Chapter 12: Mass-Storage Systems

Chapter 12: Mass-Storage Systems

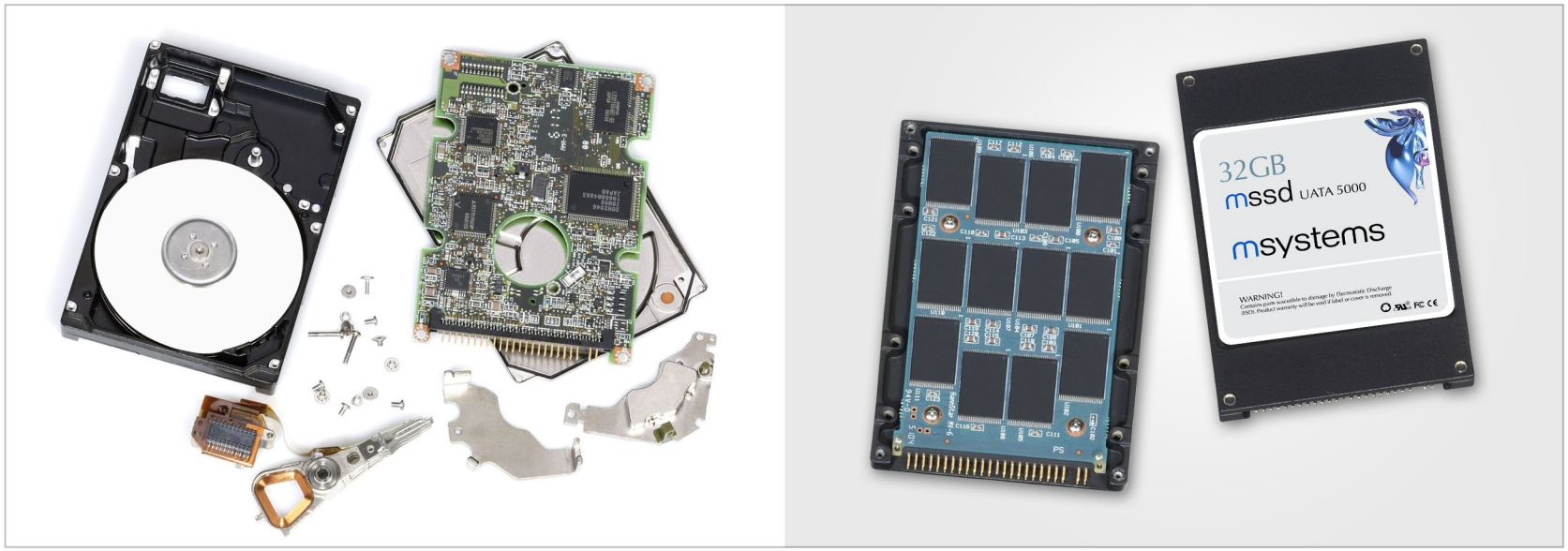
- Overview of Mass Storage Structure
- Disk Structure
- Disk Attachment
- Disk Scheduling
- Disk Management
- RAID Structure
- Operating System Issues

Objectives

- Describe the physical structure of secondary storage devices and the resulting effects on the uses of the devices
- Explain the performance characteristics of mass-storage devices
- Discuss operating-system services provided for mass storage

Overview of Mass Storage Structure

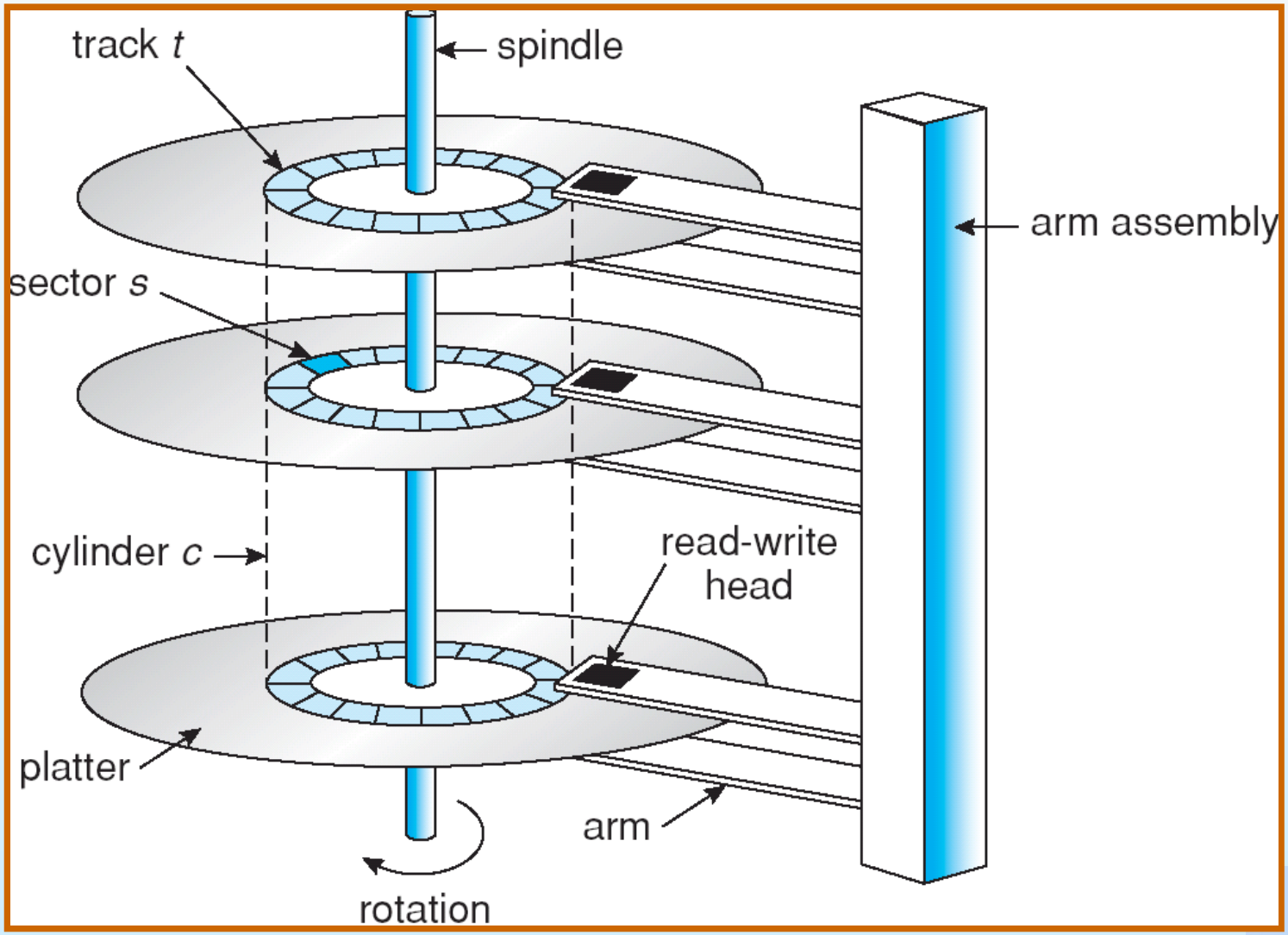
- Magnetic disks are mainstream of secondary storage
- Solid-state disks (SSD) will replace magnetic disk as the secondary storage



Overview (Cont.)

- Removable
- Drive attached to computer via I/O bus
  - Buses vary, including EIDE, ATA, SATA, USB, Fiber Channel, SCSI, IEEE 1394
  - Host controller in computer uses bus to talk to disk controller built into drive

Moving-head Disk Machanism



Disk Structure

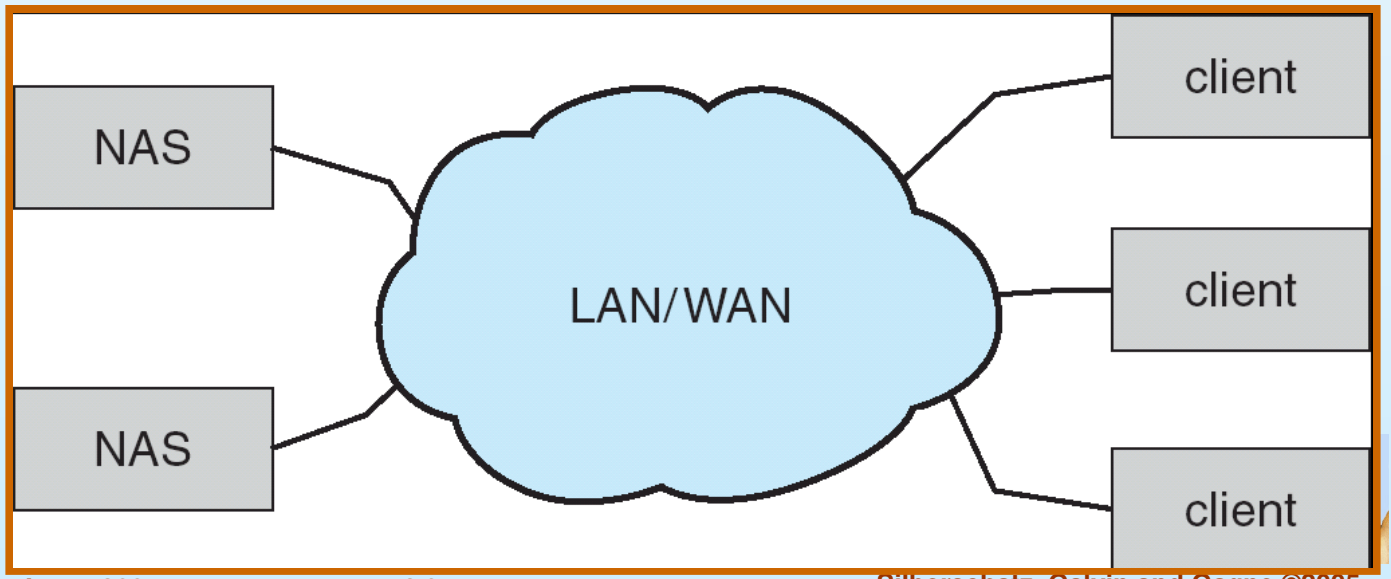
- Disk drives are addressed as large 1-dimensional arrays of *logical blocks*
  - The logical block is the smallest unit of transfer.
  - Usually 512 bytes or 1024 bytes
- The 1-dimensional array of logical blocks is mapped into the sectors of the disk sequentially.
  - Each sector is 512 bytes

Disk Attachment

- Host-attached storage accessed through I/O ports talking to I/O busses
- SCSI itself is a bus, up to 16 devices on one cable, SCSI initiator requests operation and SCSI targets perform tasks
  - Each target can have up to 8 logical units (disks attached to device controller
- FC is high-speed serial architecture
  - Can be switched fabric with 24-bit address space – the basis of storage area networks (SANs) in which many hosts attach to many storage units
  - Can be arbitrated loop (FC-AL) of 126 devices

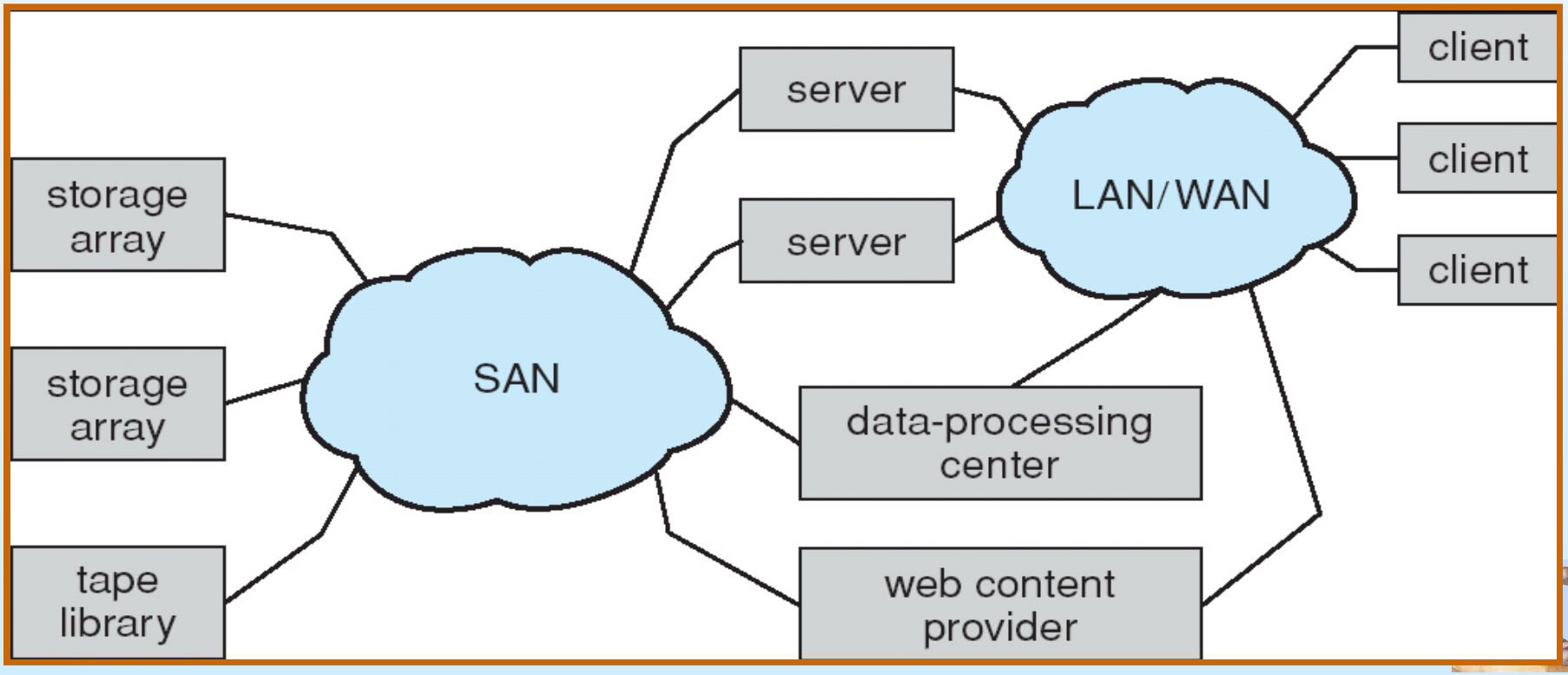
Network-Attached Storage

- Network-attached storage (NAS) is storage made available over a network
  - 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



Disk Scheduling

- The OS is responsible for using disks efficiently
  - Having a fast access time and disk bandwidth
- Access time has two major components
  - *Seek time* is the time to move the heads to the cylinder containing the desired sector.
  - *Rotational latency* is the additional time waiting for the desired sector rotated to the head.
- Minimize seek time
  - Seek time ≈ seek distance

Disk Scheduling (Cont.)

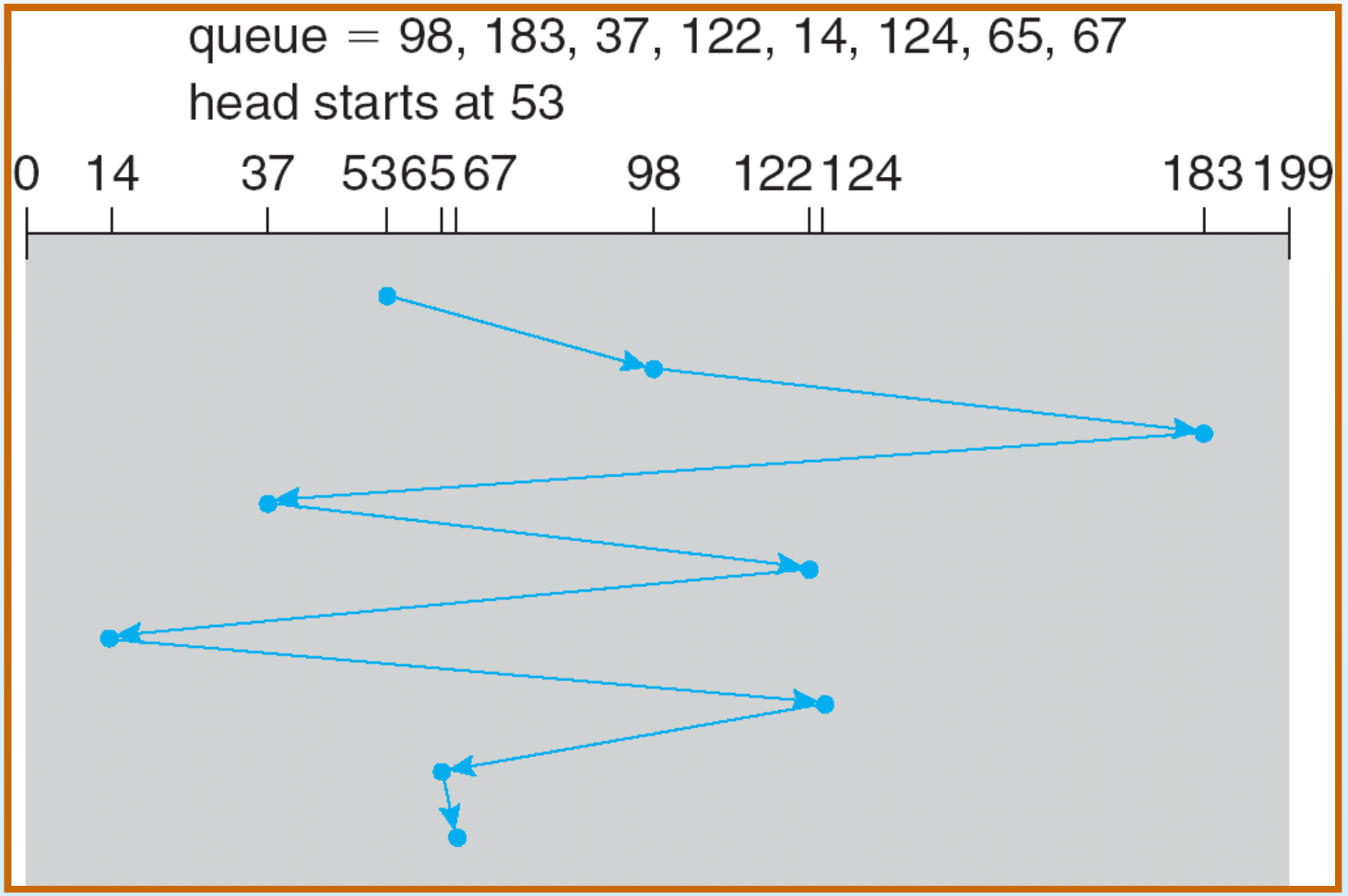
- Several algorithms exist to schedule the servicing of disk I/O requests.
- We illustrate them with a request queue (0-199).

98, 183, 37, 122, 14, 124, 65, 67

Head pointer 53

FCFS

Illustration shows total head movement of 640 cylinders.

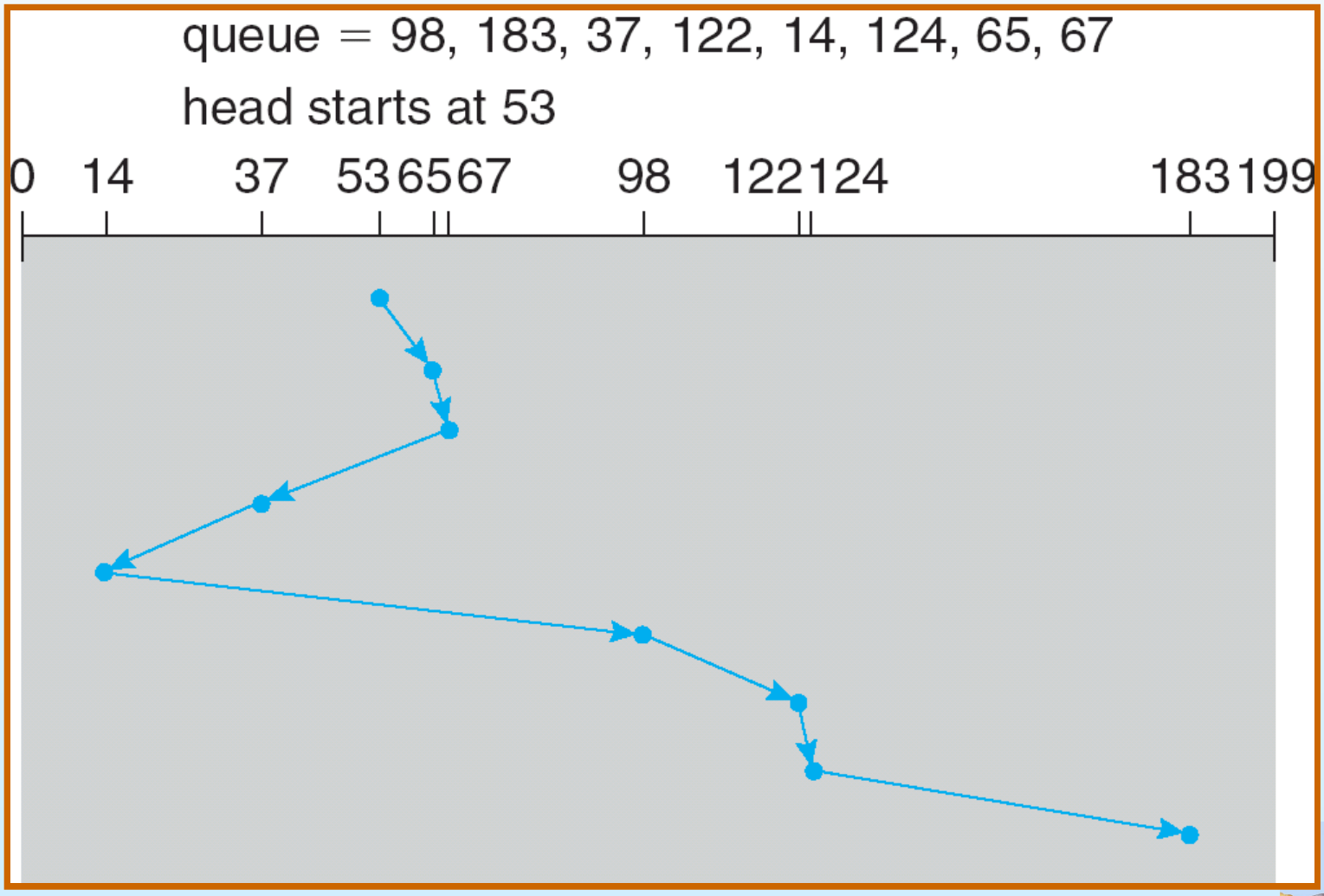


SSTF (Shortest Seek Time First)

- Selects the request with the minimum seek time from the current head position.
- SSTF scheduling is a form of SJF scheduling
  - Starvation

SSTF (Cont.)

Illustration shows total head movement of 236 cylinders.

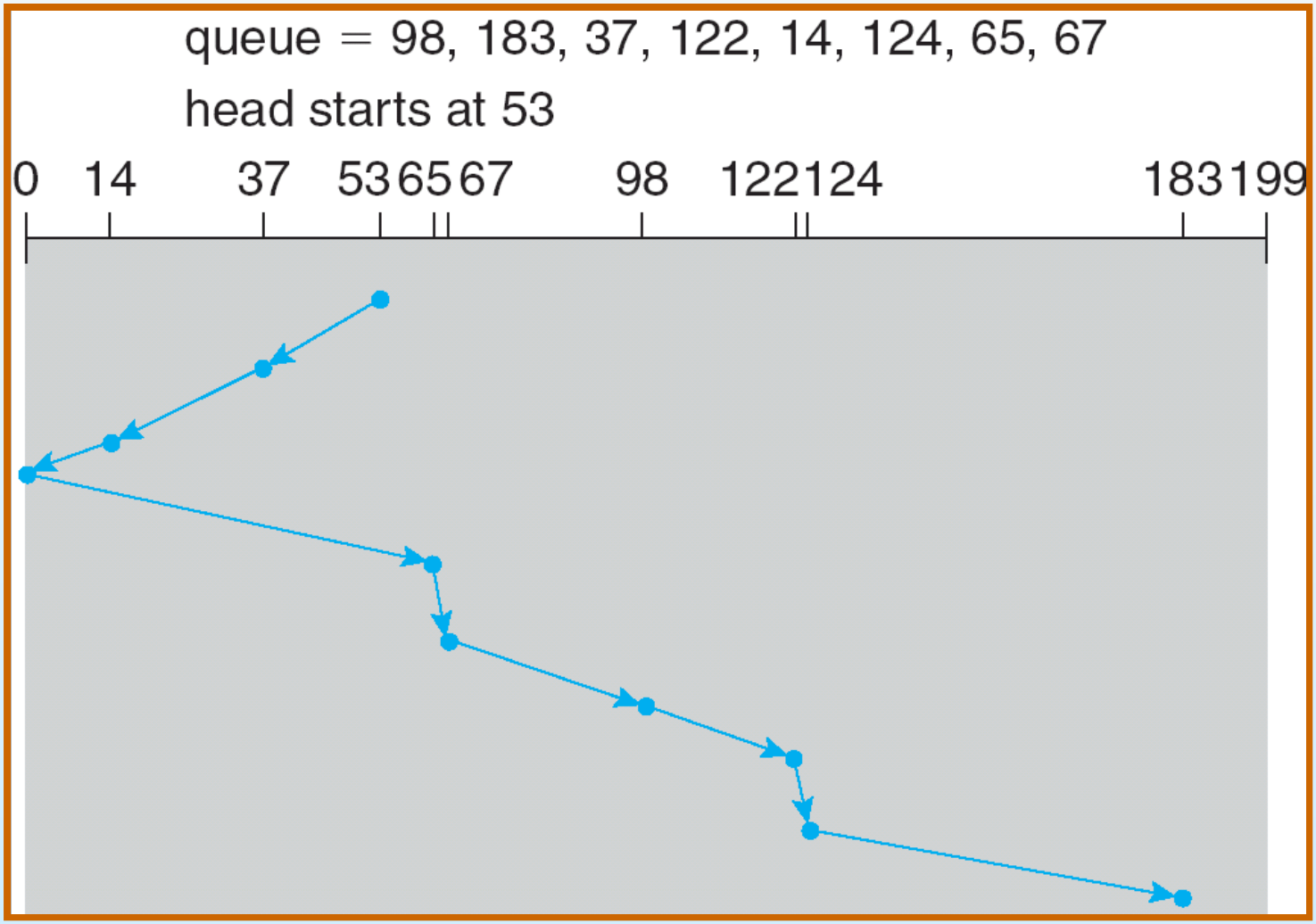


SCAN

- Algorithm
  - The disk arm starts at one end of the disk
  - Moves toward the other end
  - Servicing requests until it gets to the other end
  - The head movement is reversed and servicing continues.
- Sometimes called the *elevator algorithm*.

SCAN (Cont.)

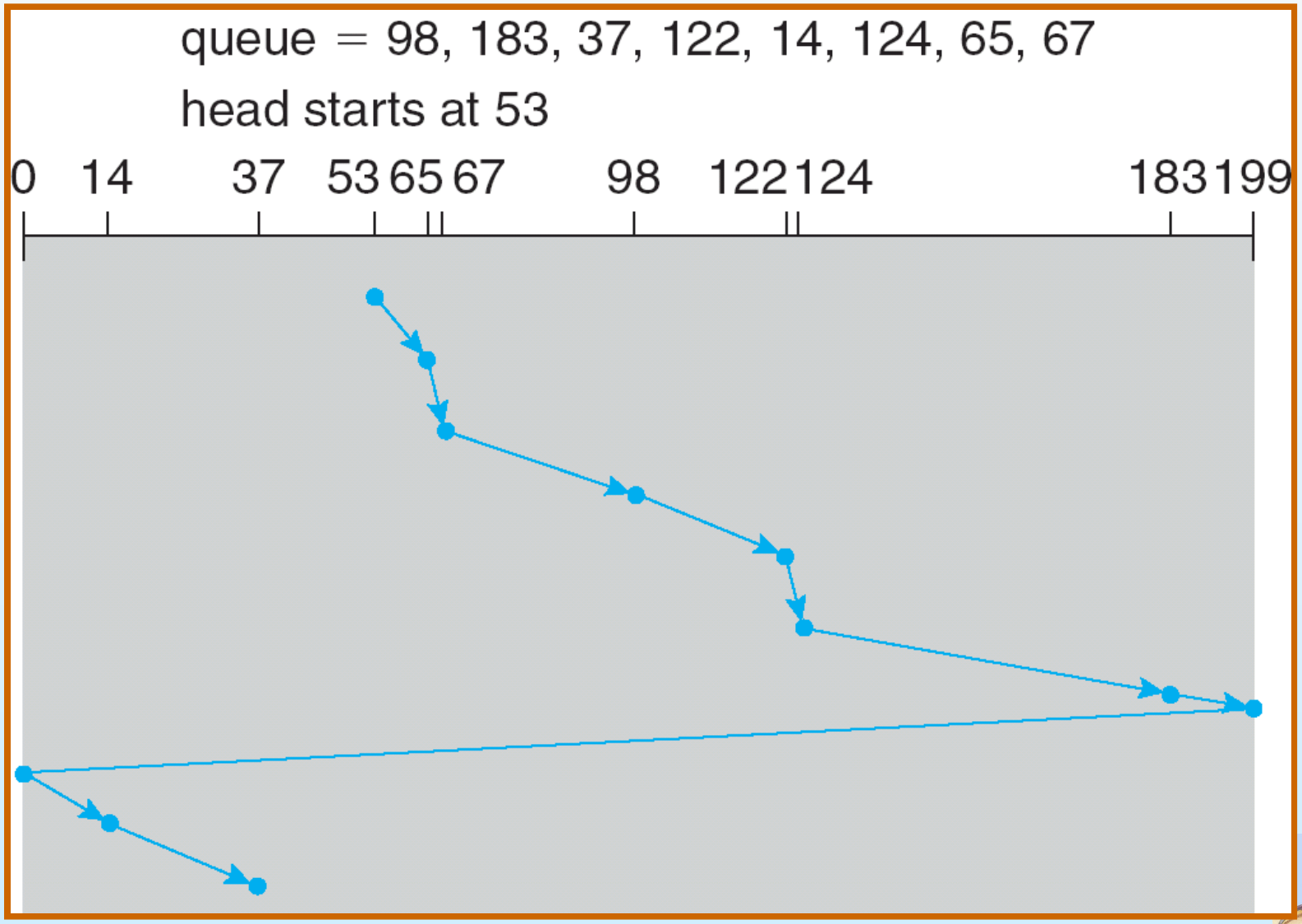
Illustration shows total head movement of 208 cylinders.



C-SCAN (Circular SCAN)

- Provides a more uniform wait time than SCAN.
  - The head moves from one end to the other
  - Servicing requests as it goes
  - When it reaches the other end, however, it immediately returns to the beginning of the disk, without servicing any requests on the return trip.

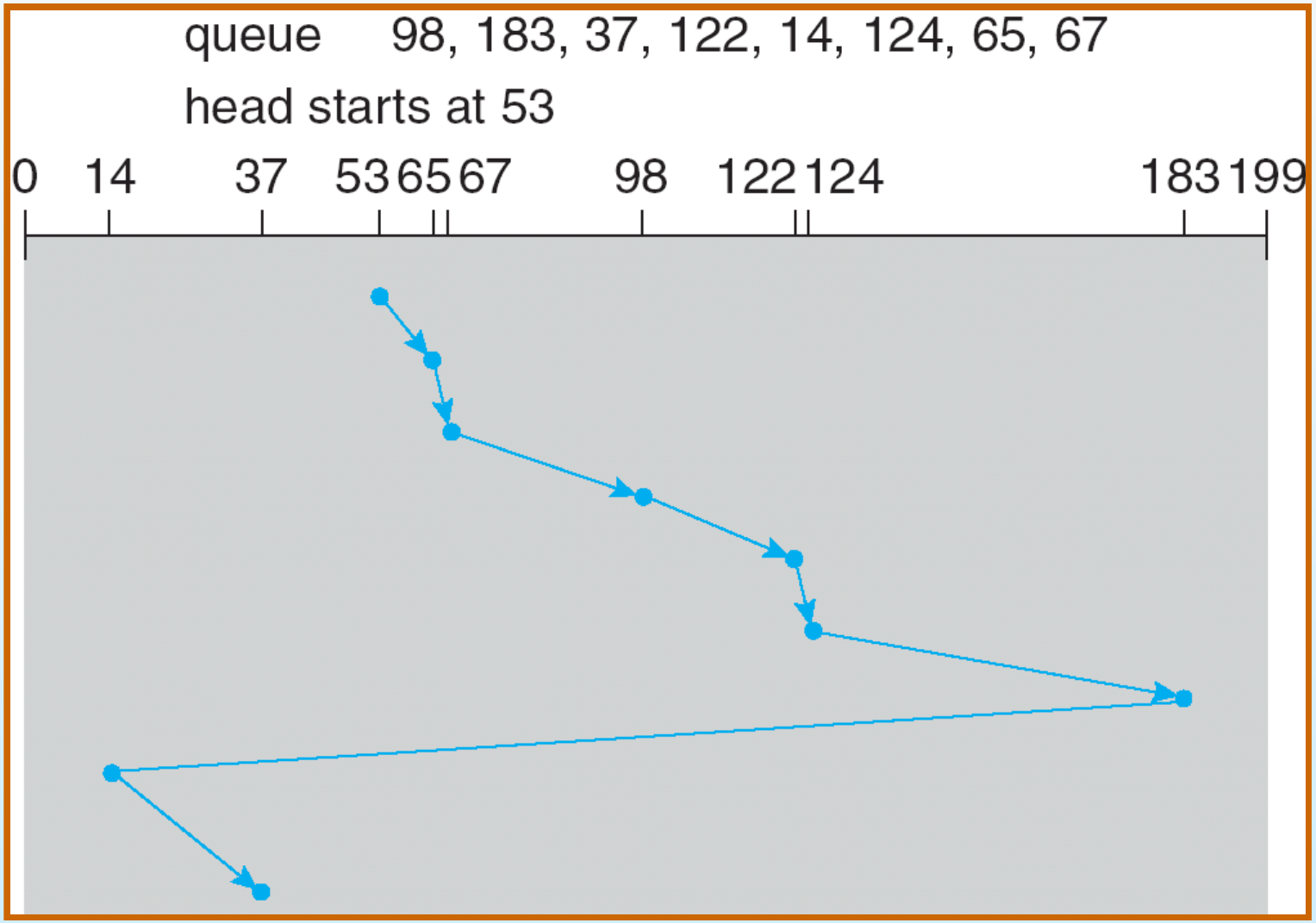
C-SCAN (Cont.)


queue = 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53
0
14
37
53
65
67
98
122
124
183
199

C-LOOK

- Version of C-SCAN
  - Arm only goes as far as the last request in each direction
  - Then reverses direction immediately
  - Without first going all the way to the end of the disk

C-LOOK (Cont.)


queue 98, 183, 37, 122, 14, 124, 65, 67
head starts at 53
0
14
37
53
65
67
98
122
124
183
199

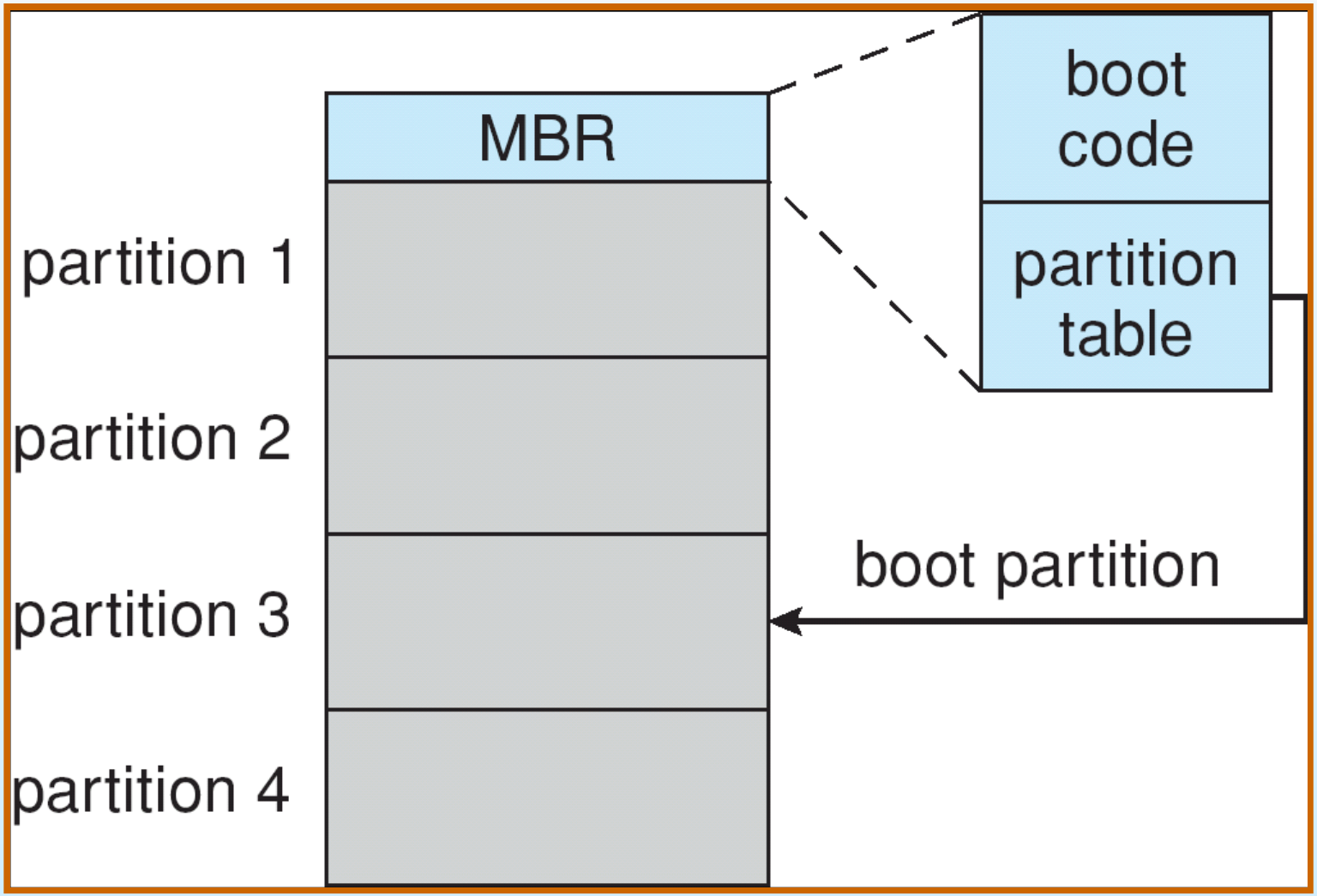
Selecting a Disk-Scheduling Algorithm

- Performance depends on the number and types of requests.
  - SSTF is common and has a natural appeal
  - SCAN and C-SCAN perform better for heavy load
- Requests for disk service can be influenced by the file-allocation method.
- The disk-scheduling algorithm should be a separate module
  - Allowing it to be replaced with a different algorithm
- Either SSTF or LOOK is a reasonable choice for the default algorithm.

Disk Management

- *Low-level formatting*, or *physical formatting*
  - Dividing a disk into sectors that the disk controller can read and write.
- Construct file system on the disk.
  - *Partition* the disk into one or more groups of cylinders.
  - *Logical formatting* or "making a file system".
- Bad blocks
  - Marked by file system
  - Sector sparing and sector slipping

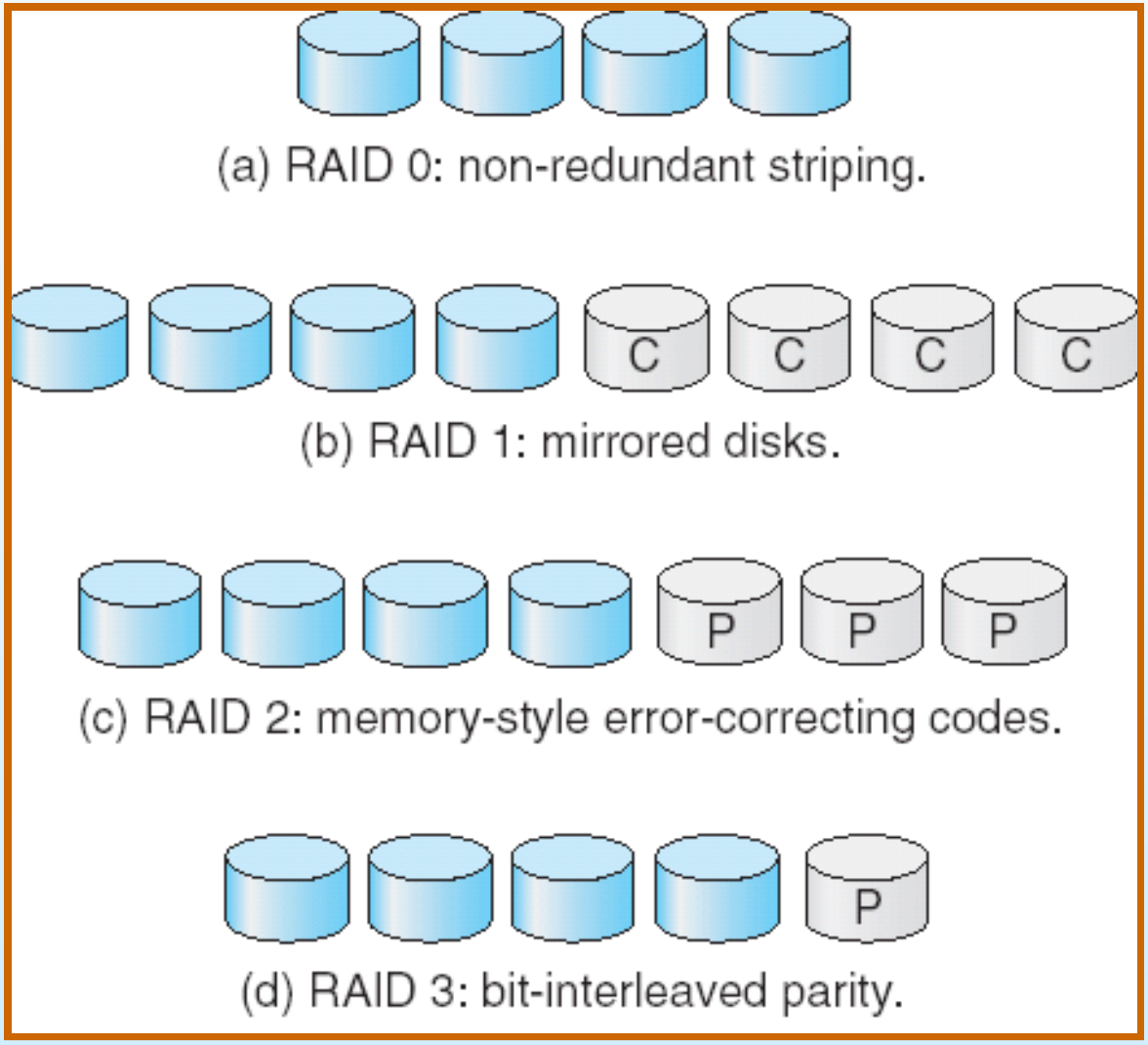
Booting from a Disk in Windows 2000



RAID Structure

- RAID
  - Redundancy Arrays of Inexpensive Disk
- Uses a group of disks as one storage unit
- Improve performance and reliability by storing redundant data.
  - *Mirroring* or *shadowing* keeps duplicate of each disk
  - *Block interleaved parity* uses much less redundancy
- RAID is arranged into six different levels

RAID Levels



(a) RAID 0: non-redundant striping.

(b) RAID 1: mirrored disks.

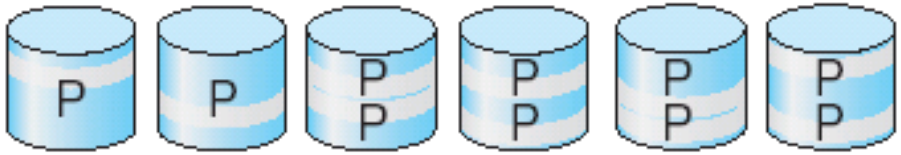
(c) RAID 2: memory-style error-correcting codes.

(d) RAID 3: bit-interleaved parity.

RAID Levels

(e) RAID 4: block-interleaved parity.

(f) RAID 5: block-interleaved distributed parity.



(g) RAID 6: P + Q redundancy.

Operating System Issues

- Major OS jobs are to manage physical devices and to present a virtual machine abstraction to applications

- For hard disks, the OS provides two abstraction:
  - Raw device – an array of data blocks.
  - File system – the OS queues and schedules the interleaved requests from several applications.

End of Chapter 12