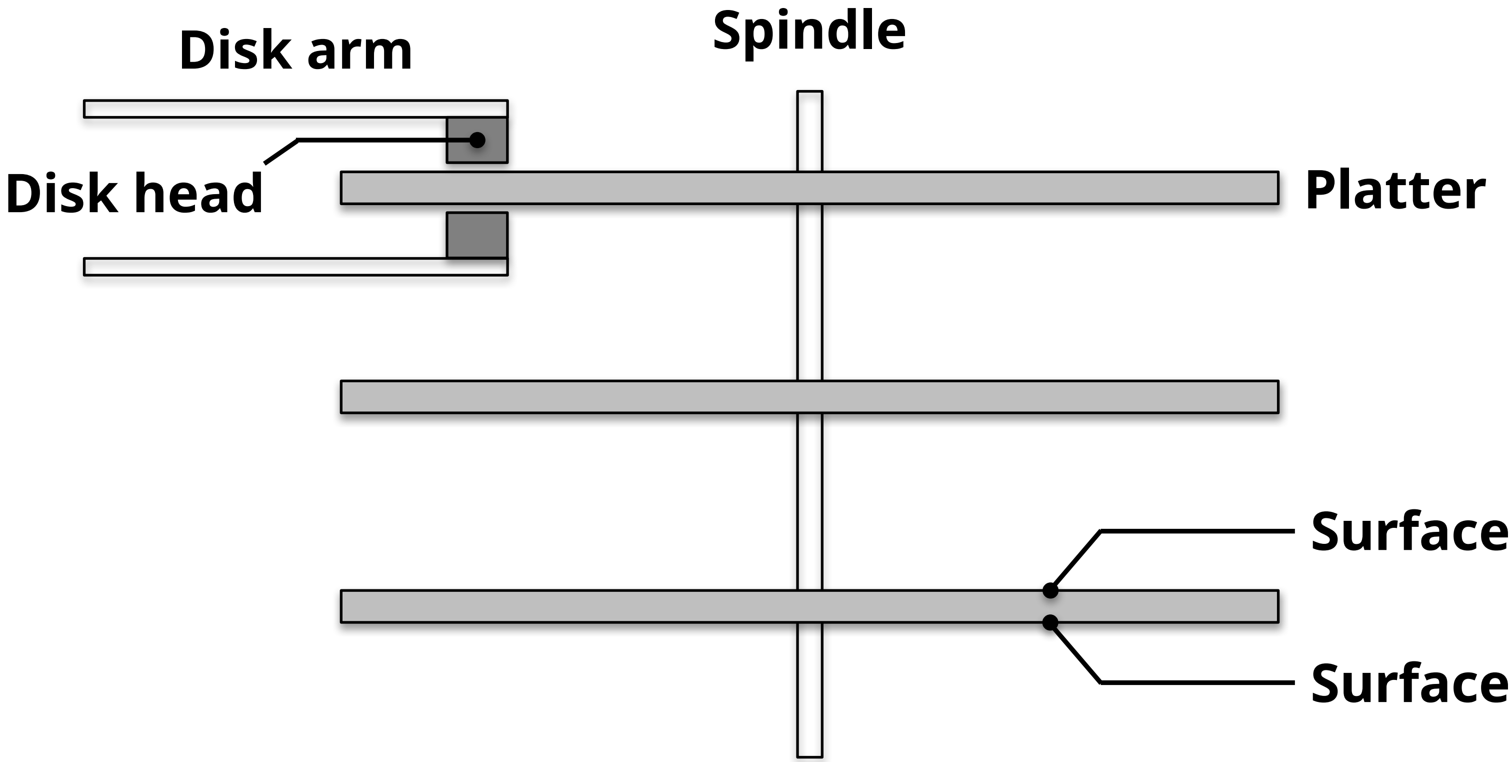
Operating Systems

Youjip Won

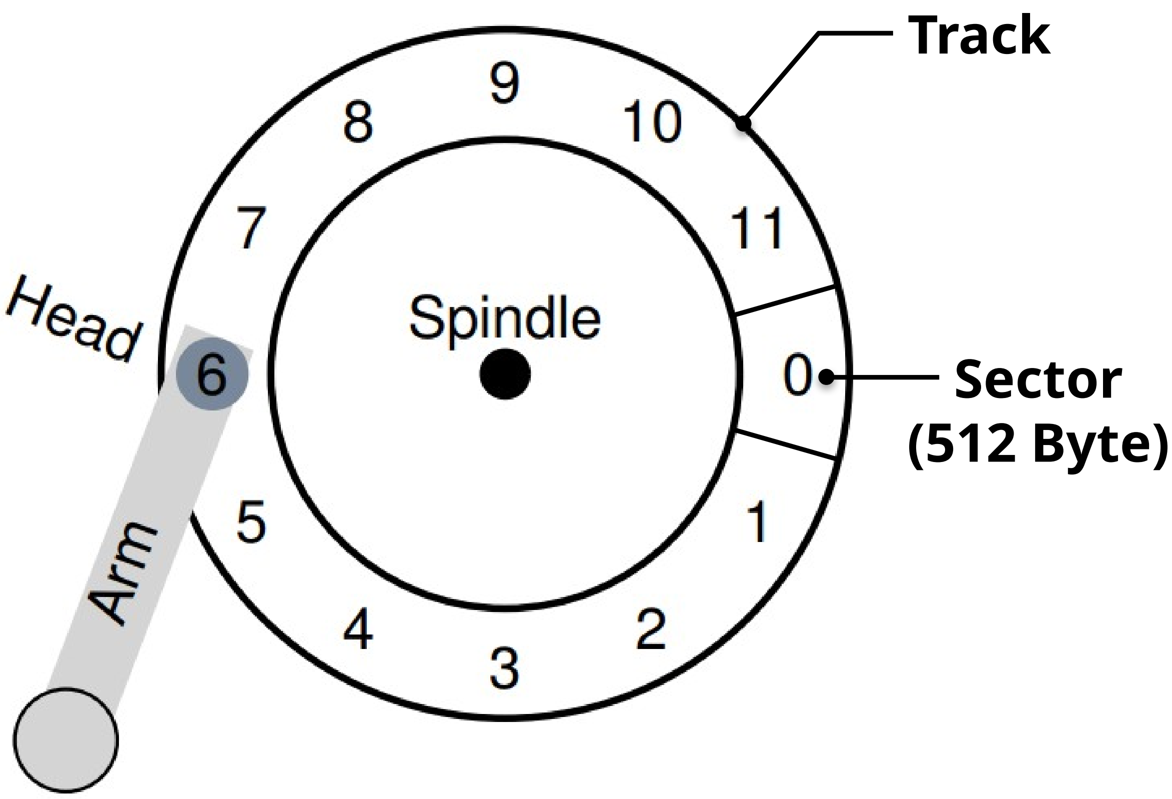
37. Hard Disk Drives

Operating System: Three Easy Pieces

Base Geometry

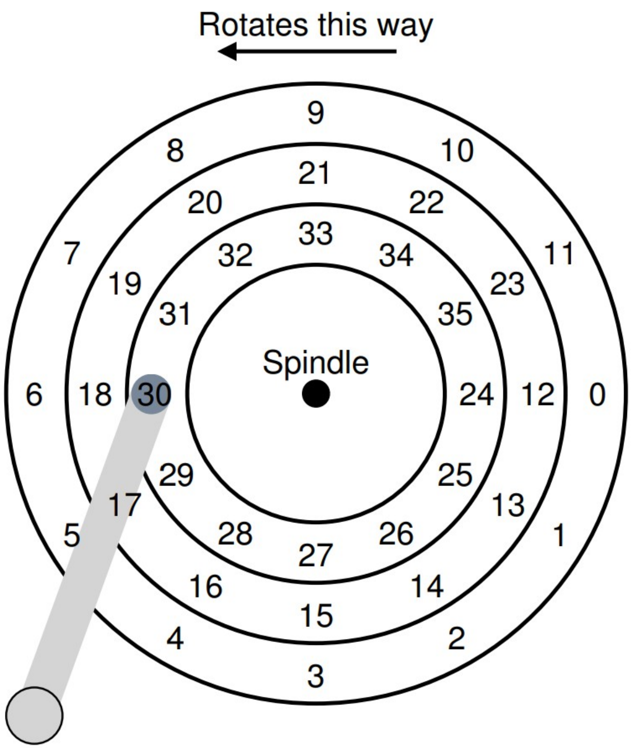


Base Geometry (Cont.)



A Simple Disk Drive

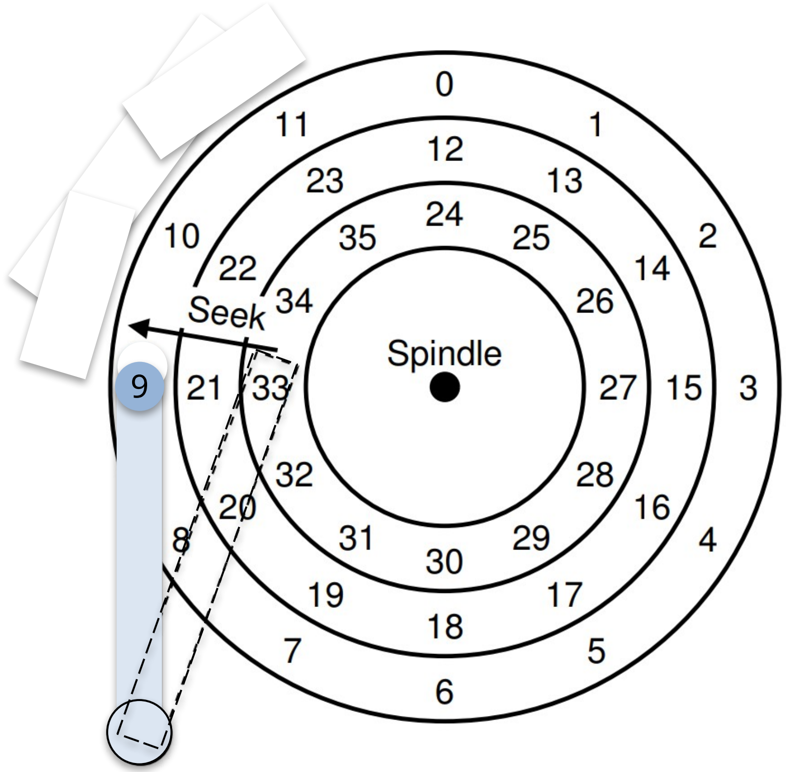
- Rotation Delay



If the full rotation delay is R,

$$Average\ turnaround\ time = \frac{10 + 20 + 30}{3} = 20\ sec$$

- Seek Time

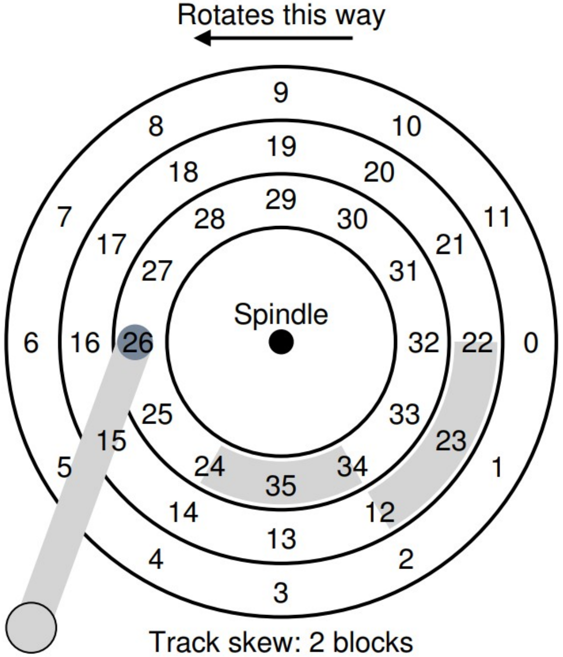


Phases of seek

Acceleration → *Coasting* → *Deceleration* → *Setting time* (about 0.5~2 ms)

A Simple Disk Drive (Cont.)

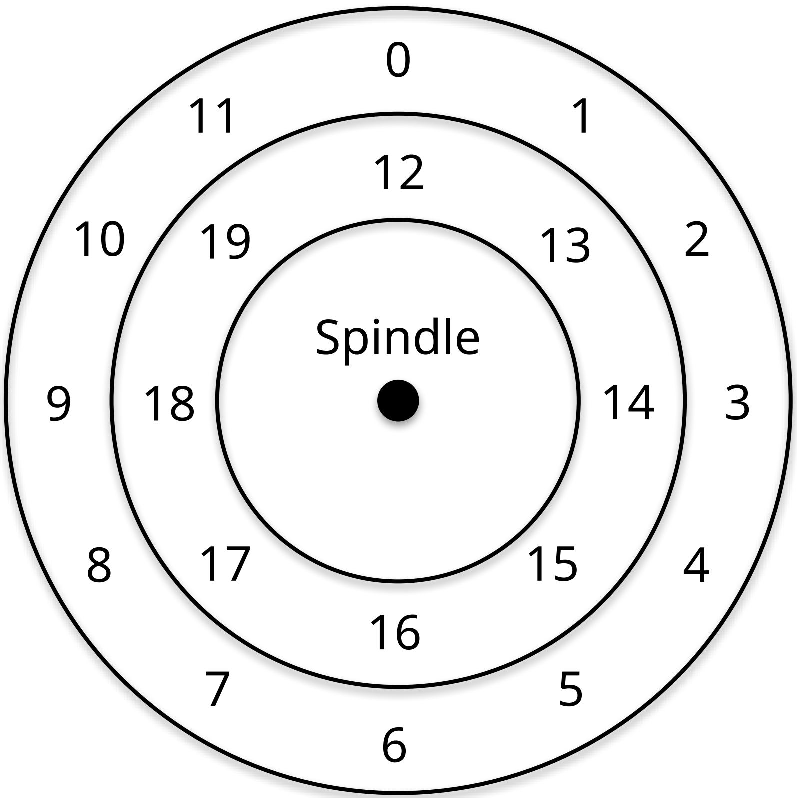
- Track skew
  - Make sure that sequential reads can be properly serviced even when crossing track boundaries
  - Without such skew, the head would be moved to the next track but the desired next block would have already rotated under the head



Track skew: 2 blocks

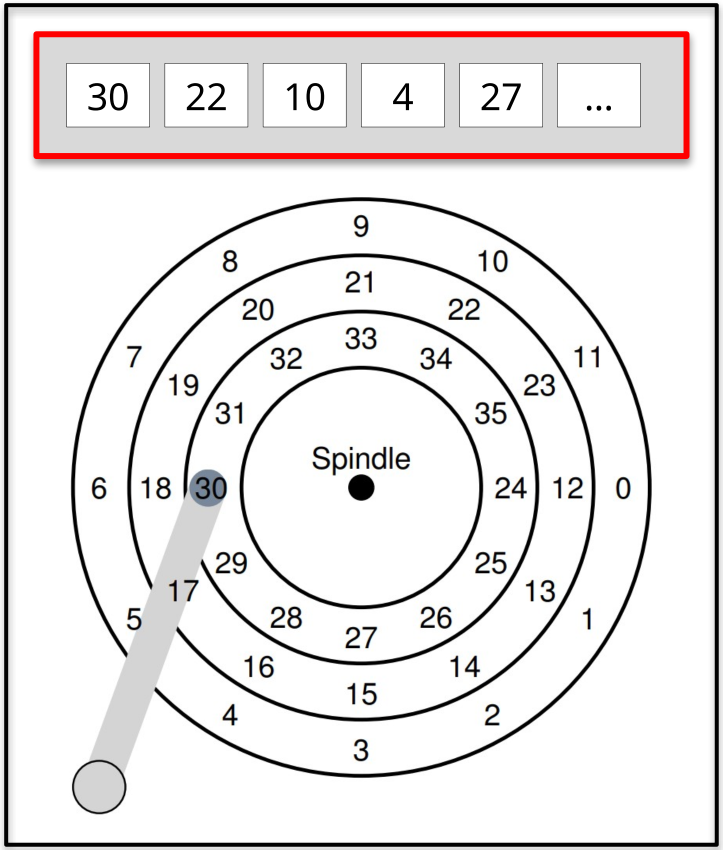
A Simple Disk Drive (Cont.)

- Multi-zoned
  - Outer tracks tend to have more sectors than inner tracks, because of result of geometry



A Simple Disk Drive (Cont.)

- Cache (Track buffer)



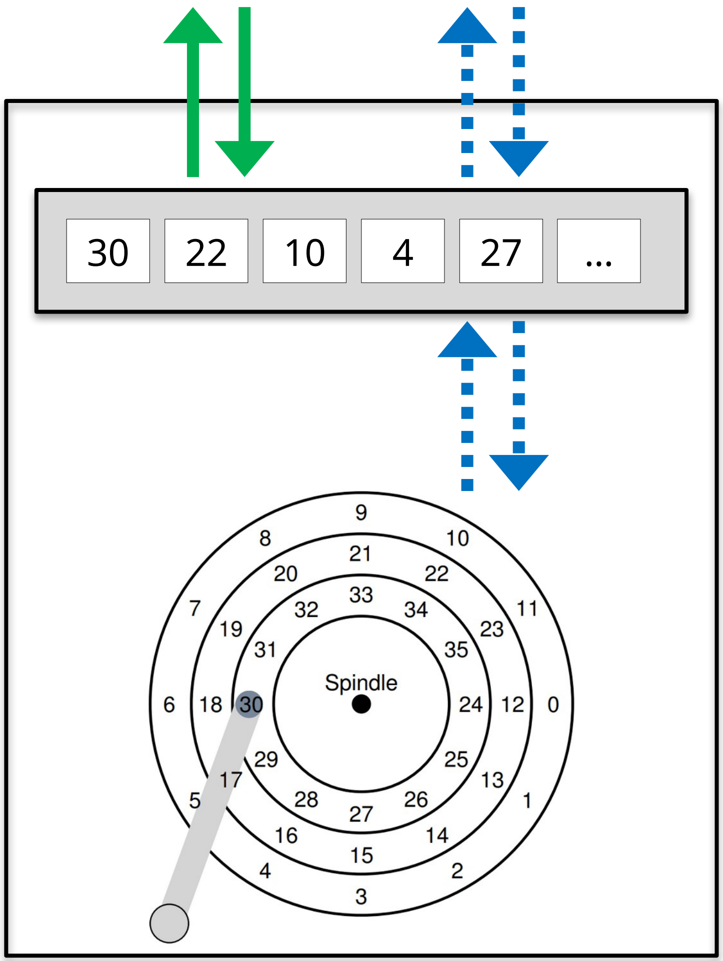
Small amount of memory
(usually around 8 or 16MB)

Hold data read from or written to the disk

Allow the drive to quickly respond to requests

A Simple Disk Drive (Cont.)

- Cache (Track buffer)



Write-Back

Acknowledge the write has completed when it has put the data in its memory

Write-Through

Acknowledge after the write has actually been written to disk

I/O Time: Doing The Math

- I/O Time

$$T_{I/O} = T_{seek} + T_{rotation} + T_{transfer}$$

- I/O Rate

$$R_{I/O} = \frac{Size_{Transfer}}{T_{I/O}}$$

I/O Time: Doing The Math (Cont.)

- 4KB Random Write Example

	Cheetah 15K.5	Barracuda
Capacity	300 GB	1 TB
RPM	15,000	7,200
Average Seek	4 ms	9 ms
Max Transfer	125 MB/s	105 MB/s
Platters	4	4
Cache	16 MB	16/32 MB
Connects via	SCSI	SATA

T_{seek} = 4ms

$T_{rotation}$ = 15,000 RPM(= 250RPS = 4ms / 1 rotation) / 2
= 2ms

$T_{transfer}$ = 4KB / 125(MB/s)
= 30us

$T_{I/O}$ = 4ms + 2ms + 30us ≒ 6ms

$R_{I/O}$ = 4KB / 6ms = 0.66MB/s

I/O Time: Doing The Math (Cont.)

- 4KB Random Write Example (Cont.)

	Cheetah 15K.5	Barracuda
Capacity	300 GB	1 TB
RPM	15,000	7,200
Average Seek	4 ms	9 ms
Max Transfer	125 MB/s	105 MB/s
Platters	4	4
Cache	16 MB	16/32 MB
Connects via	SCSI	SATA

T_{seek} = 9ms

$T_{rotation}$ = 7,200 RPM(= 120RPS = 8ms / 1 rotation) / 2
= 4ms

$T_{transfer}$ = 4KB / 105(MB/s)
= 38us

$T_{I/O}$ = 9ms + 4ms + 38us ≒ 13ms

$R_{I/O}$ = 4KB / 13ms = 0.31MB/s

I/O Time: Doing The Math (Cont.)

- Sequential Write Example

	Cheetah 15K.5	Barracuda
Capacity	300 GB	1 TB
RPM	15,000	7,200
Average Seek	4 ms	9 ms
Max Transfer	125 MB/s	105 MB/s
Platters	4	4
Cache	16 MB	16/32 MB
Connects via	SCSI	SATA

T_{seek} = 4ms

$T_{rotation}$ = 15,000 RPM(= 250RPS = 4ms / 1 rotation) / 2
= 2ms

$T_{transfer}$ = 100MB / 125(MB/s)
= 800ms

$T_{I/O}$ = 4ms + 2ms + 800ms = 806ms ≒ 800ms

$R_{I/O}$ = 100MB / 800ms = 125MB/s

I/O Time: Doing The Math (Cont.)

- Sequential Write Example (Cont.)

	Cheetah 15K.5	Barracuda
Capacity	300 GB	1 TB
RPM	15,000	7,200
Average Seek	4 ms	9 ms
Max Transfer	125 MB/s	105 MB/s
Platters	4	4
Cache	16 MB	16/32 MB
Connects via	SCSI	SATA

T_{seek} = 9ms

$T_{rotation}$ = 7,200 RPM(= 120RPS = 8ms / 1 rotation) / 2
= 4ms

$T_{transfer}$ = 100MB / 105(MB/s)
= 950ms

$T_{I/O}$ = 9ms + 4ms + 950ms = 963ms ≒ 950ms

$R_{I/O}$ = 100MB / 950ms = 105MB/s

I/O Time: Doing The Math (Cont.)

	Cheetah	Barracuda
$R_{I/O}$ Random	0.66 MB/s	0.31 MB/s
$R_{I/O}$ Sequential	125 MB/s	105 MB/s

Performance vs Capacity

	Cheetah	Barracuda
$R_{I/O}$ Random	0.66 MB/s	0.31 MB/s
$R_{I/O}$ Sequential	125 MB/s	105 MB/s

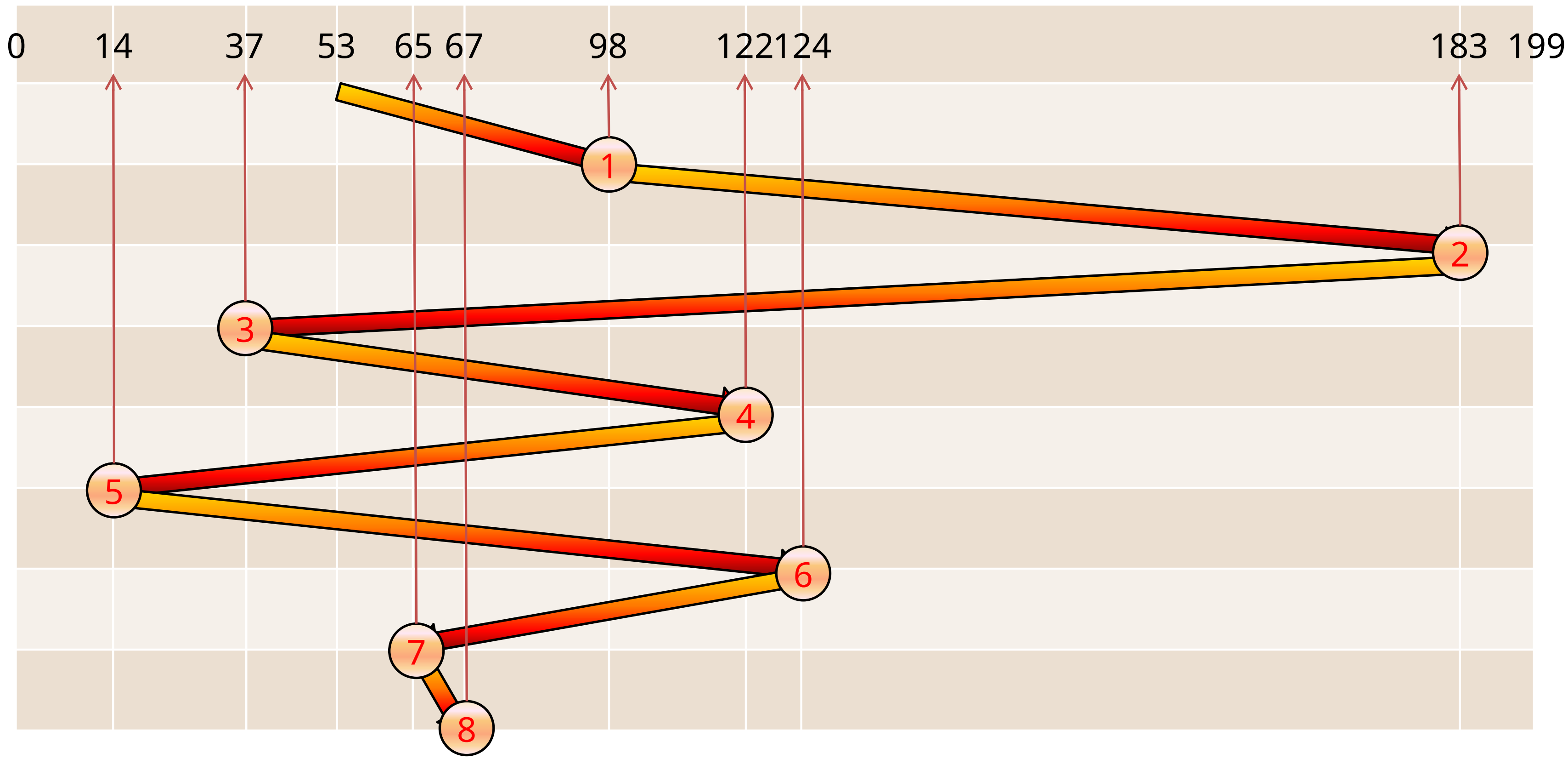
Random Write vs Sequential Write

Disk Scheduling: FCFS

How to order the services for the requests in the queue?

Illustration shows total head movement of 640 cylinders.

Queue = 98, 183, 37, 122, 14, 124, 65, 67
Head starts at 53

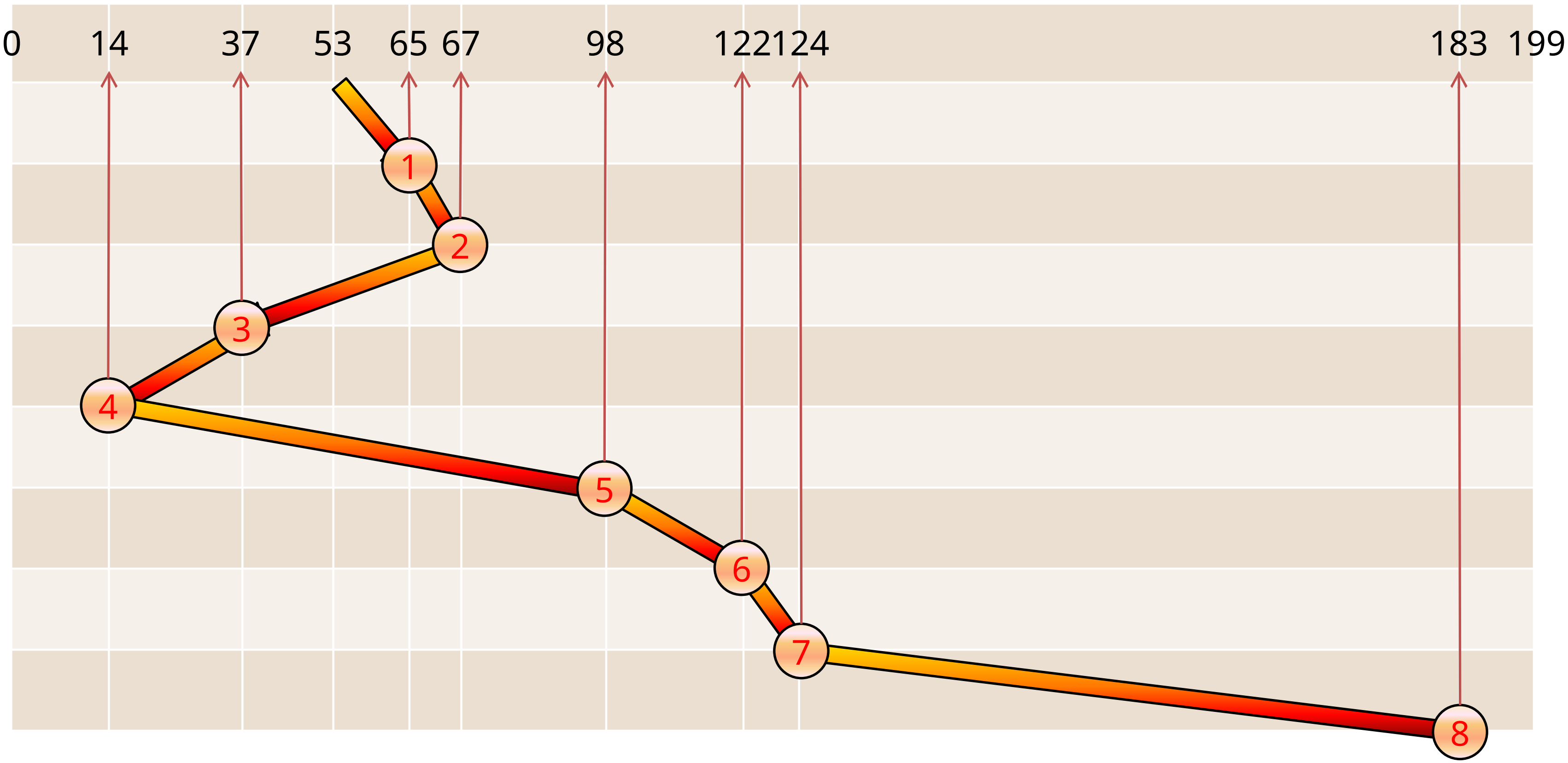


SSTF

- Selects the request with the minimum seek time from the current head position.
- SSTF scheduling is a form of SJF scheduling; may cause starvation of some requests.
- Illustration shows total head movement of 236 cylinders.

SSTF (Cont.)

Queue = 98, 183, 37, 122, 14, 124, 65, 67
Head starts at 53

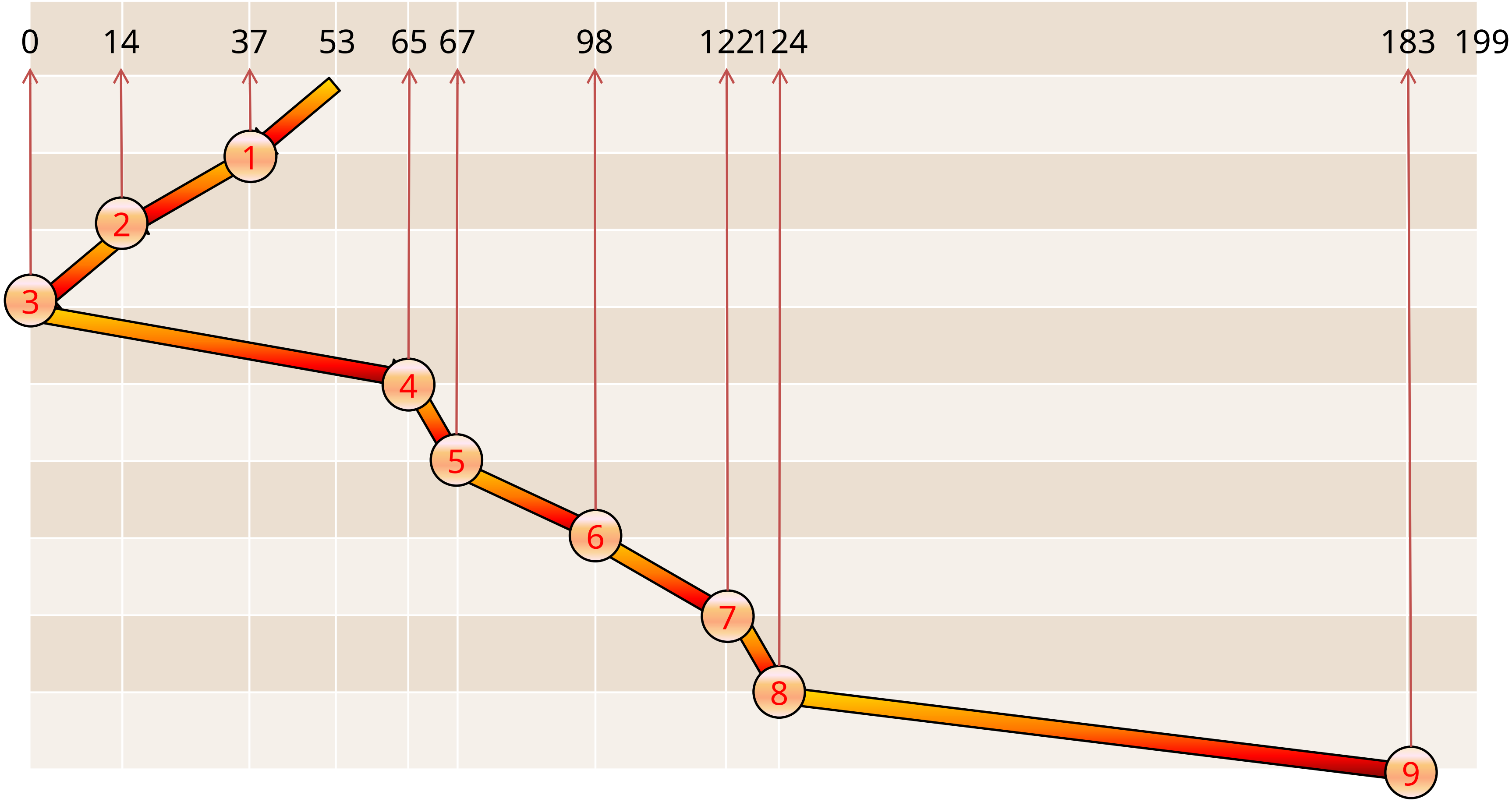


SCAN

- The disk arm starts at one end of the disk, and moves toward the other end, servicing requests until it gets to the other end of the disk, where the head movement is reversed and servicing continues.
- Sometimes called the *elevator algorithm*.
- Illustration shows total head movement of 208 cylinders.

SCAN (Cont.)

Queue = 98, 183, 37, 122, 14, 124, 65, 67
Head starts at 53

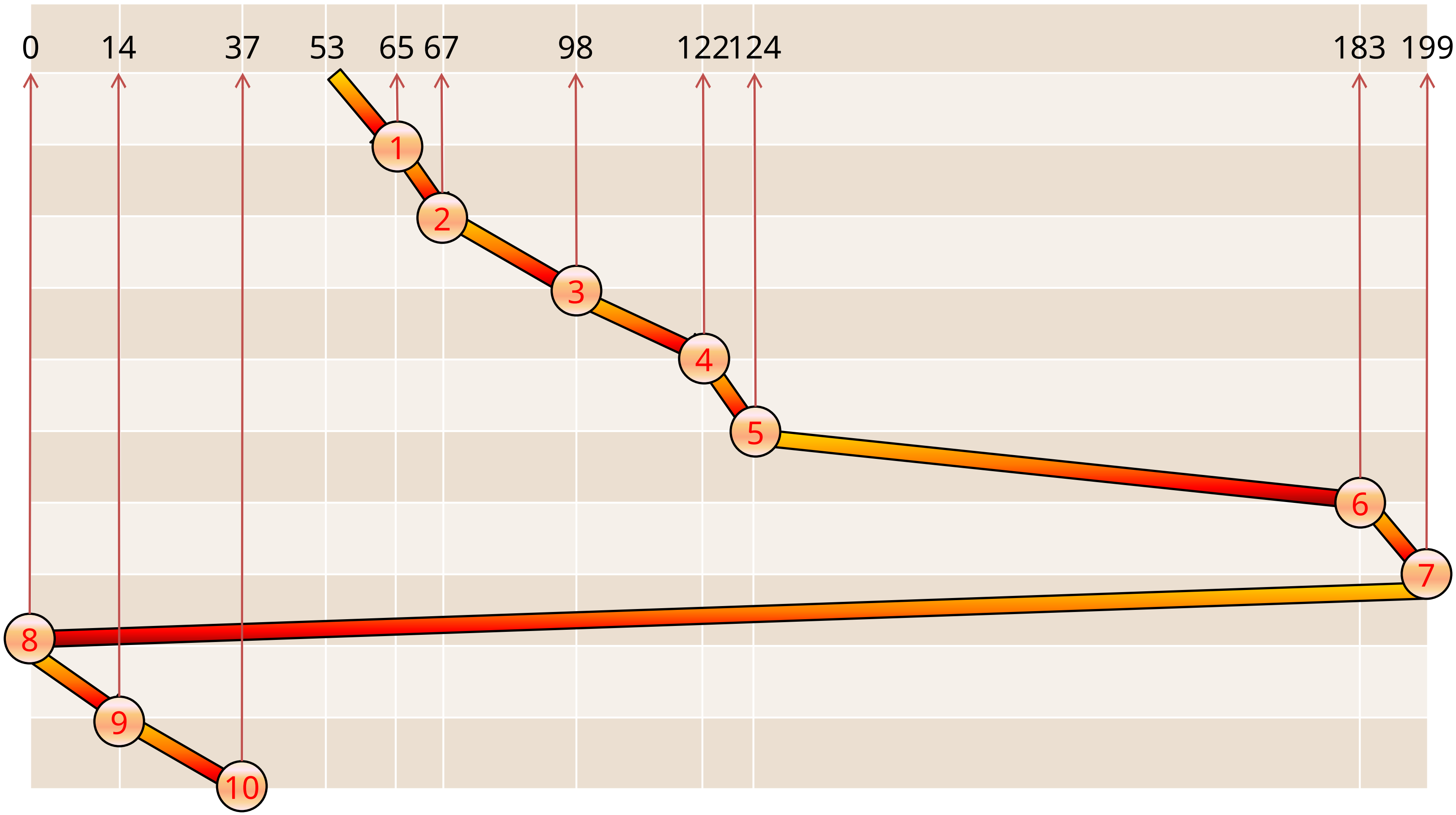


C-SCAN

- Provides a more uniform wait time than SCAN.
- The head moves from one end of the disk 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.
- Treats the cylinders as a circular list that wraps around from the last cylinder to the first one.

C-SCAN (Cont.)

Queue = 98, 183, 37, 122, 14, 124, 65, 67
Head starts at 53

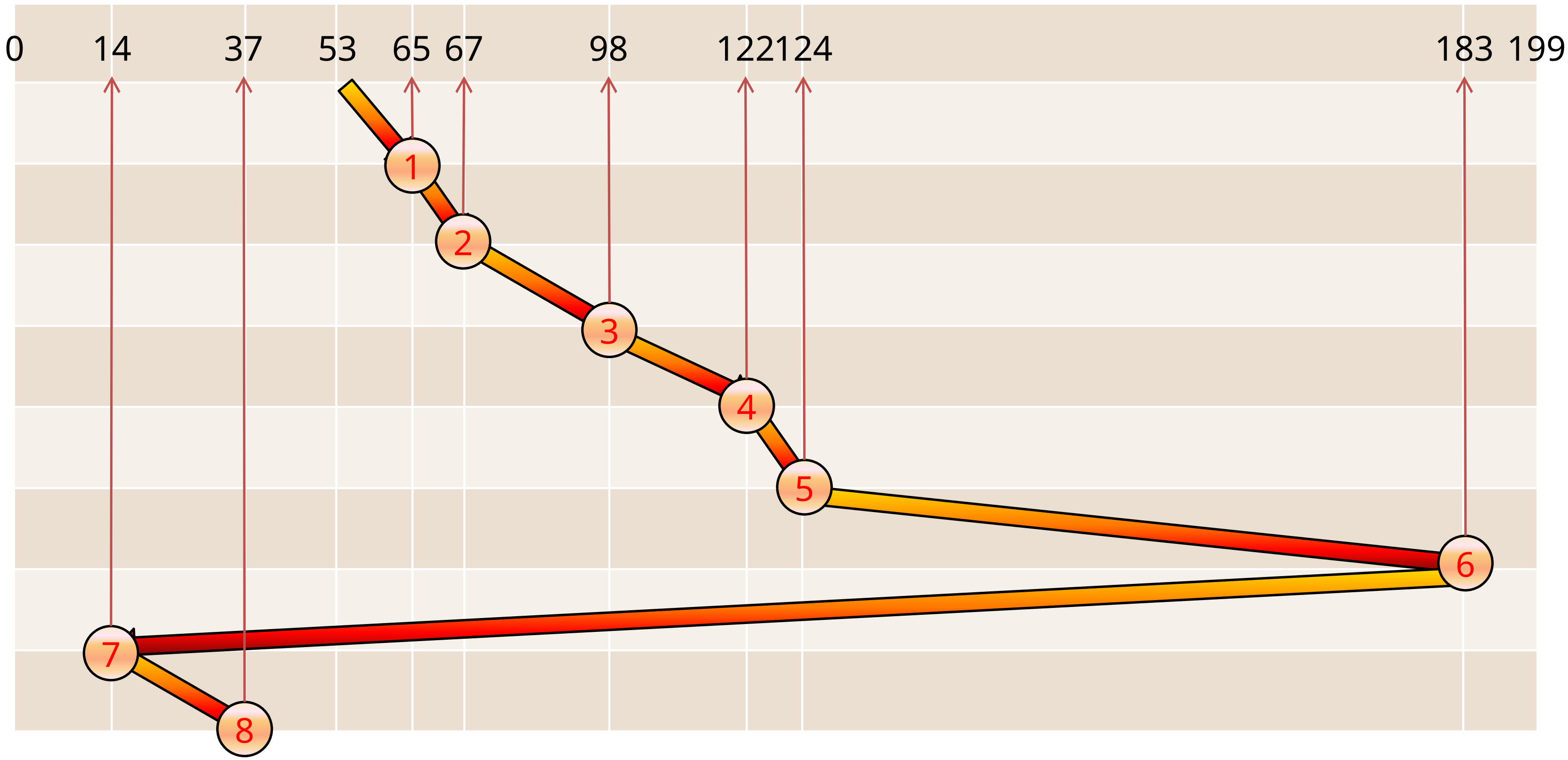


C-LOOK

- 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



Selecting a Disk-Scheduling Algorithm

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

Disk Scheduling (Cont.)

- I/O merging
  - Merge the continuous requests into a single two-block request
  - By waiting, a new and "better" request may arrive at the disk