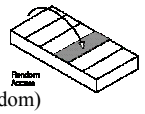
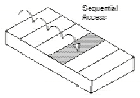
Computer Architecture 2

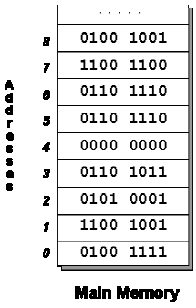
Storage

- Short-term storage
  - E.g. RAM (Random Access Memory)
  - a.k.a. : main memory, primary storage
  - Allow data to be accessed in any order (random)
  - Fast, expensive
- Long-term storage
  - E.g. Hard Disk
  - a.k.a. : mass storage, secondary storage
  - Memory can only be accessed sequentially
  - Slow, cheap





RAM

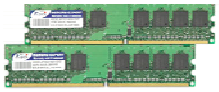


RAM

- RAM function:
  - Working area used for loading, displaying and manipulating applications and data
  - Stores intermediate processing results and output ready for transmission to a secondary storage device or other output device

RAM

- Composed of integrated circuits (vacuum tubes of today)
- Called *memory sticks* or *RAM sticks* because they are manufactured as small circuit boards with plastic packaging
- The size of a few *sticks* of gum



Cache

- Cache is very short-term memory
- Frequently stored data can be stored for rapid access
- Very small, very fast, very costly
- Level 1(L1) cache ("Internal cache"): built into the CPU
- Level 2 (L2) cache ("External secondary cache"): separate chip (approx. 1MB)

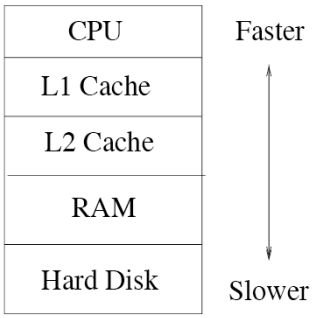
Cache

- Your web browser uses a cache
- First check the cache for data:
  - If an entry can be found with a tag matching that of the desired datum, the search stops
  - If data not found, then it must be searched for. It is then inserted into the cache.
- What to do when the cache runs out of space?

Ideas:

- replace the least recently used entry
- consider frequency of usage against the size of the stored contents

Speed



Are we missing something?

- What happens when you turn off the power?
  - ?????

Read-Only Memory (ROM)

- Computer memory on which data has been prerecorded
- ROM: retains its contents even when the computer is turned off. ROM is referred to as being *nonvolatile*
- RAM: deletes all contents when computer turned off. RAM is *volatile*
- Once data has been written onto a ROM chip, it cannot be removed and can only be read

ROM

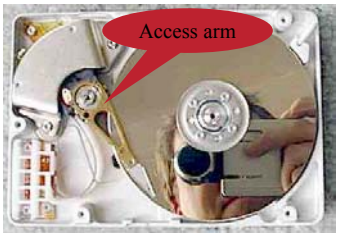
- Stores:
  - Program that boots the computer. Performs the Power-on Self Test (POST) that checks each system component for proper functioning
  - In addition, ROMs provide the basic instructions and settings for the PC's hardware: Basic Input/Output System (BIOS)
    - Keyboard, display screen, disk drive, printer

Secondary Storage

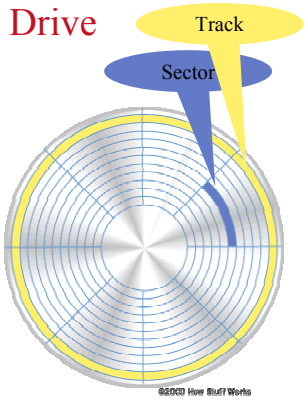
- punched paper cards
- magnetic
  - floppy disks: slow and have a small capacity, but portable, inexpensive
  - hard disks: very fast, high capacity, but also more expensive. Some hard disk systems are portable, but most are not
  - tapes : relatively inexpensive and can have very large storage capacities, but have slow access of data
- optical disks: large storage capacity, but they are not as fast as hard disks

Secondary Storage

- Uses mechanical motion to access data
- Slower than RAM which is all electronic



Hard Drive



- A sector contains a fixed number of bytes (e.g. 512)

Random or Sequential?

- Random access to sectors, sequential access within sector

Connecting it all together

- Buses are electronic highways within a computer system along which bits are transmitted
- Two parts
  - address bus: address where the data should go
  - data bus: transfers data
- The size of a bus (*width*) determines how much data can be transmitted at one time
  - e.g. 16-bit bus transmits 16 bits of data

Types of Buses

- Internal Bus: Links all the components within the CPU
- Data/Memory Bus: Links the CPU to the RAM
- Expansion (frontside) bus: extends the data bus to establish links with low-speed peripherals
- Local (frontside) bus: serves high-speed peripherals directly to the CPU
- Backside bus: connects the CPU to L1 and L2 cache

Ports

- Physical interface through which information transfers in and out
- Serial port: transfer one bit at a time
- Parallel port: transfer several bits at the same time
- USB port: Universal Serial Bus (universal protocol for data transfer) port

Video Card

- Generates and outputs images to a display
- Contains:
  - Graphics processing unit (GPU)
    - Microprocessor dedicated to manipulating and rendering 2D and 3D graphics
    - Render objects with shading, textures
    - Very powerful computationally
  - Video memory: video RAM (VRAM)
    - Used to store the display image, as well as textures and other elements
  - Video BIOS
    - Chip that contains the basic program that governs the video card's operations and provides the instructions that allow the computer and software to interface with the card

Sound Card

- Provides the audio component for multimedia applications
- Includes a digital-to-analog converter:
  - converts recorded or generated digital waveforms of sound into an analog format
  - able to play multiple digital samples at different pitches and volumes
  - if a small amount of ROM memory stores samples of standard instruments, can synthesize music

Network Card

- Implements low level protocols to allow network communication
- Every network card has a unique 48-bit serial number (MAC address)
  - written to ROM carried on the card

Motherboard

The motherboard is a single circuit board, made up of:

- CPU
- Main Memory
- Cache
- ROM (BIOS)
- Several Buses
- Video card, Sound Card, etc.

Data Organization

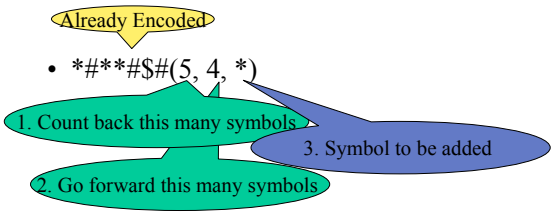
Goals:

- Speed, Cost: use the physical resources (disk space, transmitting channels) intelligently to achieve results as quickly as possible
- Accuracy: don't want to sacrifice accuracy when possible

Data Compression

- Adaptive dictionary encoding:
  - use a collection of building blocks from which the message to be compressed is constructed
  - collection changes during the encoding process
- Lempel-Ziv encoding: dictionary at any point consists of those patterns that have already been encoded.

Lempel-Ziv Encoding



- *#**#$#(5, 4, *)
- *#**#$#**#$*

Lempel-Ziv Encoding

To compress a message using this encoding:

1. Quote the beginning of the message
2. Search through the pattern quoted to find the longest segment that agrees with the remaining message to be compressed (the pattern of the first triple).
3. Repeat search

Data Compression

- When is this encoding useful?
- When you have repeating patterns of data

File Formats

- This kind of encoding scheme constitutes a particular file format (*set of rules describing how the binary information in the file should be interpreted*)
- File formats are needed:
  - to exchange information between two parties
  - to be able to efficiently store and retrieve information
  - to compress information for easy transfer

One More Example

- An appointment book

Tuesday, September 19, 2006
8:35 am COMP 102 Lecture
12:00 pm Lunch
3:00 pm Meeting

Wed., September 20, 2006
8:35 am Sleep
11:00 am Wake Up
11:00 pm Do COMP 102 Homework

- Encoding

|{TU091906}8:351<COMP 102 Lecture>12:002<Eat Lunch>3:002<Meeting>|

|{W092106}8:351<Sleep>11:001<Wake Up>11:002<Do COMP 102 Homework>|