

UNIVERSITY OF MASSACHUSETTS
Dept. of Electrical & Computer Engineering

Computer Architecture
ECE 568

Part 16

Input/Output

Israel Koren
Fall 2011

ECE568/Koren Part.16 .1

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

Motivation: Why Care About I/O?

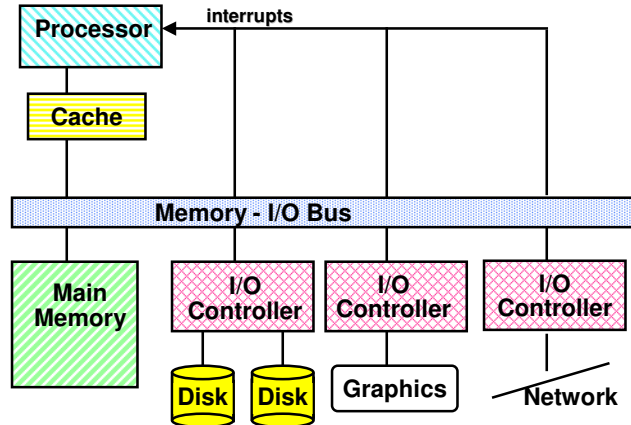
- ◆ CPU Performance: 50% per year
- ◆ I/O system performance limited by *mechanical* delays (e.g., disk I/O)
 - < 10% per year (IO per sec)
- ◆ Amdahl's Law: system speed-up limited by the slowest part
- ◆ I/O bottleneck:
 - Diminishing fraction of time in CPU
 - Diminishing value of faster CPUs

ECE568/Koren Part.16 .2

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

I/O Systems



Disks:

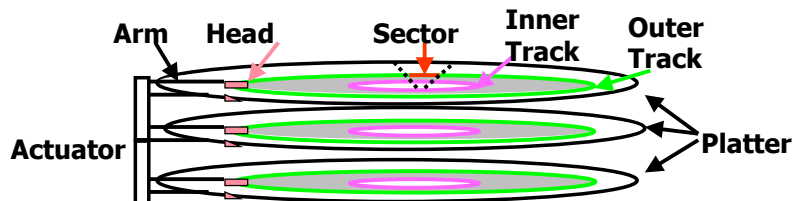
Long-term, nonvolatile storage
 Large, inexpensive, also serve as slow level in the storage hierarchy

ECE568/Koren Part.16.3

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

Disk Terminology



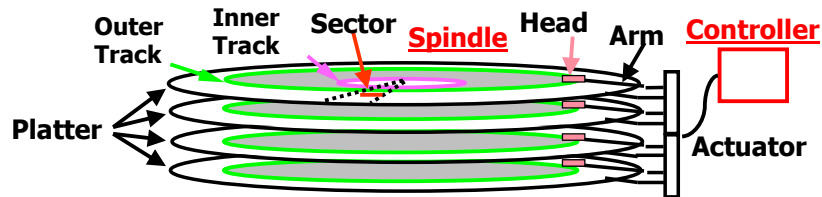
- ◆ Several platters, with information recorded magnetically on both surfaces (usually)
- ◆ Bits recorded sequentially in tracks, which in turn divided into sectors (e.g., 512 Bytes)
- ◆ Actuator moves head (end of arm, 1/surface) over track ("seek"), select surface, wait for sector rotate under head, then read or write
 - "Cylinder": all tracks under heads

ECE568/Koren Part.16.4

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

Disk Device Performance



- ◆ **Disk Latency = Seek Time + Rotation Time + Transfer Time + Controller Overhead**
- ◆ **Seek Time:** depends on number of tracks arm moves
- ◆ **Rotation Time:** depends on speed disk rotates, how far sector is from head's current position
- ◆ **Transfer Time:** depends on data rate (bandwidth) of disk (bit density), size of request

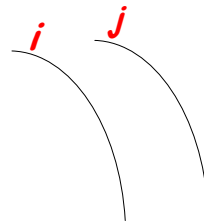
ECE568/Koren Part.16.5

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

Disk Device Performance

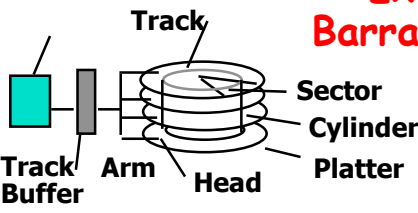
- ◆ **Rotation Time:** Avg. distance to sector from head
 - 1/2 time of a rotation
 - 7200 Revolutions Per Minute \Rightarrow 120 Rev/sec
 - 1 revolution = 1/ 120 sec \Rightarrow 8.33 milliseconds
 - 1/2 rotation (revolution) \Rightarrow 4.17 ms
- ◆ **Seek Time:** Average number of tracks arm moves
 - Sum all possible seek distances from all possible tracks / Total_#_ops
 - » Assumes random seek distance
 - Disk industry standard benchmark
 - Typical: ~8 ms
- ◆ **Transfer rate**
 - 10-40 MByte/sec
- ◆ **Capacity: 100s Gigabytes**
 - Quadruples every 2 years




ECE568/Koren Part.16.6

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann



Example: Barracuda 180



- 181.6 GB, 3.5 inch disk
- 12 platters, 24 surfaces
- 24,247 cylinders
- 7,200 RPM; (4.2 ms avg. latency)
- 7.4 ms avg. seek
- 65 MB/s
- 0.1 ms controller time

- ◆ Calculate time to read 64 KB (128 sectors) using specs
- ◆ Disk latency = avg. seek time + avg. rotational delay + transfer time + controller overhead

$$= 7.4 \text{ ms} + 0.5 * 1/(7200 \text{ RPM}) + 64 \text{ KB}/(65 \text{ MB/s}) + 0.1 \text{ ms}$$

$$= 7.4 + 4.2 + 1.0 + 0.1 \text{ ms} = 12.7 \text{ ms}$$

source: www.seagate.com

ECE568/Koren Part.16 .7
Adapted from UCB and other sources
Copyright UCB & Morgan Kaufmann

Disk Performance Example Recalculated

- ◆ Calculate again using 1/3 quoted seek time (not random, mostly to adjacent tracks), 3/4 of internal bandwidth (check bits and gaps between sectors)
- ◆ Disk latency = average seek time + average rotational delay + transfer time + controller overhead

$$= (0.33 * 7.4 \text{ ms}) + 0.5 * 1/(7200 \text{ RPM}) + 64 \text{ KB} / (0.75 * 65 \text{ MB/s}) + 0.1 \text{ ms}$$

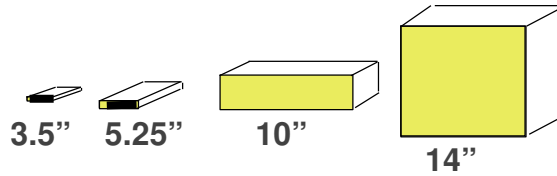
$$= 2.5 \text{ ms} + 0.5 / (7200 \text{ RPM}/(60000\text{ms/M})) + 64 \text{ KB} / (47 \text{ KB/ms}) + 0.1 \text{ ms}$$

$$= 2.5 + 4.2 + 1.4 + 0.1 \text{ ms} = 8.2 \text{ ms (64\% of 12.7)}$$

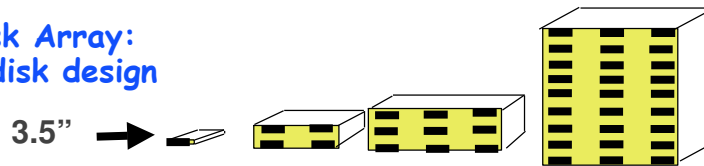
ECE568/Koren Part.16 .8
Adapted from UCB and other sources
Copyright UCB & Morgan Kaufmann

Arrays of Small Disks

Conventional:
4 disk designs



Disk Array:
1 disk design



Advantages of Small Form-factor Disk Drives:

Low cost/MB; High MB/volume; High MB/watt

ECE568/Koren Part.16 .9

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

Replace Large Disks with arrays of many Small Disks

Disk Arrays have potential for better performance
but what about reliability?

- Reliability_{disk} = $\exp(-\lambda t)$ where λ is the failure rate
- $\lambda_{\text{array}} = N \lambda$
- Reliability of N disks = $\exp(-\lambda_{\text{array}} t) = \exp(-N\lambda t)$
- MTTF = Mean Time to Failure = $1 / \lambda$
- MTTF_{array} = $1 / N \lambda = (1 / \lambda) / N$
 - = 50,000 Hours ÷ 70 disks = 700 hours
 - MTTF: Drops from 6 years to 1 month!
- Arrays (without redundancy) too unreliable

ECE568/Koren Part.16 .10

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

Array Reliability

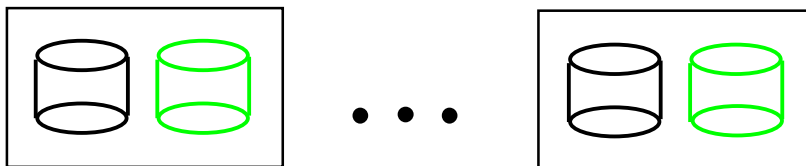
- ◆ **Solution: Redundant Arrays of (Inexpensive/ independent) Disks - RAID**
- ◆ Files are "striped" across multiple disks
- ◆ Redundancy yields high data **availability**
 - service still provided to user, even if some components fail
- ◆ Upon failure: Contents reconstructed from data redundantly stored in the array
 - ⇒ Capacity penalty to store redundant info
 - ⇒ Bandwidth penalty to update redundant info

ECE568/Koren Part.16 .11

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

RAID 1: Disk Mirroring/Shadowing



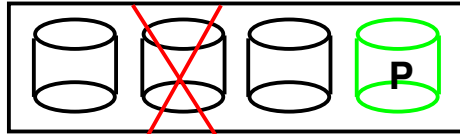
- Each disk is fully duplicated onto its "**mirror**"
Very high availability
- Bandwidth sacrifice on write:
Logical write = two physical writes
 - Reads may be optimized
- Most expensive solution: 100% capacity overhead
- (RAID 2 not interesting, so skip)

ECE568/Koren Part.16 .12

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

RAID 3: Parity Disk



Striped physical record

- $P = \text{sum mod } 2 \text{ of other disks (parity)}$
- If disk fails, subtract P from sum of other disks to find missing information
- Capacity overhead
- Wider arrays reduce capacity overhead, but decrease reliability

1	1	1	1
0	1	0	1
1	0	1	0
0	0	0	0
0	1	0	1
1	1	0	0
0	0	1	1
1	1	1	1

ECE568/Koren Part.16 .13

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

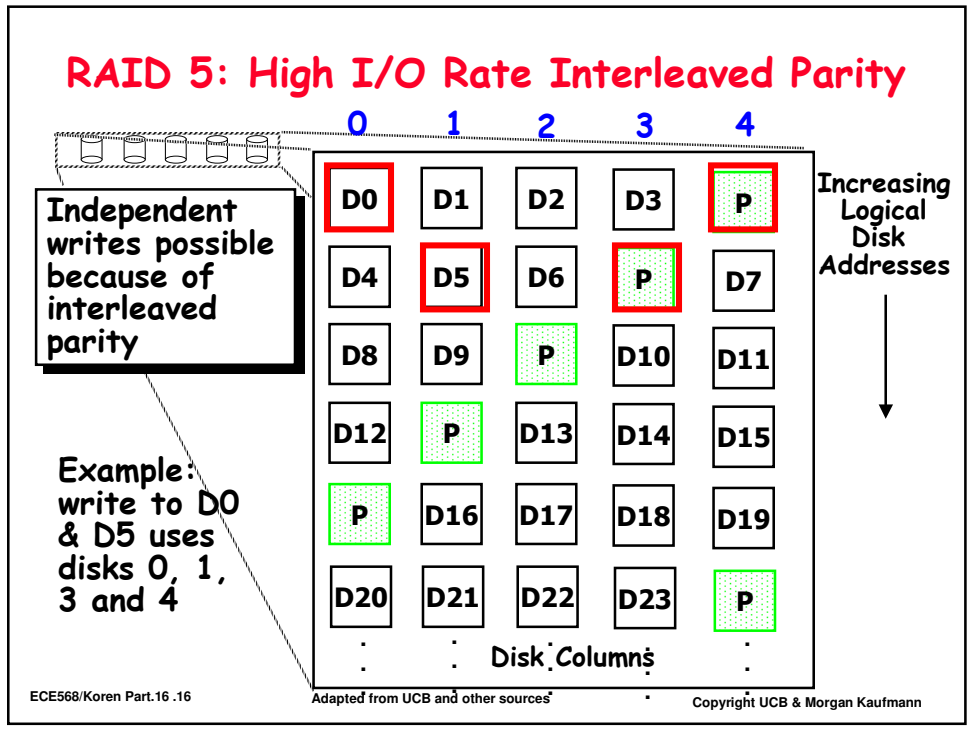
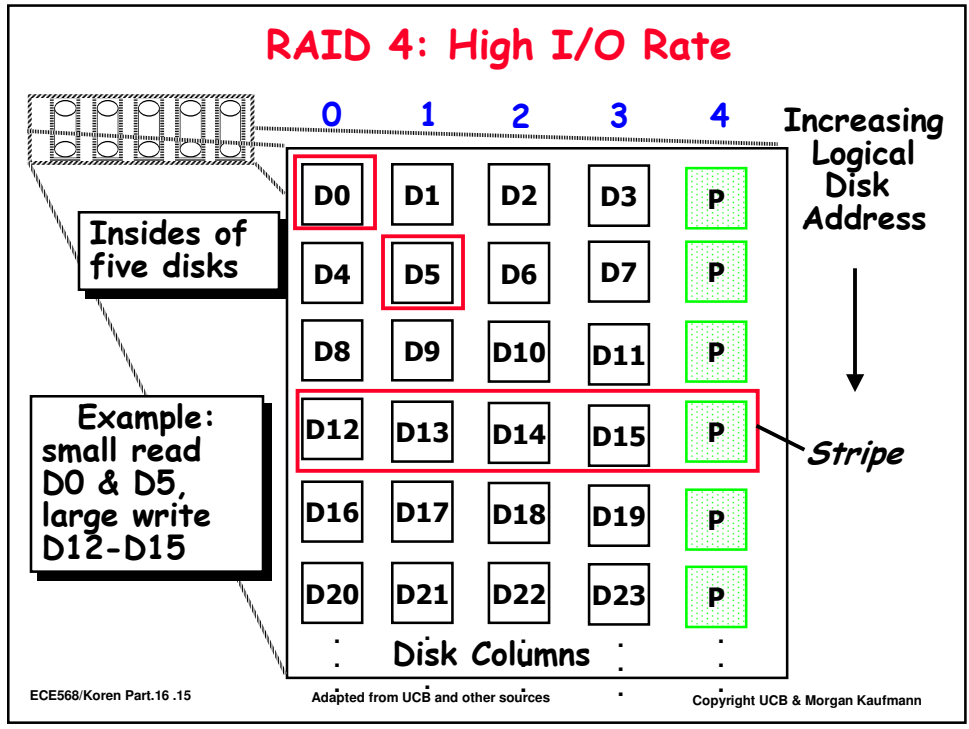
Inspiration for RAID 4

- ◆ RAID 3 relies on parity disk to detect and correct errors on Read
 - Must read from all disks every time
- ◆ But every sector has an error check field
- ◆ Rely on error check field to catch errors on read, not on the parity disk
- ◆ Use parity disk only for complete disk failure

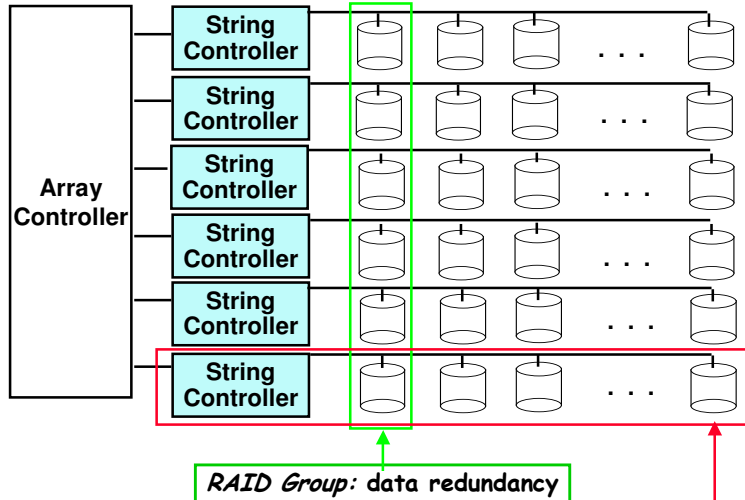
ECE568/Koren Part.16 .14

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann



System Availability: Orthogonal RAIDs



ECE568/Koren Part.16 .17

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann

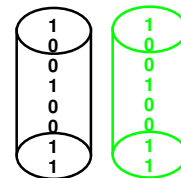
Summary: RAID Techniques

- *Disk Mirroring, Shadowing (RAID 1)*

Each disk is fully duplicated

Logical write = two physical writes

100% capacity overhead

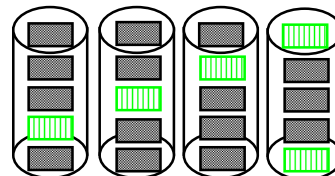


- *High I/O Rate Parity Array (RAID 5)*

Interleaved parity blocks

Independent reads and writes

Logical write = 2 reads + 2 writes



ECE568/Koren Part.16 .18

Adapted from UCB and other sources

Copyright UCB & Morgan Kaufmann