Operating System

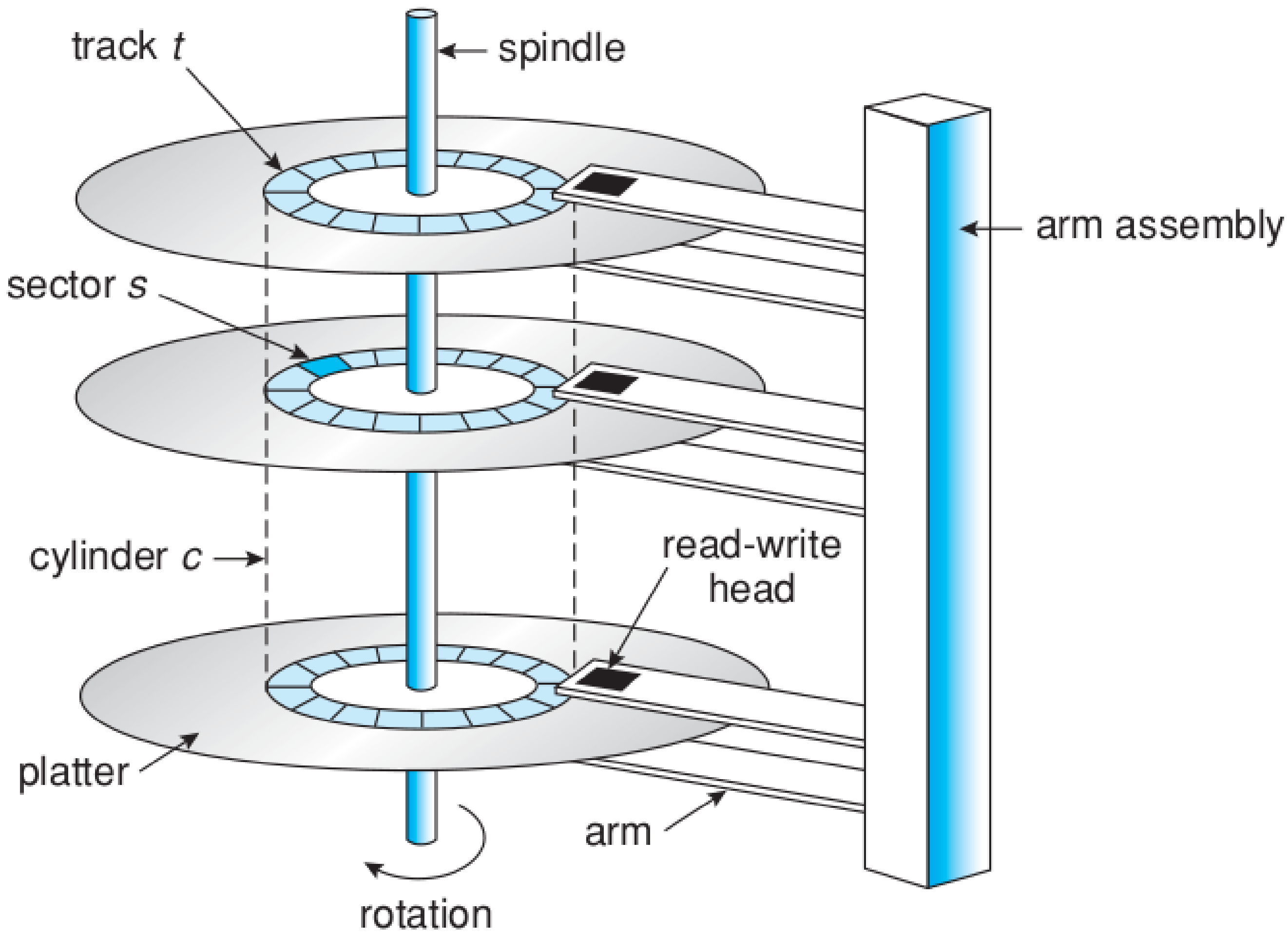
Disk management

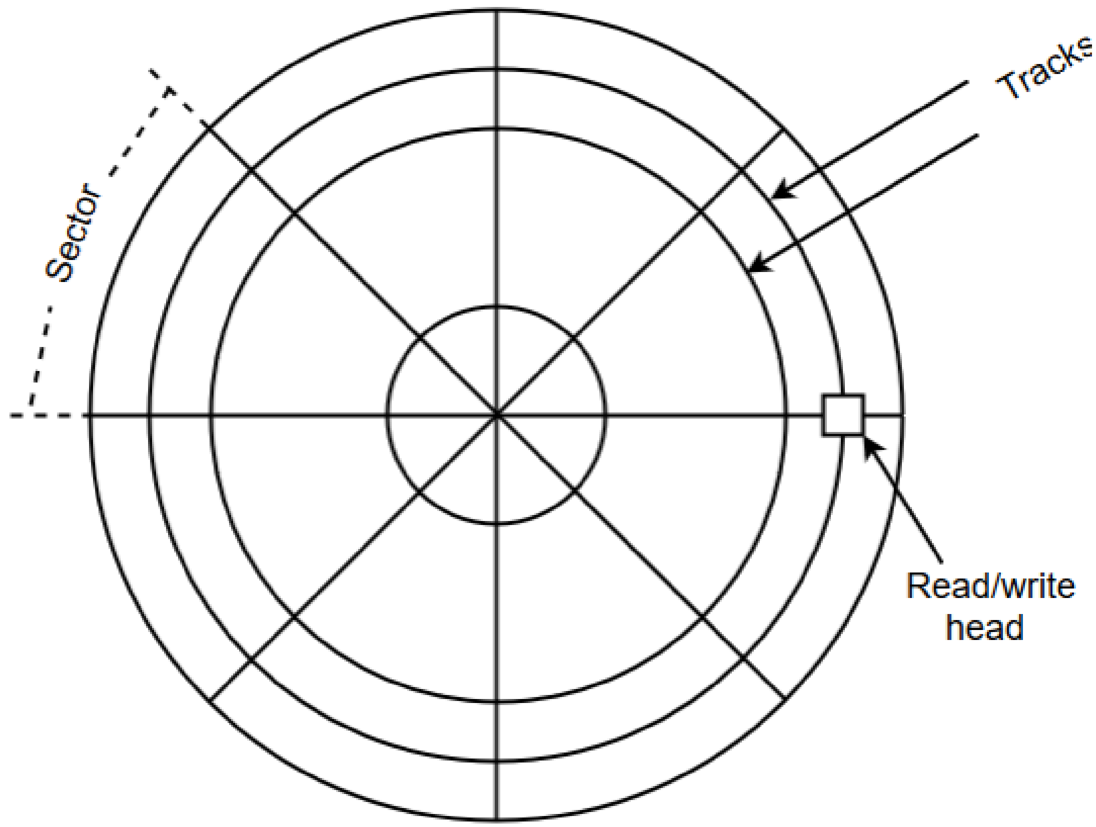
Ahmad Yoosofan

Spring 2023

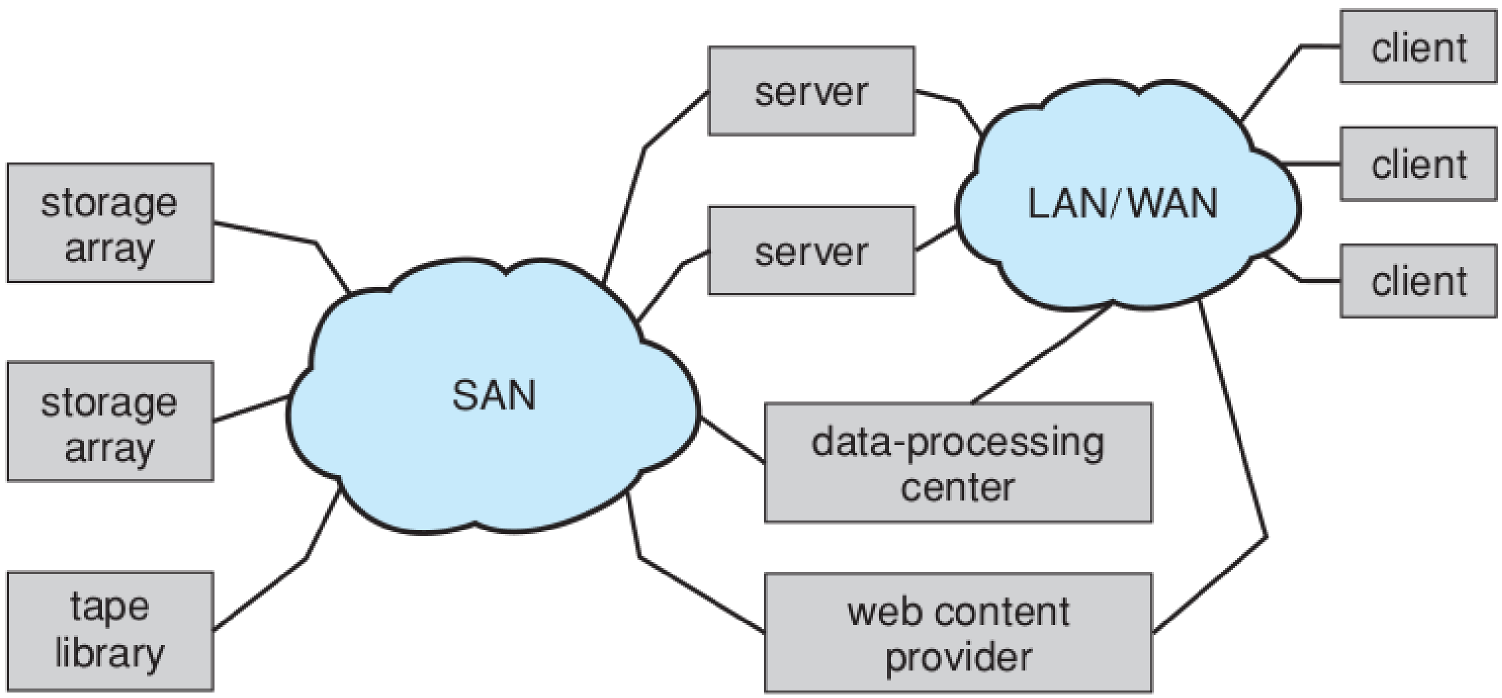
Disk

Disk structure




Tracks
Sector
Read/write
head

SStorrage Areaa Network(SAN)



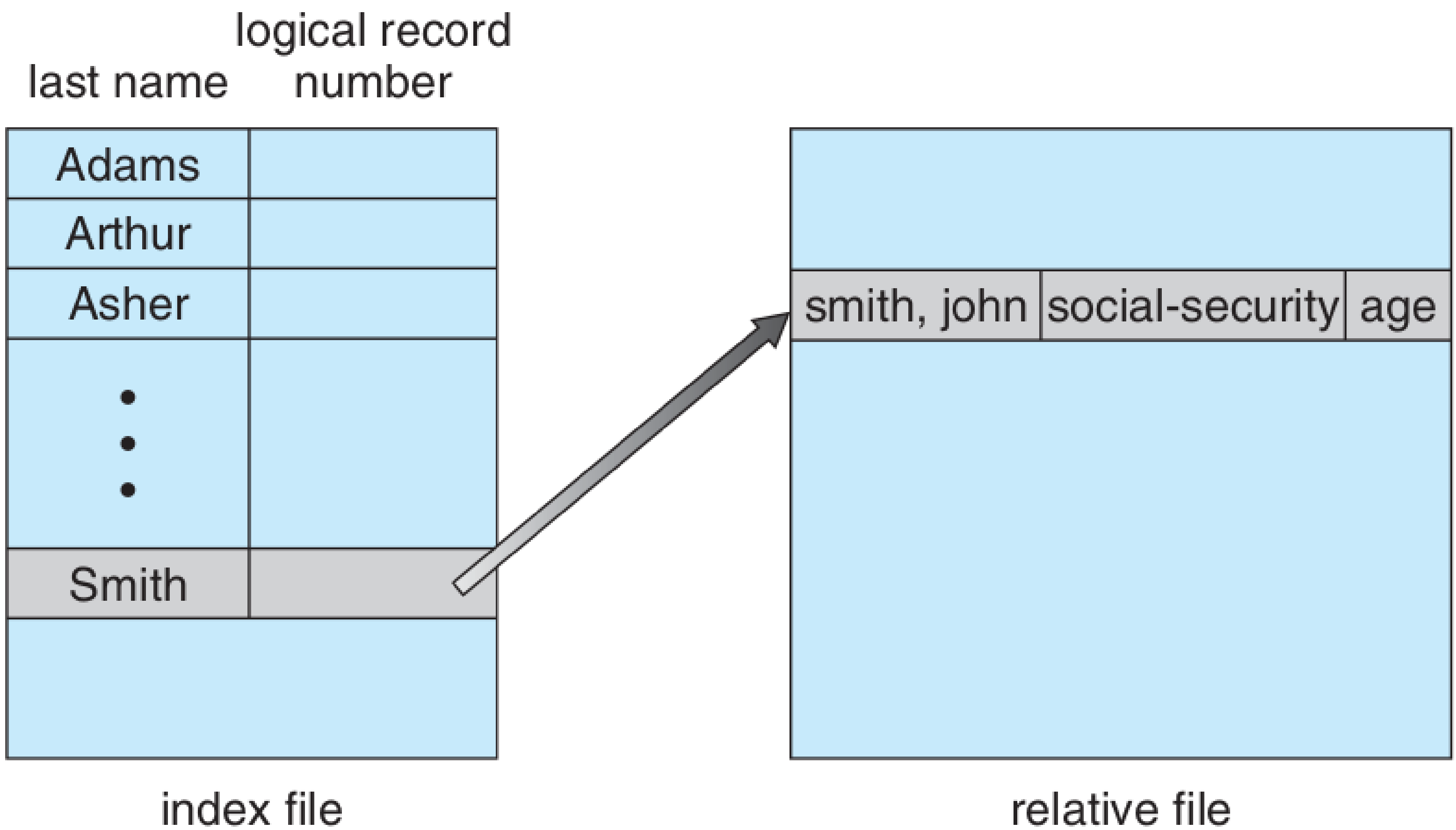
File Management System

.

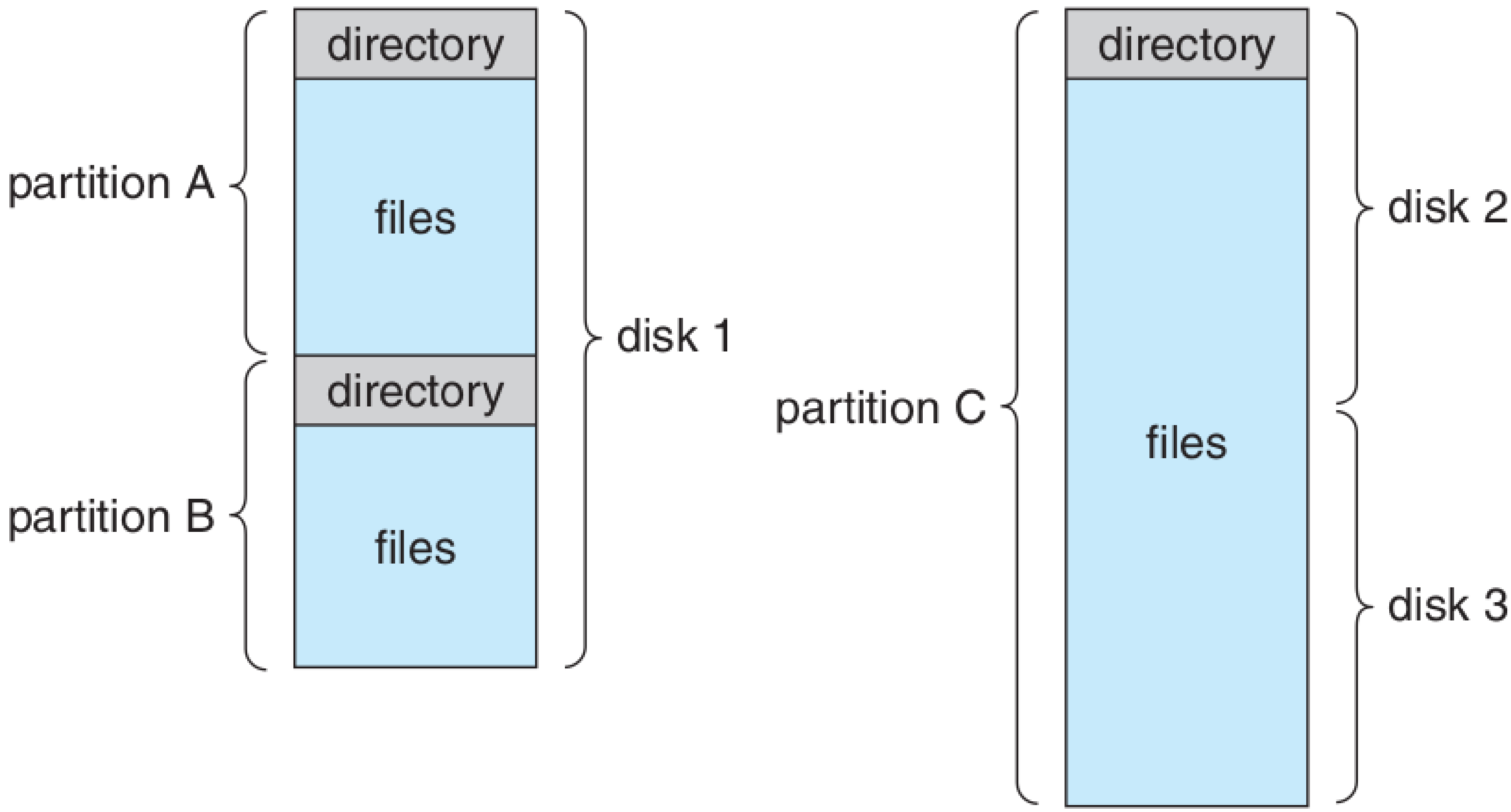
File Access Methods

- Sequential Access
- Random Access

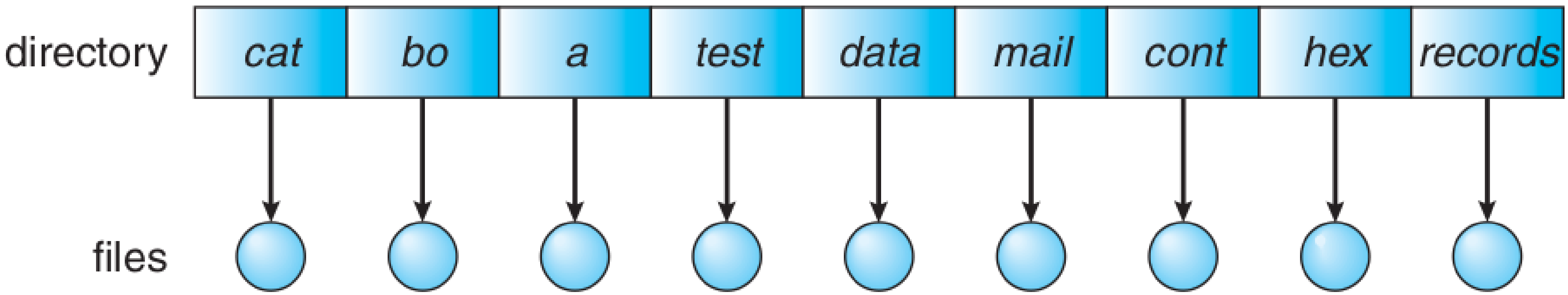
Simple Access Disk



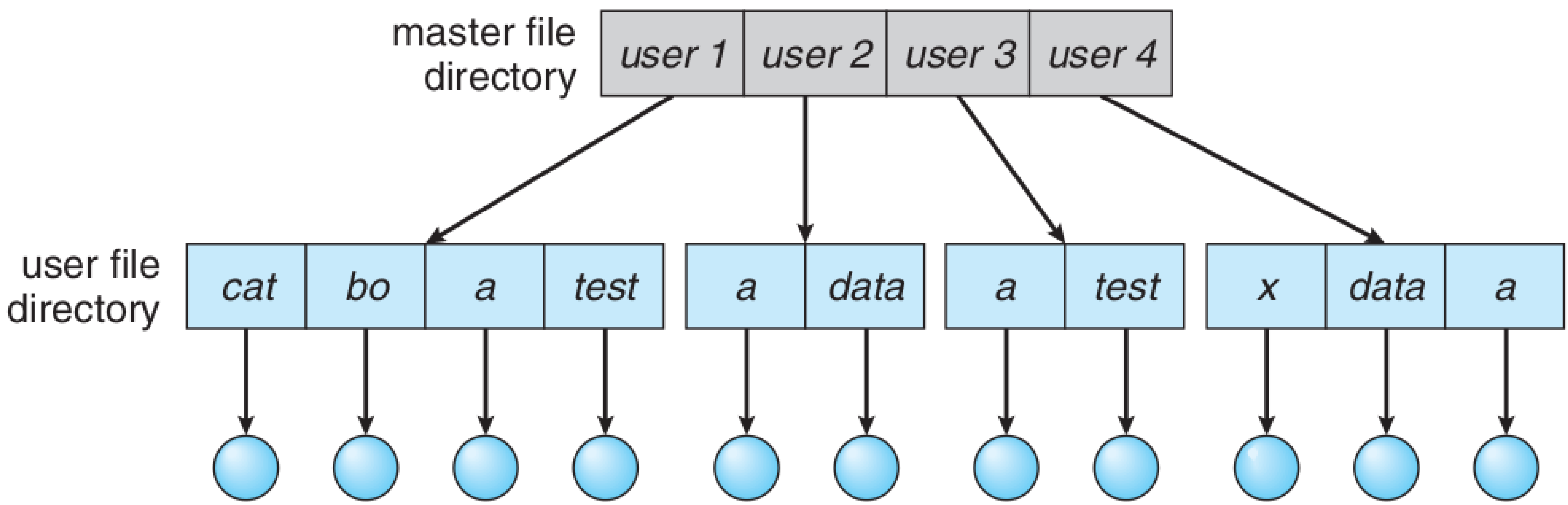
Simple Disk Organization



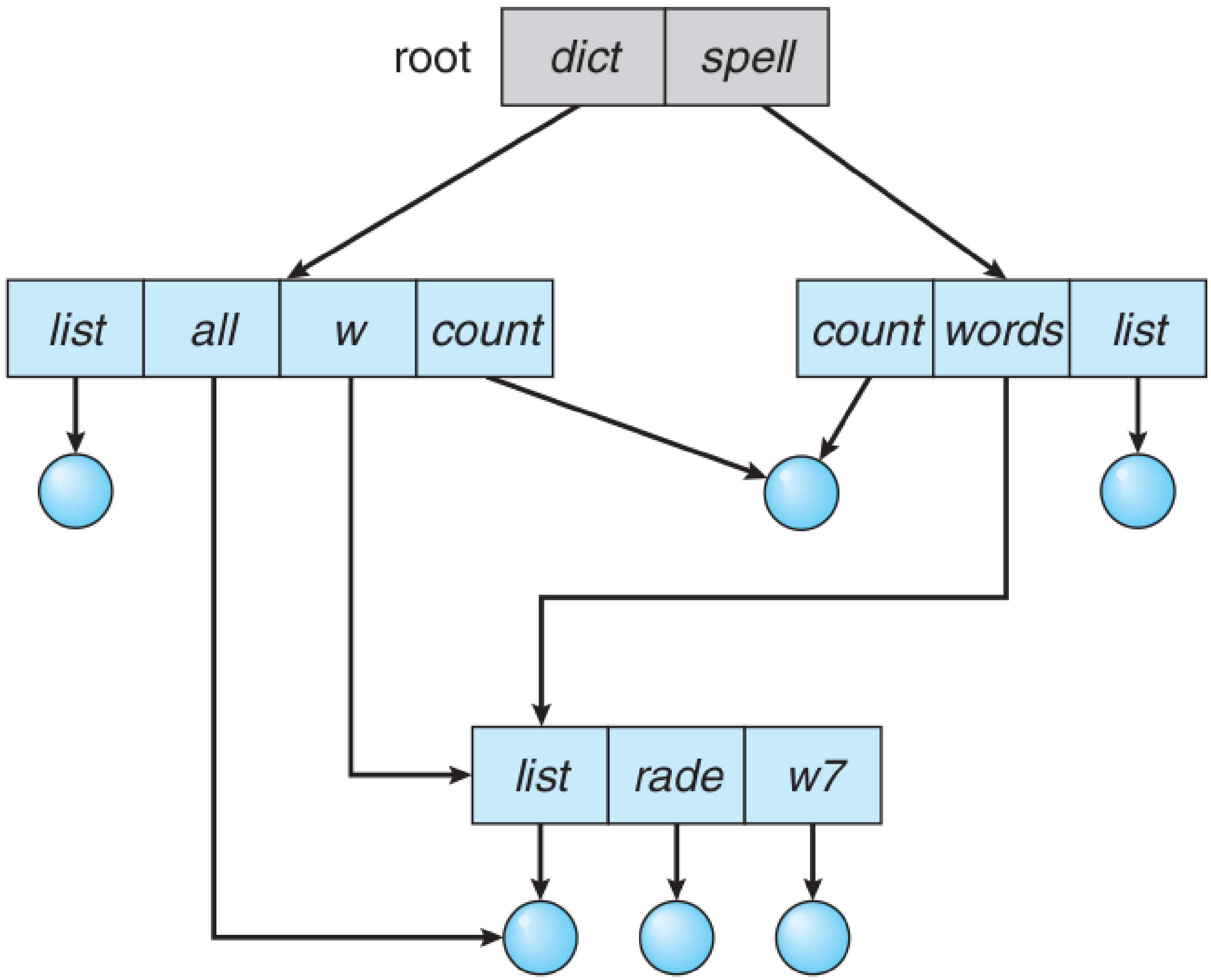
One Level Structure of Folders



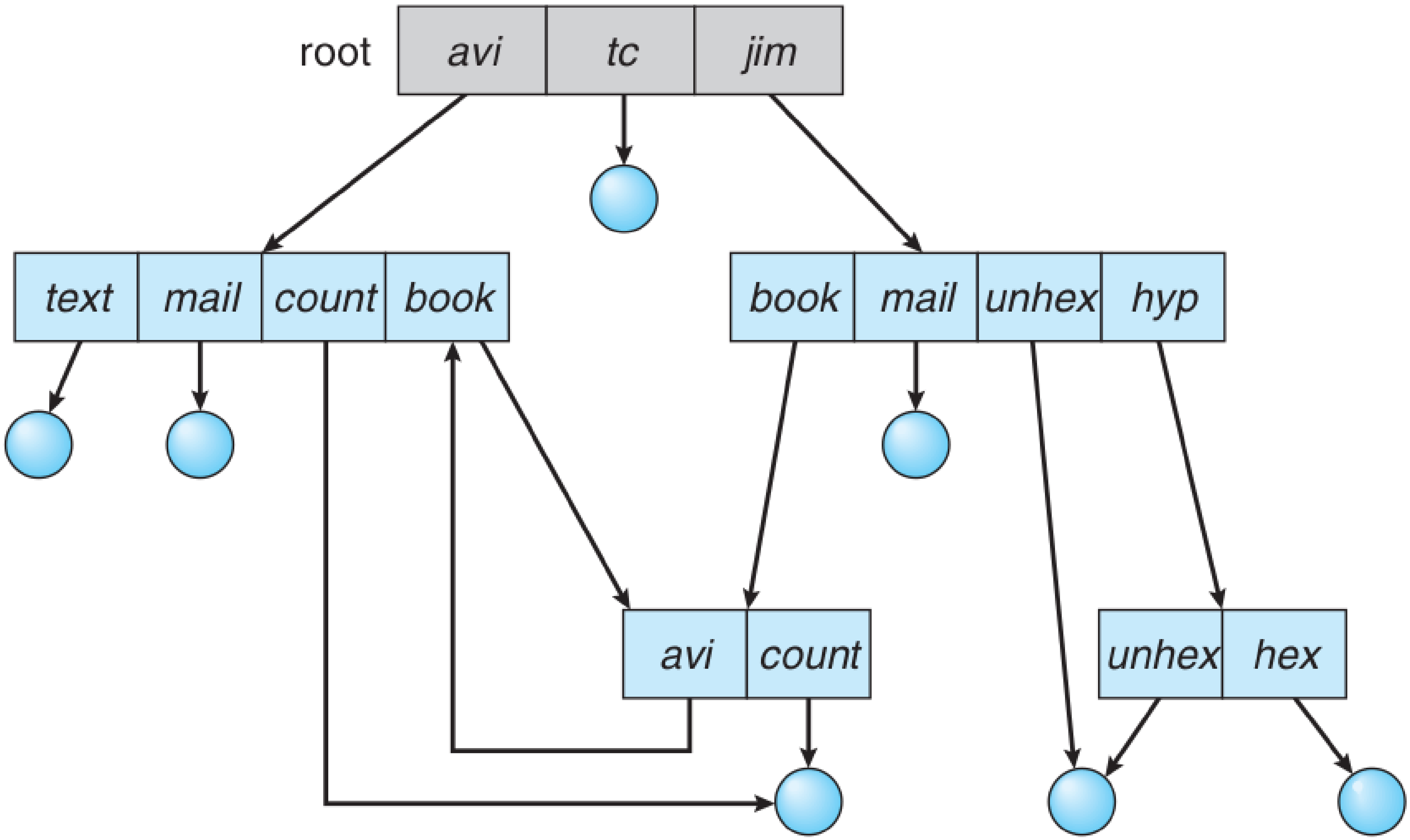
Two Level Structure of Folders



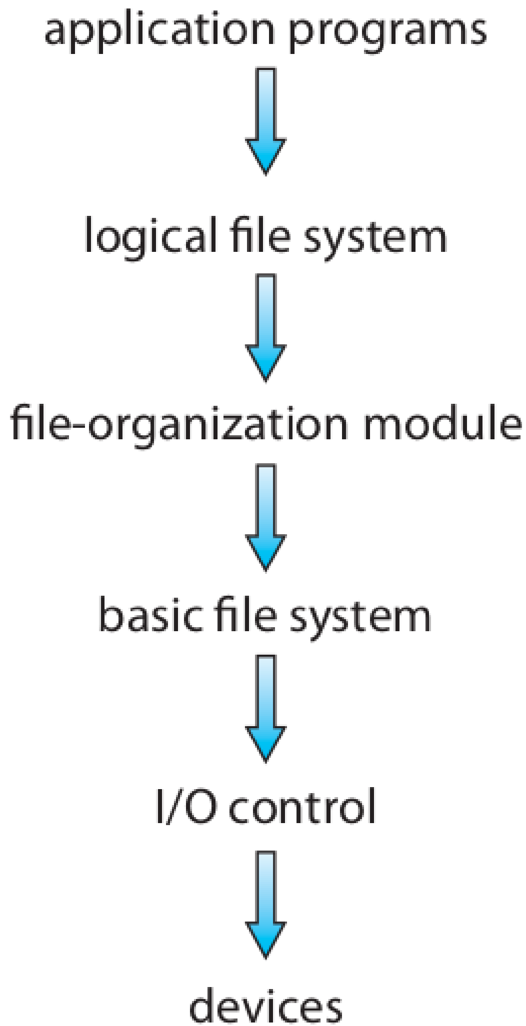
Acyclic Graph based Folder Structure



General Graph based Folder Structure



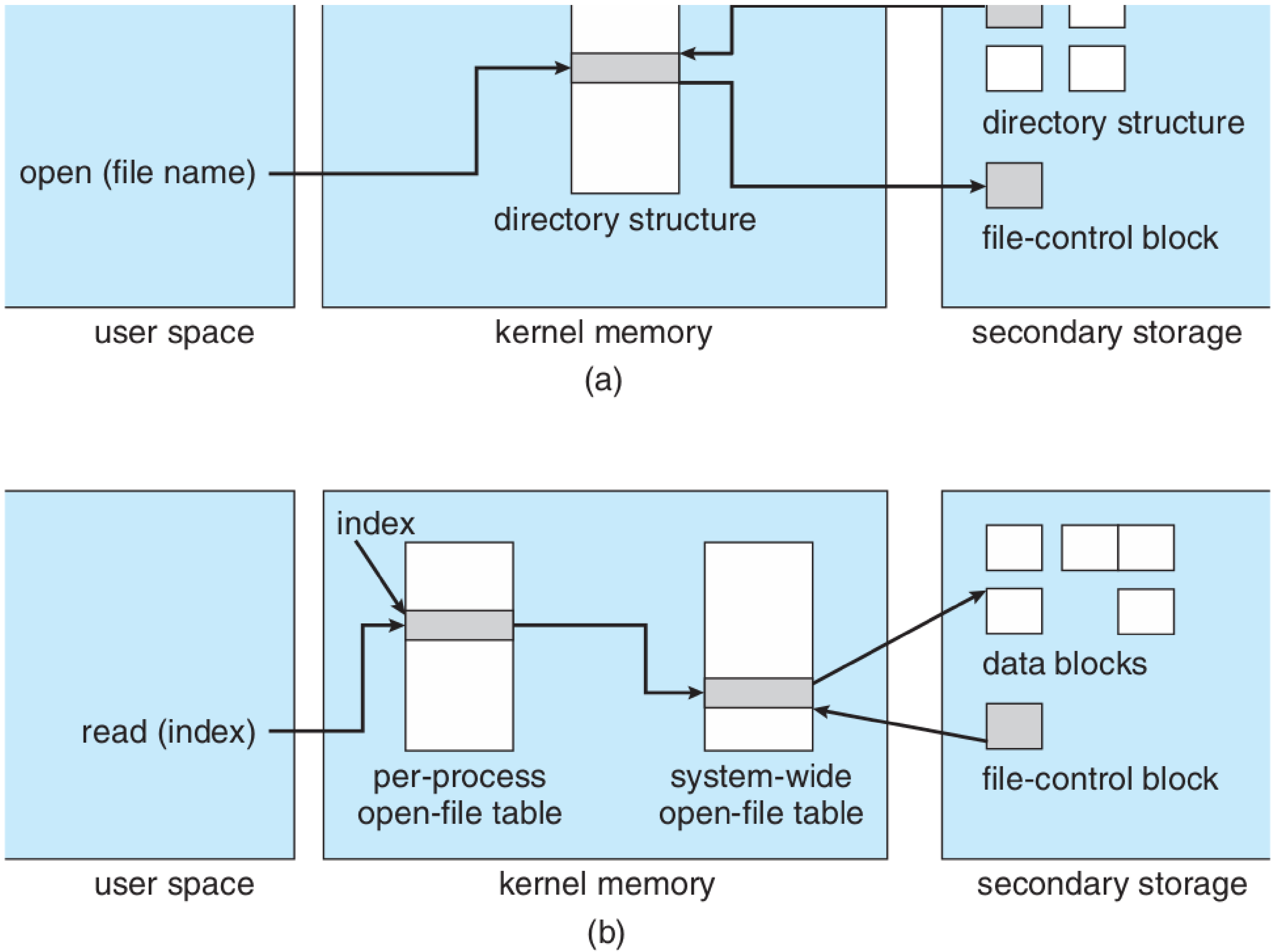
Multi Level File Structure



Simple Control Block of a File

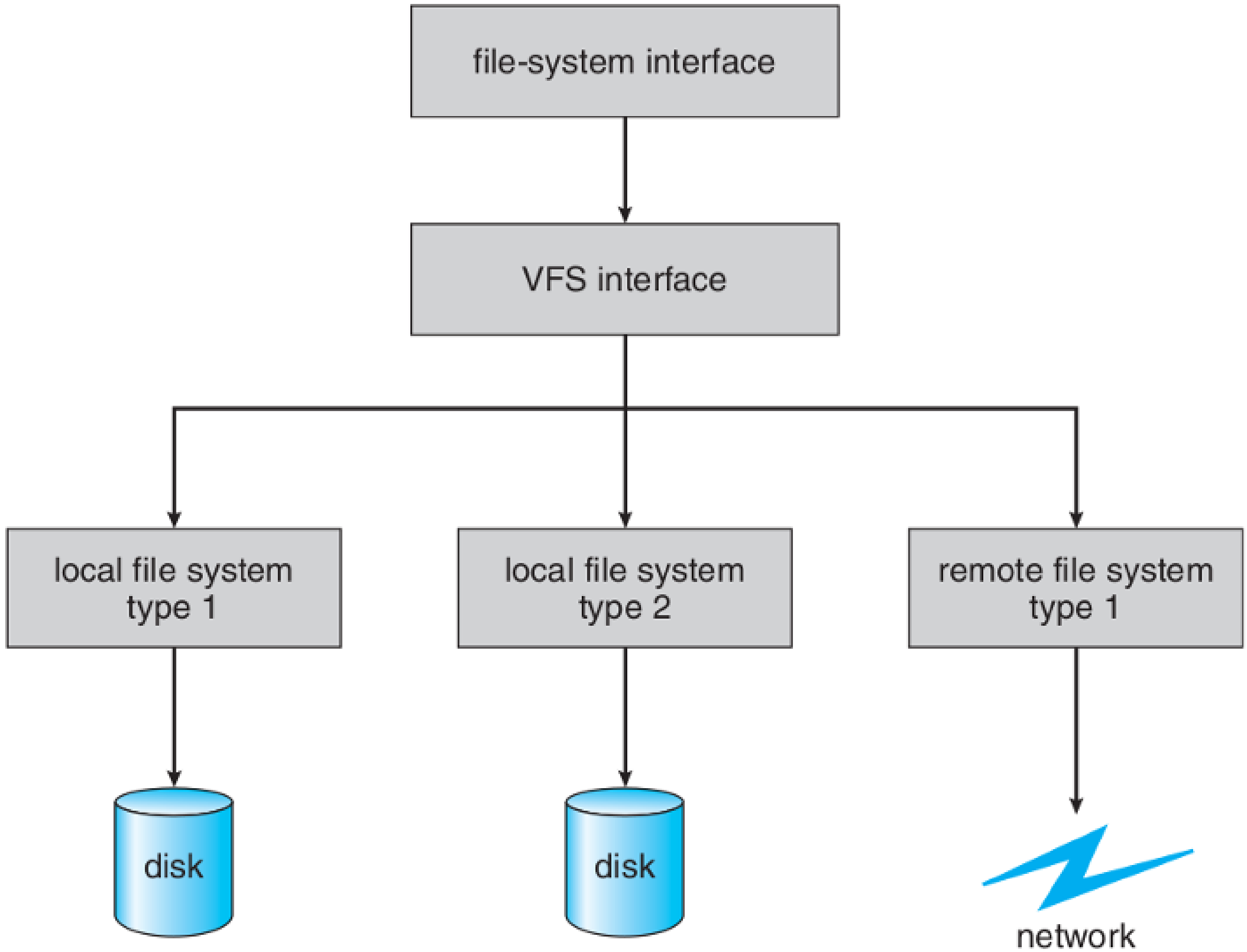
file permissions
file dates (create, access, write)
file owner, group, ACL
file size
file data blocks or pointers to file data blocks

File struture in Main Memory


open (file name)
directory structure
directory structure
file-control block
user space
kernel memory
secondary storage
(a)
index
read (index)
per-process
open-file table
system-wide
open-file table
data blocks
file-control block
user space
kernel memory
secondary storage
(b)

-
  1. open file
-
  1. close file

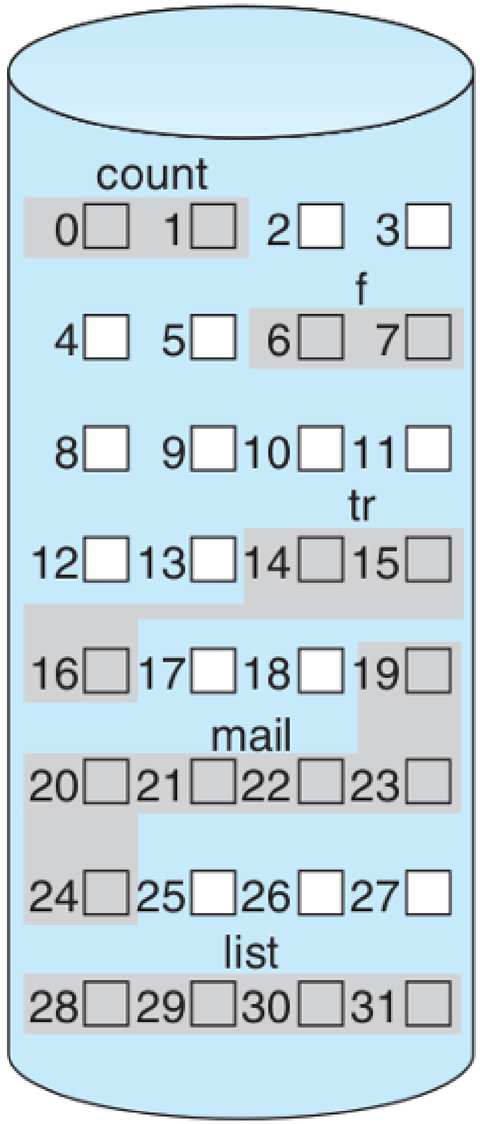
Schematic View of Virtual File System



Allocation Files Methods

.

Contiguous Block Allocation

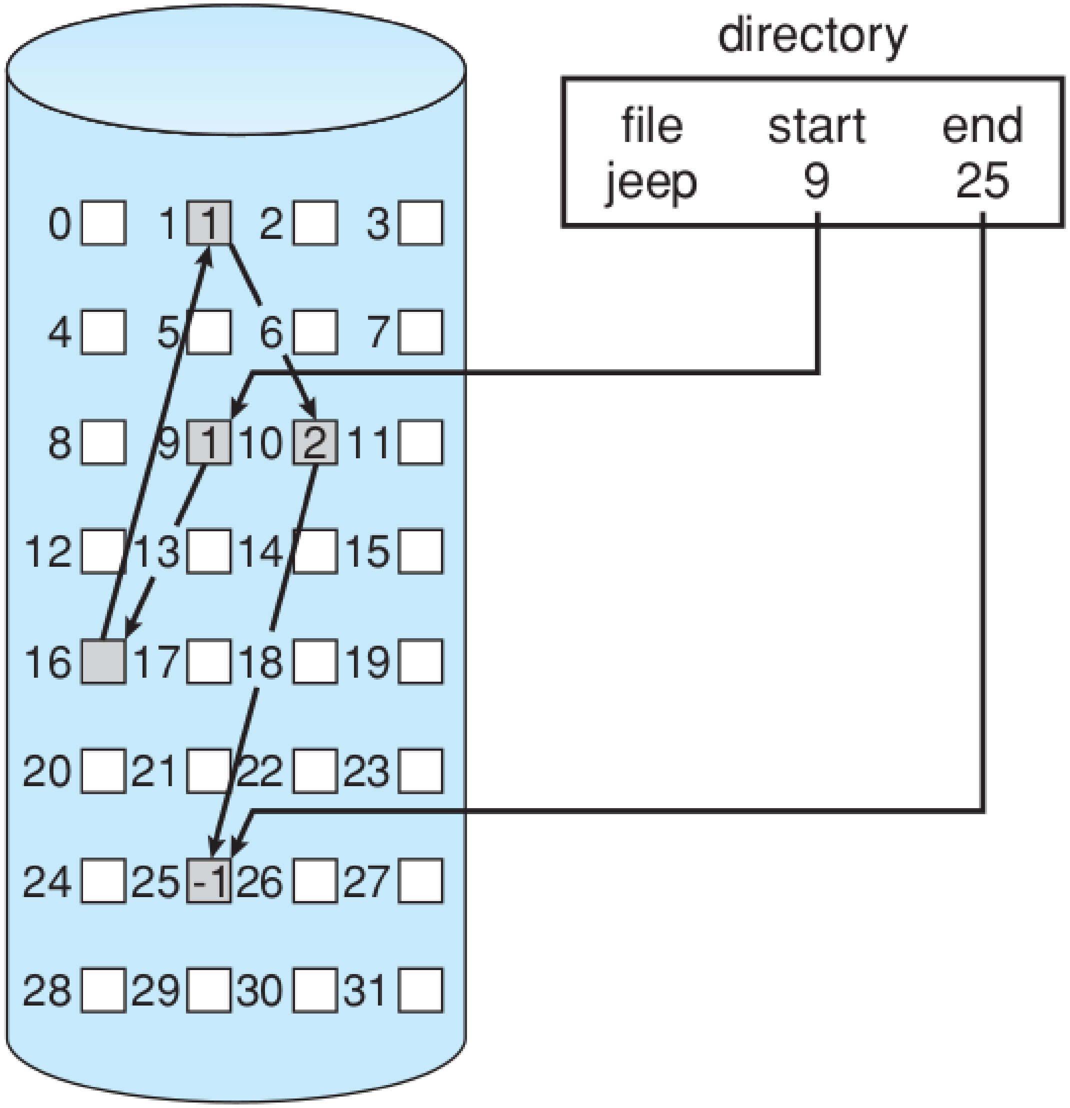


directory

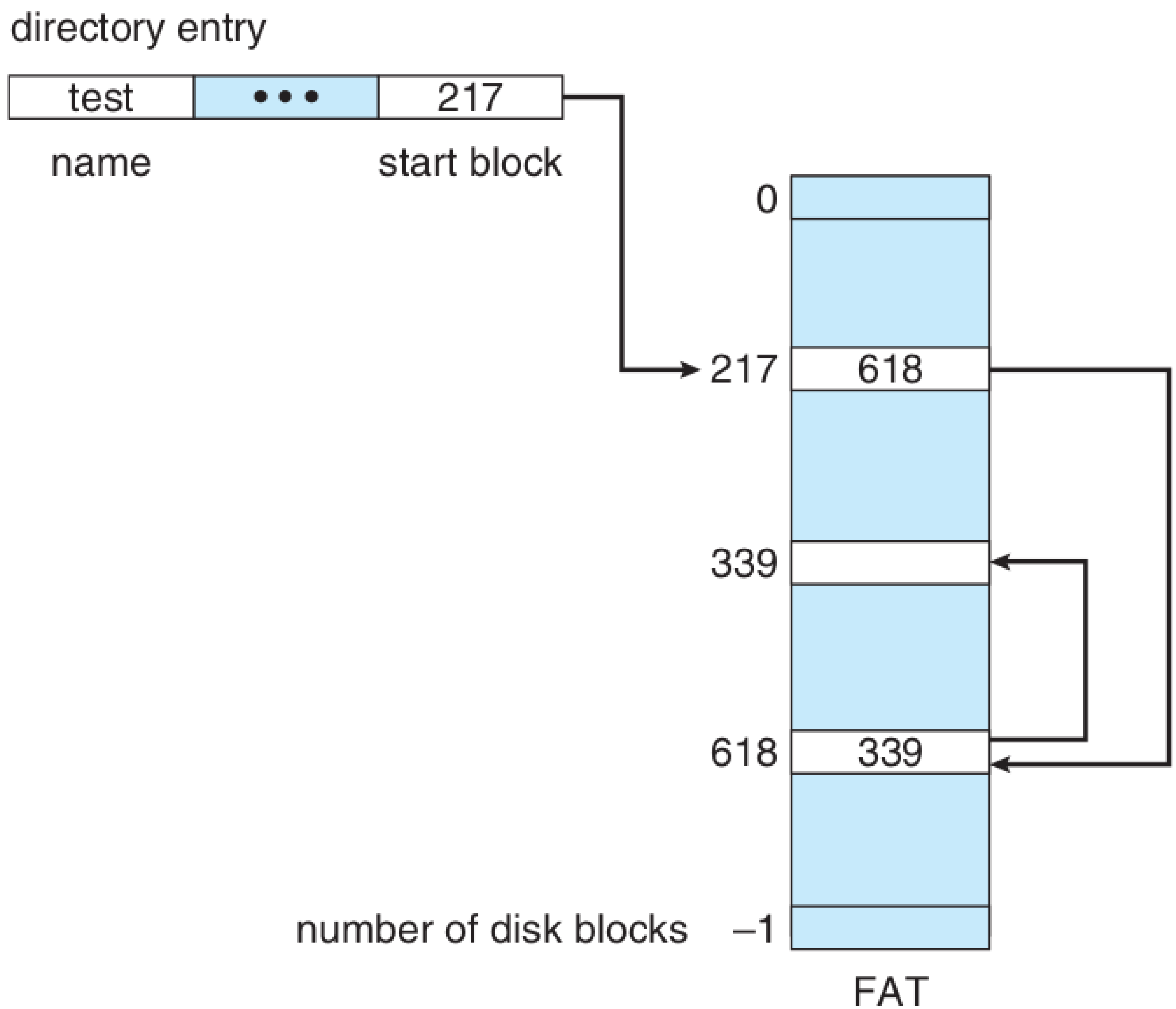
file	start	length
count	0	2
tr	14	3
mail	19	6
list	28	4
f	6	2

- External Fragmentation
- Adding Block to File
- Speed

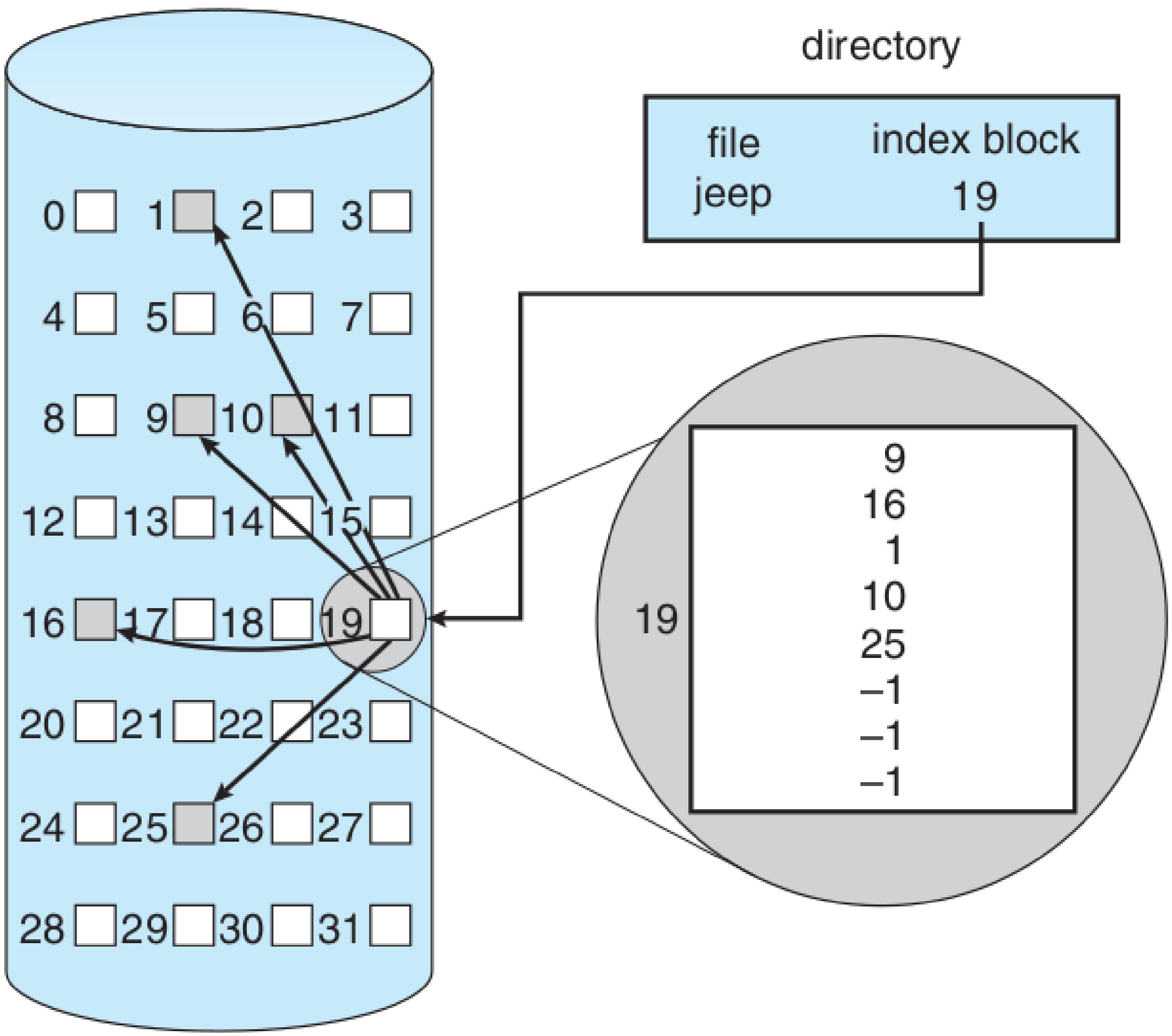
List Block Allocation



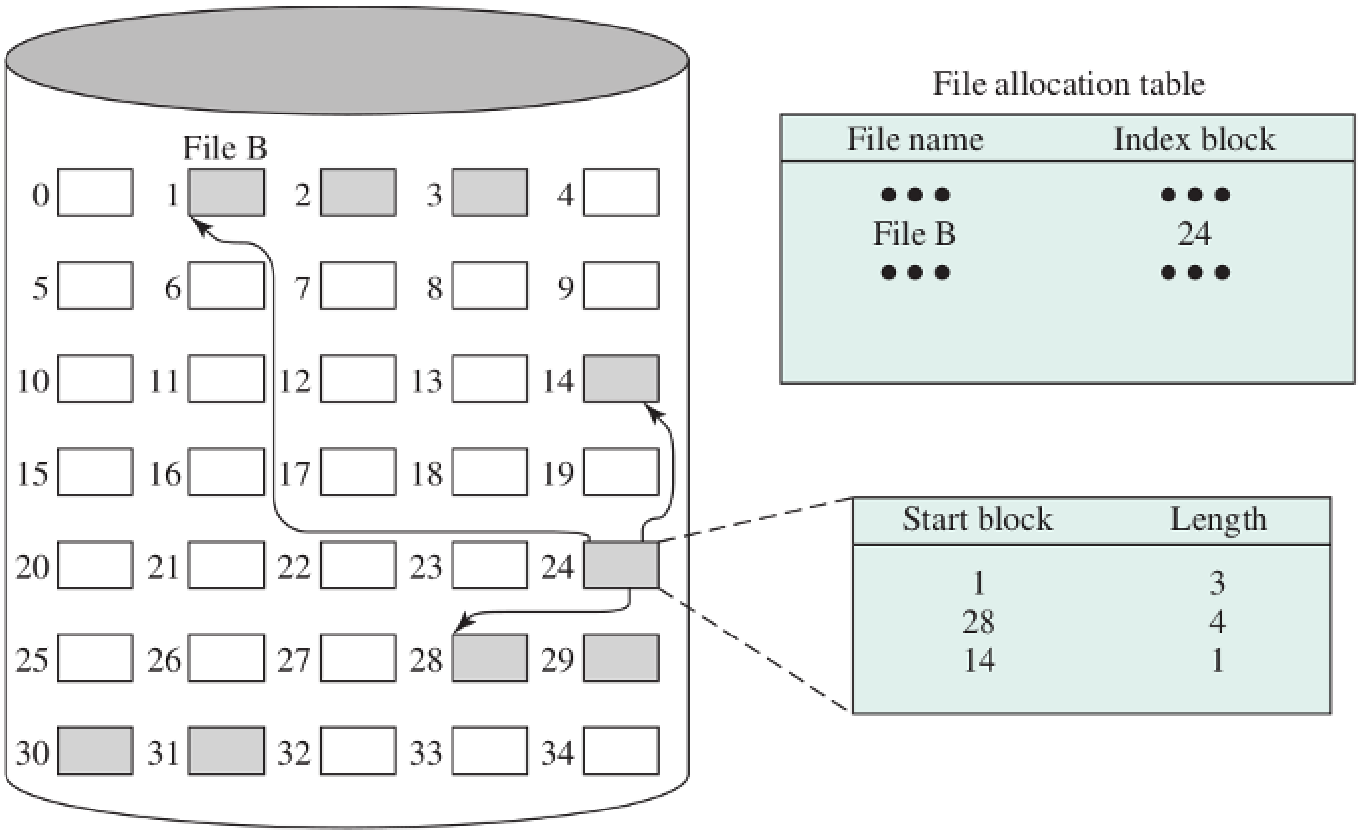
FAT Structure



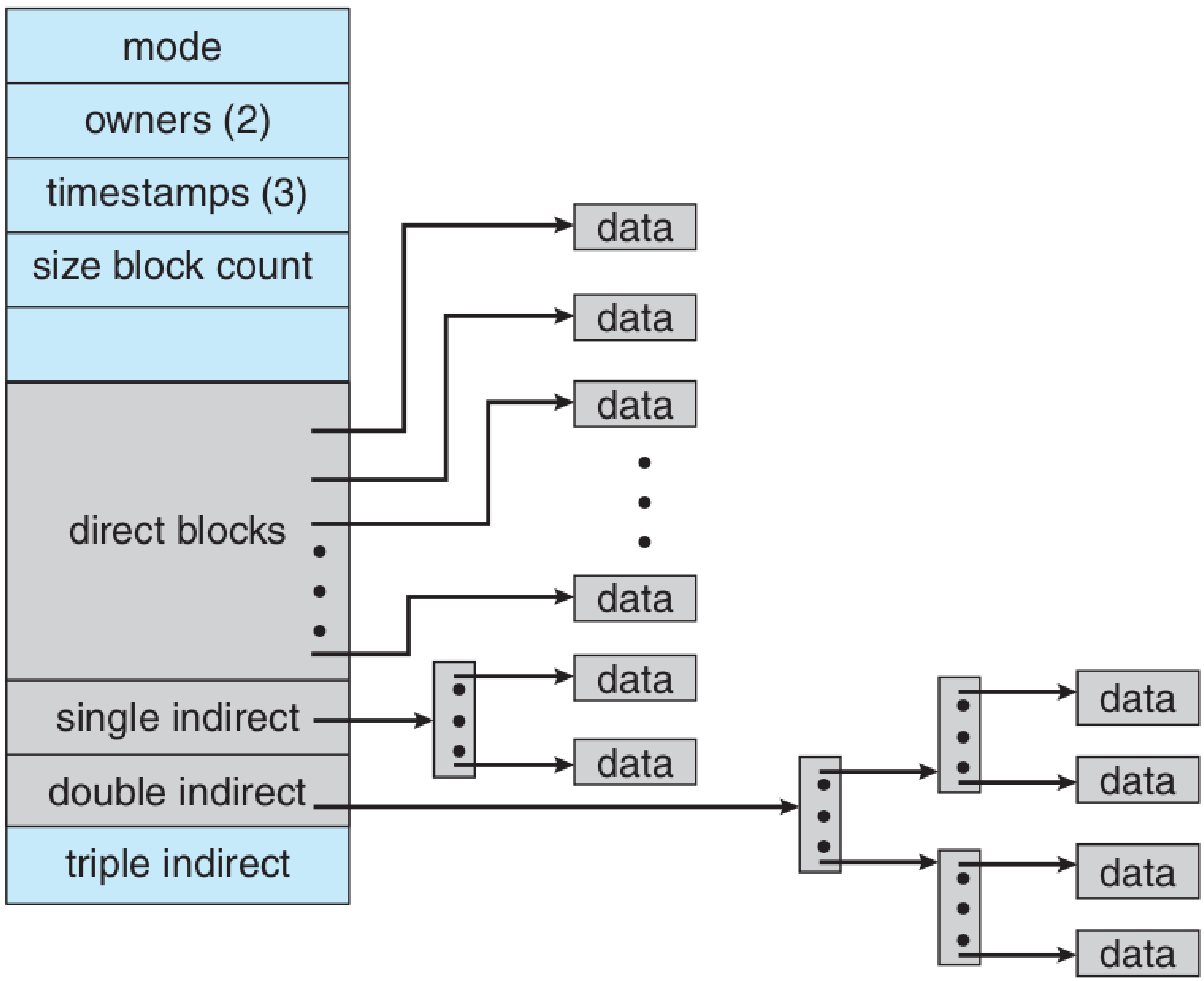
Index Based Block Allocation



Contiguous Index Based Block Allocation



File Structure in Unix



RAM in DISK

Results of performance test

HDD vs. SSD vs. RAM upgrade

	HDD, 2GB DDR3	HDD, 8GB DDR3	SSD, 2GB DDR3	SSD, 8GB DDR3
boot-up	01:11:03	00:52:95 **+25%**	00:32:36 **+54%**	00:30:19 **+57%**
Photoshop	00:15:71	00:14:58 **+2%**	00:04:07 **+72%**	00:03:97 **+73%**
1GB file	00:53:35	00:33:30 **+37%**	00:21:45 **+59%**	00:19:45 **+63%**

System tested: MacBook Pro (15-inch, Late 1008) 2,4 GHz Intel Core 2 Duo, NVIDIA GeForce 9400M, 2GB 1066MHz DDR3 SDRAM, 250GB 5400-rpm HDD (filled to 78%, about 250,000 folders and 933,000 files), Mac OS X 10.6.8 Snow Leopard.

Percentage values (rounded) refer to the performance boost compared to a factory built MacBook Pro 15 inch, late 2008). The colored label indicates the highest boost in this test category.

File Allocation Methods Comparison

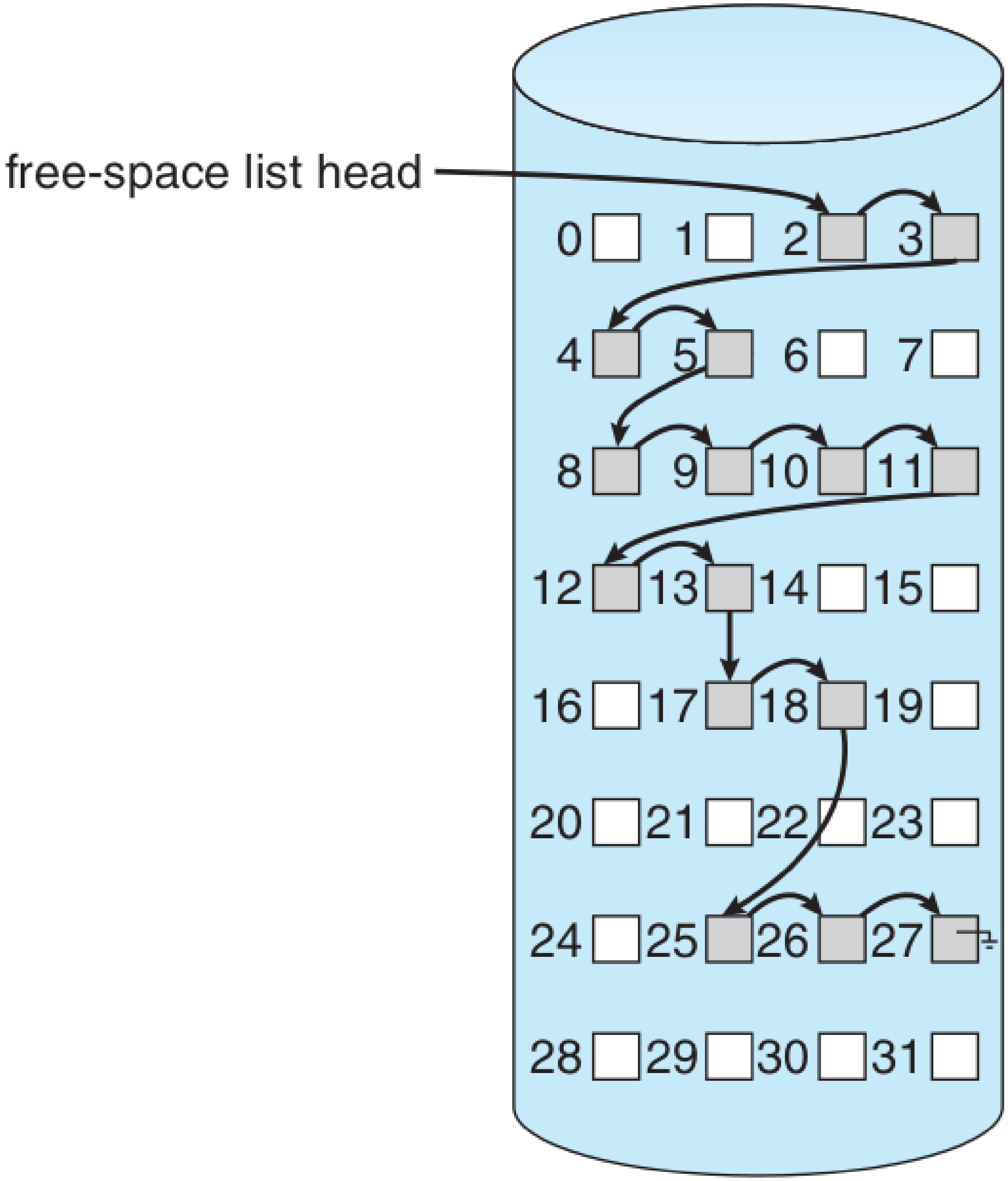
Method	Contiguous	Chained	Indexed	
Preallocation?	Necessary	Possible	Possible	
Fixed or Variable Size Portions?	Variable	Fixed blocks	Fixed blocks	Variable
Portion Size	Large	Small	Small	Medium
Allocation Frequency	Once	Low To High	High	Low
Time to Allocate	Medium	Long	Short	Medium
File Allocation Table Size	One Entry	One Entry	Large	Meduim

Free Space Management

- How do we keep track free blocks on a disk?
- A free-list is maintained. When a new block is requested, we search this list to find one.

- The following are commonly used techniques:
  - Bit Vector
  - Linked List
  - Linked List + Grouping
  - Linked List+Address+Count

List Based



free-space list head
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Bit Vector

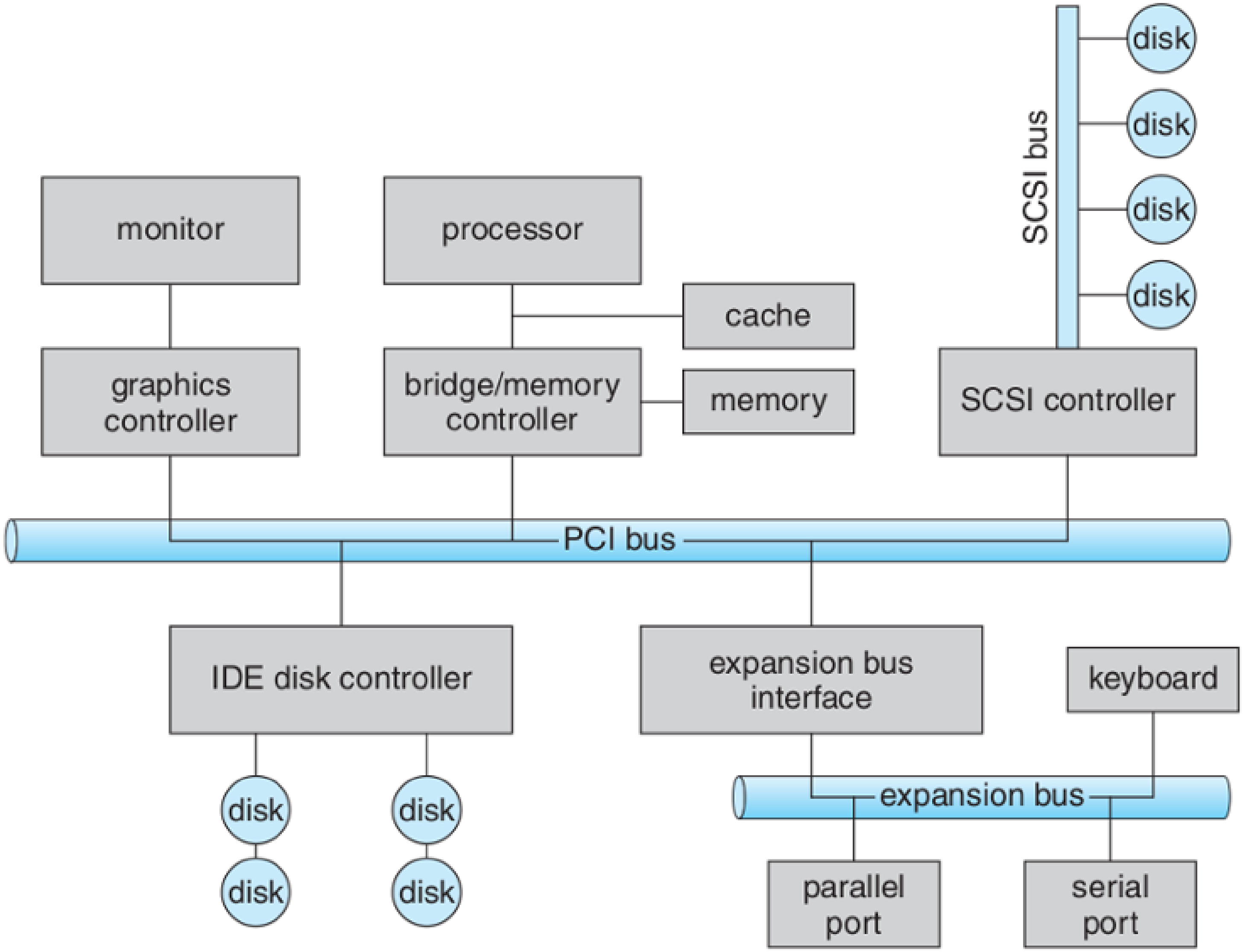
- Each block is represented by a bit in a table. Thus, if there are " n " disk blocks, the table has " n " bits.
- If a block is free, its corresponding bit is 1.
- When a block is needed, the table is searched. If a 1 bit is found in position " k ", block " k " is free.
- If the disk capacity is small, the whole bit vector can be stored in memory. For a large disk, this bit vector will consume too much memory.
- We could group a few blocks into a clusterand allocate " clusters ". This saves space and may cause internal fragmentation.
- Another possibility is the use of a " summary table ".

Input Output Structure

. .. :

https://en.wikipedia.org/wiki/I/O_scheduling
https://computationstructures.org/lectures/interrupts/interrupts.html

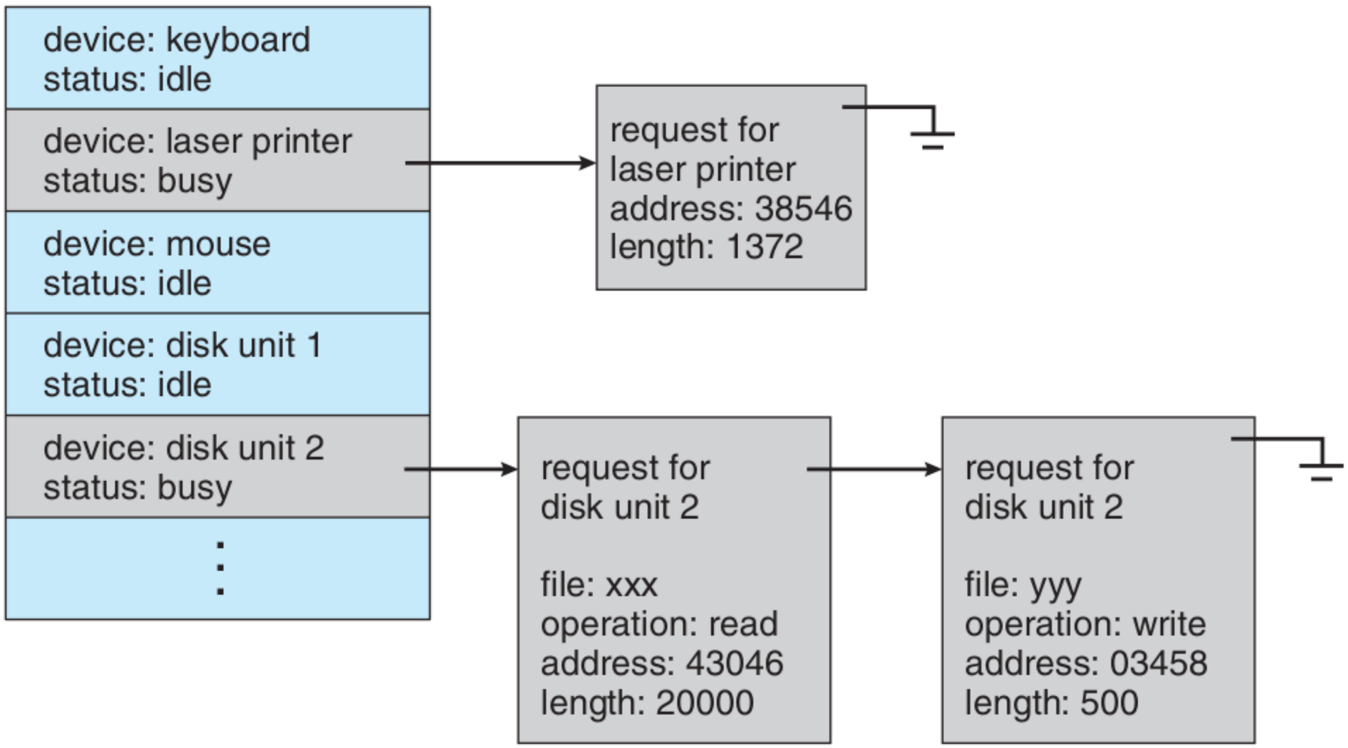
Computer Bus



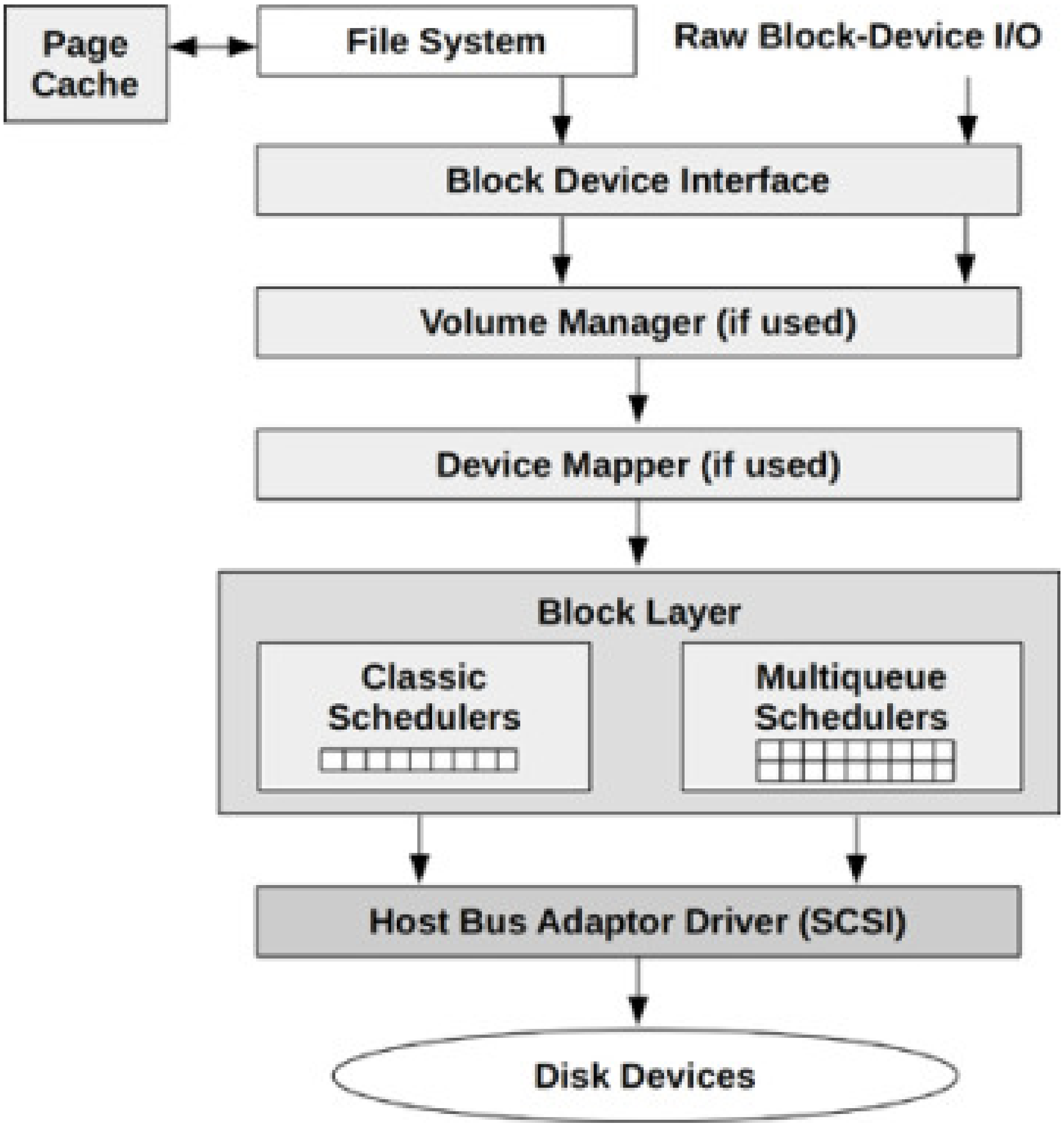
Hardware Ports

I/O address range (hexadecimal)	device
000–00F	DMA controller
020–021	interrupt controller
040–043	timer
200–20F	game controller
2F8–2FF	serial port (secondary)
320–32F	hard-disk controller
378–37F	parallel port
3D0–3DF	graphics controller
3F0–3F7	diskette-drive controller
3F8–3FF	serial port (primary)

Status of Disk Requests

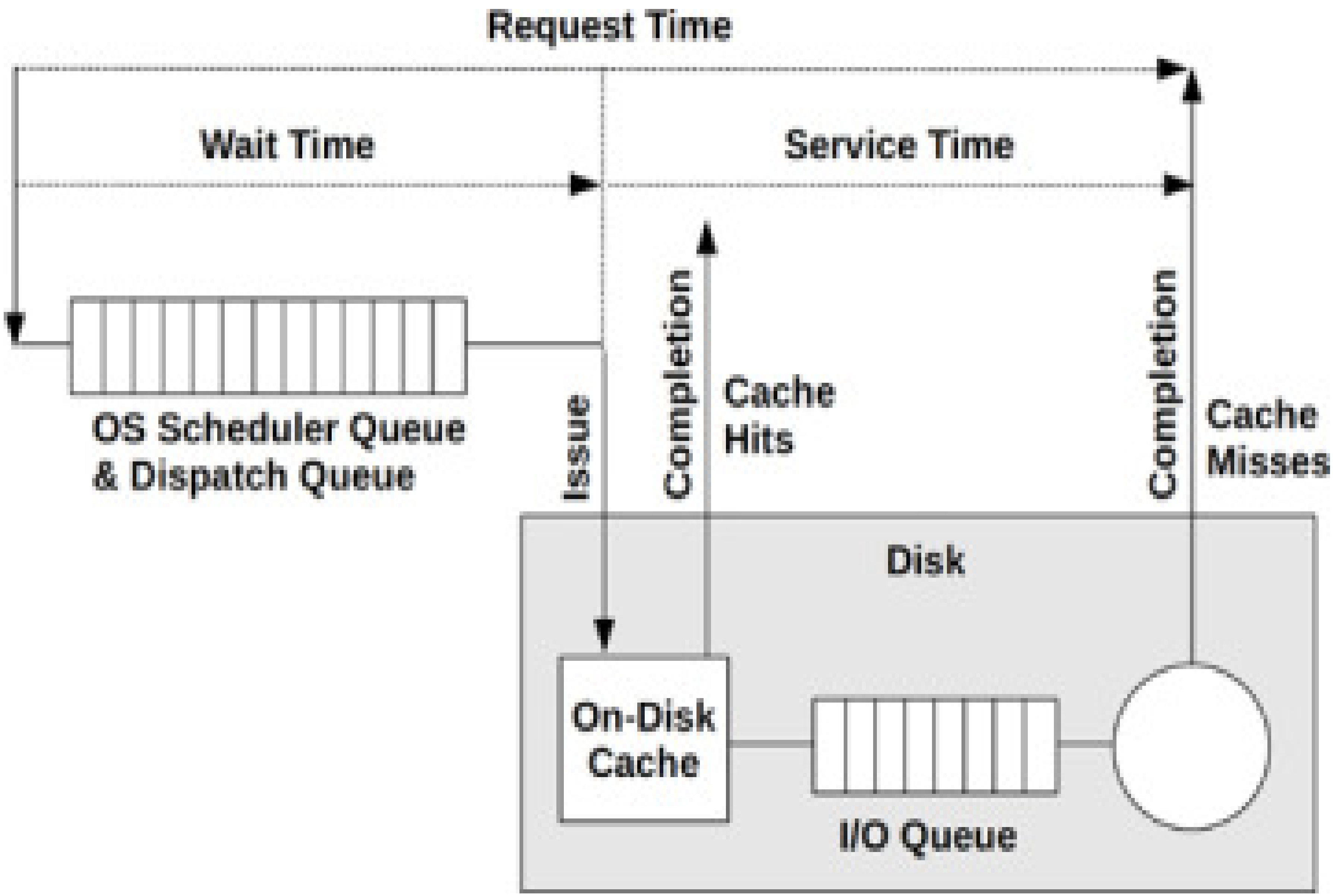


Linux Block I/O Stack



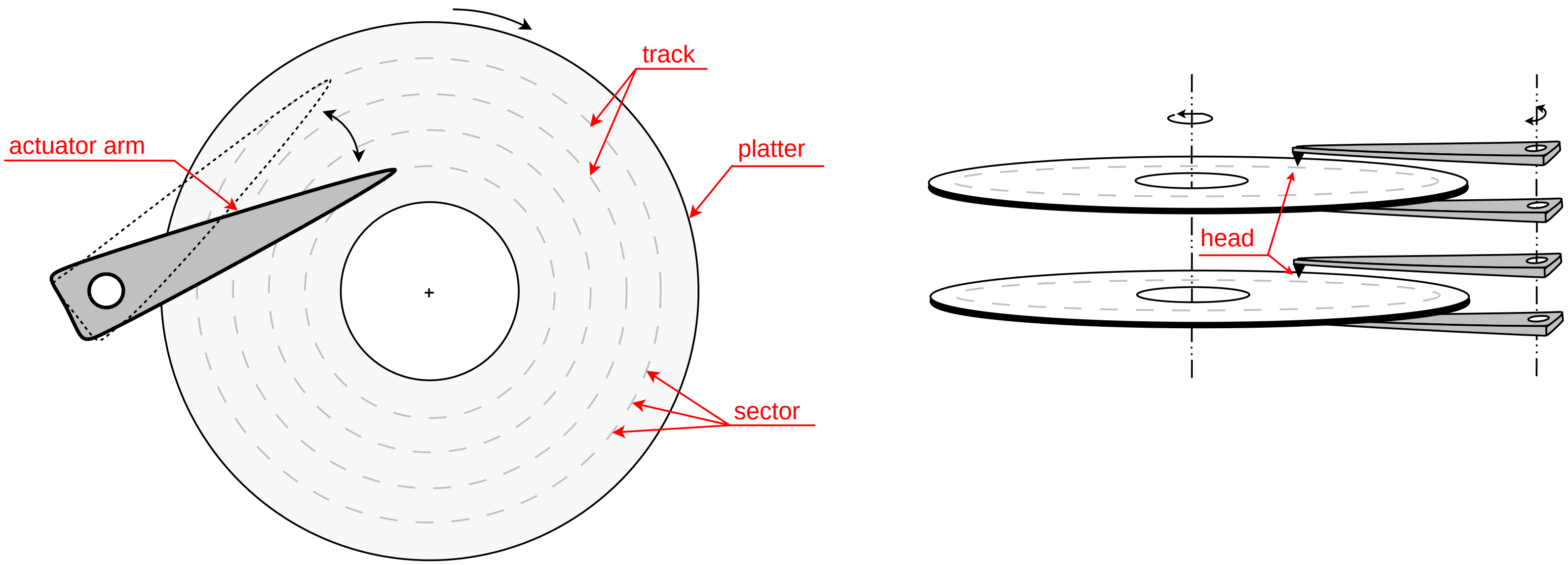
[informit]()

Disk I/O OS Terminology



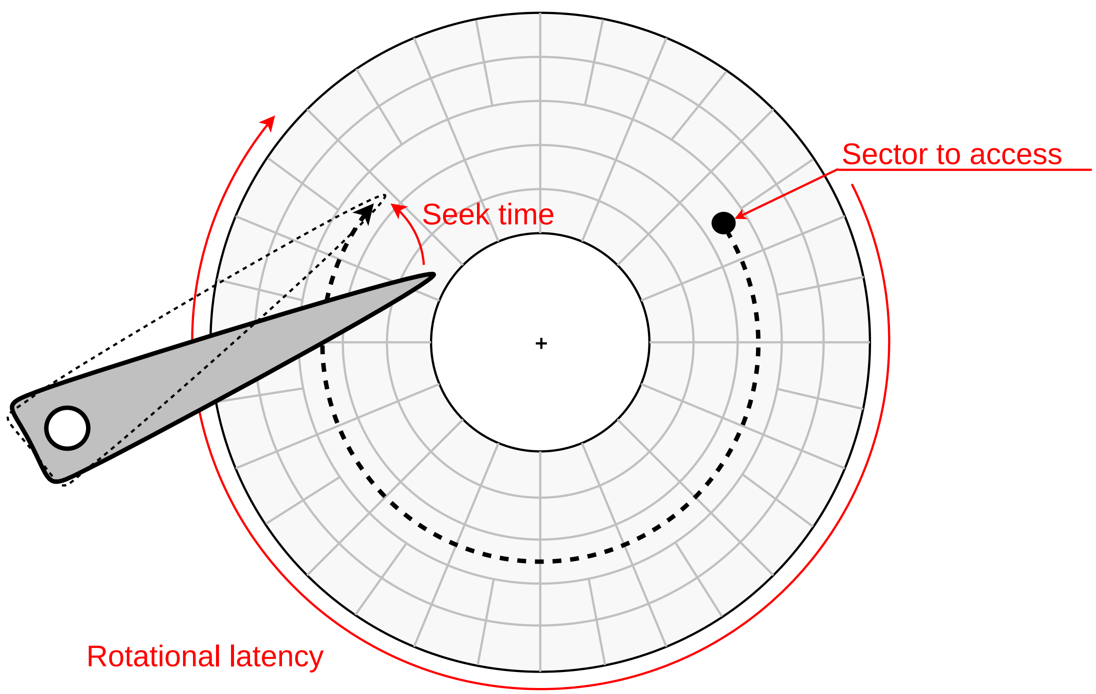
informit

Disk Arm Scheduling Algorithms



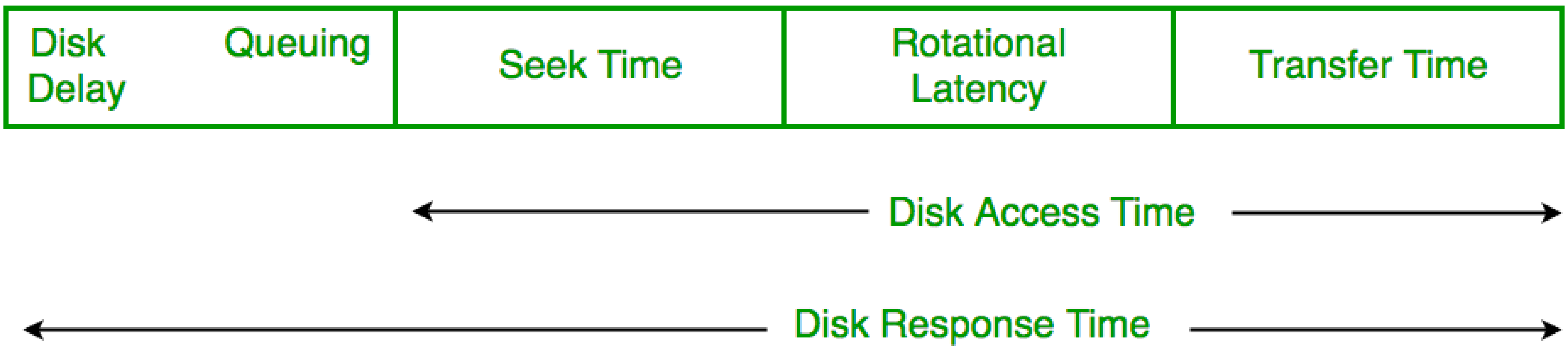
http://gudok.xyz/sspar/

Operations performed during random access



http://gudok.xyz/sspar/

Important Disk Parameters



- Seek Time
- Rotational Latency
- Transfer Time
- Disk Access Time
- Disk Response Time

Imporatnce of Track or Cylinder

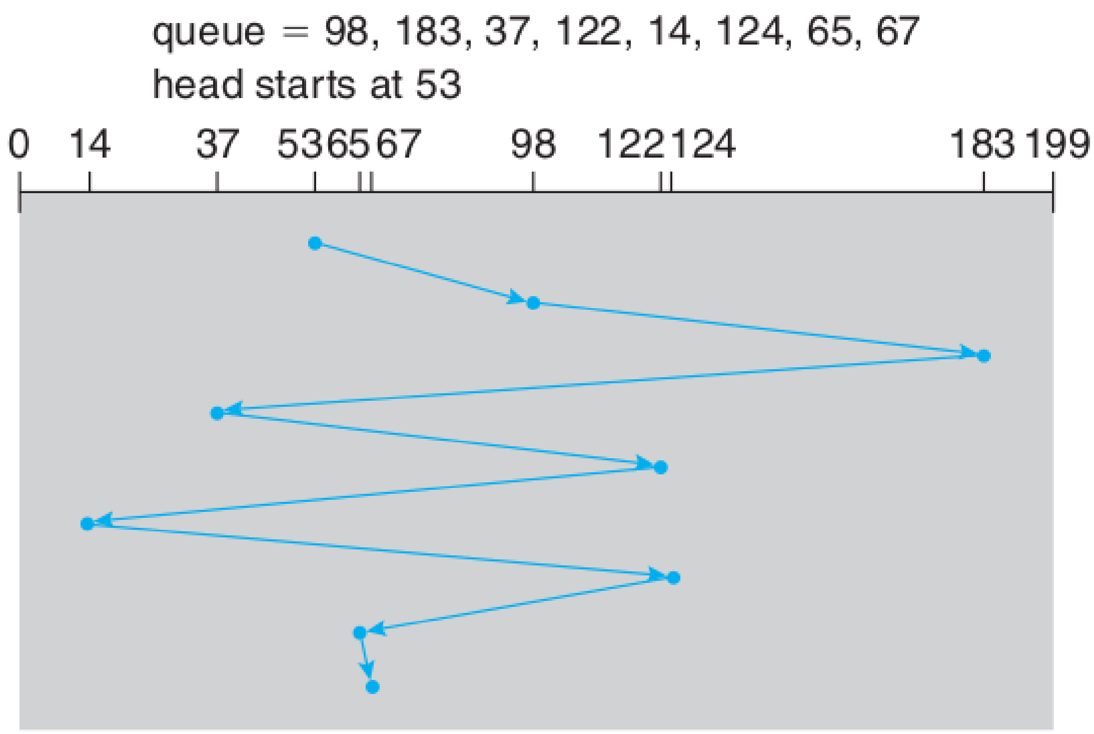
زمان پیگیرد یا زمان جستجو

Seek time

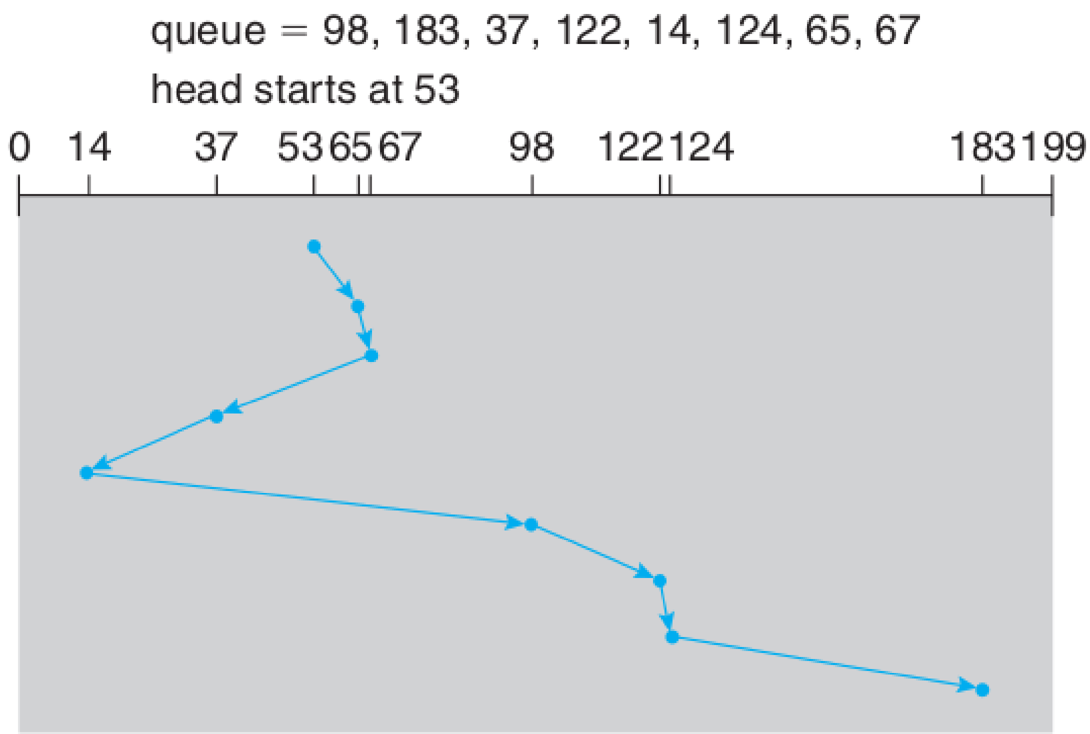
- درخواست‌های فرآیندها به بخش‌های مختلف دیسک اشاره می‌کند که مهم‌ترین پارامتر رفتن به سیلندر مورد نظر است.
- پس سیستم عامل در پایین‌ترین سطح کار خود با دیسک دنباله‌ای از درخواست‌های به سیلندرهای گوناگون را زمان‌بندی می‌کند.

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

FCFS


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

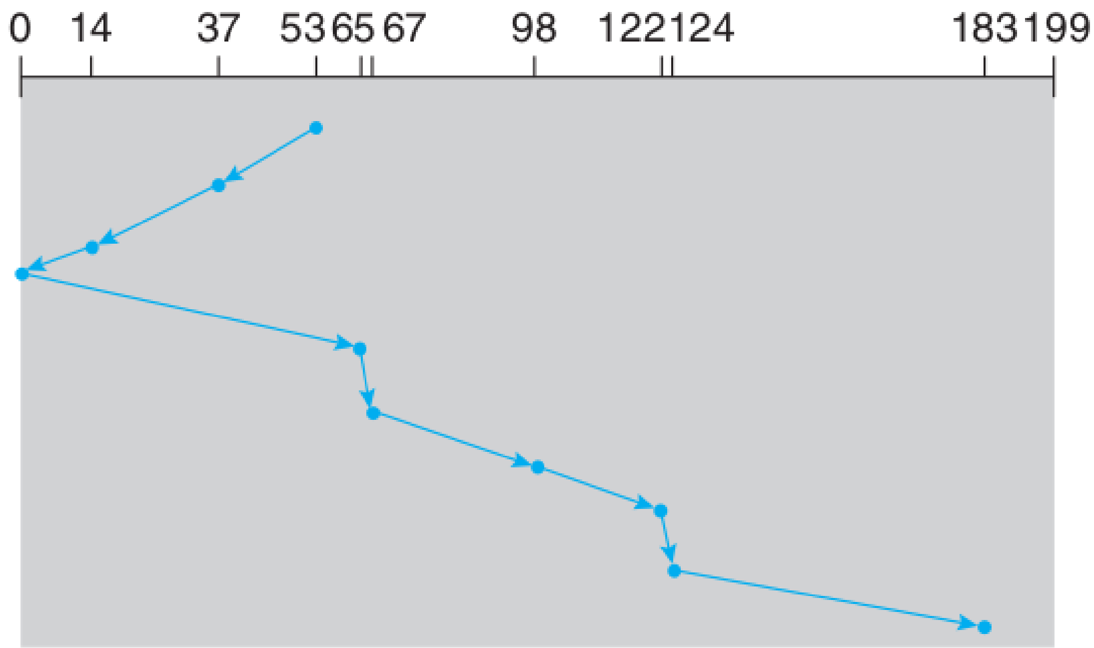
SSTF


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

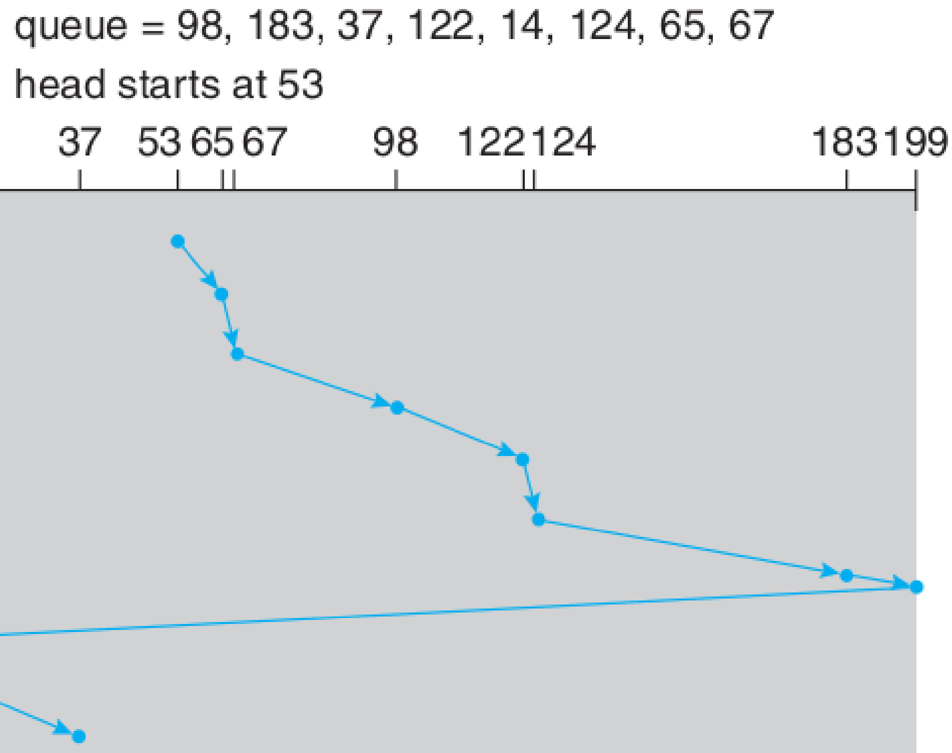
SCAN

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

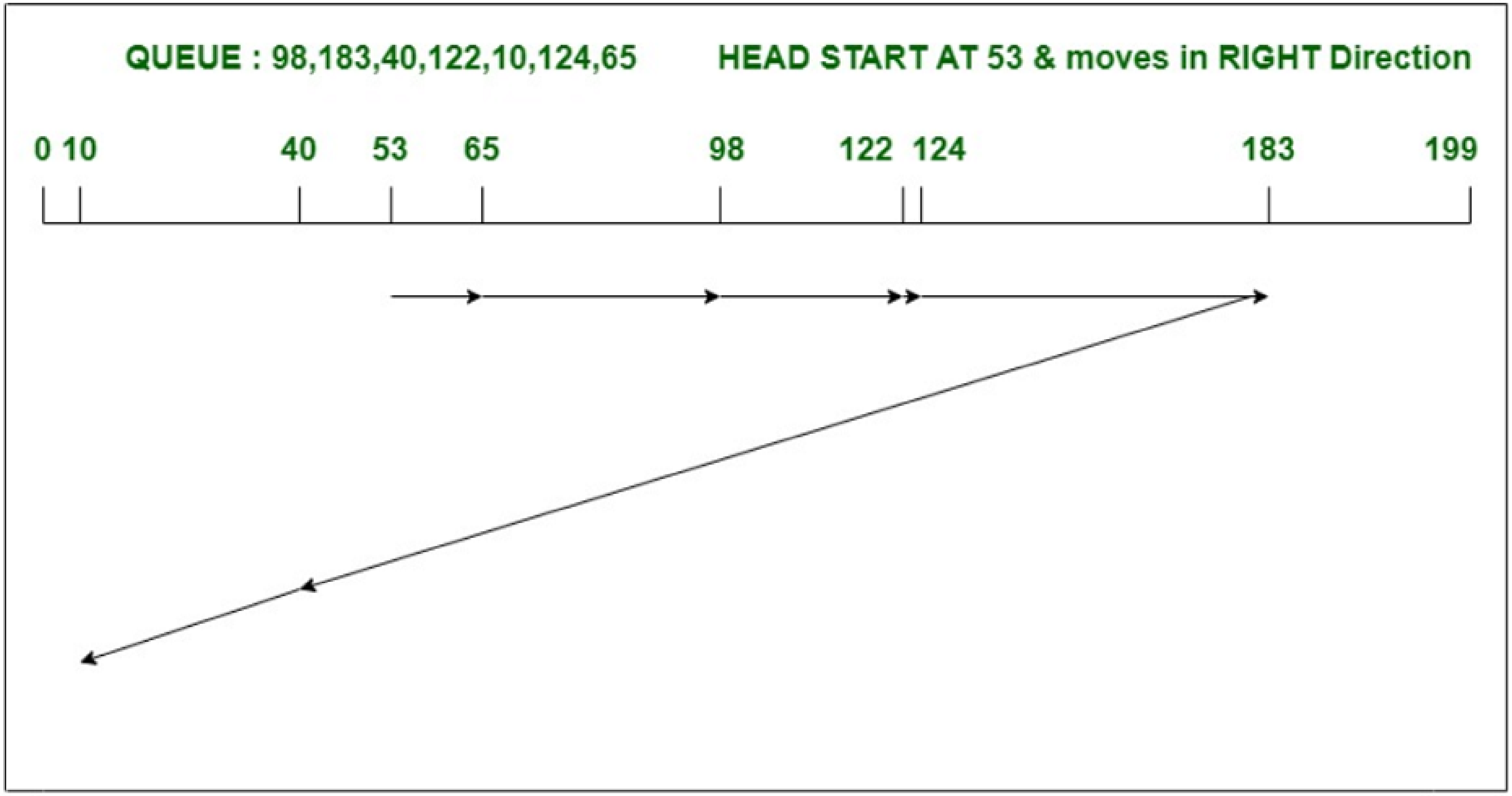
head starts at 53



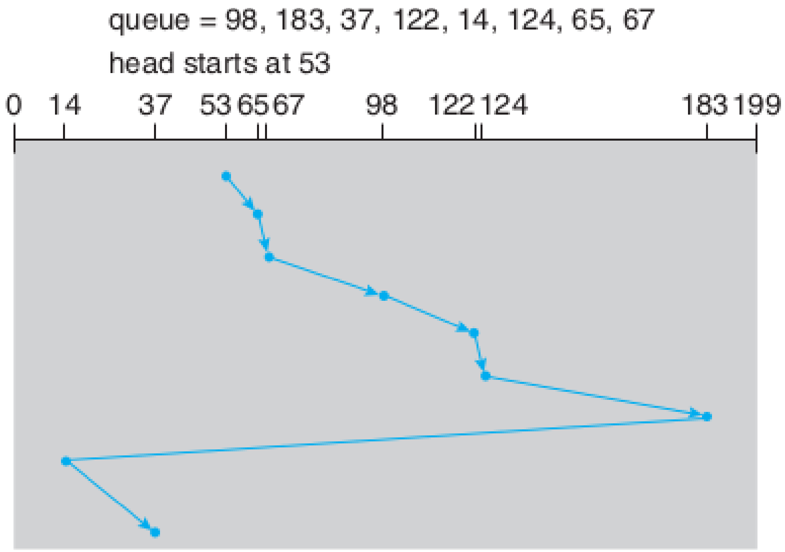
C-SCAN


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

LOOK


QUEUE : 98,183,40,122,10,124,65
HEAD START AT 53 & moves in RIGHT Direction
0
10
40
53
65
98
122
124
183
199

C-LOOK


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

F-SCAN

.

N-Step Scan

.

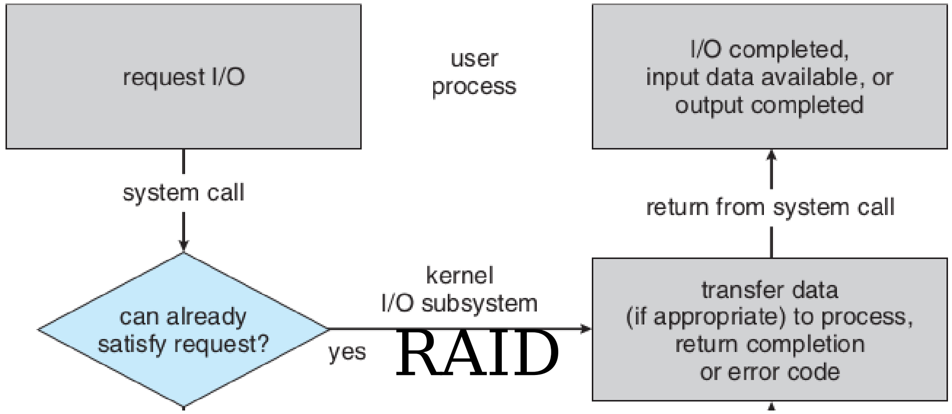
چند الگوریتم زمان‌بندی دیسک

Name	Description	Remarks
Selection according to requestor		
Random	Random scheduling	For analysis and simulation
FIFO	First in first out	Fairest of them all
PRI	Priority by process	Control outside of disk queue management
LIFO	Last in first out	Maximize locality and resource utilization
Selection according to requested item		
SSTF	Shortest-service-time first	High utilization, small queues
SCAN	Back and forth over disk	Better service distribution
C-SCAN	One way with fast return	Lower service variability
N-step-SCAN	SCAN of *N* records at a time	Service guarantee
FSCAN	*N*-step-SCAN with *N* = queue size at beginning of SCAN cycle	Load sensitive

نمونه‌ای از مقایسهٔ چند الگوریتم

(a) FIFO (starting at track 100)		**(b) SSTF** (starting at track 100)		**(c) SCAN** (starting at track 100, in the direction of increasing track number)		**(d) C-SCAN** (starting at track 100, in the direction of increasing track number)	
Next track accessed	**Number of tracks traversed**	**Next track accessed**	**Number of tracks traversed**	**Next track accessed**	**Number of tracks traversed**	**Next track accessed**	**Number of tracks traversed**
55	45	90	10	150	50	150	50
58	3	58	32	160	10	160	10
39	19	55	3	184	24	184	24
18	21	39	16	90	94	18	166
90	72	38	1	58	32	38	20
160	70	18	20	55	3	39	1
150	10	150	132	39	16	55	16
38	112	160	10	38	1	58	3
184	146	184	24	18	20	90	32
Average seek length	55.3	**Average seek length**	27.5	**Average seek length**	27.8	**Average seek length**	35.8

انجام ورودی/خروجی‌ها



RAID

Table 11.4 RAID Levels

Category	Level	Description	Disks Required	Data Availability	Large I/O Data Transfer Capacity	Small I/O Request Rate
Striping	0	Nonredundant	N	Lower than single disk	Very high	Very high for both read and write
Mirroring	1	Mirrored	$2N$	Higher than RAID 2, 3, 4, or 5; lower than RAID 6	Higher than single disk for read; similar to single disk for write	Up to twice that of a single disk for read; similar to single disk for write
Parallel access	2	Redundant via Hamming code	$N+m$	Much higher than single disk; comparable to RAID 3, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
	3	Bit-interleaved parity	$N+1$	Much higher than single disk; comparable to RAID 2, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
Independent access	4	Block-interleaved parity	$N+1$	Much higher than single disk; comparable to RAID 2, 3, or 5	Similar to RAID 0 for read; significantly lower than single disk for write	Similar to RAID 0 for read; significantly lower than single disk for write
	5	Block-interleaved distributed parity	$N+1$	Much higher than single disk; comparable to RAID 2, 3, or 4	Similar to RAID 0 for read; lower than single disk for write	Similar to RAID 0 for read; generally lower than single disk for write
	6	Block-interleaved dual distributed parity	$N+2$	Highest of all listed alternatives	Similar to RAID 0 for read; lower than RAID 5 for write	Similar to RAID 0 for read; significantly lower than RAID 5 for write

Note: N, number of data disks; m, proportional to log N.