CS354: Machine Organization and Programming

Lecture 24
Wednesday the October 28th 2015

Section 2
Instructor: Leo Arulraj

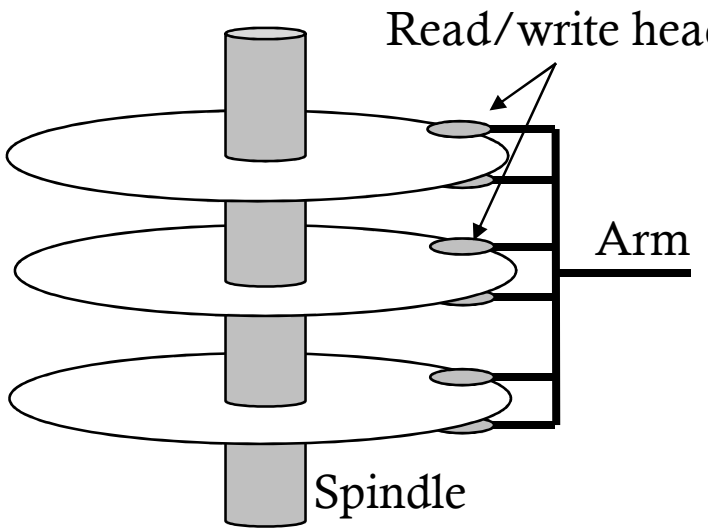
Class Announcements

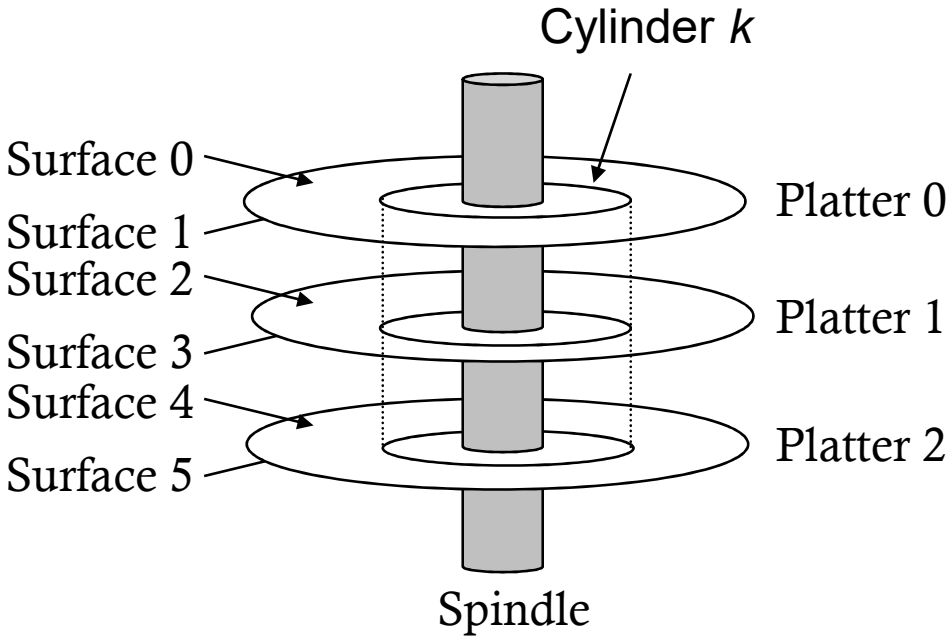
1. Grades for Programming Assignment 2 have been released in learn@uw. Contact us during office hours if you need clarifications about your grade.

2. You are allowed to bring a single sheet of paper (not unreasonably large – "letter" is considered as a reasonable size) as cheat sheet for the Midterm Exam 2

Lecture Overview

1. Magnetic Disks
2. Solid State Disks
3. Storage Trends

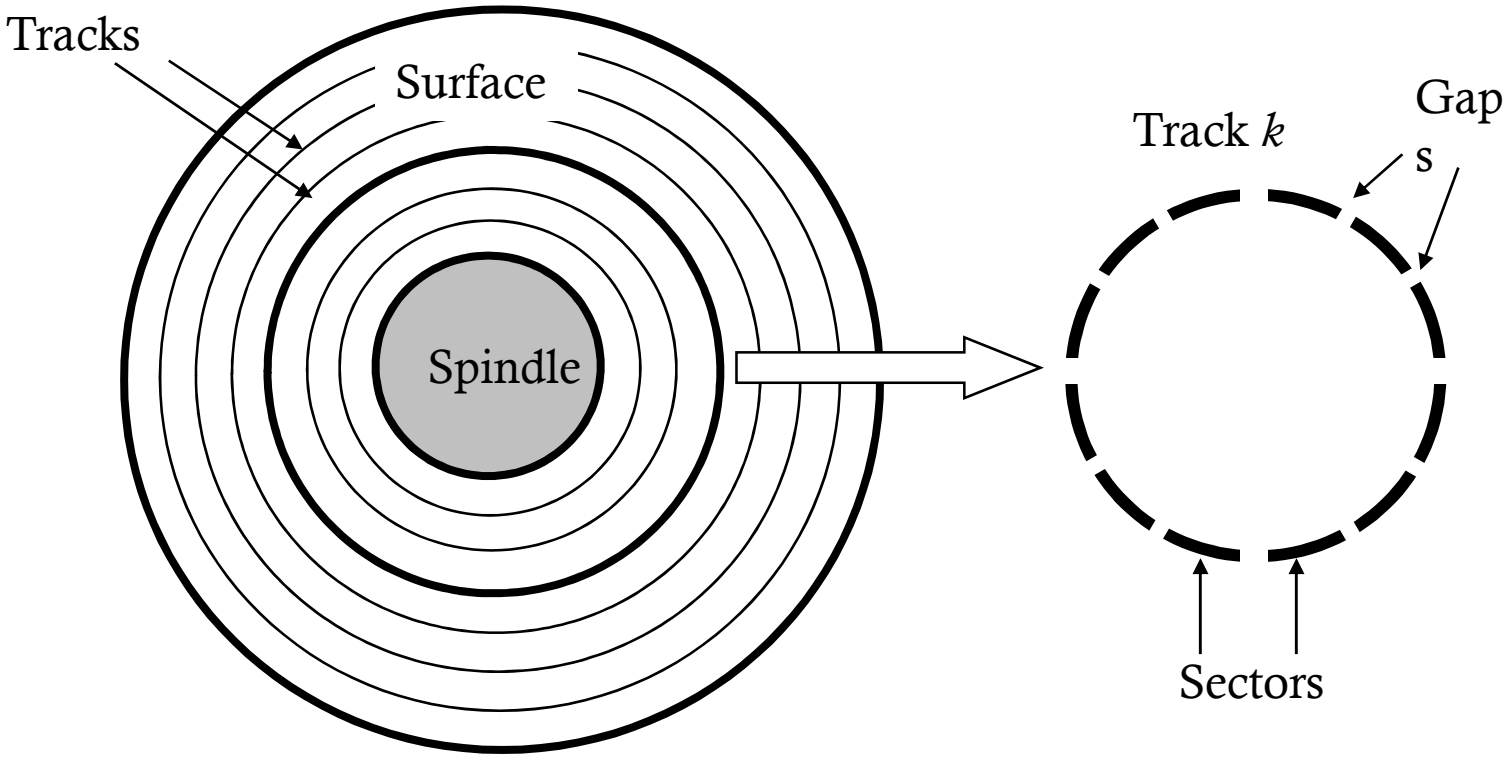




Hard Disk Overview

Platters: Consists of two sides or surfaces coated with magnetic recording material

Spindle: Spins the platter at fixed rotational rate (5400 to 15K revolutions per minute - RPM)



Track: A ring on the magnetic surface made up of sectors.

Sectors: Each track is partitioned into a collection of Sectors. Gaps in between sectors store formatting bits that identify sectors.

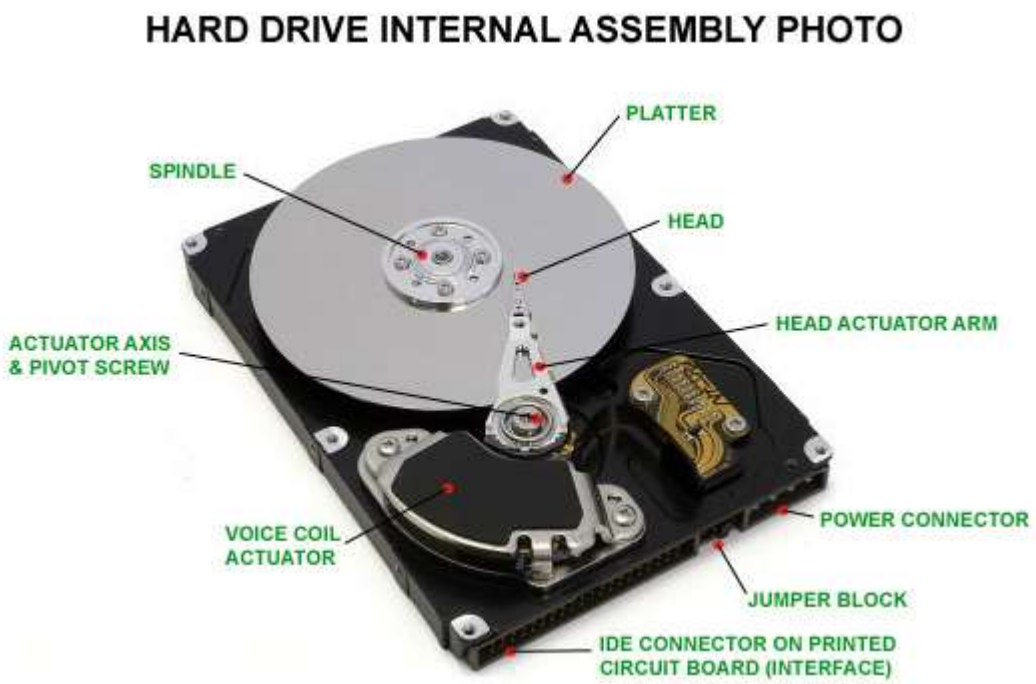
Cylinder: Collection of tracks on all surfaces that are equidistant from the center of the spindle.

Disk Capacity

Disk Capacity is the maximum number of bits that can be recorded by a disk.

Disk capacity = (#Bytes/Sector) * (Avg#Sectors/Track) * (#Tracks/surface) * (#surfaces/platter) * (#platters/disk)

Parts of a Real Disk



Slow motion video of hard disk driver operation:

Access Time for a Sector

$$T_{I/O} = T_{seek} + T_{rot} + T_{xfr}$$

Seek Time: Time taken to position the head over the track that contains the target sector.

Rotational Latency: Time taken for the first bit of the target sector to pass under the head.

Transfer time: Time taken to read or write the contents once the first bit of the target sector is under the head.

Access Time for a Sector

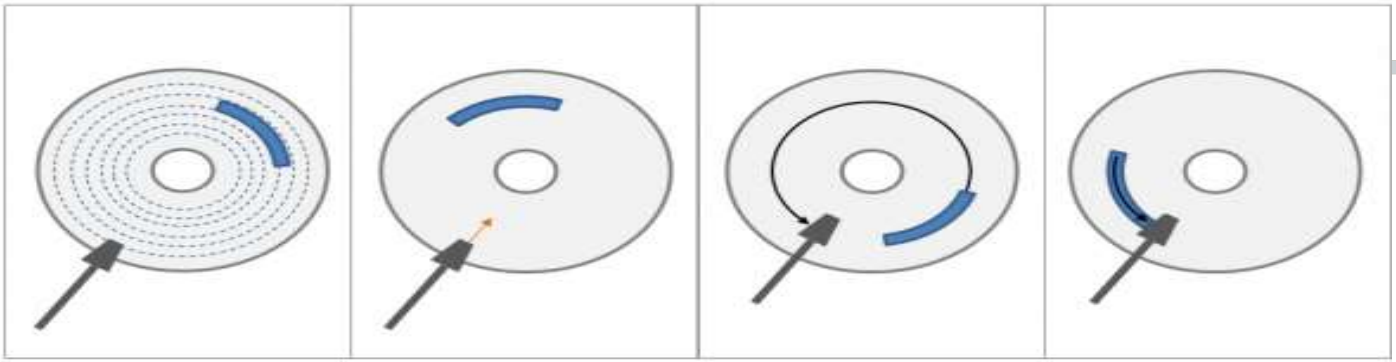
$$T_{I/O} = T_{seek} + T_{rot} + T_{xfr}$$

Seek Time: Time taken to position the head over the track that contains the target sector.

Rotational Latency: Time taken for the first bit of the target sector to pass under the head.

Transfer time: Time taken to read or write the contents once the first bit of the target sector is under the head.

Access Time for a Sector



Average Seek time is approx. the time taken to seek through one third of the tracks.
Average Rotational latency is approx. half of a full rotation's time.

Modern disk controllers take in multiple outstanding requests and follow certain scheduling algorithms to choose which one to service first.
Some examples of disk scheduling algorithms that can be used in both the operating system disk driver and disk drive firmware are :
First Come-First Serve (FCFS) , Shortest Seek Time First (SSTF)

A Real Disk from Seagate

Geometric Attributes:

Platters: 4
Surfaces: 8
Surface Diameter: 3.5in
Sector size: 512 bytes
Zones: 15
Cylinders: 50,864
Recording density(max): 628,000 bits/in
Track density: 85,000 tracks/in
Areal density(max): 53.4 Gbits/sq.in
Formatted Capacity 146.8 GB

A Real Disk from Seagate

Performance Attributes:

15000 RPM ➔ Rot. Latency of 60000/15000 = 4 ms
Avg. Rot. Latency = 2 ms
Avg. Seek Latency = 4 ms
Sustained transfer rate = 56-96 MB/s

Random I/O = 667 KB/s

What buffer size to use in order to achieve 90% of Sustained Transfer Rate approx.100 MB/s?

Logical Disk Blocks

Disks export a simpler interface with B sector-sized logical blocks numbered 0 to B-1 and hide the actual disk geometry.

The firmware in the disk called the disk controller maintains the mapping between logical block numbers and actual physical disk sectors.

Disks also have a cache inside them. Disk controllers do prefetching in order to speed up reads and do write back in order to speed up writes.

Accessing Disks

Memory mapped I/O: A block of addresses in the address space is reserved for communicating with the I/O devices.

Each of these addresses is called an I/O Port.

Every device attached to a bus will have a corresponding I/O Port.

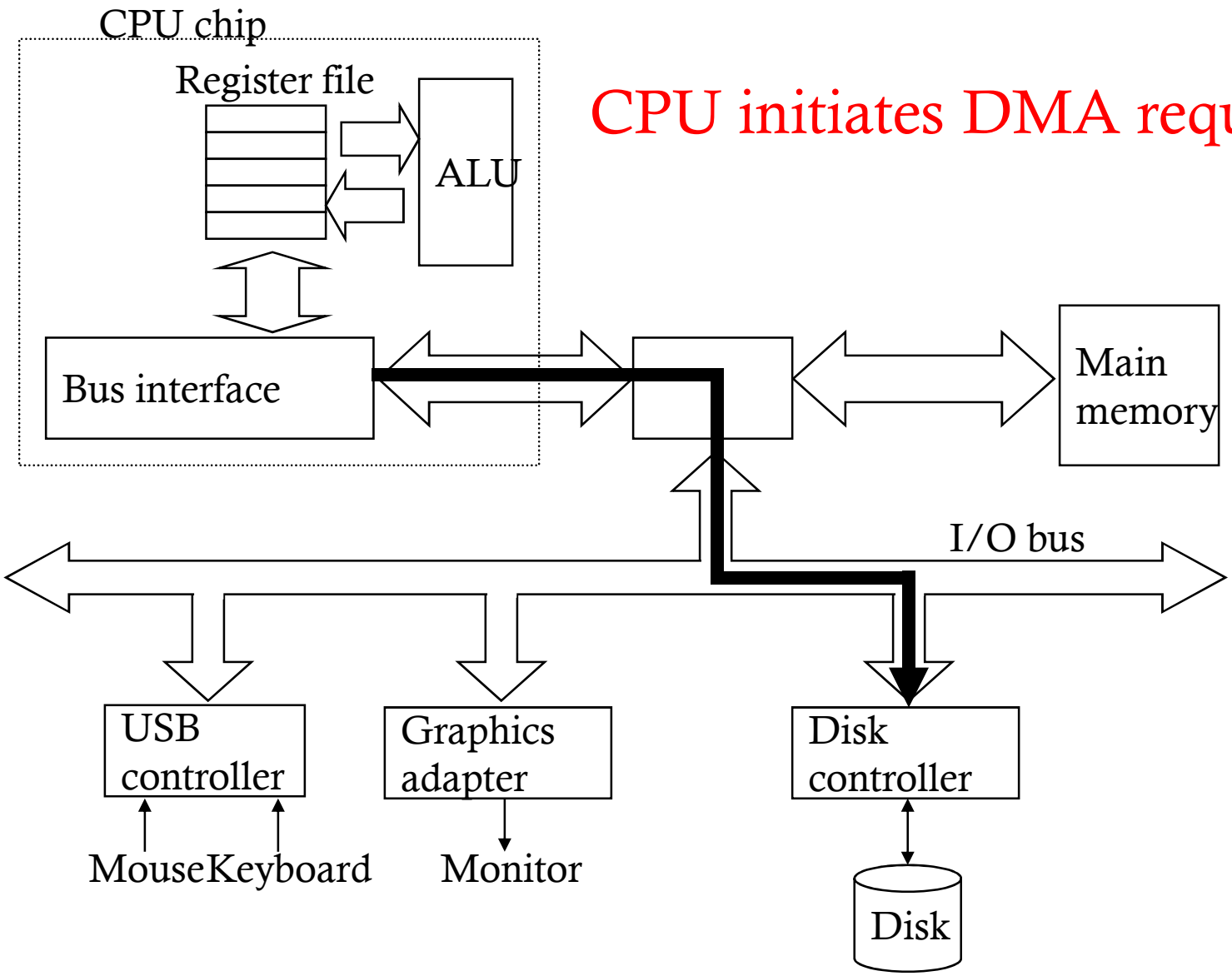
Small Communication with the I/O devices happens through the I/O port.

Textbook Example

Suppose magnetic disk is mapped to I/O port 0xa0

CPU initiates disk read by 3 store instructions:

1) Send a command word indicating to perform a read along with other parameters like interrupt on completion of DMA.
2) Send the logical block number that must be read.
3) Indicate the main memory address where the contents of the disk sector should be stored.

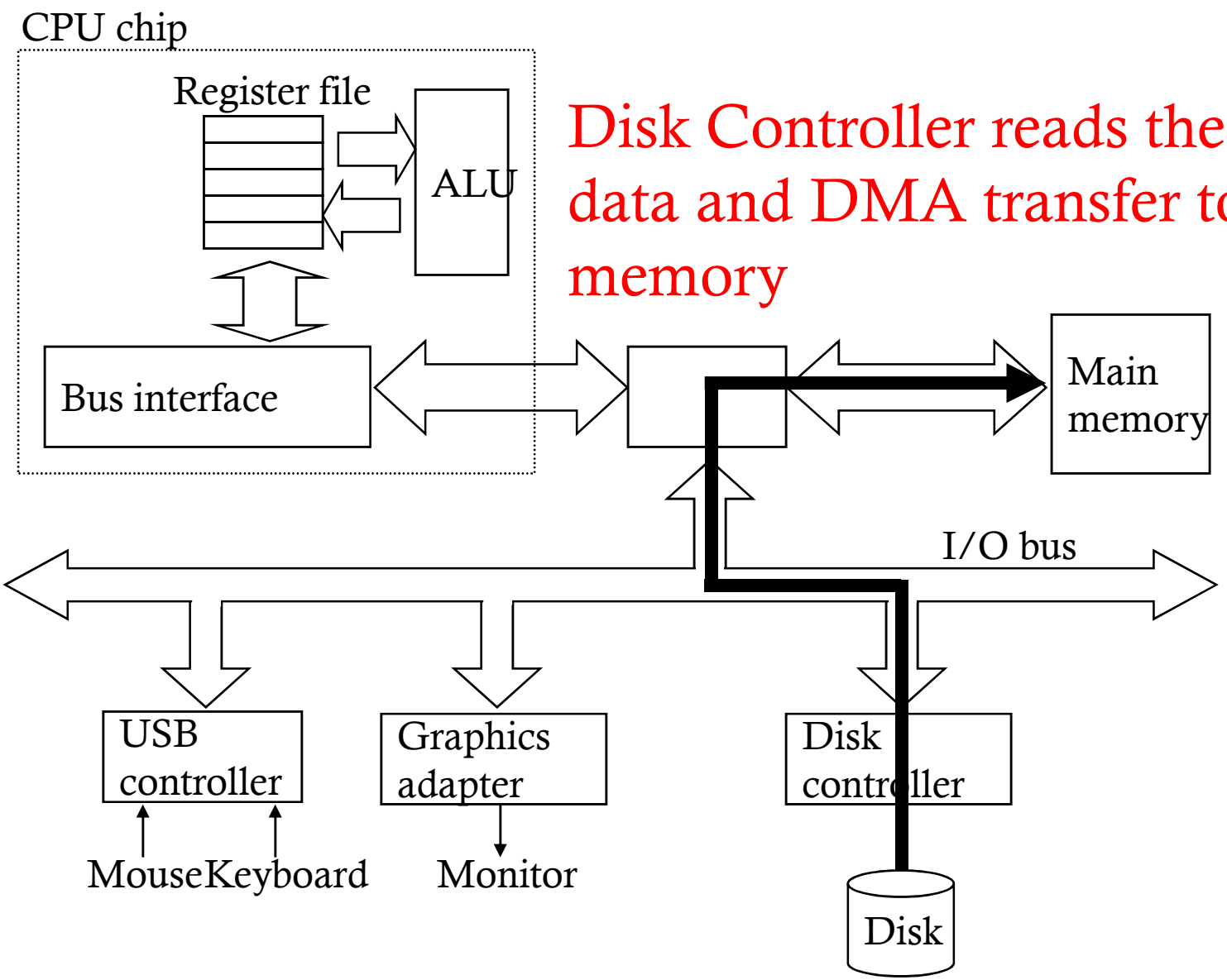

CPU chip
Register file
ALU
CPU initiates DMA request
Bus interface
Main memory
I/O bus
USB controller
Graphics adapter
Disk controller
Mouse
Keyboard
Monitor
Disk

Direct Memory Access

After issuing the request, CPU executes other instructions. Note: A 3GHz processor with a 0.33 ns clock cycle can execute 9 million instructions in 3ms.

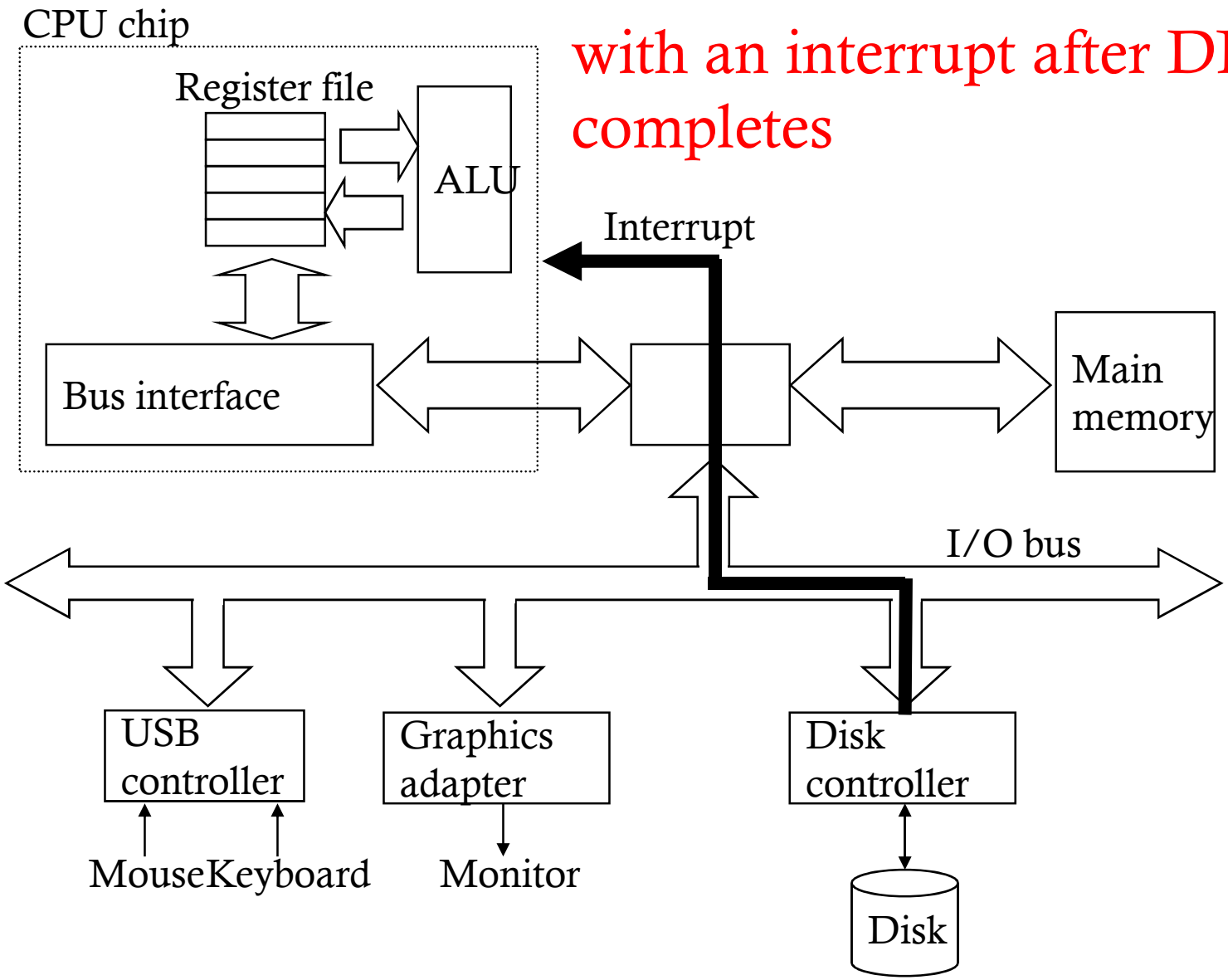
After receiving the read command from the CPU, the disk controller fetches data from the right sector and transfers it directly to the memory without the involvement of the CPU.

After DMA is complete, disk controller notifies by sending interrupt signal to the CPU.



Disk Controller reads the data and DMA transfer to memory

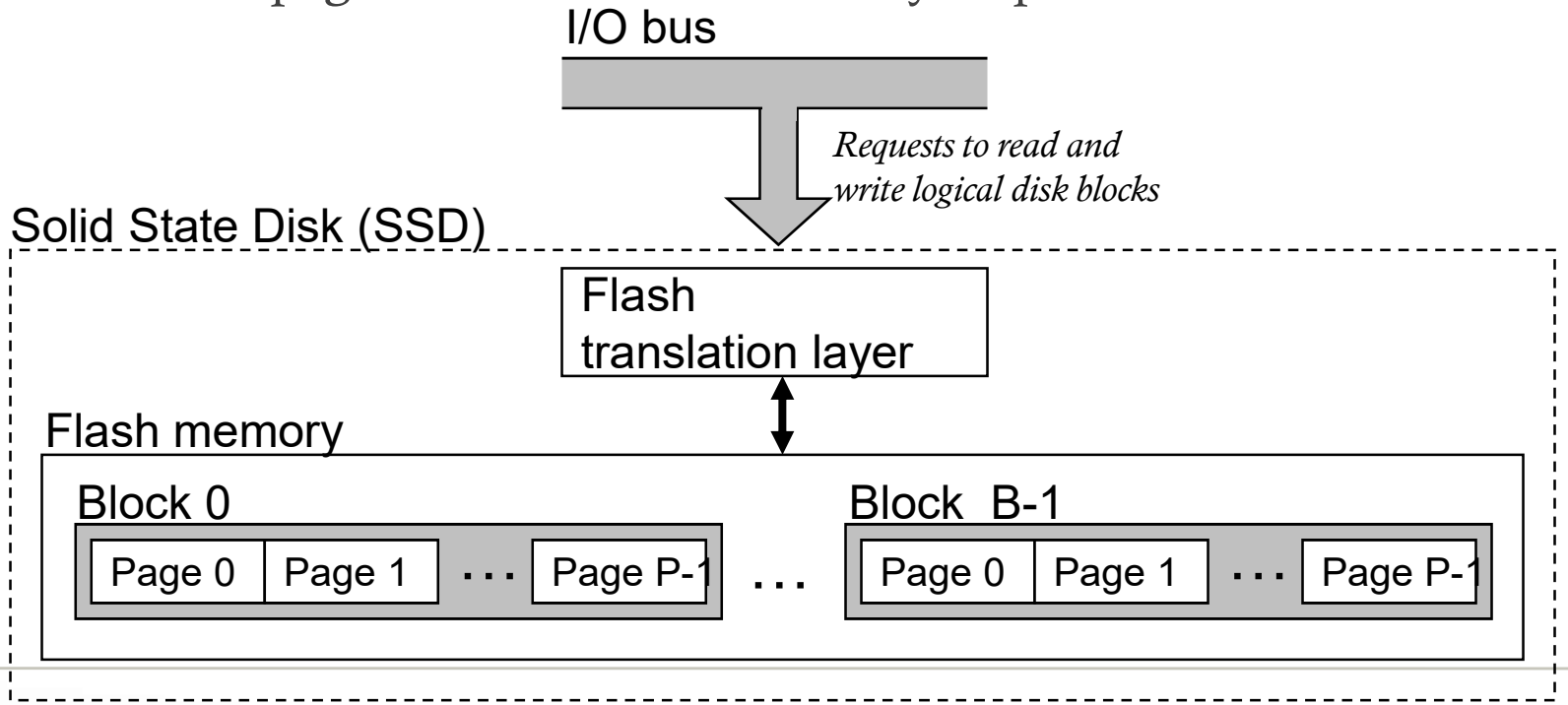
Disk Controller notifies CPU with an interrupt after DMA completes



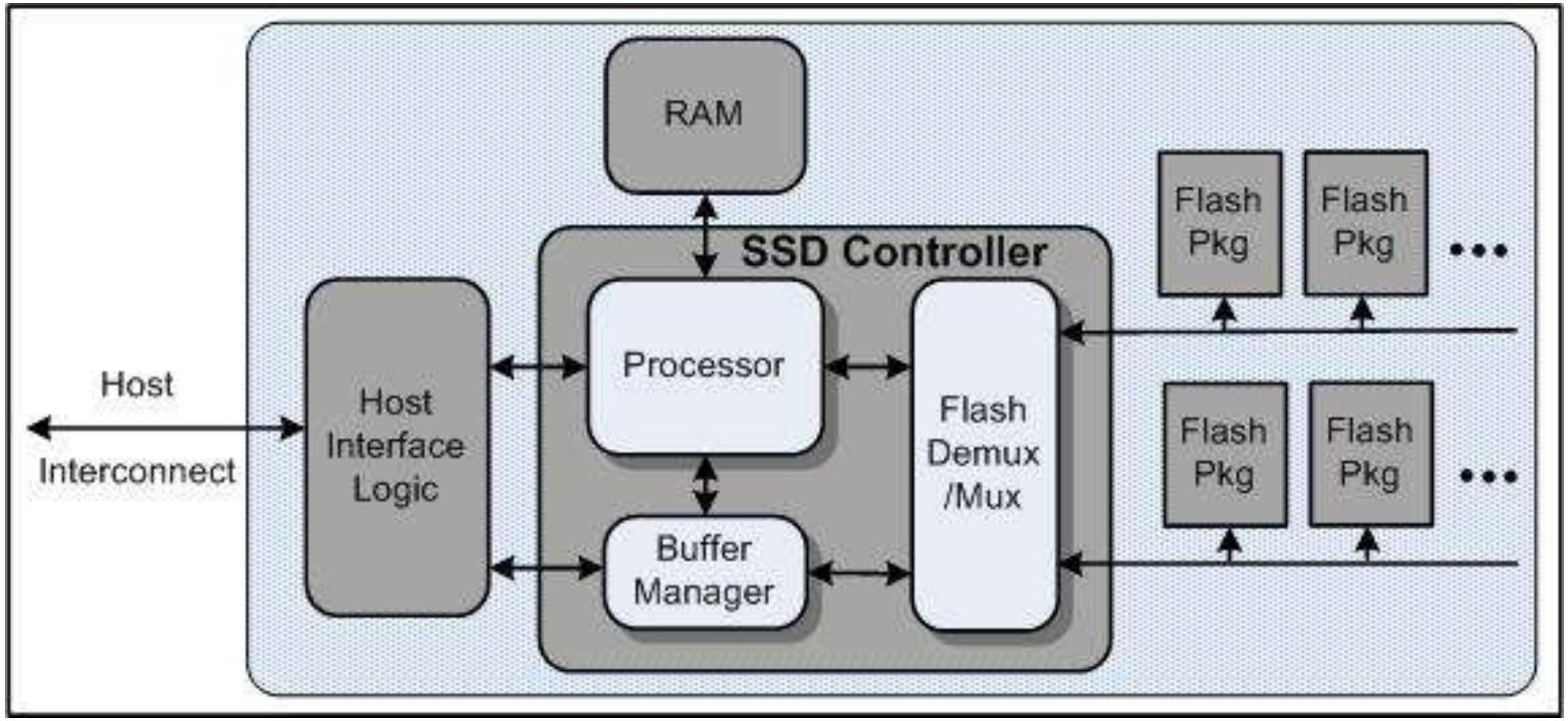
Solid State Disks

Made up of flash memory chips that store data instead of the magnetic surfaces in a conventional disk.

A Flash Translation Layer translates logical block addresses to accesses to the right block and page within the flash memory chip.



Architecture of a Solid State Disk



SSD vs. Magnetic Disk

SSD vs. Magnetic Disks

Advantages: SSDs have no moving parts and hence are more quiet during operation, faster, need less power and are more rugged.

Disadvantages:

1. Possibility of wear out after several program-erase cycles. This is mitigated by the flash translation layer.

2. More expensive than magnetic disks.

Solid State Disks

A Flash Translation Layer translates logical block addresses to accesses to the right block and page within the flash memory chip.

A page can be (re)written only after the entire block to which it belongs has been erased.

Once a block is erased all pages within the block can be written once without erasing.

-Erasing a block is a relatively long operation.
-When writing to a page with existing data, all the other pages containing existing data in the same block need to be copied along with the page being written to into a new erased block.

Performance Characteristics: Solid State Disks (from textbook)

Read Page: Approx. 10s of micro secs

Erase Block: Approx. 1 milli secs

Blocks wears out after 100K repeated writes.

Program Page: 10 micro secs to 100 micro secs

	Reads	Writes
Sequential throughput	250 MB/s	170 MB/s
Random throughput	140 MB/s	14 MB/s
Random read access time	30 micro secs	300 micro secs

Solid State Disks

Random Writes are slow for two reasons:

1. Erasing a block takes a relatively long time
2. To modify a page p that contains existing data, other pages containing existing data in the same block must be copied to a new erased block

FTL has sophisticated logic to amortize the high cost of erasing blocks and to minimize the number of internal copies on writes.

Solid State Disks

Wear levelling: Blocks wear out after repeated writes. Flash controller tries to keep all blocks at approximately the same number of erase/write cycles.

This way a scenario won't arise when only a certain subset of the blocks in the chip will not end up being worn out earlier than the rest.

Flash controller use many other sophisticated techniques to map logical blocks to flash blocks and improve the overall performance and lifetime of the disk.

Storage Trends

DRAM and disk performance are lagging behind CPU performance. Though SRAM performance also lags, it is roughly keeping up.