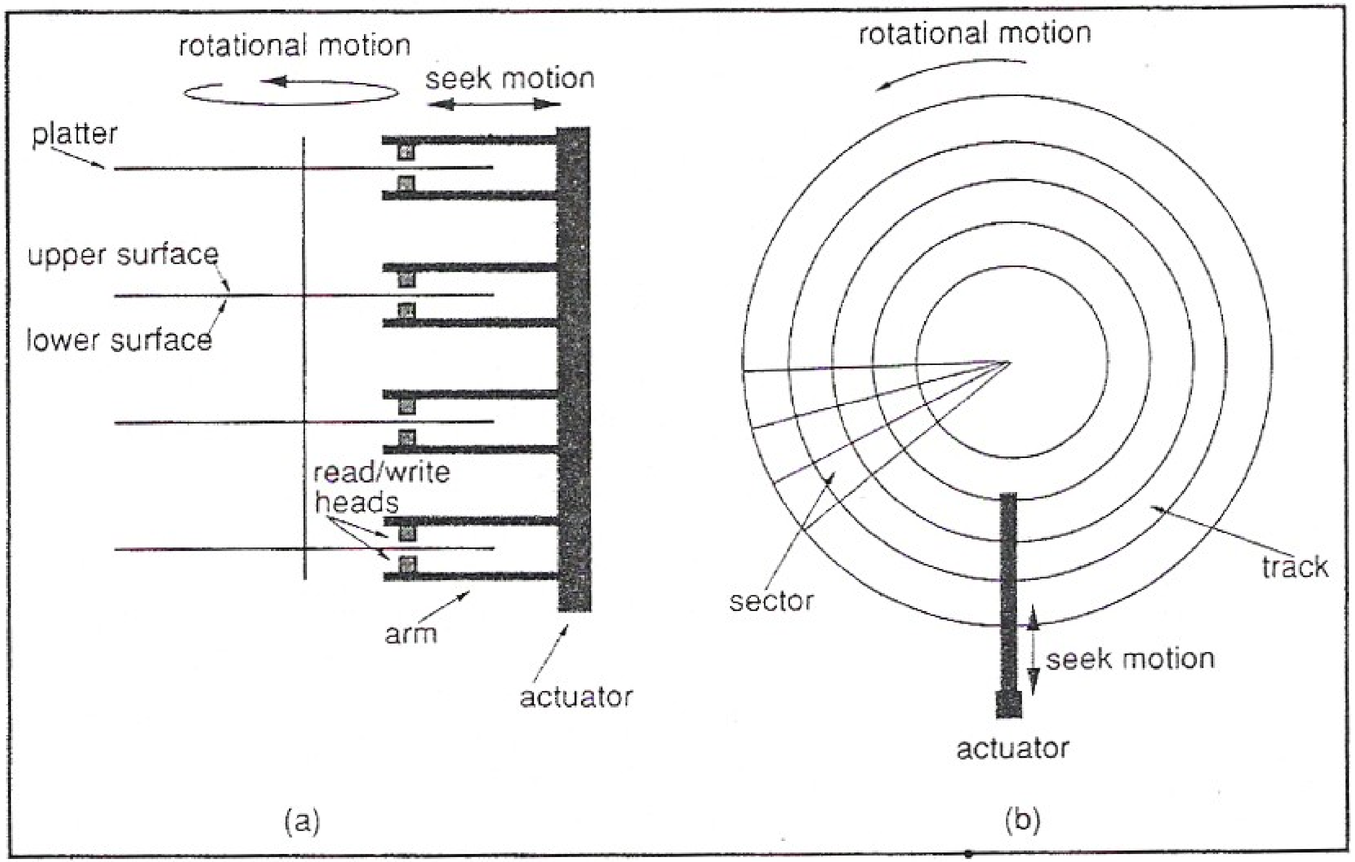
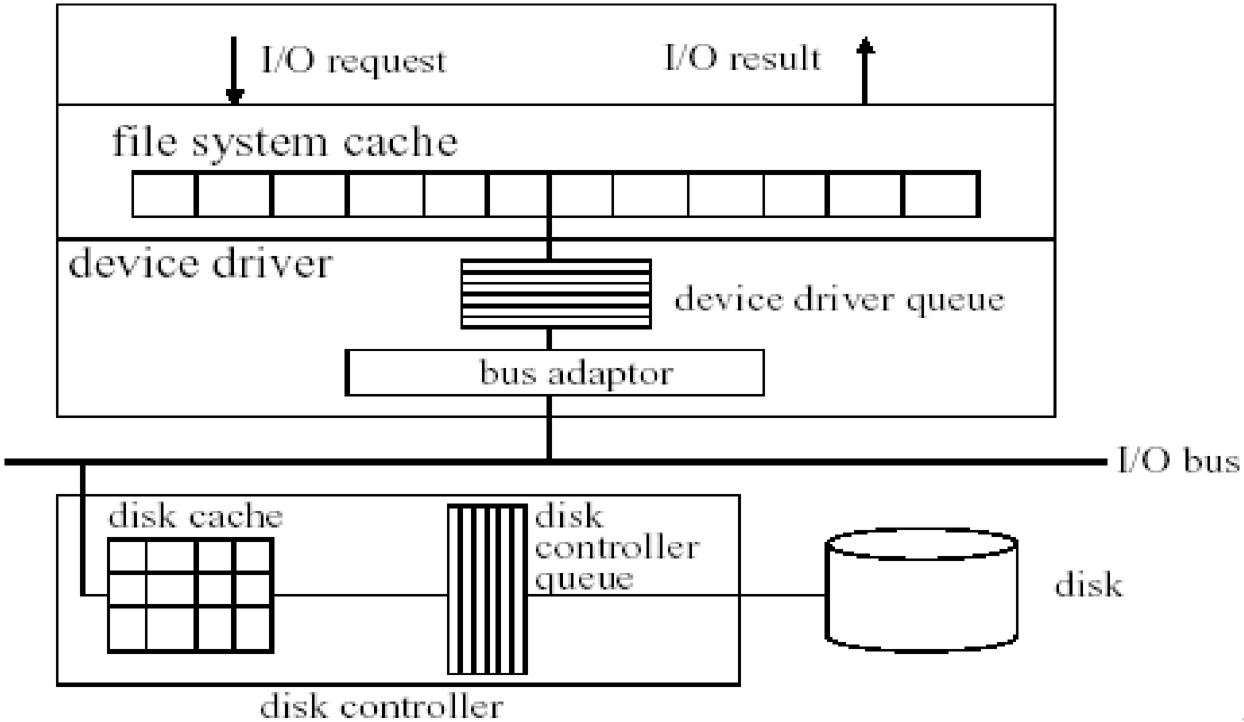

SAMSUNG
28171

Single Disk



(a) Side view of a magnetic disk.
(b) Top view of a magnetic disk.

Computing Disk Service Time



Architecture of a typical I/O subsystem

Nomenclature

- $\overline{S}_d$: average time, in seconds, spent at the controller plus disk to access a block from a disk
- SeekTime: average seek time in seconds. That is, the average time to position the arm at the proper cylinder
- $\text{Seek}_{\text{rand}}$: average seek time, in seconds, for a request to a random cylinder, provided by the disk manufacturers. Sometimes, the average seek times for read and write requests are slightly different
- DiskSpeed: disk rotation speed, in revolutions per minute, provided by the disk manufacturer
- DiskRevolutionTime: time, in seconds, for a disk to complete a full revolution, equal to 60/DiskSpeed
- RotationalLatency: average rotational latency in seconds; i.e., the average time elapsed since the end of the seek until data transfer starts. This time is spent waiting for the disk to rotate until the desired sector lies under the read/write head
- BlockSize: block size in bytes
- TransferRate: rate at which data is transferred to/from a disk, in MB/sec
- TransferTime: time, in seconds, to transfer a block from the disk to the disk controller
- ControllerTime: time, in seconds, spent at the controller processing an I/O request. This includes the time to check the cache plus the time to read/write a block from/to the cache
- P_{miss}: probability that the desired block is not in the disk cache

Computing Disk Service Time

$$s_d = ContrTime + P_{miss}(Seek + Latency + TransferT)$$

$$TransferT = \frac{BlockSize}{TransferRate}$$

Computing Disk Service Time

Types of workloads

Random Workload:
10, 201, 15, 1023, 45, 39, 782

Sequential Workload:

$$4,\ \underbrace{102,\ 103,\ 104,\ 105,\ 106}_{\text{run length}= 5},\ 25,\ 88,\ \underbrace{32,\ 33,\ 34,\ 35,\ 36,\ 37,\ 38}_{\text{run length}= 7},\ 29,\ 15$$

Computing Disk Service Time

Random Workload:

$$P_{miss} = 1$$

$$RunLength = 1$$

$$SeekTime = S_{rand}$$

$$Latency = 1/2 \times RevolutionTime$$

Computing Disk Service Time

Sequential Workload:

$$P_{miss} = 1 / RunLength$$

$$\mathrm{SeekTime} = \mathrm{Seek}_{\mathrm{rand}}/\mathrm{RunLength}$$

$$Latency = \frac{1/2 + (RunLength - 1)[(1 + U_d)/2]}{RunLength} \times$$

$$\mathrm{Re}\,volutionTime$$

Computing Disk Service Time

$$\text{RunLength} = \text{NumberRequests}/\text{NumRuns}$$

$$\begin{aligned}\text{SeekTime} &= \frac{\text{NumRuns} \times \text{Seek}_{\text{rand}}}{\text{NumberRequests}} \\ &= \frac{\text{Seek}_{\text{rand}}}{\text{NumberRequests}/\text{NumRuns}} \\ &= \text{Seek}_{\text{rand}}/\text{RunLength}.\end{aligned}$$

Disk service times equations

$$\text{TransferTime} = \frac{\text{BlockSize}}{10^6 \times \text{TransferRate}} \tag{3.3.19}$$

Random workloads:

$$\overline{S}_d = \text{ControllerTime} + \text{Seek}_{\text{rand}} + \frac{\text{DiskRevolutionTime}}{2} + \text{TransferTime} \tag{3.3.20}$$

Sequential workloads:

$$\begin{aligned}\overline{S}_d \;=\; & \text{ControllerTime} + \frac{\text{Seek}_{\text{rand}}}{\text{RunLength}} + \\ & \frac{[1/2 + (\text{RunLength} - 1)(1 + U_d)/2] \;\times\; \text{DiskRevolutionTime}}{\text{RunLength}} + \\ & \text{TransferTime}/\text{RunLength}\end{aligned} \tag{3.3.21}$$

Example: a DB Server

Arrival rate: 20 req/sec

Workload:
 20% random
 80% sequential
 avarege run length = 24

Block size: 2,048 bytes
Disk Revolution Time: 7,200 RPM
Seek Time: 9 msec
Transfer Rate: 20 MB/sec
Controller Time: 0.1 msec

$\Rightarrow$ Latency = ½ x 60/7,200 x 1,000 = 4.17 msec

$\Rightarrow$ TransferTime = 2,048/(10^6 x 20) x 1,000 = 0.1 msec

$\Rightarrow$ **Average Service Time (Random Request)** = 0.1+ 9 +4.17 + 0.1 =13.4 msec

U_d = ArrivalRate x (RandomSeek + AverageLatency + AverageTransferTime)

$\Rightarrow$ U_d = 0.020 x (9+4.17+0.1) = 0.27

$\Rightarrow$ P_{miss} = 1/24 = 4.2 %

$\Rightarrow$ AverageSeekTime = 9/24 = 0.38 msec

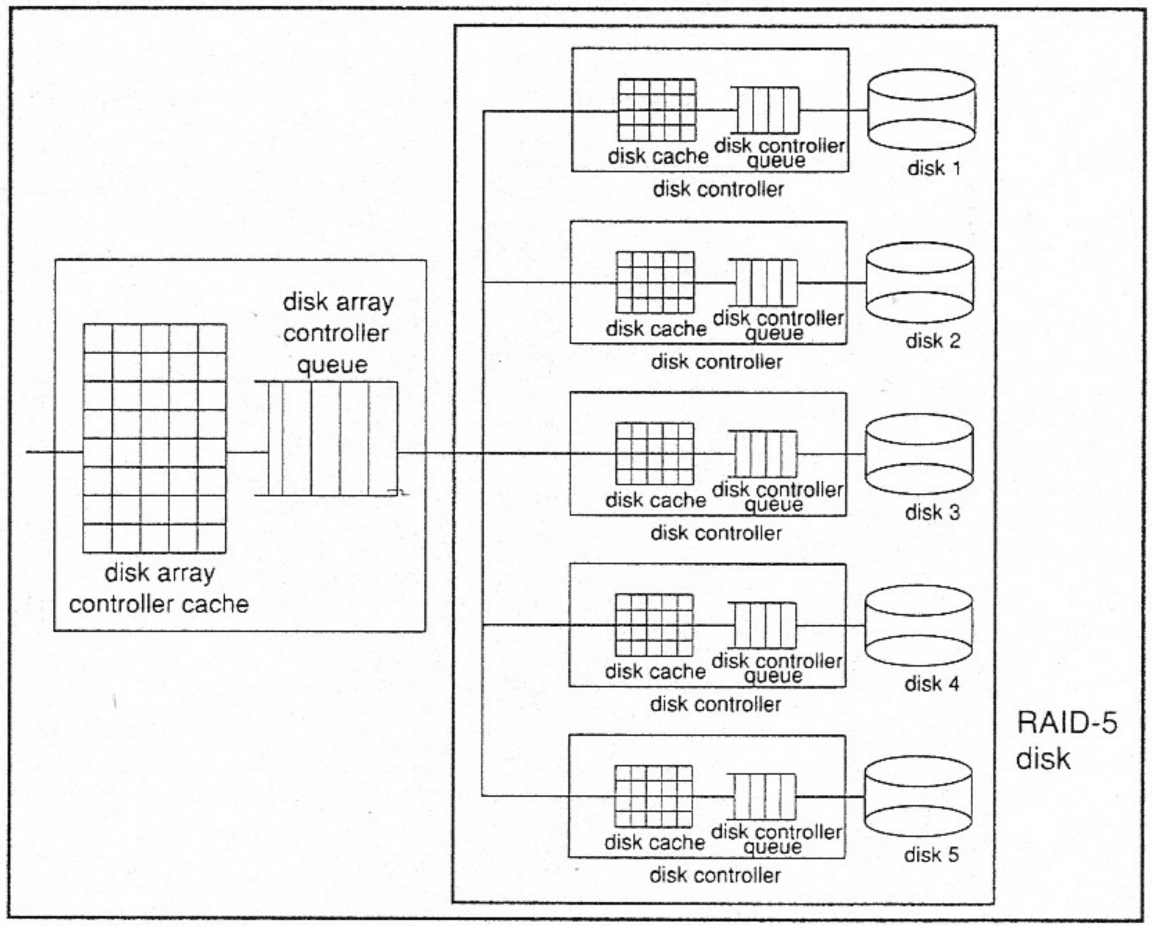
$\Rightarrow$ AverageRotationalLatency =

$$\frac{\frac{1}{2}+23(1+0.27)/2}{24} \times \frac{60 \times 1{,}000}{7{,}200} = 5.25 \text{ msec}$$

⇒ **Average Service Time (Sequential Request)** = 0.1+ 0.38 +5.25 + 0.042x0.1 =5.73 msec

⇒ **Average Service Time** = 7.25 msec

Disk Arrays

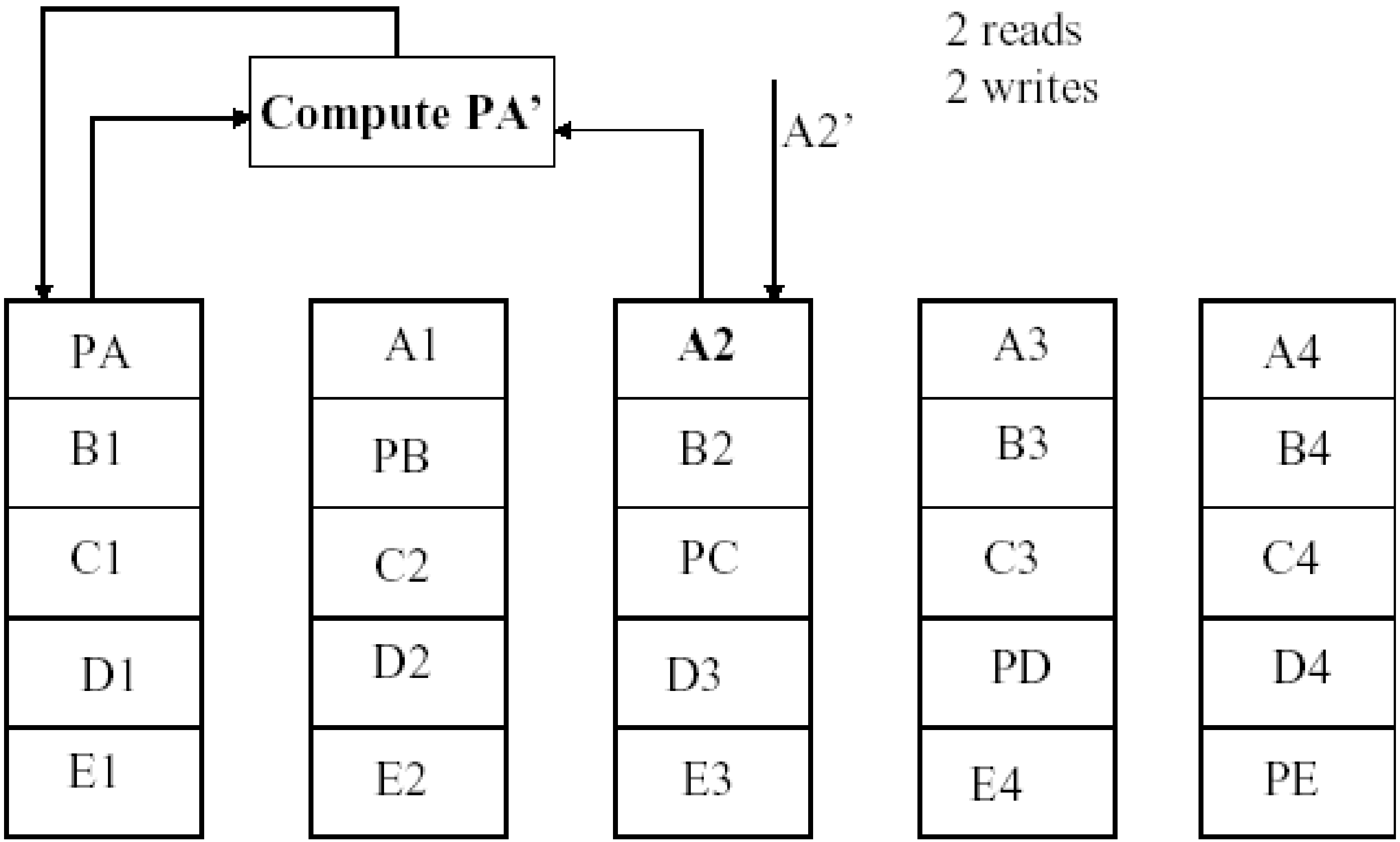


Disk array

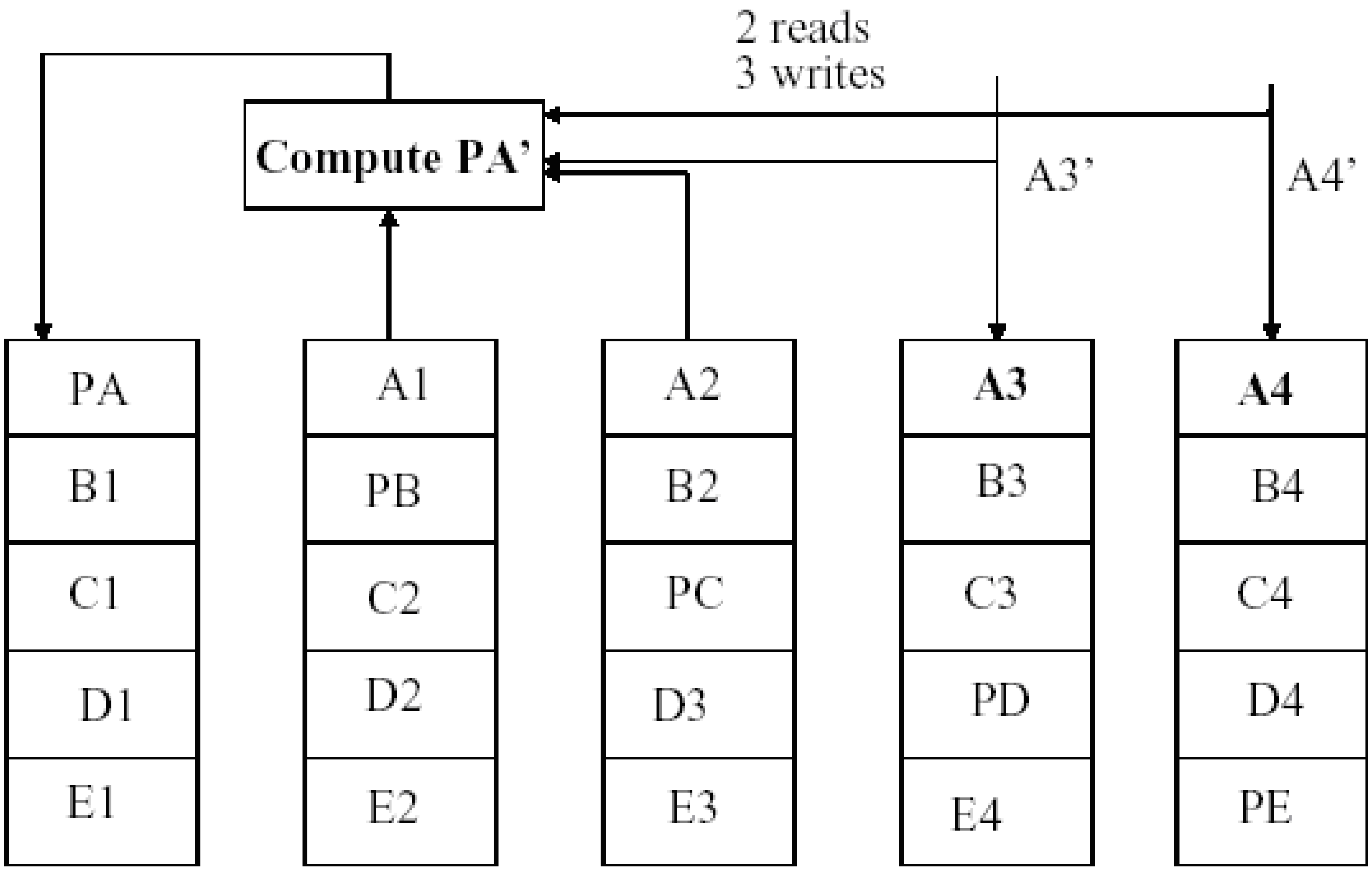
Disk Arrays

PA	A1	A2	A3	A4
B1	PB	B2	B3	B4
C1	C2	PC	C3	C4
D1	D2	D3	PD	D4
E1	E2	E3	E4	PE

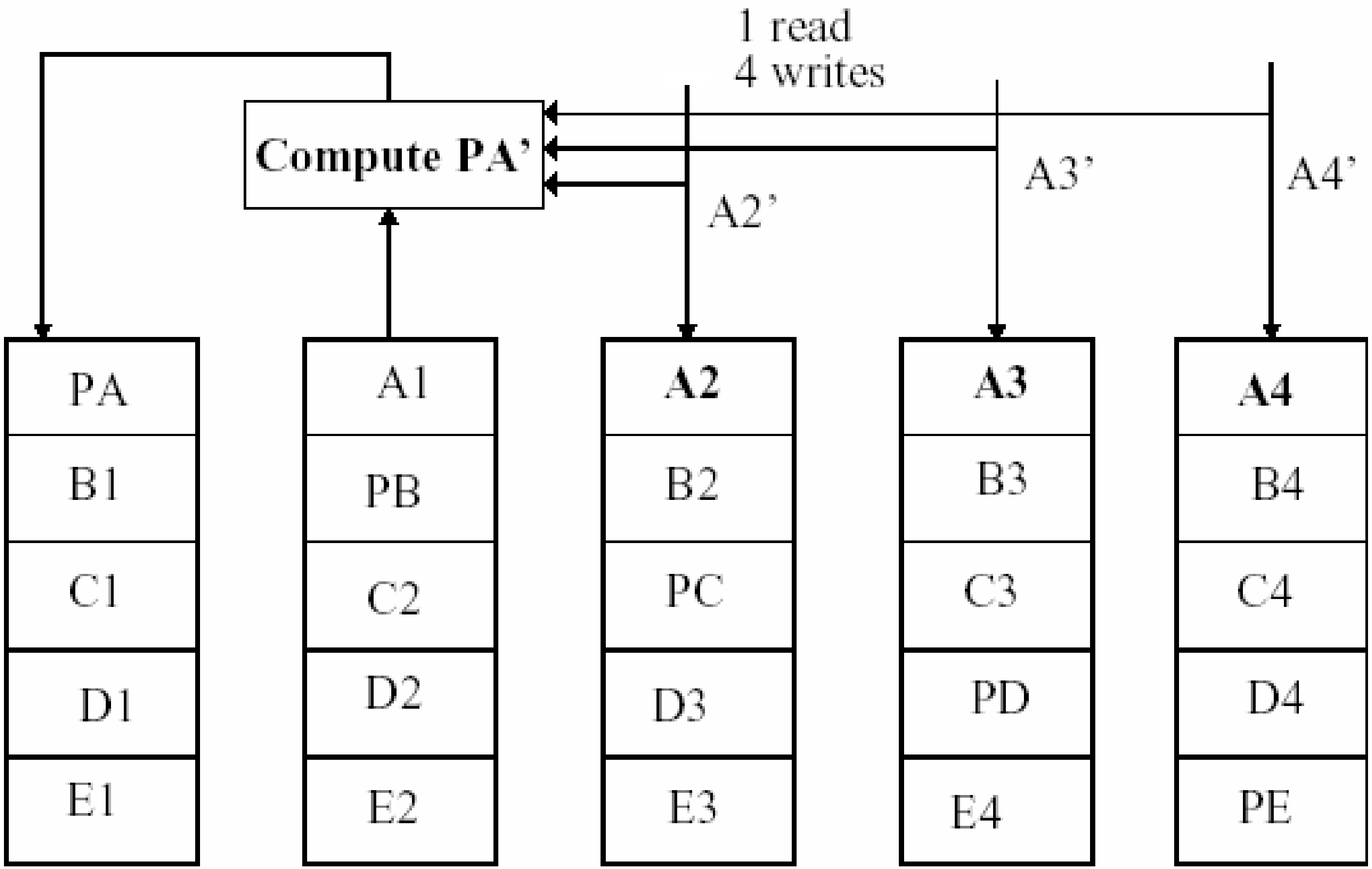
Disk Arrays - Write One Stripe Unit



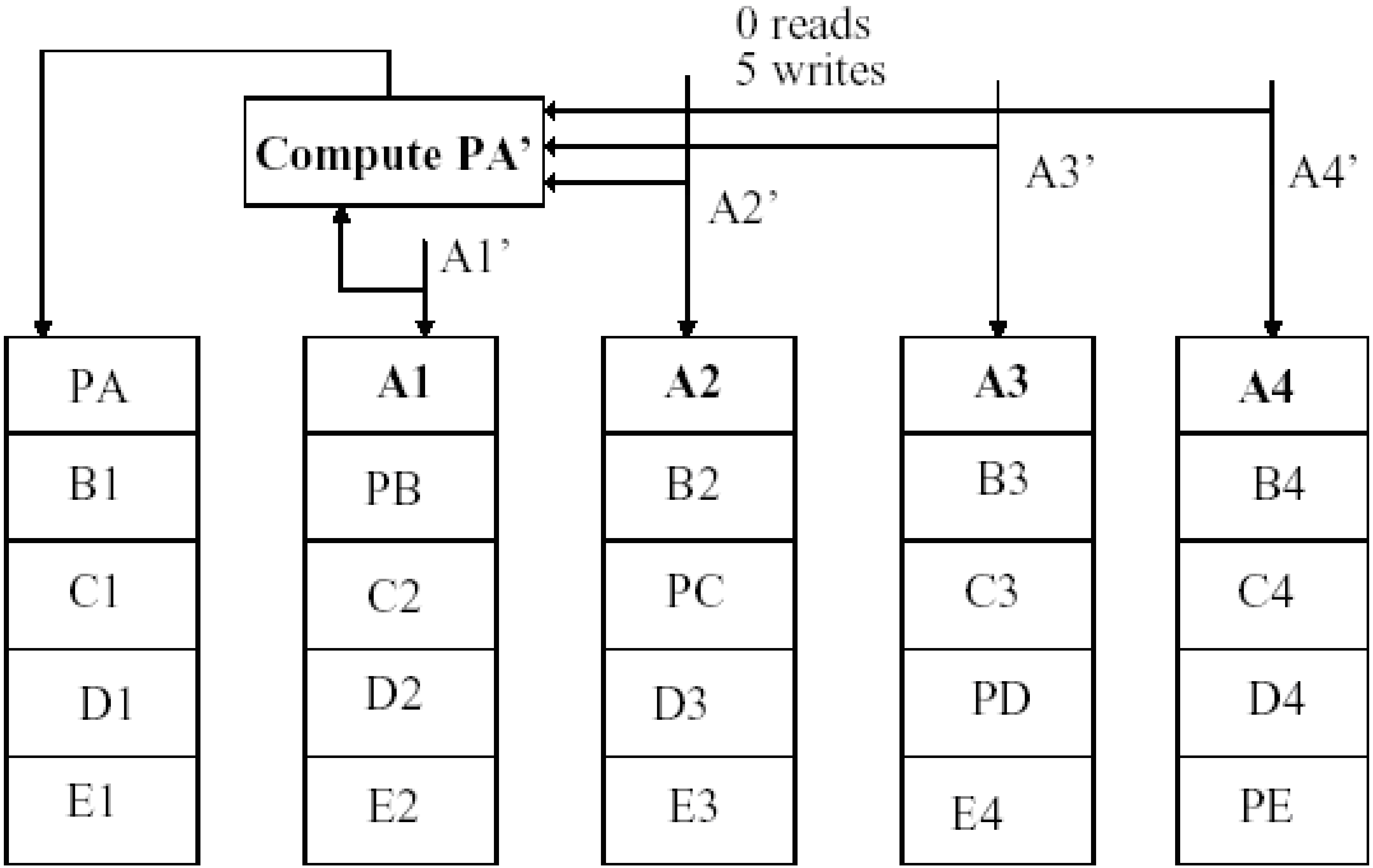
Disk Arrays - Write Two Stripe Units



Disk Arrays - Write Three Stripe Units



Disk Arrays - Write Four Stripe Units



Nomenclature

- n_r : number of stripe units reads by a read request
- n_w : number of stripe units modified by a write request
- λ^r_{array}: arrival rate of read requests to the disk array
- λ^w_{array}: arrival rate of write requests to the disk array
- λ^r_{disk}: arrival rate of read requests to any of the disk in the array
- λ^w_{disk}: arrival rate of write requests to any of the disk in the array
- N: number of physical disks in the array
- S^r_{array}: average service time at a disk array for read
- S^w_{array}: average service time at a disk array for write

- Fraction of read requests to the array that goes to any of the disks (uniform distribution to all disks):

$$\frac{n_r}{N}$$

- Arrival rate of disk requests at a disk:

$$\lambda_{disk}^{r} = \frac{n_r}{N} \cdot \lambda_{array}^{r} + \frac{rw(n_w)}{N} \cdot \lambda_{array}^{w}$$

where $rw(n_w)$ is the number of read request to a disk as a result of n_w write requests

- Arrival rate of write requests to any disk in the array:

$$\lambda_{disk}^{w} = \frac{n_w + 1}{N} \cdot \lambda_{array}^{w}$$

$\Rightarrow$ Average service time of read request at disk array:

$$S_{array}^{r} = \max_{i=1}^{n_r} \left\{ R_{disk\,i}^{r} \right\}$$

where $R^r{}_{disk\ i}$ is the average response time of read requests at disk i :

$\Rightarrow$ Average service time of write request at disk array:

$$S^w_{array} = \max_{i=1}^{rw(n_w)} \left\{ R^r_{disk\,i} \right\} + \max_{i=1}^{n_w+1} \left\{ R^w_{disk\,i} \right\}$$

where $R^w{}_{disk\ i}$ is the average response time of write requests at disk i.