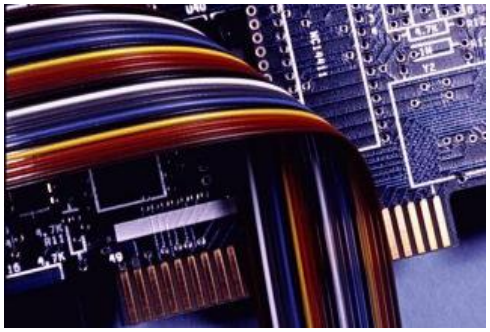

computer
architecture

computer architecture – Learning Objectives

- Understand the 'Three-box model'
  - Processor
  - Main memory
  - I/O
  - Bus system
- Be familiar with the Internal and External components – CPU and peripherals
- Understand Memory and the Stored Program Concept

computer architecture



simplicity is the ultimate sophistication

Leonardo da Vinci

what is a computer made of?

a **computer system** is made up of various parts:

- one (or more) processors
- veritable oceans of memory
- input devices
- storage devices of different types
- output devices

- and the user(s)

architecture

a style and method of design and construction

dictionary.com

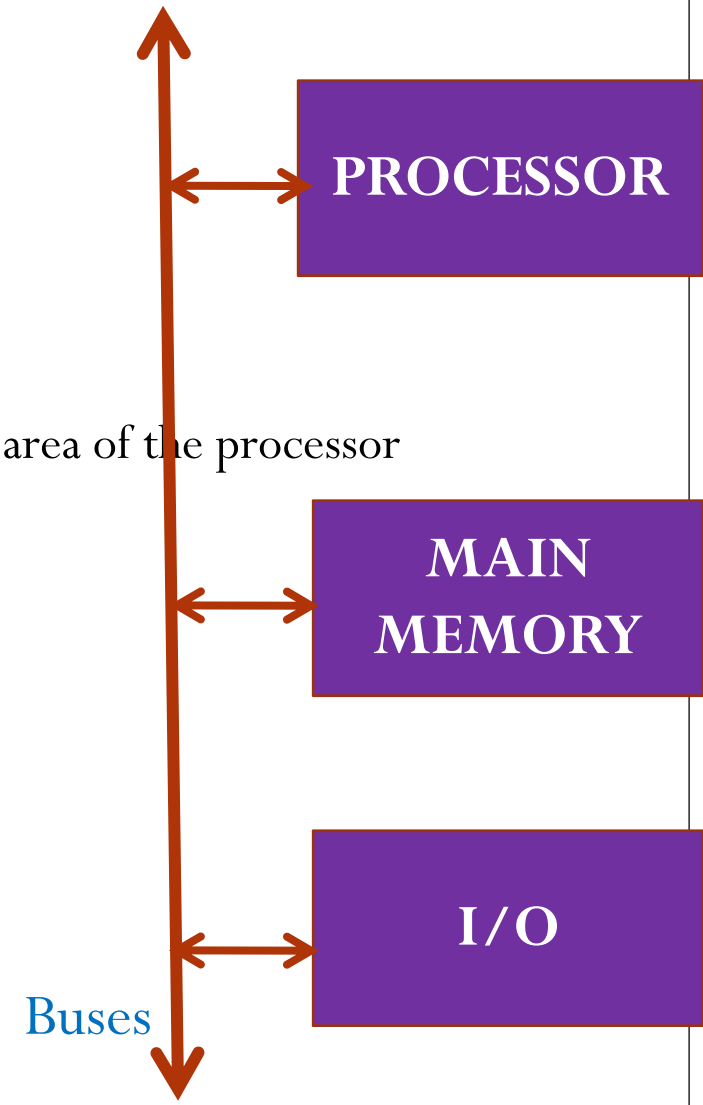
computer architecture

- design and structural organisation
- the internal structure of a microprocessor
- the relationships between the parts of a computer
- functional behaviour of those parts and the computer

computer architecture – The Three Box Model

- Contains:
  - The processor
    - Control Unit
    - Arithmetic and Logic Unit (ALU)
    - Registers (storage Units) – faster access, working area of the processor
    - System clock
    - Interconnections
  - Main memory
  - I/O controllers (input, output, or both)
  - Buses

Draw this in your notes!



constituent parts of the architecture

- system clock
- processor
- registers
- memory
- bus system
- addressing modes
- interrupts
- fetch-execute cycle

the system clock

- computers are digital as they operate by changing from one **discrete** state to another

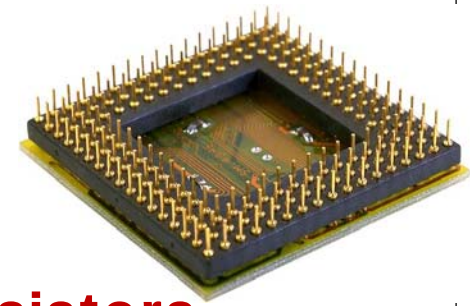
- the **change** between the discrete states is orchestrated by the **system clock**

- the system clock produces a continual stream of **ON and OFF pulses** at an incomprehensible speed

- each **ON** pulse tells the computer it is time to change to the next state

the processor

a single **microchip** that contains all the basic elements required to implement a computer system



The brain of the system.

A typical processor contains **100,000,000+ transistors**

Moore's Law says that the number of transistors that can be integrated into a chip doubles about every 12-24 months!

Intel Itanium 2 processor has nearly 1 billion transistors!

the processor

five functional units:

- control unit
- arithmetic and logic unit
- input unit
- output unit
- memory

its structure is independent of the problems it solves

programs and data are stored in memory

memory is a sequence of uniquely identifiable identical storage locations

control unit

acts under the control of the **system clock**

manages the fetch-execute cycle

- **determines** the internal paths to get data from its source location to its intended target location
- fetches, decodes and causes the execution of each and every **instruction**
- **synchronises** the parts of the processor so that the correct data is in the correct locations when the next instruction is activated

arithmetic and logic unit

carries out the program **instructions**

there are two main families of instructions

arithmetic

- simple arithmetic
- simple storage tasks

logic

- comparisons between two values
- altering the logic flow

a question of speed

factors

- the clock speed
- the word size
- the size of the processor
- the available power supply

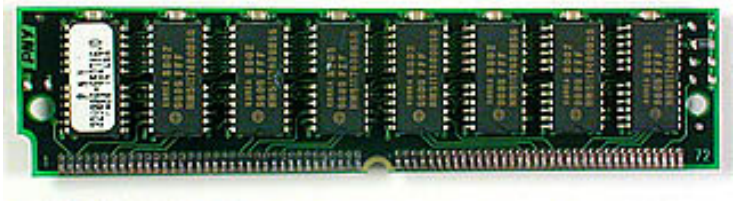
main memory

is the memory that can be **directly addressed** by the processor

also called IAS (immediate access store)

each byte goes into a separate area of the chip called **memory location**

memory



SIMM

DIMM

no moving parts, so very fast

assured integrity using **parity checking**

connected to the processor via a **bus system**

provides mass storage for the processor to use

memory

one basic **function**: storing the program instructions **and** the data upon which those instructions act

memory can be seen as a massive set of storage spaces where each space is uniquely identifiable

like pigeonholes

"640K ought to be enough for anybody."
Bill Gates, 1981

memory

unused memory is stored in a pool for later use (heap of memory)

each process requiring an amount of memory is **allocated** from this pool

-> using the smallest unused chunk of memory in the pool

sufficient memory must be allocated to the process before it can begin execution

the memory is **returned** to the pool **when no longer required** by the process

memory

Research (textbook, internet) in brief the following:

- RAM
- ROM
- EEPROM
- DRAM
- SDRAM

RAM

random access memory, **volatile**, content is erased on shutdown

the 'notepad' for the computer, being used to store values that represent anything and everything

RAM **holds**

- program instructions
- part of the operating system
- data being processed by those programs

ROM

read only memory

non-volatile or **permanent** memory

content remains even if power is no longer available

stores **system boot program**, the essential parts of the operating system (**kernel**)

used extensively in embedded microprocessors and control systems (e.g. microwave ovens and missiles)

effects of memory

more RAM means

- more instructions and/or data can be held and accessed without the need to read a disk
- more tasks can be processed simultaneously
- less memory swaps to backing store means less degradation to system performance

access speed is affected by size of data bus (word size)
-> each word is transferred at the same speed regardless of the amount of memory available or being used

common types of RAM

DRAM	dynamic RAM
SRAM	static RAM
EDO RAM	extended data out RAM
SDRAM	synchronous RAM
burst EDO RAM	burst EDO RAM
VRAM	video RAM
DDR	double data rate SDRAM
PCMCIA	personal computer memory card international association

(**research**)

the structure of memory

consists of **memory cells** (tiny circuits)

each cell holds one binary digit, either a zero or one

eight of these cells are required to make a byte

each byte could be thought of as a pigeonhole

memory consists of many millions of bytes

memory then becomes a massive set of pigeonholes

each with a **unique address**

Address	Contents
000000	LOAD #12
000001	ADD 6
000010	STORE 35

memory management system

accepts address from the processor via the **address bus**

data can be placed in or retrieved from any cell

memory manager can jump to any cell at will

hence the name **random access memory**

effects of electricity

... it needs electricity to maintain the values held

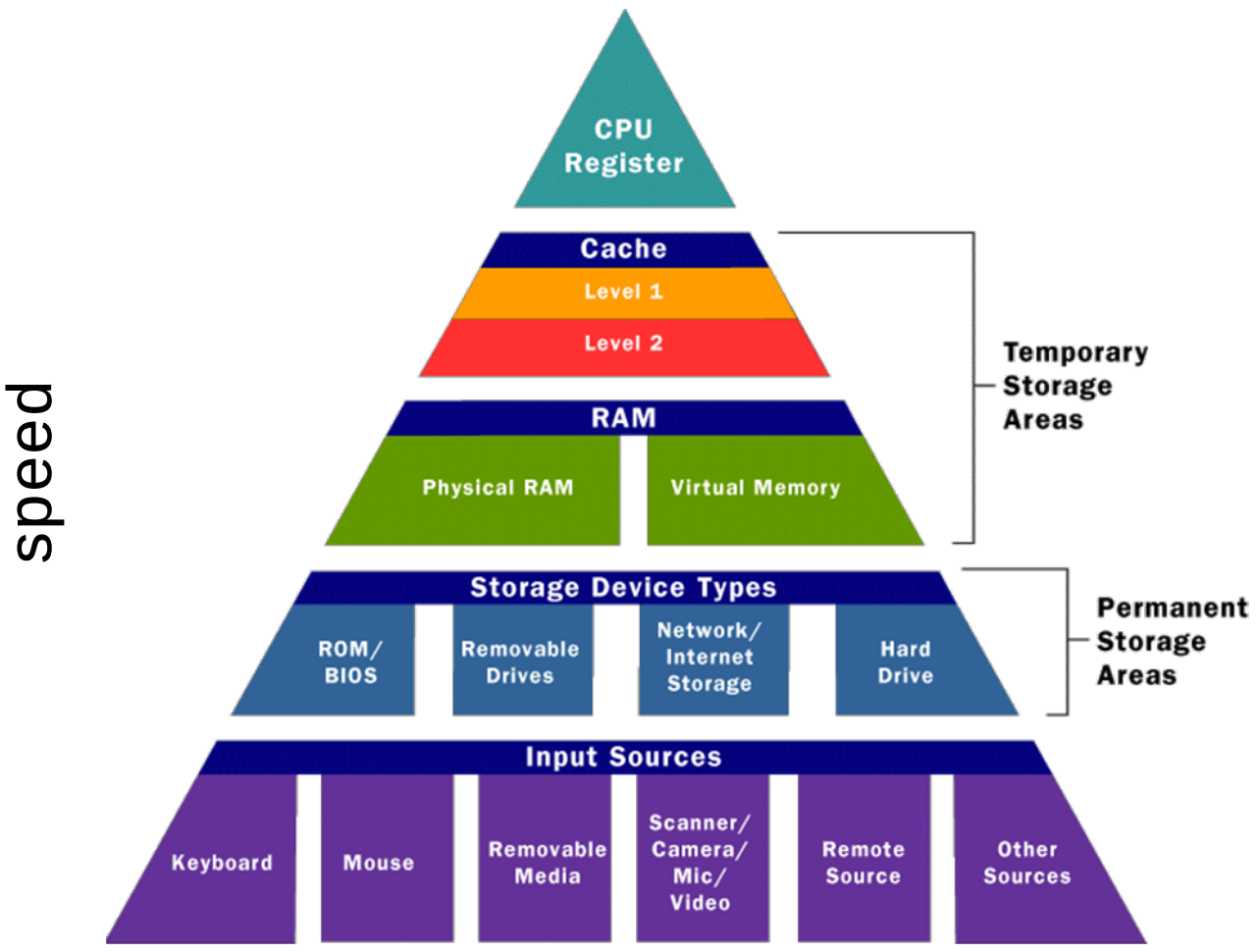
removing the supply of electricity ...

... loses the values stored in RAM

some form of **permanent storage** is required to overcome this loss when a computer is powered off ...

... which is provided by **secondary storage**

memory hierarchy



cache memory

purpose is to **improve system's overall performance**

much **faster** than standard RAM

system looks in cache **before** it looks in RAM

holds recently accessed instructions and/or data

the bus system

communication pathways between different components of a computer system

a system of parallel wires along which data/addresses can travel; a shared transmission medium

each wire holds one bit of data

the number of wires is called the **bus width**

the more wires the more efficient the bus

the more wires the greater the cost of manufacture

address bus

carries the **addresses** of memory locations

When the processor reads a word of data (8, 16, 32 bits) from memory, it first puts the address of the desired word on the address bus.

'**one way**' - from processor to memory manager

the number of wires (**width** of the address bus) determines the amount address that can be transmitted;

it also determines the maximum possible memory capacity of the system

Eg if the address bus has only 8 lines, then the max address it could transmit would be 11111111 (binary system) or 255 (as we know numbers, in denary system), with a max memory capacity of 256 (which is 2 by the power of_____).

data bus

carries **data** between the processor and memory

bi-directional (read **from** and write **to** memory)

the number of wires determines the amount of data that can be processed in one operation

the width of the data bus is usually taken as the **word size** of the computer

sometimes referred to as the **memory buffer register**

data bus

Normally the data bus is matched to the **size of a word** (location/register) in the processor.

The size of the processor determines the size of a data item that can be processed in one go.

E.g. if the data bus is 8 bits, and the instruction is 16 bits long, then the processor must access the memory twice just to fetch the instruction.

control bus

system management function, carrying

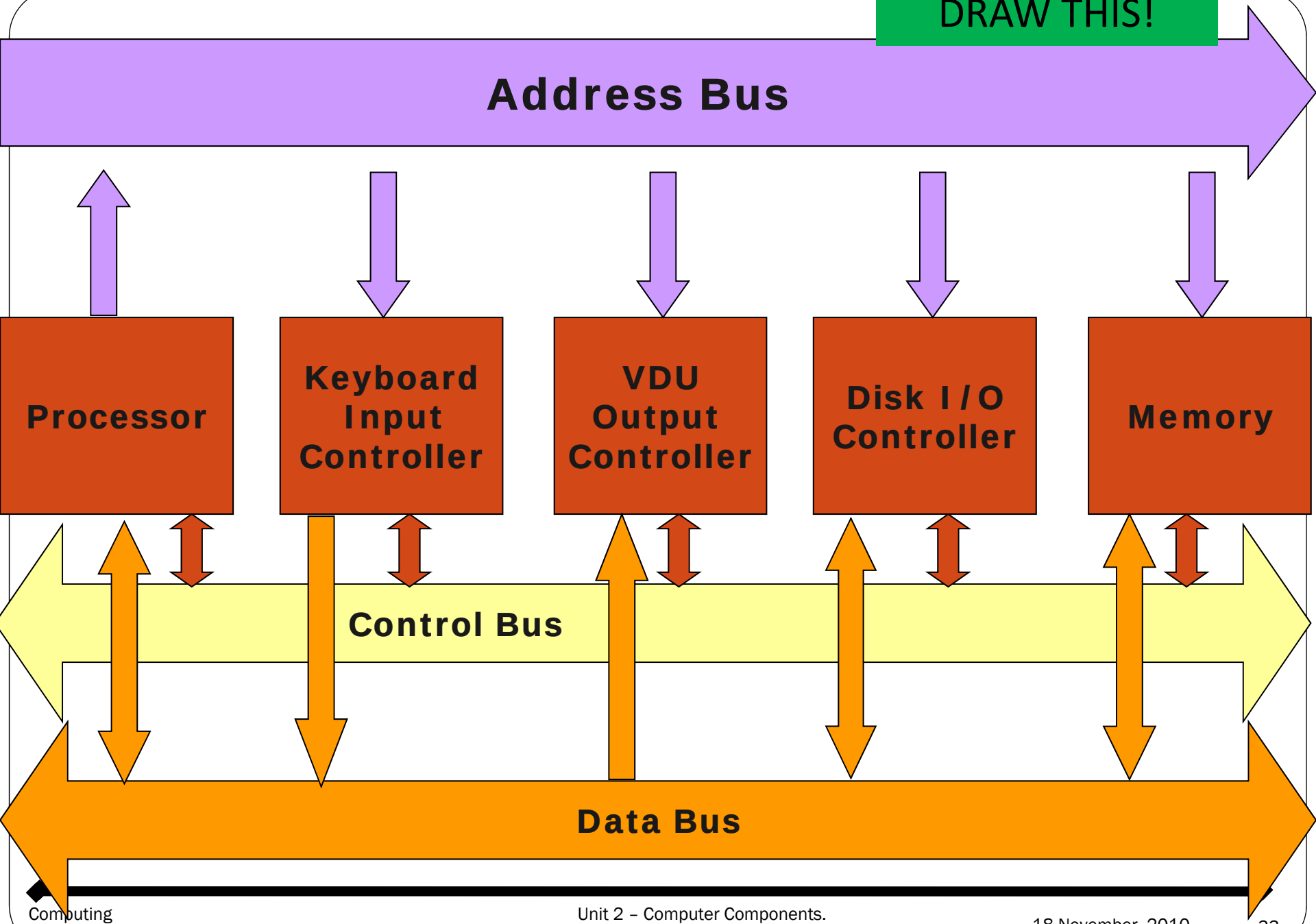
signals to inform memory manager to:

- **read** the value in memory at the location given by the value on the address bus and to **send** it to the processor
- **store** the value on the data bus in memory at the location given by the value on the address bus

interrupt signals

Other control lines are: Memory read/write; I/O read/write; Transfer ACK – indicates that data have been accepted from or placed on the data bus; Bus request / grant etc.

DRAW THIS!


Address Bus
Processor
Keyboard Input Controller
VDU Output Controller
Disk I/O Controller
Memory
Control Bus
Data Bus

I/O

CPU needs to communicate with **external components** – peripherals

They can be

- Input
- Output
- Storage

The processor commands the **I/O controllers** to read / write data from / to devices.

What is an I/O controller? Research!

TASK: Textbook – page 148, draw Fig.7.2.2

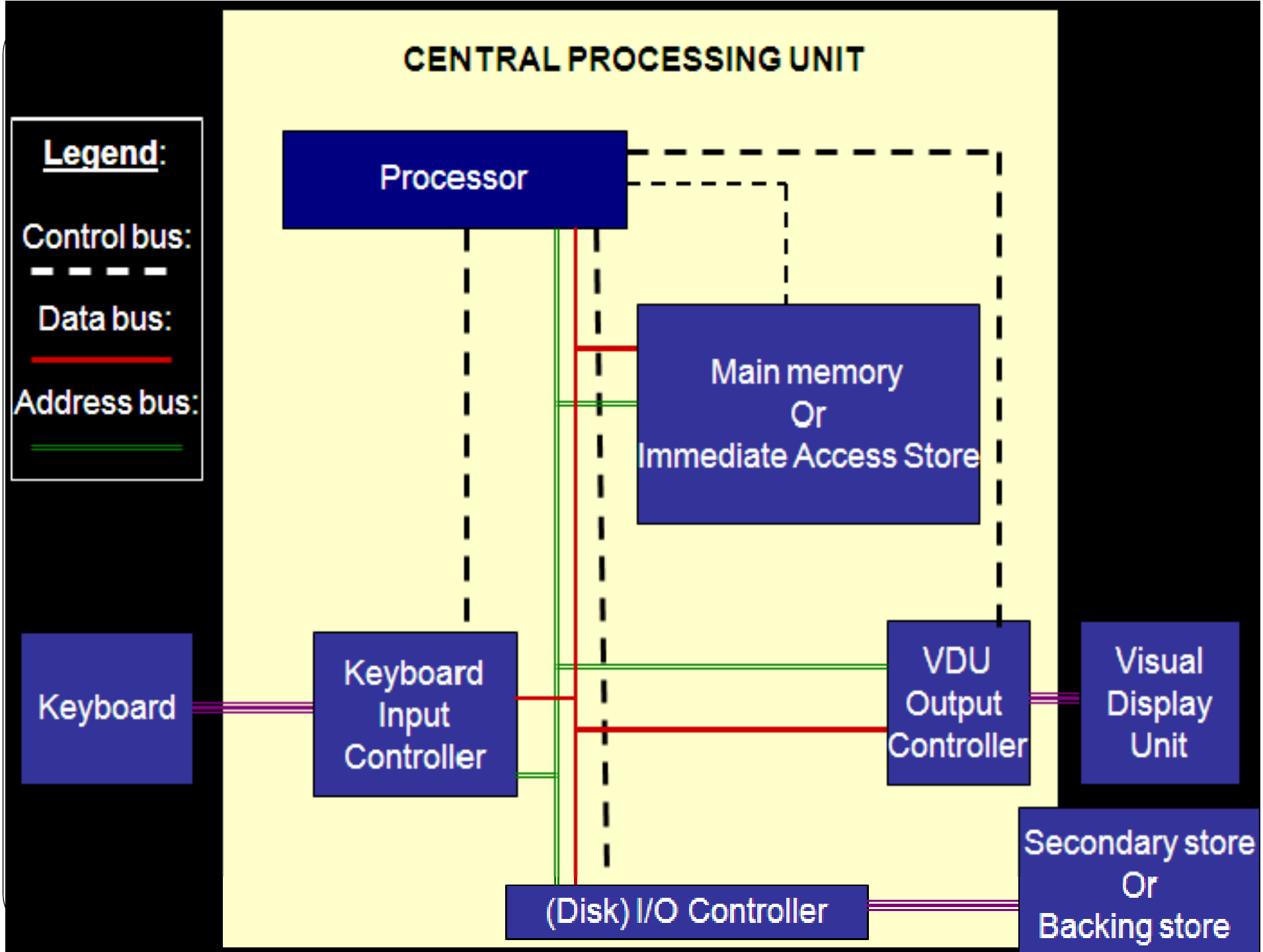
characteristics

- used to **connect peripheral devices to the processor**
- some can operate both input and output transfers of bits – e.g. floppy disk controller
- others operate in one direction only, either as input or as output – e.g. VDU controller

What is it anyway?

An electronic circuit board consisting of 3 parts:

- Interface allowing connection to the system (I/O bus)
- A set of data, command and status registers
- Interface enabling connection to the cable which links the device to the computer


CENTRAL PROCESSING UNIT
Legend:
Control bus:
Data bus:
Address bus:
Processor
Main memory
Or
Immediate Access Store
Keyboard
Keyboard
Input
Controller
VDU
Output
Controller
Visual
Display
Unit
(Disk) I/O Controller
Secondary store
Or
Backing store

computer architecture – Spot Check

- Name the three boxes in the three-box model
- What is a computer bus?
- Name the three sub-buses that make up the system bus
- What is an I/O controller?
- Give a definition for:
  - Main memory
  - Secondary storage
  - Peripheral

computer architecture – Memory and the Stored Program Concept

- **Memory** – collection of bits / bytes (8 bits = 1 byte)
  - How does the computer know which contents to retrieve from which location?
  - ANSWER: **memory address** – unique numeric code which represents each memory location
- How does it work?
  - Computer needs a main memory location
  - Computer puts its unique address on the _____ bus
  - Processor asserts, over the _____ bus, if READ or WRITE to this location
  - Processor uses the _____ bus to transfer the content FROM / TO this address

computer architecture – Memory and the Stored Program Concept

First computers – built in 1940s (Alan Turing, John von Neumann)

- The **von Neumann** architecture = **holds in a single store (main memory) both the instructions and the data, and shared bus**.
  - This is the "stored program concept" = computer acts upon internally stored instructions
- The **Harvard** architecture = **has separate instruction and data memories**, separate instruction & data buses
- QUESTION: which of the above is more efficient (greater processing speed) and why?

We covered: control unit; ALU.

Registers:

storage areas located within the processor

provide instantaneous access to values held

they can be **DEDICATED** (assigned a specific role) or **GENERAL PURPOSE** (available to the programmer for temporary storage of data)

most have a single purpose (e.g. **program counter**)

some (e.g. the **accumulator**) are used as a general-purpose work space

computer architecture – **task (10 minutes)**

1. Take notes about the general purpose & dedicated registers.
2. Read & take notes about **what they are** and **how**:
   a) System clock & clock speed
   b) Word length
   c) Bus width

affect a computer's **performance**.

Textbook.

Page 152-157

computer architecture – **plenary**

1. What is the effect on processor performance of increasing:

   a) Clock speed

   b) Word length

   c) Bus width

2. Why is there a limit on clock speed?

3. Why are microprocessors being designed with multiple cores?

computer architecture – Learning Objectives reviewed

- Understand the 'Three-box model'
  - Processor
  - Main memory
  - I/O
  - Bus system
- Be familiar with the Internal and External components – CPU and peripherals
- Understand Memory and the Stored Program Concept