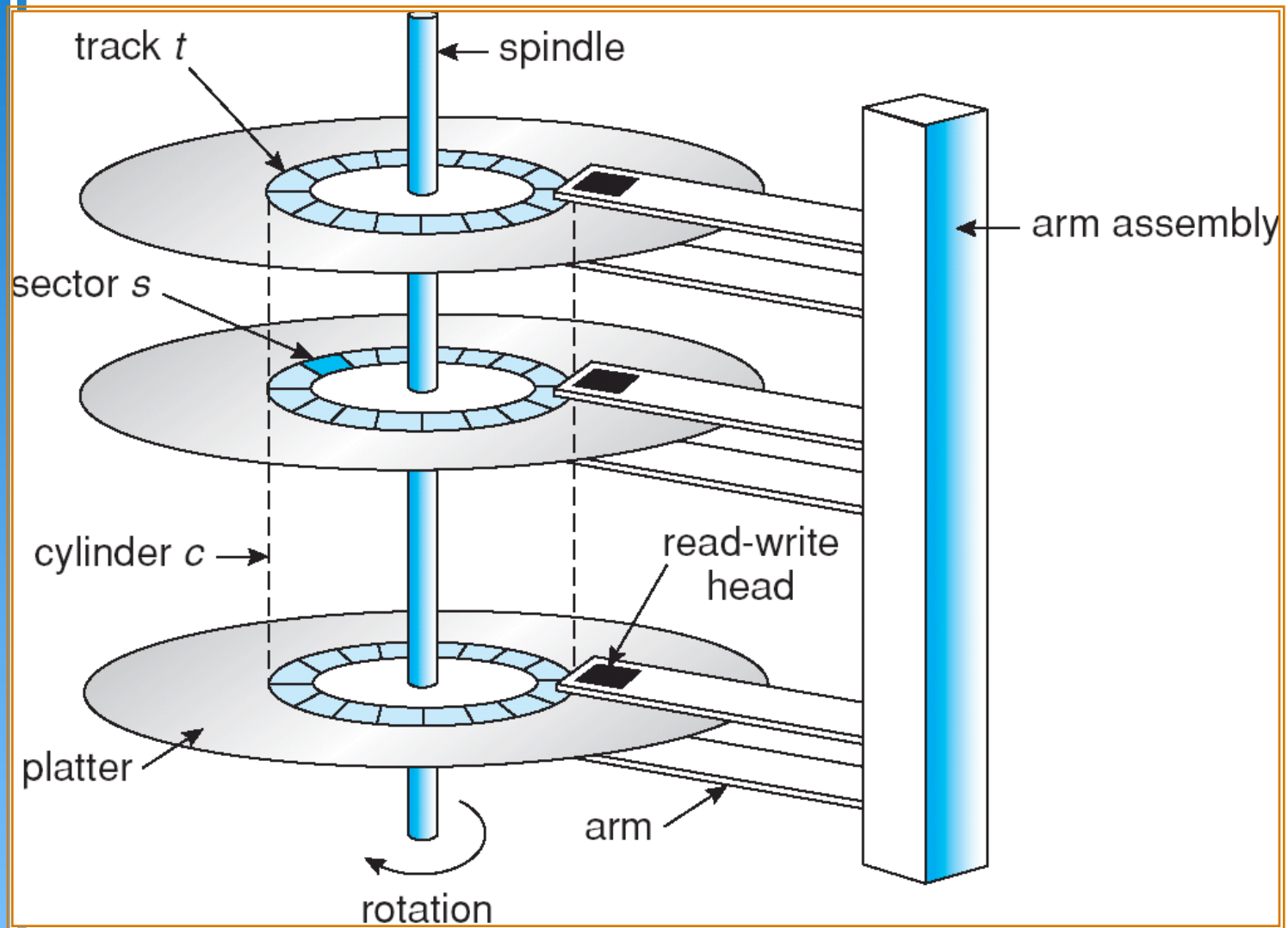


Overview of Mass Storage Structure

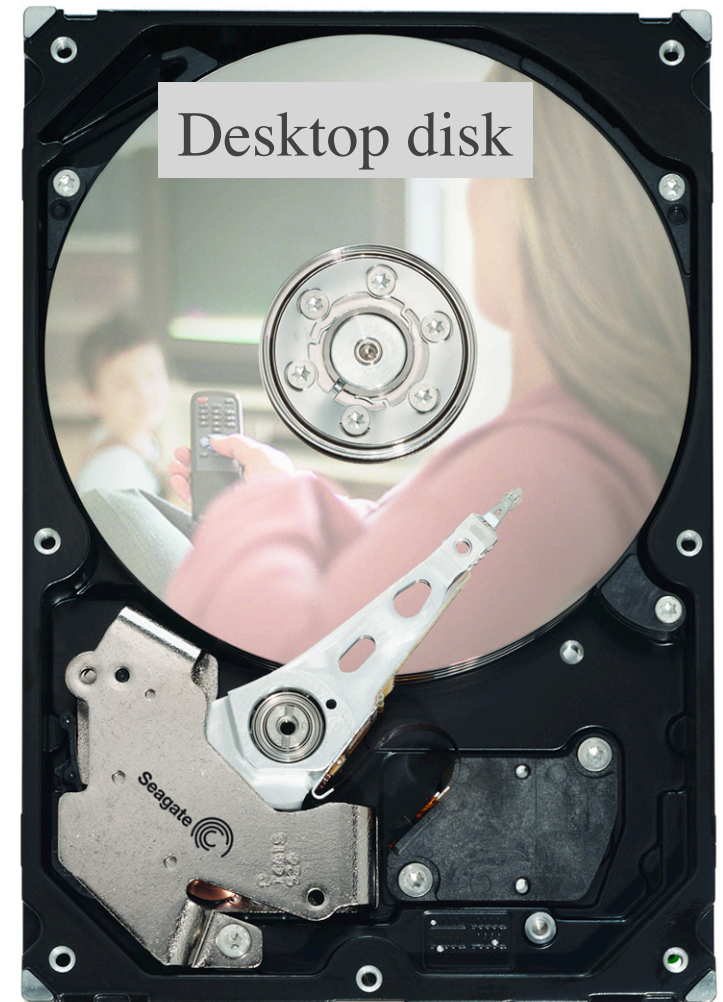
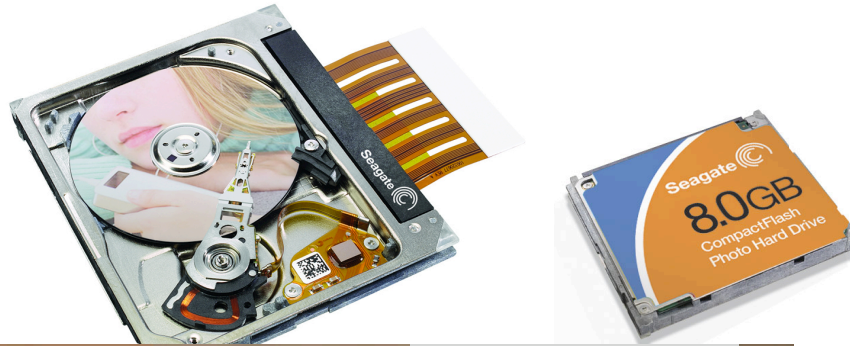
- ▶ Magnetic disks provide bulk of secondary storage
 - Drives rotate at 70 to 250 times per second
 - Ipod disks: 4200 rpm
 - Laptop disks: 4200, 5400 rpm or 7200 rpm
 - Desktop disks: 7200 rpm
 - Server disks: 10000 rpm or 15000 rpm
 - **Transfer rate** is rate at which data flow between drive and computer
 - **Positioning time (random-access time)** is time to move disk arm to desired cylinder (**seek time**) and time for desired sector to rotate under the disk head (**rotational latency**)
 - **Head crash** results from disk head contacting disk surface
 - That's bad
- ▶ Disks can be removable
- ▶ Drive attached to computer via **I/O bus**
 - Busses vary, including **EIDE, ATA, SATA, Firewire, USB, Fibre Channel, SCSI**
 - **Host controller** in computer uses bus to talk to **disk controller** built into drive or storage array



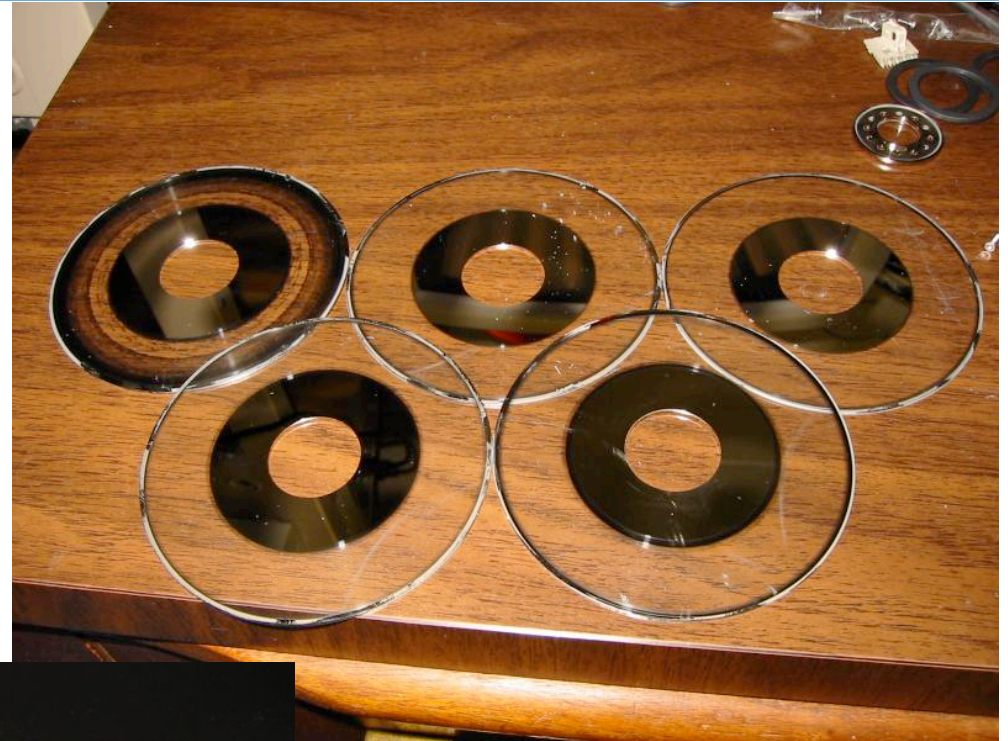
Moving-head Disk Mechanism



Disk drives



Hard disk head, platter and disk crash



Disk Structure

- ▶ Disk drives are addressed as large 1-dimensional arrays of *logical blocks*, where the logical block is the smallest unit of transfer.
- ▶ The 1-dimensional array of logical blocks is mapped into the sectors of the disk sequentially.
 - Sector 0 is the first sector of the first track on the outermost cylinder.
 - Mapping proceeds in order through that track, then the rest of the tracks in that cylinder, and then through the rest of the cylinders from outermost to innermost.



Magnetic tape

- ▶ Was early secondary-storage medium
- ▶ Relatively permanent and holds large quantities of data
- ▶ Access time slow
- ▶ Random access ~1000 times slower than disk
- ▶ Mainly used for backup, storage of infrequently-used data, transfer medium between systems
- ▶ Kept in spool and wound or rewound past read-write head
- ▶ Once data under head, transfer rates comparable to disk
- ▶ 20-200GB typical storage
- ▶ Common technologies are 4mm, 8mm, 19mm, LTO-2 and SDLT



Tape pictures



Tape Drives

- ▶ The basic operations for a tape drive differ from those of a disk drive.
- ▶ **locate** positions the tape to a specific logical block, not an entire track (corresponds to **seek**).
- ▶ The **read position** operation returns the logical block number where the tape head is.
- ▶ The **space** operation enables relative motion.
- ▶ Tape drives are “append-only” devices; updating a block in the middle of the tape also effectively erases everything beyond that block.
- ▶ An EOT mark is placed after a block that is written.



Application Interface

- ▶ Most OSs handle removable disks almost exactly like fixed disks — a new cartridge is formatted and an empty file system is generated on the disk.
- ▶ Tapes are presented as a raw storage medium, i.e., and application does not open a file on the tape, it opens the whole tape drive as a raw device.
- ▶ Usually the tape drive is reserved for the exclusive use of that application.
- ▶ Since the OS does not provide file system services, the application must decide how to use the array of blocks.
- ▶ Since every application makes up its own rules for how to organize a tape, a tape full of data can generally only be used by the program that created it.



Tertiary Storage Devices

- ▶ Low cost is the defining characteristic of tertiary storage.
- ▶ Generally, tertiary storage is built using *removable media*
- ▶ Common examples of removable media are floppy disks and CD-ROMs; other types are available.



Removable Disks

- ▶ Floppy disk — thin flexible disk coated with magnetic material, enclosed in a protective plastic case.
 - Most floppies hold about 1 MB; similar technology is used for removable disks that hold more than 1 GB.
 - Removable magnetic disks can be nearly as fast as hard disks, but they are at a greater risk of damage from exposure.



Removable Disks (Cont.)

- ▶ A magneto-optic disk records data on a rigid platter coated with magnetic material.
 - Laser heat is used to amplify a large, weak magnetic field to record a bit.
 - Laser light is also used to read data (Kerr effect).
 - The magneto-optic head flies much farther from the disk surface than a magnetic disk head, and the magnetic material is covered with a protective layer of plastic or glass; resistant to head crashes.
- ▶ Optical disks do not use magnetism; they employ special materials that are altered by laser light.



WORM Disks

- ▶ The data on read-write disks can be modified over and over.
- ▶ WORM (“Write Once, Read Many Times”) disks can be written only once.
- ▶ Thin aluminum film sandwiched between two glass or plastic platters.
- ▶ To write a bit, the drive uses a laser light to burn a small hole through the aluminum; information can be destroyed but not altered.
- ▶ Very durable and reliable.
- ▶ *Read Only* disks, such as CD-ROM and DVD, come from the factory with the data pre-recorded.



Speed

- ▶ Two aspects of speed in tertiary storage are bandwidth and latency.
- ▶ Bandwidth is measured in bytes per second.
 - Sustained bandwidth – average data rate during a large transfer; # of bytes/transfer time.
Data rate when the data stream is actually flowing.
 - Effective bandwidth – average over the entire I/O time, including **seek** or **locate**, and cartridge switching.
Drive's overall data rate.



Speed (Cont.)

- ▶ Access latency – amount of time needed to locate data.
 - Access time for a disk – move the arm to the selected cylinder and wait for the rotational latency; < 35 ms
 - Access on tape requires winding the tape reels until the selected block reaches the tape head; 10s or 100s of secs.
 - random access within a tape cartridge is about a thousand times slower than random access on disk.
- ▶ Low cost of tertiary storage is a result of having many cheap cartridges share a few expensive drives
- ▶ A removable library is best devoted to the storage of infrequently used data, because the library can only satisfy a relatively small number of I/O requests per hour



Reliability

- ▶ A fixed disk drive is likely to be more reliable than a removable disk or tape drive.
- ▶ An optical cartridge is likely to be more reliable than a magnetic disk or tape.
- ▶ A head crash in a fixed hard disk generally destroys the data, whereas the failure of a tape drive or optical disk drive often leaves the data cartridge unharmed.

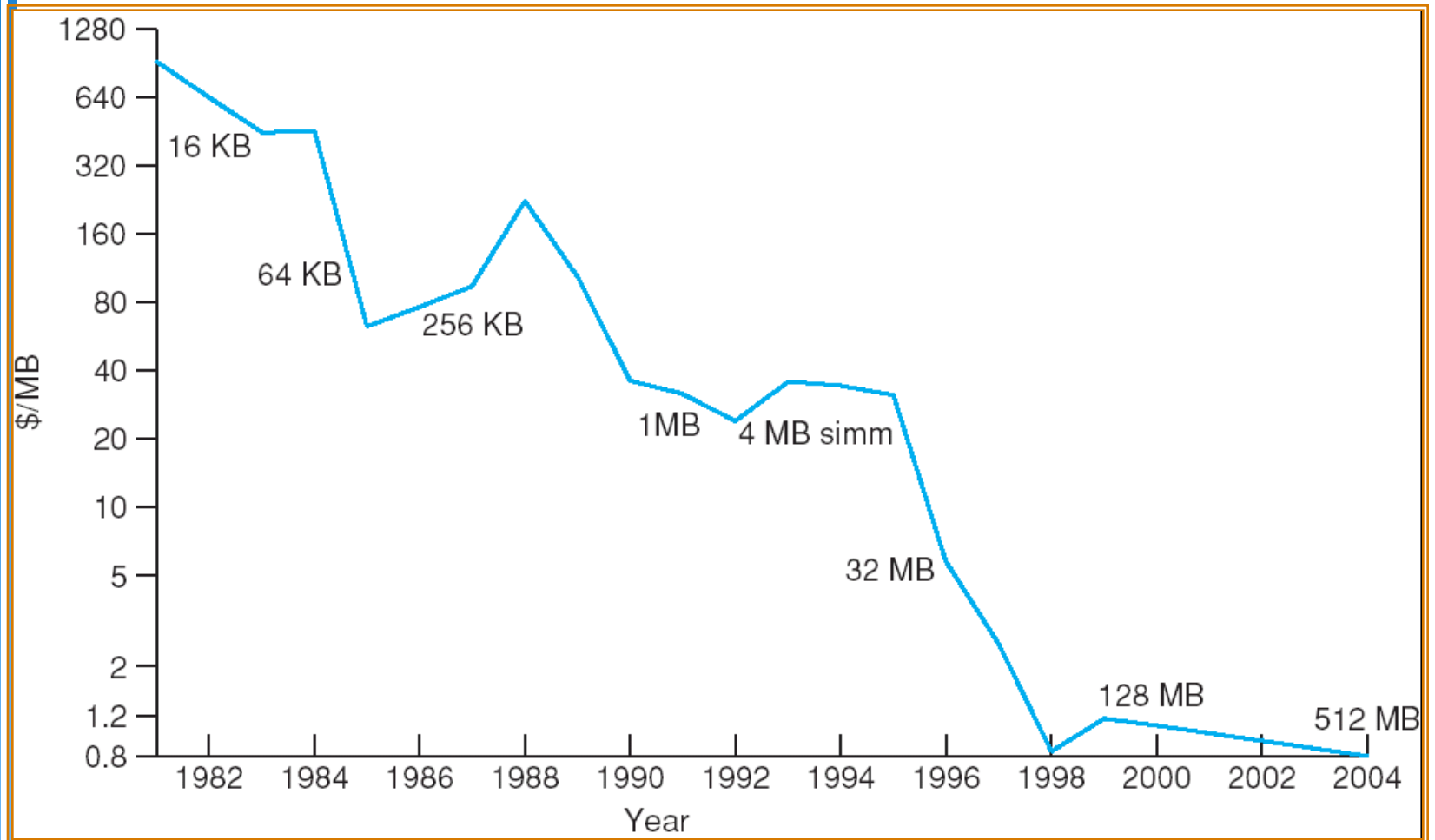


Cost

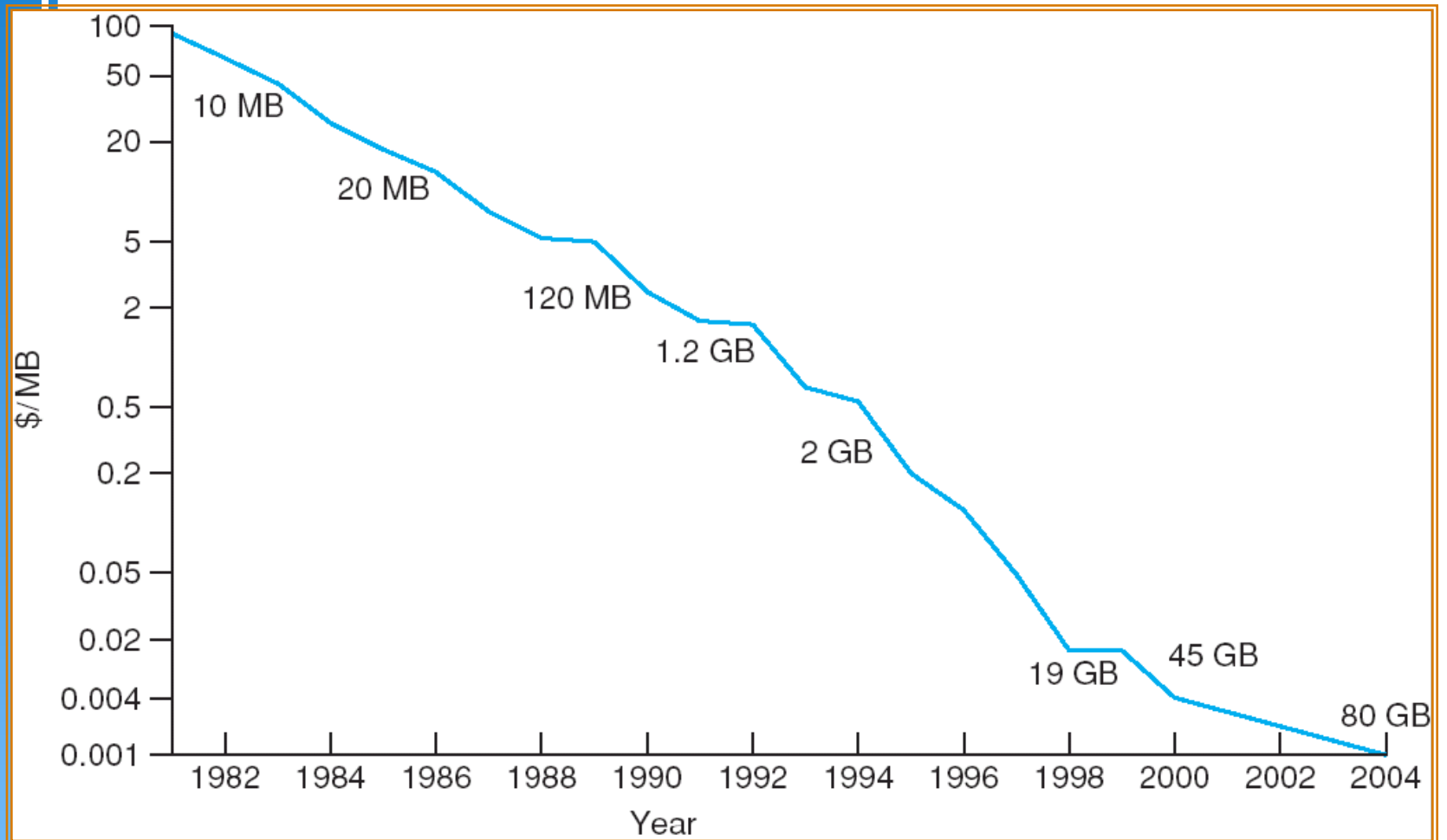
- ▶ Main memory is much more expensive than disk storage
- ▶ The cost per megabyte of hard disk storage is competitive with magnetic tape if only one tape is used per drive.
- ▶ The cheapest tape drives and the cheapest disk drives have had about the same storage capacity over the years.
- ▶ Tertiary storage gives a cost savings only when the number of cartridges is considerably larger than the number of drives.



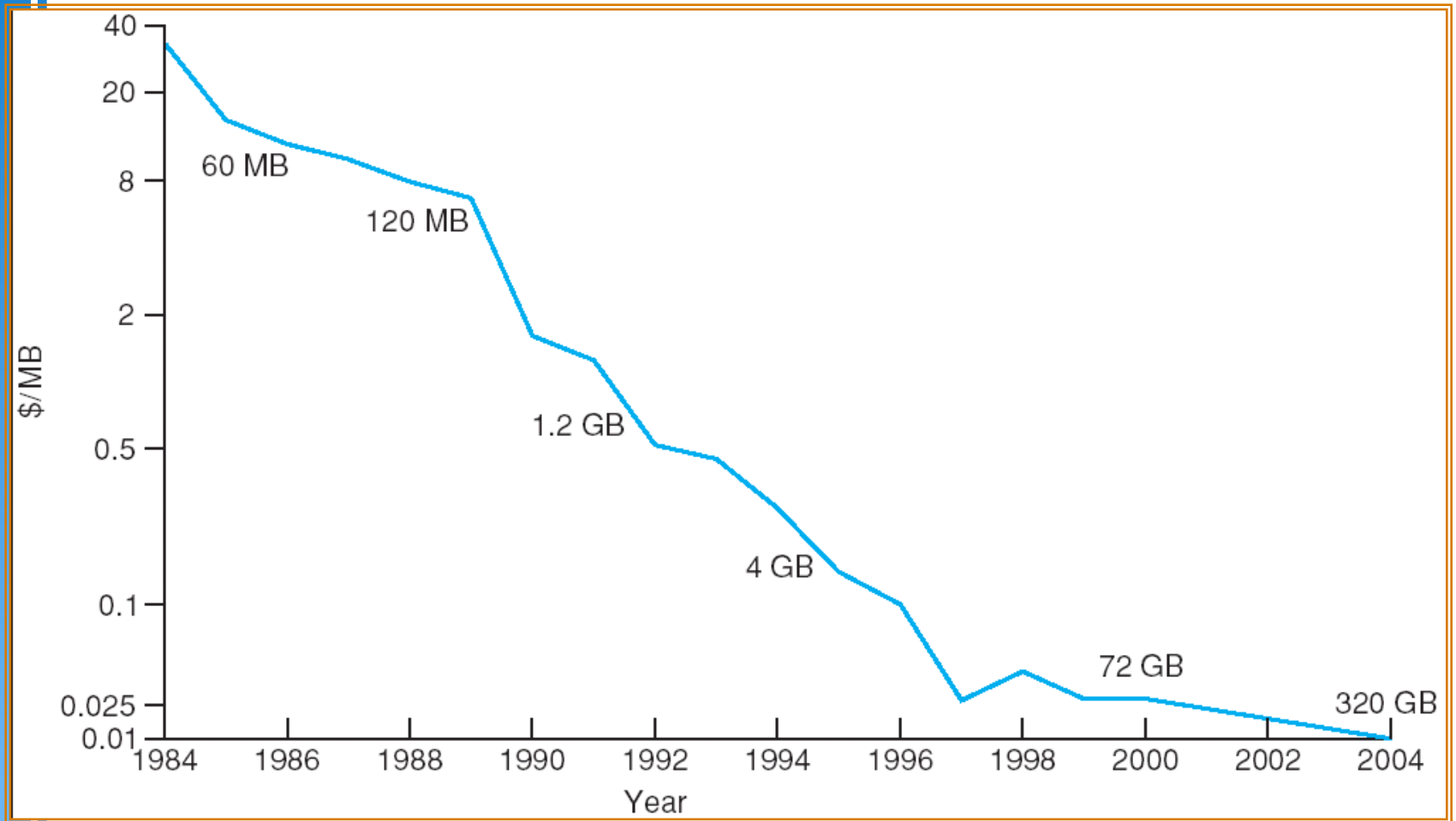
Price per Megabyte of DRAM, From 1981 to 2004



Price per Megabyte of Magnetic Hard Disk, From 1981 to 2004

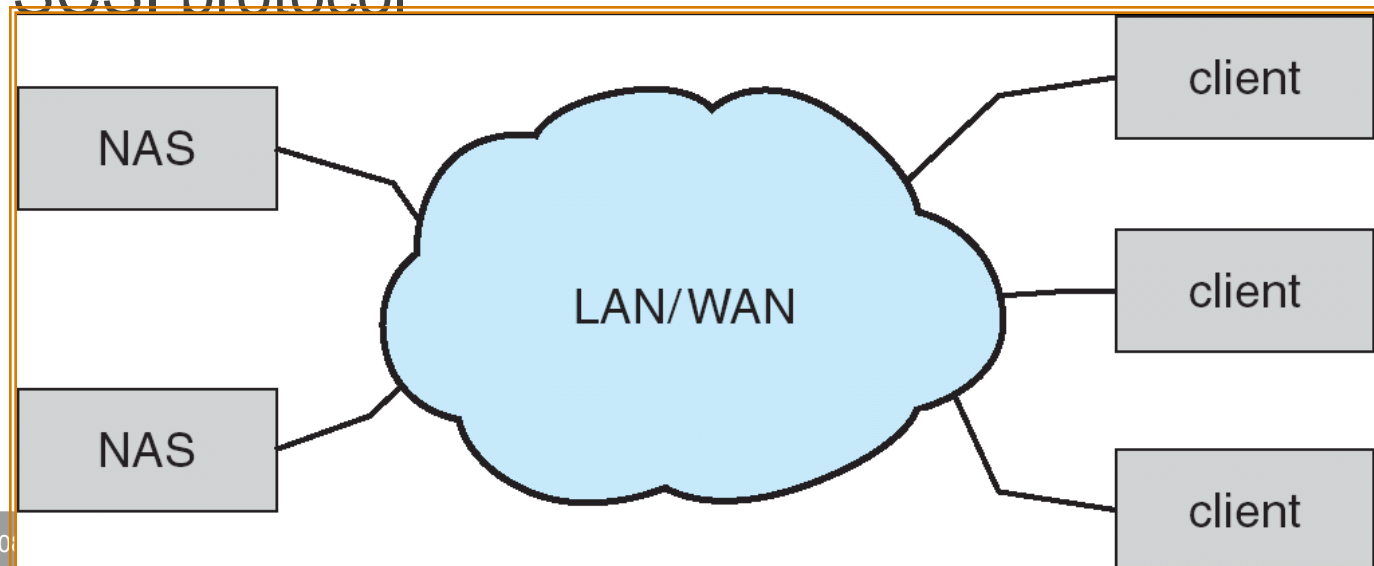


Price per Megabyte of a Tape Drive, From 1984-2000



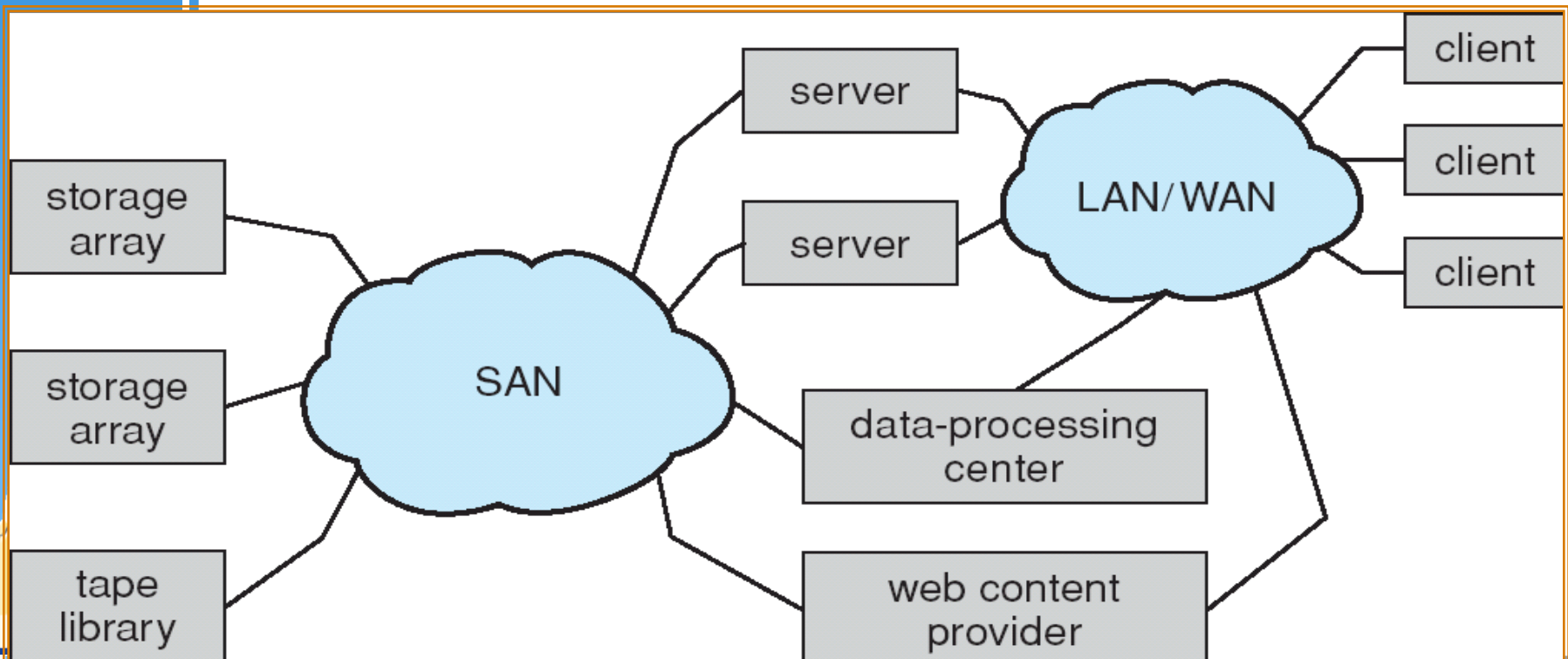
Network-Attached Storage

- ▶ Network-attached storage (**NAS**) is storage made available over a network rather than over a local connection (such as a bus)
- ▶ NFS and CIFS are common protocols
- ▶ Implemented via remote procedure calls (RPCs) between host and storage
- ▶ New iSCSI protocol uses IP network to carry the SCSI protocol



Storage Area Network

- ▶ Common in large storage environments (and becoming more common)
- ▶ Multiple hosts attached to multiple storage arrays - flexible



Hierarchical Storage Management (HSM)

- ▶ A hierarchical storage system extends the storage hierarchy beyond primary memory and secondary storage to incorporate tertiary storage — usually implemented as a jukebox of tapes or removable disks.
- ▶ Usually incorporate tertiary storage by extending the file system.
 - Small and frequently used files remain on disk.
 - Large, old, inactive files are archived to the jukebox.
- ▶ HSM is usually found in supercomputing centers and other large installations that have enormous volumes of data.

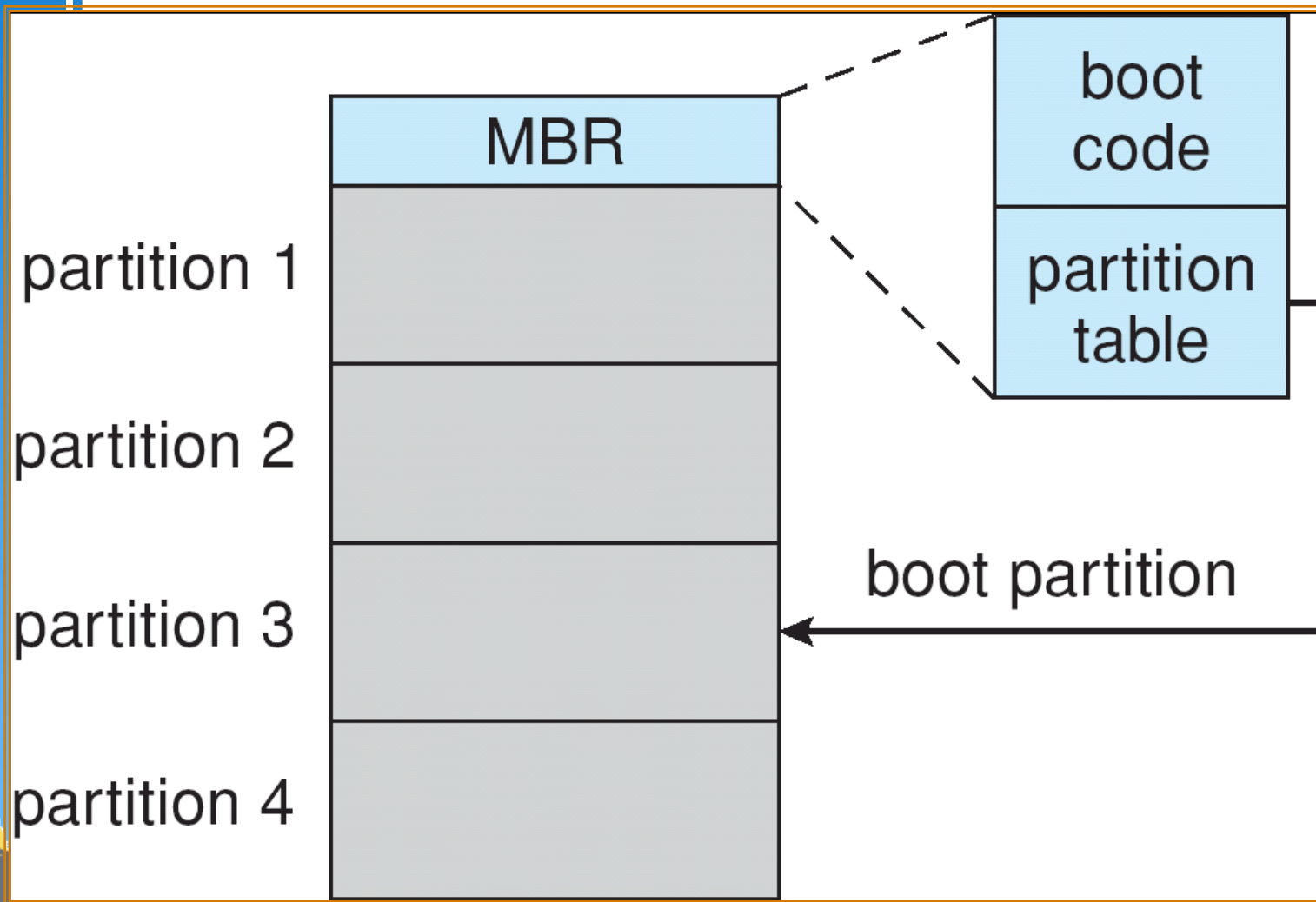


Disk Management

- ▶ *Low-level formatting, or physical formatting* — Dividing a disk into sectors that the disk controller can read and write.
- ▶ To use a disk to hold files, the operating system still needs to record its own data structures on the disk.
 - *Partition* the disk into one or more groups of cylinders.
 - *Logical formatting* or “making a file system”.
- ▶ Boot block initializes system.
 - The bootstrap is stored in ROM.
 - *Bootstrap loader* program.
- ▶ Methods such as *sector sparing* used to handle bad blocks.



Booting from a Disk in Windows 2000



Swap-Space Management

- ▶ Swap-space — Virtual memory uses disk space as an extension of main memory.
- ▶ Swap-space can be carved out of the normal file system, or, more commonly, it can be in a separate disk partition.
- ▶ Swap-space management
 - 4.3BSD allocates swap space when process starts; holds *text segment* (the program) and *data segment*.
 - Kernel uses *swap maps* to track swap-space use.
 - Solaris 2 allocates swap space only when a page is forced out of physical memory, not when the virtual memory page is first created.

