3. *Direct Memory Access* (DMA) offloads I/O processing to a special-purpose chip that takes care of the details.
4. *Channel I/O* uses dedicated I/O processors.

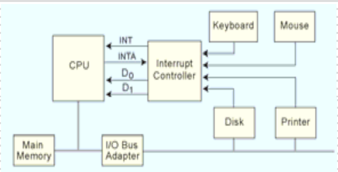
7.4.1 Programmed (Polled) I/O

- Reserves a register for each I/O device
- *Polling* -- CPU continually monitors each register waiting for data
- If data ready condition is true, CPU processes the data
- Pros:
  - We have programmatic control over the behavior of each I/O device
  - Adjustments can be made to the number and types of devices, polling priorities and intervals.
- Con:
  - Constant register polling -- CPU is in a busy wait state
  - How often to poll
- Best suited for special-purpose systems
  - Automated teller machines and systems
  - Monitoring or control systems

7.4.1 Interrupt-Driven I/O

- This is an idealized I/O subsystem that uses interrupts.
- Each device connects its interrupt line to the interrupt controller.

The controller signals the CPU when any of the interrupt lines are asserted.

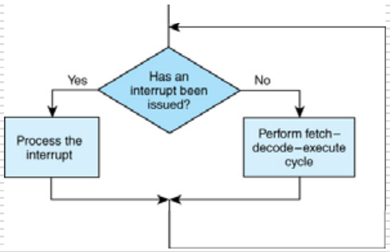


7.4.1 Interrupt-Driven I/O

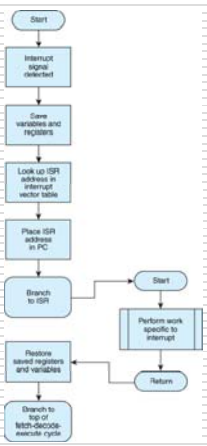
- Recall from Chapter 4 that in a system that uses interrupts, the status of the interrupt signal is checked at the top of the fetch-decode-execute cycle.
- The particular code that is executed whenever an interrupt occurs is determined by a set of addresses called interrupt vectors that are stored in low memory.
- The system state is saved before the interrupt service routine is executed and is restored afterward.

7.4.1 Interrupt-Driven I/O

- I/O device send a request (interrupt) to CPU for servicing
- *Interrupt flag* – a bit in the flag register to signal to CPU
- Interrupt vectors – service routines
- Service routines can be modified.



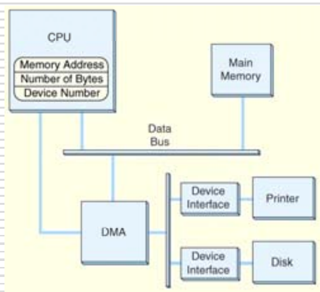
The Fetch-Decode-Interrupt Cycle with Interrupt Checking



Processing an Interrupt

7.4.1 Direct Memory Access (DMA)

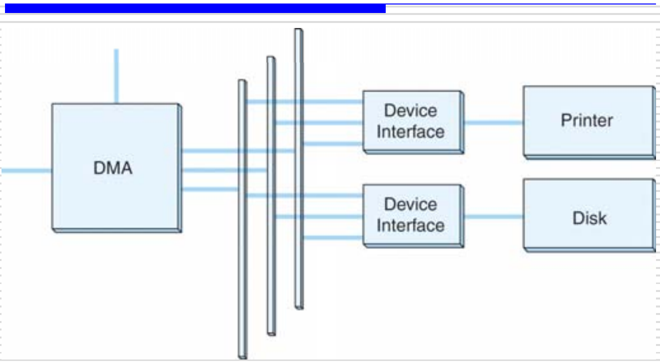
- Notice that the DMA and the CPU share the bus.
- CPU provides DMA with location of the bytes, # bytes to be transferred, destination device or memory address
- The DMA runs at a higher priority and steals memory cycles from the CPU.



7.4.1 Direct Memory Access (DMA)

```
WHILE More – input AND NOT Error
    ADD 1 TO Byte-count
    IF Byte-count > Total-bytes-to-be-transferred THEN
        EXIT
    ENDIF
    Place byte in destination buffer
    Raise Byte-ready signal
    Initialize timer
    REPEAT
        WAIT
    UNTIL Byte-acknowledged, Timeout, OR Error
ENDWHILE
```


7.4.1 Direct Memory Access (DMA)



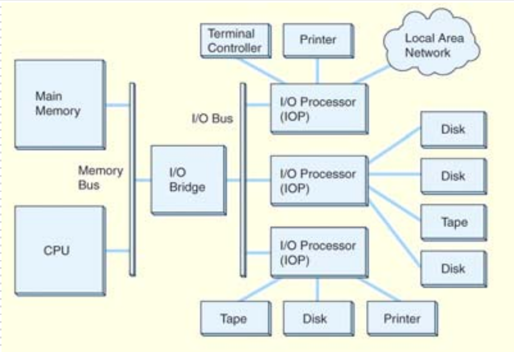
7.4.1 Channel I/O

- Very large systems employ channel I/O.
- Channel I/O consists of one or more I/O processors (IOPs) that control various channel paths.
- Slower devices such as terminals and printers are combined (multiplexed) into a single faster channel.
- On IBM mainframes, multiplexed channels are called multiplexor channels, the faster ones are called selector channels.

7.4.1 Channel I/O

- Channel I/O is distinguished from DMA by the intelligence of the IOPs.
- The IOP negotiates protocols, issues device commands, translates storage coding to memory coding, and can transfer entire files or groups of files independent of the host CPU.
- The host has only to create the program instructions for the I/O operation and tell the IOP where to find them.

7.4.1 Channel I/O Configuration



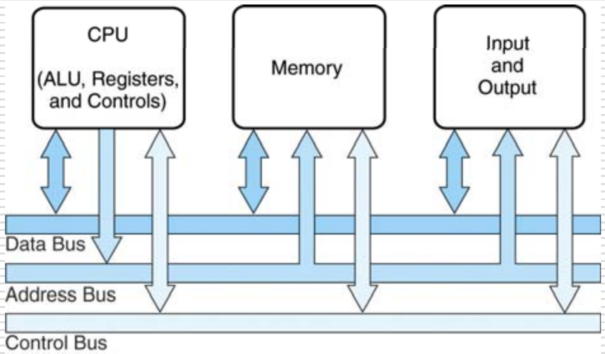
7.4.2 Character I/O Versus Block I/O

- Character I/O devices process one byte (or character) at a time.
  - Examples include modems, keyboards, and mice.
  - Keyboards are usually connected through an interrupt-driven I/O system.
- Block I/O devices handle bytes in groups.
  - Most mass storage devices (disk and tape) are block I/O devices.
  - Block I/O systems are most efficiently connected through DMA or channel I/O.

7.4.3 I/O Bus Operations

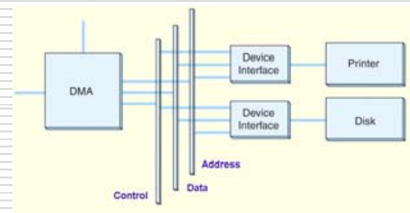
- I/O buses, unlike memory buses, operate asynchronously. Requests for bus access must be arbitrated among the devices involved.
- Bus control lines activate the devices when they are needed, raise signals when errors have occurred, and reset devices when necessary.
- The number of data lines is the width of the bus.
- A bus clock coordinates activities and provides bit cell boundaries.

7.4.3 System Bus

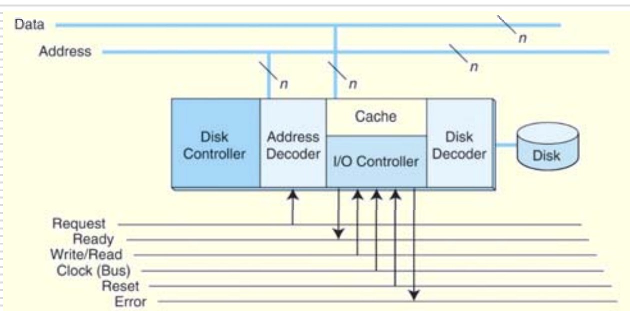


7.4.3 System Bus

- This is a generic DMA configuration showing how the DMA circuit connects to a data bus.



7.4.3 I/O Bus Connections

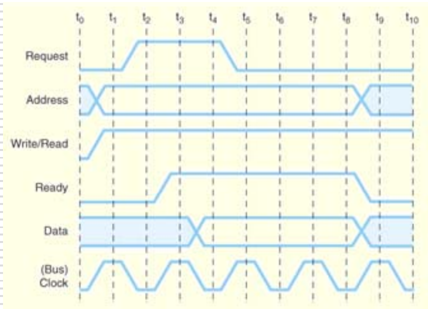


7.4.3 Write Data to Disk

1. The DMA circuit places the address of the disk controller on the address lines, and raises (asserts) the **Request** and **Write** signals.
2. With the **Request** signal asserted, decoder circuits in the controller interrogate the address lines.
3. Upon sensing its own address, the decoder enables the disk control circuits. If the disk is available for writing data, the controller asserts a signal on the **Ready** line. At this point, the handshake between the DMA and the controller is complete. With the **Ready** signal raised, no other devices may use the bus.
4. The DMA circuits then place the data on the lines and lower the **Request** signal.
5. When the disk controller sees the **Request** signal drop, it transfer the byte from data lines to the disk buffer, and then lowers its **Ready** signal.

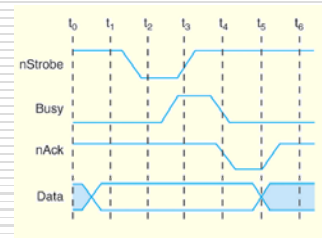
7.4.3 Bus Timing Diagram

Timing diagrams define bus operation in detail.



7.5 Data Transmission Modes

- Bytes can be conveyed from one point to another by sending their encoding signals simultaneously using parallel data transmission or by sending them one bit at a time in serial data transmission
  - Parallel data transmission for a printer resembles the signal protocol of a memory bus:



7.5 Data Transmission Modes

- In parallel data transmission, the interface requires one conductor for each bit.
- Parallel cables are fatter than serial cables.
- Compared with parallel data interfaces, serial communications interfaces:
  - Require fewer conductors.
  - Are less susceptible to attenuation.
  - Can transmit data farther and faster.

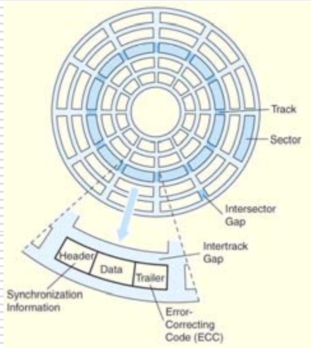
Serial communications interfaces are suitable for time-sensitive (*isochronous*) data such as voice and video.

7.6 Magnetic Disk Technology

- Magnetic disks offer large amounts of durable storage that can be accessed quickly.
- Disk drives are called random (or direct) access storage devices, because blocks of data can be accessed according to their location on the disk.
  - This term was coined when all other durable storage (e.g., tape) was sequential.
- Magnetic disk organization is shown on the following slide.

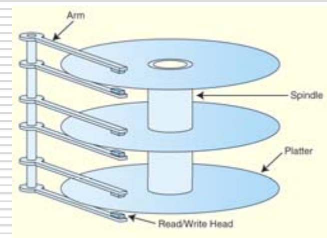
7.6 Magnetic Disk Technology

- Disk tracks are numbered from the outside edge, starting with zero.



7.6.1 Rigid Disk Drives

- Hard disk platters are mounted on spindles.
- Read/write heads are mounted on a comb that swings radially to read the disk.
- The rotating disk forms a logical cylinder beneath the read/write heads.
- Data blocks are addressed by their cylinder, surface, and sector.



7.6.1 Rigid Disk Drives

- There are a number of electromechanical properties of hard disk drives that determine how fast its data can be accessed.
- Seek time is the time that it takes for a disk arm to move into position over the desired cylinder.
- Rotational delay is the time that it takes for the desired sector to move into position beneath the read/write head.
- Seek time + rotational delay = access time.

7.6.1 Rigid Disk Drives

- Transfer rate gives us the rate at which data can be read from the disk.
- Average latency is a function of the rotational speed:

$$\frac{\dfrac{60 \text{ seconds}}{\text{disk rotation speed}} \times \dfrac{1000 \text{ ms}}{\text{second}}}{2}$$

- Mean Time To Failure (MTTF) is a statistically-determined value often calculated experimentally.
  - It usually doesn't tell us much about the actual expected life of the disk. Design life is usually more realistic.

7.6.1 Rigid Disk Drives

CONFIGURATION:	
Formatted Capacity, MB	1340
Integrated Controller	SCSI
Encoding Method	RLL 1,7
Buffer Size	64K
Platters	3
Data Surfaces	5
Tracks per Surface	3,196
Track Density	5,080 tpi
Recording Density	92.2 Kbpi
Bytes per Block	512
Sectors per Track	132
PHYSICAL:	
Height	12.5mm
Length	100mm
Width	70mm
Weight	170g
Temperature (C°)	
Operating	5°C to 55°C
Non-operating/Storage	40°C to 71°C
Relative Humidity	5% to 95%
Acoustic Noise	33dBA, idle

RELIABILITY AND MAINTENANCE:	
MTTF	300,000 hours
Start/Stop Cycles	50,000
Design Life	5 years (minimum)
Data Errors	
(non-recoverable)	<1 per 10^{13} bits read
PERFORMANCE:	
Seek times	
Track to Track	4.5 ms
Average	14 ms
Average Latency	6.72 ms
Rotational Speed	
(+/–0.20%)	4,464 rpm
Controller Overhead	<200 μSec
Data Transfer Rate:	
To/from Media	6.0 MB/Sec
To/from Host	11.1 MB/Sec
Start Time	
(0 – Drive Ready)	5 sec

POWER REQUIREMENTS

Mode	+5VDC +5% – 10%	Power +5.0VDC
Spin-up	1000mA	5000mW
Idle	190mA	950mW
Standby	50mA	250mW
Sleep	6mA	30mW

Problem 18 page 382

18. Suppose a disk drive has the following characteristics:
    - 4 surfaces
    - 1024 tracks per surface
    - 128 sectors per track
    - 512 bytes/sector
    - Track-to-track seek time of 5 milliseconds
    - Rotational speed of 5000 RPM.

    a. What is the capacity of the drive?

    b. What is the access time?

Problem 19 page 382

19. Suppose a disk drive has the following characteristics:
    - 5 surfaces
    - 1024 tracks per surface
    - 256 sectors per track
    - 512 bytes/sector
    - Track-to-track seek time of 8 milliseconds
    - Rotational speed of 7500 RPM.

    a. What is the capacity of the drive?

    b. What is the access time?

    c. Is this disk faster than the one described in question 18? Explain.

7.6.2 Flexible (Floppy) Disks

- Floppy (flexible) disks are organized in the same way as hard disks, with concentric tracks that are divided into sectors.
- Physical and logical limitations restrict floppies to much lower densities than hard disks.
- A major logical limitation of the DOS/Windows floppy diskette is the organization of its file allocation table (FAT).
  - The FAT gives the status of each sector on the disk: Free, in use, damaged, reserved, etc.

7.6.2 Flexible (Floppy) Disks

- Sector 0 is the boot sector of the disk.
- On a standard 1.44MB floppy, the FAT is limited to nine 512-byte sectors.
  - There are two copies of the FAT.
- There are 18 sectors per track and 80 tracks on each surface of a floppy, for a total of 2880 sectors on the disk. So each FAT entry needs at least 14 bits.
  - FAT entries are actually 16 bits, and the organization is called FAT16.

7.6.2 Flexible (Floppy) Disks

- The disk directory associates logical file names with physical disk locations.
- Directories contain a file name and the file's first FAT entry.
- If the file spans more than one sector (or cluster), the FAT contains a pointer to the next cluster (and FAT entry) for the file.
- The FAT is read like a linked list until the <EOF> entry is found.

7.6.2 Flexible (Floppy) Disks

- A directory entry says that a file we want to read starts at sector 121 in the FAT fragment shown below.

FAT Index →	120	121	122	123	124	125	126	127
FAT Contents	97	124	<EOF>	1258	126	<BAD>	122	577

  - Sectors 121, 124, 126, and 122 are read. After each sector is read, its FAT entry is to find the next sector occupied by the file.
  - At the FAT entry for sector 122, we find the end-of-file marker <EOF>.

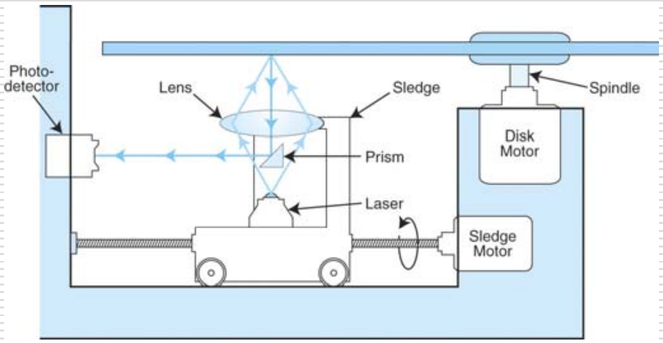
7.7 Optical Disks

- Optical disks provide large storage capacities very inexpensively.
- They come in a number of varieties including CD-ROM, DVD, and WORM.
- Many large computer installations produce document output on optical disk rather than on paper. This idea is called COLD-- Computer Output Laser Disk.
- It is estimated that optical disks can endure for a hundred years. Other media are good for only a decade-- at best.

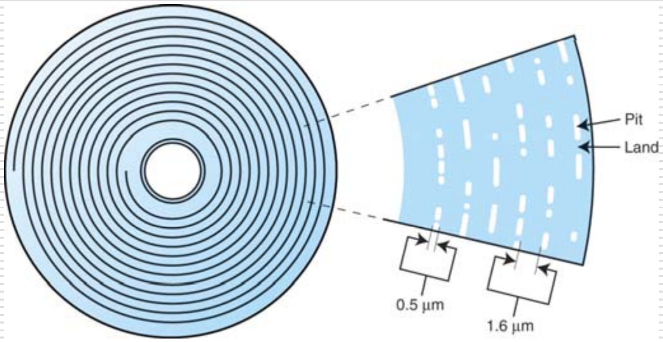
7.7.1 CD-ROM

- CD-ROMs were designed by the music industry in the 1980s, and later adapted to data.
- This history is reflected by the fact that data is recorded in a single spiral track, starting from the center of the disk and spanning outward.
- Binary ones and zeros are delineated by bumps in the polycarbonate disk substrate. The transitions between pits and lands define binary ones.
- If you could unravel a full CD-ROM track, it would be nearly five miles long!

7.7.1 CD-ROM

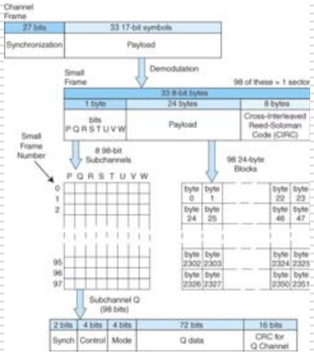


7.7.1 CD-ROM

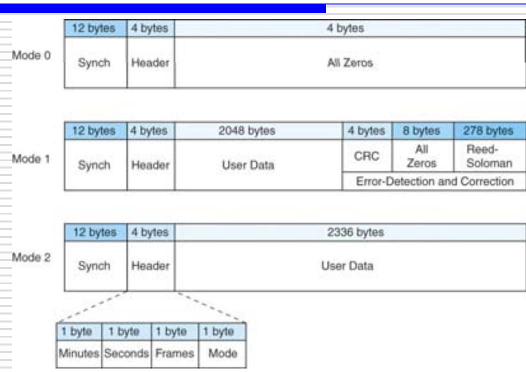


7.7.1 CD-ROM

- Data stored in 2352-byte chunk -- sectors
- 98 588 bit sectors -- *channel frames*



7.7.1 CD-ROM



7.7.1 CD-ROM

- The logical data format for a CD-ROM is much more complex than that of a magnetic disk.
- Different formats are provided for data and music.
- Two levels of error correction are provided for the data format.
- Because of this, a CD holds at most 650MB of data, but can contain as much as 742MB of music.

7.7.2 DVD

- DVDs can be thought of as quad-density CDs.
  - Varieties include single sided, single layer, single sided double layer, double sided double layer, and double sided double layer.
- Where a CD-ROM can hold at most 650MB of data, DVDs can hold as much as 17GB.
- One of the reasons for this is that DVD employs a laser that has a shorter wavelength than the CD's laser.
- This allows pits and land to be closer together and the spiral track to be wound tighter.
- It is possible that someday DVDs will make CDs obsolete.

7.7.2 DVD

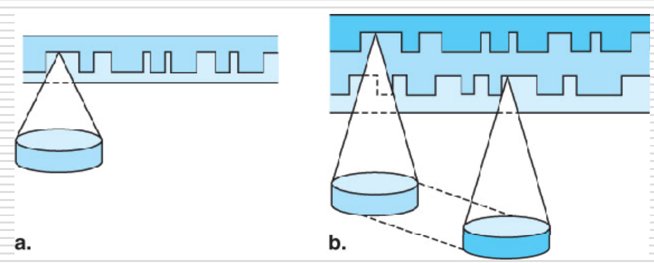


Figure 07.20
A Laser Focusing on: a) a Single-Layer DVD and b) a Double-Layer DVD one Layer at a Time

7.7.2 DVD

- A shorter wavelength light can read and write bytes in greater densities than can be done by a longer wavelength laser.
- This is one reason that DVD's density is greater than that of CD.
- The manufacture of blue-violet lasers can now be done economically, bringing about the next generation of laser disks.
- Two incompatible formats, HD-CD and Blu-Ray, are competing for market dominance.

7.7.3 Blue-Violet Laser Disks

- Blu-Ray was developed by a consortium of nine companies that includes Sony, Samsung, and Pioneer.
  - Maximum capacity of a single layer Blu-Ray disk is 25GB.
- HD-DVD was developed under the auspices of the DVD Forum with NEC and Toshiba leading the effort.
  - Maximum capacity of a single layer HD-DVD is 15GB.
- The big difference between the two is that HD-DVD is backward compatible with red laser DVDs, and Blu-Ray is not.

7.7.3 Blue-Violet Laser Disks

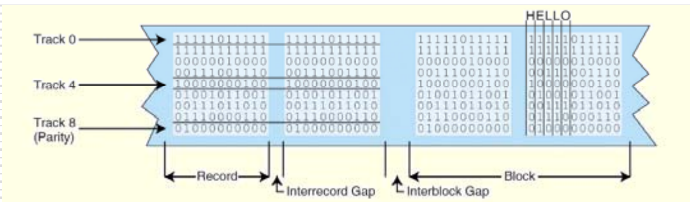
- Blue-violet laser disks have also been designed for use in the data center.
- The intention is to provide a means for long term data storage and retrieval.
- Two types are now dominant:
  - Sony's Professional Disk for Data (PDD) that can store 23GB on one disk and
  - Plasmon's Ultra Density Optical (UDO) that can hold up to 30GB.
- It is too soon to tell which of these technologies will emerge as the winner.

7.7.4 Optical Disk Recording Methods

- Ablative: a high-powered laser melts a pit in a reflective metal coatings sandwiched between the protective layers of the disk.
- Bimetallic Alloy: Two metallic layers are encased between protective coatings on the surfaces of the disk. Laser light fuses the two metallic layers together, causing a reflectance change in the lower metallic layer.
- Bubble-Forming: A single layer of thermally sensitive material is pressed between two plastic layers. When hit by high-powered laser light, bubbles form in the material, causing a reflectance change.

7.8 Magnetic Tape

- First-generation magnetic tape was not much more than wide analog recording tape, having capacities under 11MB.
- Data was usually written in nine vertical tracks:

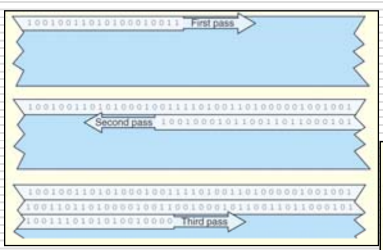


7.8 Magnetic Tape

- Today's tapes are digital, and provide multiple gigabytes of data storage.
- Two dominant recording methods are serpentine and helical scan, which are distinguished by how the read-write head passes over the recording medium.
- Serpentine recording is used in digital linear tape (DLT) and Quarter inch cartridge (QIC) tape systems.
- Digital audio tape (DAT) systems employ helical scan recording.

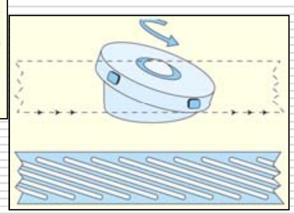
These two recording methods are shown on the next slide.

7.8 Magnetic Tape



← Serpentine

Helical Scan →



7.8 Magnetic Tape

- LTO, as the name implies, is a linear digital tape format.
- The specification allowed for the refinement of the technology through four "generations."
- Generation 3 was released in 2004.
  - Without compression, the tapes support a transfer rate of 80MB per second and each tape can hold up to 400GB.
- LTO supports several levels of error correction, providing superb reliability.
  - Tape has a reputation for being an error-prone medium.

7.8 Magnetic Tape

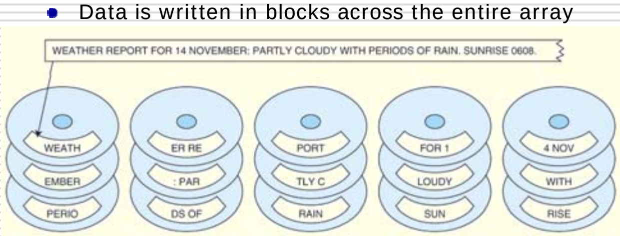
- Numerous incompatible tape formats emerged over the years.
  - Sometimes even different models of the same manufacturer's tape drives were incompatible!
- Finally, in 1997, HP, IBM, and Seagate collaboratively invented a best-of-breed tape standard.
- They called this new tape format Linear Tape Open (LTO) because the specification is openly available.

7.9 RAID

- RAID, an acronym for Redundant Array of Independent Disks was invented to address problems of disk reliability, cost, and performance.
- In RAID, data is stored across many disks, with extra disks added to the array to provide error correction (redundancy).
- The inventors of RAID, David Patterson, Garth Gibson, and Randy Katz, provided a RAID taxonomy that has persisted for a quarter of a century, despite many efforts to redefine it.

7.9.1 RAID Level 0

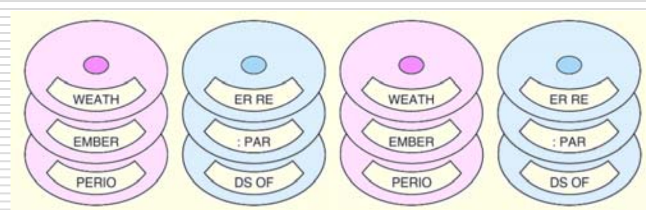
- RAID Level 0, also known as drive spanning, provides improved performance, but no redundancy.
  - Data is written in blocks across the entire array



  - The disadvantage of RAID 0 is in its low reliability.

7.9.2 RAID Level 1

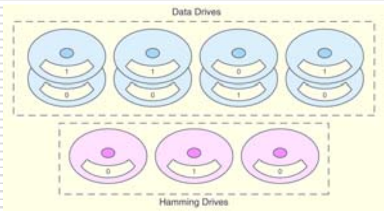
- RAID Level 1, also known as disk mirroring, provides 100% redundancy, and good performance.
  - Two matched sets of disks contain the same data.



  - The disadvantage of RAID 1 is cost.

7.9.3 RAID Level 3

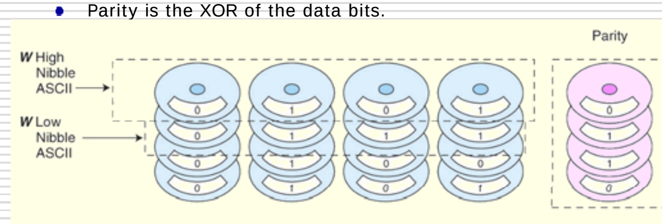
- A RAID Level 2 configuration consists of a set of data drives, and a set of Hamming code drives.
  - Hamming code drives provide error correction for the data drives.



  - RAID 2 performance is poor and the cost is relatively high.

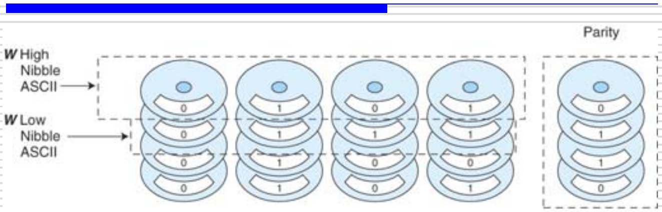
7.9.4 RAID Level 3

- RAID Level 3 stripes bits across a set of data drives and provides a separate disk for parity.
  - Parity is the XOR of the data bits.



  - RAID 3 is not suitable for commercial applications, but is good for personal systems.

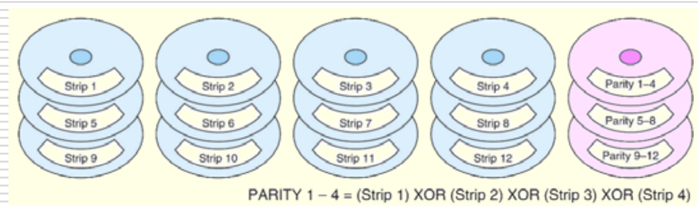
7.9.4 RAID Level 3



Letter	ASCII	Parity(even) High Nibble	Parity(even) Low Nibble
W	0101 0111	0	1
E	0100 0101	1	0
A	0100 0001	1	1
T	0101 0100	0	1
H	0100 1000	1	1
E	0100 0101	1	0
R	0101 0010	0	1

7.9.5 RAID Level 4

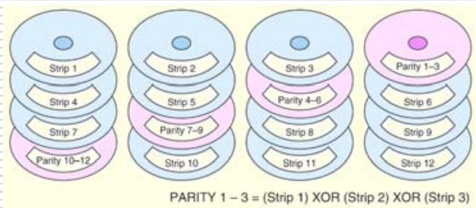
- RAID Level 4 is like adding parity disks to RAID 0.
  - Data is written in blocks across the data disks, and a parity block is written to the redundant drive.



  - RAID 4 would be feasible if all record blocks were the same size.

7.9.6 RAID Level 5

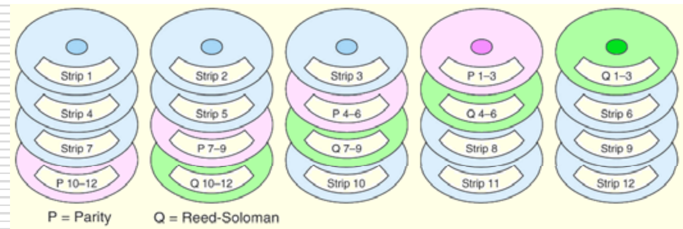
- RAID Level 5 is RAID 4 with distributed parity.
  - With distributed parity, some accesses can be serviced concurrently, giving good performance and high reliability.



  - RAID 5 is used in many commercial systems.

7.9.7 RAID Level 6

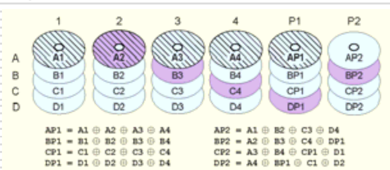
- RAID Level 6 carries two levels of error protection over striped data: Reed-Soloman and parity.
  - It can tolerate the loss of two disks.



  - RAID 6 is write-intensive, but highly fault-tolerant.

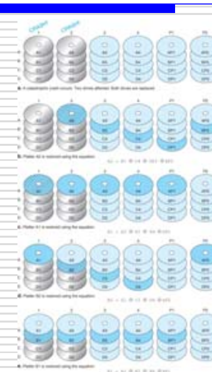
7.9.8 RAID DP

- Double parity RAID (RAID DP) employs pairs of over- lapping parity blocks that provide linearly independent parity functions.



Error Recovery Pattern for RAID DP
The recovery of A2 is provided by the overlap of equations AP1 and BP2.

Restoring Two Crashed Spindles using RAID DP



7.9.8 RAID DP

- Like RAID 6, RAID DP can tolerate the loss of two disks.
- The use of simple parity functions provides RAID DP with better performance than RAID 6.
- Of course, because two parity functions are involved, RAID DP's performance is somewhat degraded from that of RAID 5.
  - RAID DP is also known as EVENODD, diagonal parity RAID, RAID 5DP, advanced data guarding RAID (RAID ADG) and-- erroneously-- RAID 6.

7.9.9 Hybrid RAID Systems

- Large systems consisting of many drive arrays may employ various RAID levels, depending on the criticality of the data on the drives.
  - A disk array that provides program workspace (say for file sorting) does not require high fault tolerance.
- Critical, high-throughput files can benefit from combining RAID 0 with RAID 1, called RAID 10.
- Keep in mind that a higher RAID level does not necessarily mean a "better" RAID level. It all depends upon the needs of the applications that use the disks.

Summary of RAID Capabilities

RAID Level	Description	Reliability	Throughput	Pro and con
0	Block interleave data striping	Worse than single disk	Very good	Least cost, no protection
1	Data mirrored on second identical set	Excellent	Better than single disk on reads, worse on writes	Excellent protection, high cost
2	Bit interleave data striping with Hamming code	Good	Very good	Good performance, high cost, not used in practice
3	Bit interleave data striping with parity disk	Good	Very good	Good performance, reasonable cost
4	Block interleave data striping with one parity disk	Very good	Much worse on writes as single disk, very good on reads	Reasonable cost, poor performance, not used in practice
5	Block interleave data striping with distributed parity	Very good	On writes not as good as single disk, very good on reads	Good performance, reasonable cost
6	Block interleave data striping with dual error protection	Excellent	On writes much worse than single disk, very good on reads	Good performance, reasonable cost, complex to implement
10	Mirrored disk striping	Excellent	Better than single disk on reads, not as good as single disk on writes	Good performance, high cost, excellent protection
DP	Block interleave data striping with dual parity disks	Excellent	Better than single disk on reads, not as good as single disk on writes	Good performance, reasonable cost, excellent protection

Problem 30 Page 383

A particular high-performance computer system has been functioning as an e-business server on the Web. This system supports $10,000 per hour in gross business volume. It has been estimated that the net profit per hour is $1,200. In other words, if the system goes down, the company will lose $1,200 every hour until repairs are made. Furthermore, any data on the damaged disk would be lost. Some of this data could be retrieved from the previous night's backups, but the rest would be gone forever. Conceivably, a poorly-timed disk crash could cost your company hundreds of thousands of dollars in immediate revenue loss, and untold thousands in permanent business loss. The fact that this system is not using any type of RAID is disturbing to you.

Although your chief concern is data integrity and system availability, others in your group are obsessed with system performance. They feel that more revenue would be lost in the long run if the system slows down after RAID is installed. They have stated specifically that a system with RAID performing at half the speed of the current system would result in gross revenue dollars per hour declining to $5,000 per hour.

In total, 80% of the system e-business activity involves a database transaction. The database transactions consist of 60% reads and 40% writes. On average, disk access time is 20ms.

The disks on this system are nearly full and are nearing the end of their expected life, so new ones must be ordered soon. You feel that this is a good time to try to install RAID, even though you'll need to buy extra disks. The disks that are suitable for your system cost $2000 for each 10 gigabyte spindle. The average access time of these new disks is 15ms with a MTTF of 20,000 hours and a MTTR of 4 hours. You have projected that you will need 60 gigabytes of storage to accommodate the existing data as well as the expected data growth over the next 5 years. (All of the disks will be replaced.)

Problem 30 Page 383

a. Are the people who are against adding RAID to the system correct in their assertion that 50% slower disks will result in revenues declining to $5,000 per hour? Justify your answer.

No. Only 80% of the system activity involves the database. Proportionately then, doubling the disk access time would affect only 80% of the system activity. Access time would still be a lot slower according to their thinking:

(% of transaction on disk * half of the throughput)
= 0.8 * 0.5 * 10,000 = 4000
(% of transaction in CPU * 10,000) = 0.2 * 10,000 = 2000
Gross Volume= 4000 + 2000 = $6,000 per hour.

So, using their assumptions, revenues would decline by only $4,000 not $5,000!

Problem 30 Page 383

b. What would be the average disk access time on your system if you decide to use RAID-1?

In RAID-1, it takes twice as long to do a write as a read, because data has to be written twice. However, access time for a read is half of what we would expect from a system not using RAID-1, assuming that the disk arms are 180 degrees offset from one another.

Average Access Time
= 0.4 * (15 ms / 2) + 0.6 * (15 ms * 2) = 21 ms.

Problem 30 Page 383

c. What would be the average disk access time on your system using a RAID-5 array with two sets of 4 disks if 25% of the database transactions must wait behind one transaction for the disk to become free?

Average Access Time
= 0.75 * 15ms + 0.25 * 30ms = 18.75 ms.

Problem 30 Page 383

d. Which configuration has a better cost-justification, RAID-1 or RAID-5? Explain your answer.

Both RAID solutions will offer database response time comparable to what is currently offered by the system. The RAID-1 system will require 2*N disks while the 4-disk RAID-5 solution will require 133% of the number of disks. That is, RAID-1 will cost $24,000 and RAID-5 will cost $16,000. The cost of the disks isn't the big issue here, however. What matters most is system availability. With 8 disks each with a MTTF of 20,000 hours, we can expect a failure of at least two of the disks to fail within 20,000/8 hours, or 2,500 hours. So at least twice a year, we could expect a disk failure that will last 4 hours. If RAID-1 is used, the system will continue to function, while the RAID-5 system will be down, costing roughly $4,800 in lost revenue during each outage. (No data would be lost, though!)

Cost of RAID-1: $24,000;
Cost of RAID-5: $16,000 + $9,600 revenue loss = $25,600.

The RAID-1 is therefore more economical. Note: We have not included loss of goodwill and permanent business loss in the RAID-5 figure. This tilts the balance greatly in favor of the RAID-1 solution.

7.10 The Future of Data Storage

- Advances in technology have defied all efforts to define the ultimate upper limit for magnetic disk storage.
  - In the 1970s, the upper limit was thought to be around 2Mb/in2.
  - Today's disks commonly support $20Gb/in^2$.
- Improvements have occurred in several different technologies including:
  - Materials science
  - Magneto-optical recording heads.
  - Error correcting codes.

7.10 The Future of Data Storage

- As data densities increase, bit cells consist of proportionately fewer magnetic grains.
- There is a point at which there are too few grains to hold a value, and a 1 might spontaneously change to a 0, or vice versa.
- This point is called the superparamagnetic limit.
  - In 2006, the superparamagnetic limit is thought to lie between $150Gb/in^2$ and $200Gb/in^2$.
- Even if this limit is wrong by a few orders of magnitude, the greatest gains in magnetic storage have probably already been realized.

7.10 The Future of Data Storage

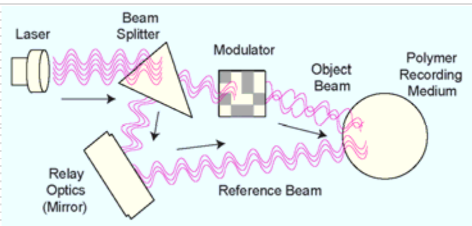
- Future exponential gains in data storage most likely will occur through the use of totally new technologies.
- Research into finding suitable replacements for magnetic disks is taking place on several fronts.
- Some of the more interesting technologies include:
  - Biological materials
  - Holographic systems and
  - Micro-electro-mechanical devices.

7.10 The Future of Data Storage

- Present day biological data storage systems combine organic compounds such as proteins or oils with inorganic (magentizable) substances.
- Early prototypes have encouraged the expectation that densities of $1Tb/in^2$ are attainable.
- Of course, the ultimate biological data storage medium is DNA.
  - Trillions of messages can be stored in a tiny strand of DNA.
- Practical DNA-based data storage is most likely decades away.

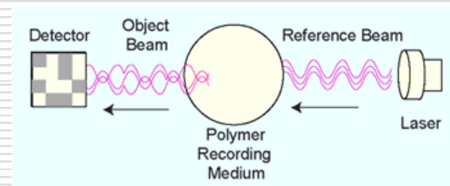
7.10 The Future of Data Storage

- Holographic storage uses a pair of laser beams to etch a three-dimensional hologram onto a polymer medium.



7.10 The Future of Data Storage

- Data is retrieved by passing the reference beam through the hologram, thereby reproducing the original coded object beam.



7.10 The Future of Data Storage

- Because holograms are three-dimensional, tremendous data densities are possible.
- Experimental systems have achieved over 30Gb/in^2, with transfer rates of around 1GBps.
- In addition, holographic storage is content addressable.
  - This means that there is no need for a file directory on the disk. Accordingly, access time is reduced.
- The major challenge is in finding an inexpensive, stable, rewriteable holographic medium.

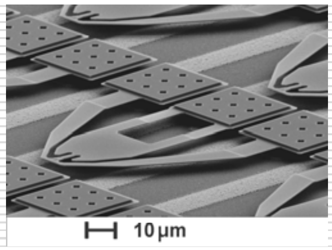
7.10 The Future of Data Storage

- Micro-electro-mechanical storage (MEMS) devices offer another promising approach to mass storage.
- IBM's Millipede is one such device.
- Prototypes have achieved densities of 100Gb/in^2 with 1Tb/in^2 expected as the technology is refined.

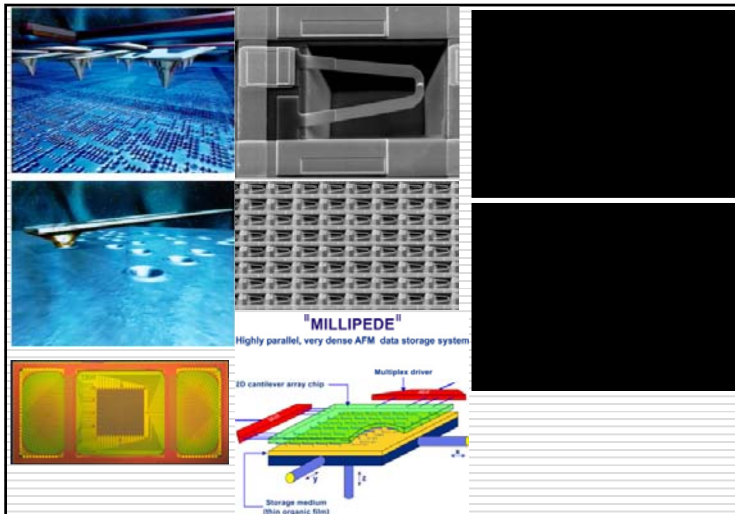
A photomicrograph of Millipede is shown on the next slide.

7.10 The Future of Data Storage

- Millipede consists of thousands of cantilevers that record a binary 1 by pressing a heated tip into a polymer substrate.
  - The tip reads a binary 1 when it dips into the imprint in the polymer



Photomicrograph courtesy of the IBM Corporation.
© 2005 IBM Corporation



Chapter 7 Conclusion

- I/O systems are critical to the overall performance of a computer system.
- Amdahl's Law quantifies this assertion.
- I/O systems consist of memory blocks, cabling, control circuitry, interfaces, and media.
- I/O control methods include programmed I/O, interrupt-based I/O, DMA, and channel I/O.
- Buses require control lines, a clock, and data lines. Timing diagrams specify operational details.

Chapter 7 Conclusion

- Magnetic disk is the principal form of durable storage.
- Disk performance metrics include seek time, rotational delay, and reliability estimates.
- Optical disks provide long-term storage for large amounts of data, although access is slow.
- Magnetic tape is also an archival medium. Recording methods are track-based, serpentine, and helical scan.

Chapter 7 Conclusion

- RAID gives disk systems improved performance and reliability. RAID 3 and RAID 5 are the most common.
- RAID 6 and RAID DP protect against dual disk failure, but RAID DP offers better performance.
- Any one of several new technologies including biological, holographic, or mechanical may someday replace magnetic disks.
- The hardest part of data storage may be end up be in locating the data after it's stored.