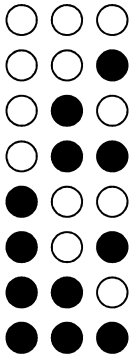
Performance

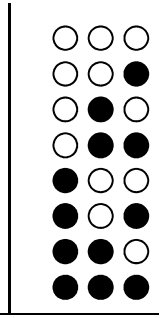
Level 3 Extended Diploma

Unit 19

Computer Systems Architecture

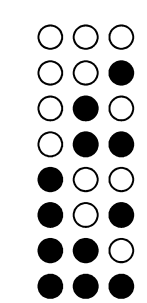


Performance issues

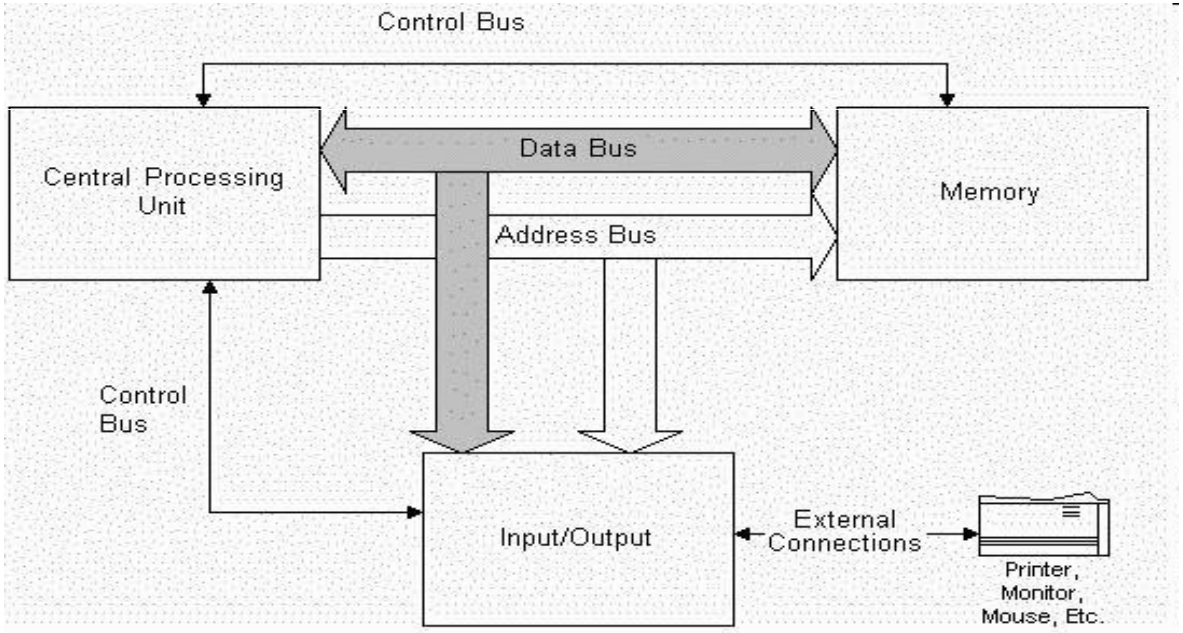


- We want the CPU to be working as efficiently as possible
- We want to get data to and from the CPU as fast as possible

Transferring data

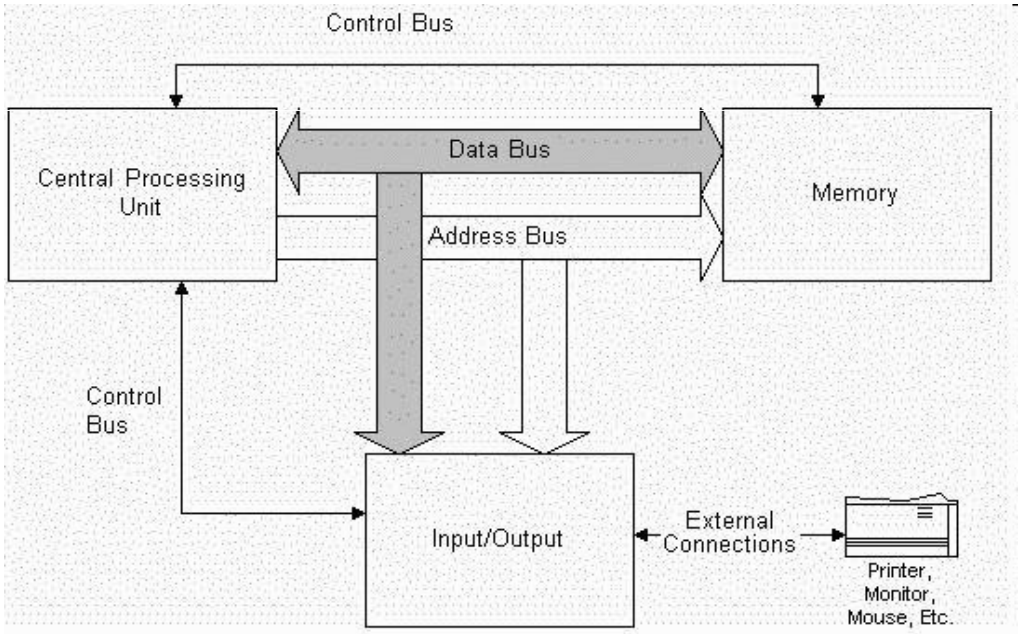


- So far we have shown the CPU handling all data transfers



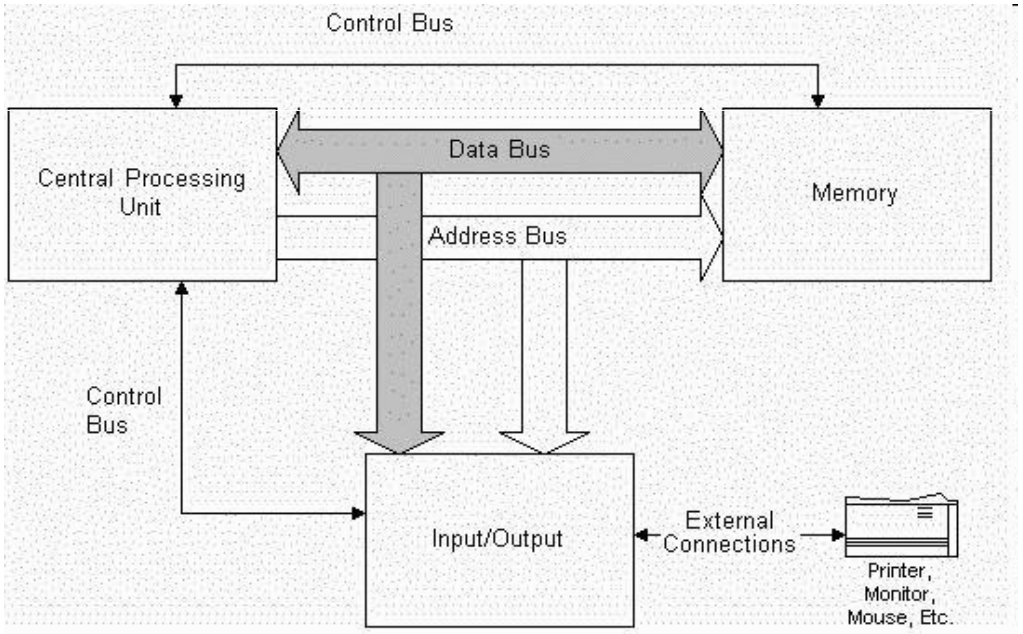
Transferring data

- What is the sequence to read data into memory from an input device?

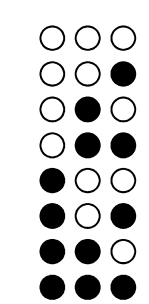


Transferring data

- CPU puts input device address on data bus
- CPU reads data into register from device using data bus
- CPU puts memory address on address bus
- CPU writes data from register to memory

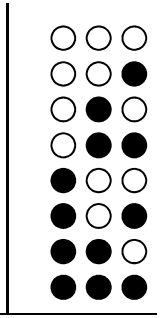


DMA – direct memory access



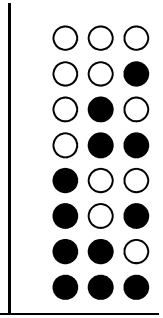
- Transferring large amounts of data will use a lot of CPU time
- This overhead can be reduced using DMA
  - Devices are fitted with their own controllers
  - DMA controllers allow devices to transfer data without involving the CPU
    - Disk drives
    - Graphics cards
    - Sound cards
    - Network cards

How does DMA work?



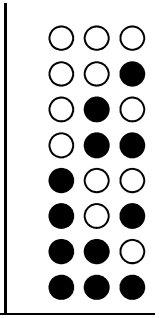
- The CPU will start a transfer by issuing a DMA read or write command to the device
  - Must include the start address and the amount of data to be transferred
- The device takes control of the bus
  - Transfers the data directly
  - Raises an interrupt when it is complete

Effect on CPU

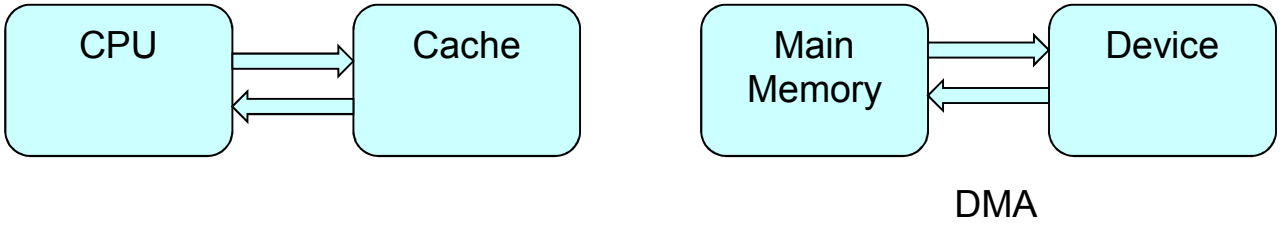


- The CPU can do other work when a DMA transfer occurs
- But if the DMA device has control of the address and data busses, what can the CPU work on?

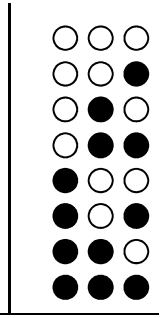
Storage hierarchy



- The CPU can work on data in the cache

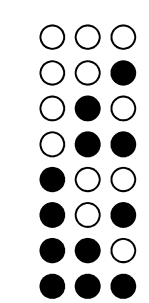


Memory performance



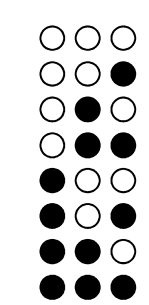
- Factors affecting memory performance
  - Width of the address bus
  - Width of the data bus
  - Memory speed
  - Memory transfers per clock cycle

SDR (single data rate RAM)



- Let us start our analysis with:
  - Bus speed = 133 MHz
  - Processor = 32 bit
  - Address bus = 32 bit
  - Data bus = 64 bit
- With SDR RAM
  - 64 bits are transferred each clock cycle
  - Transfer rate = 64 x 133 = 8512/8 = 1064 MB/s

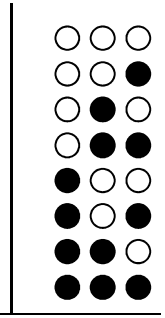
DDR (double data rate RAM)



- With DDR RAM
  - 64 bits are transferred twice each clock cycle
  - 64 x 133 x 2 = 17024/8 = 2128 MB/s

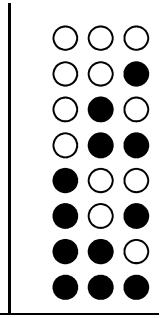
- Double the theoretical performance of SDR

DDR2



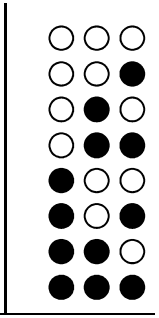
- With DDR 2 RAM
  - 64 bits are transferred four times each clock cycle
  - 64 x 133 x 4 = 34048/8 = 4256 MB/s
- Double the theoretical performance of DDR

Dual channel DDR2



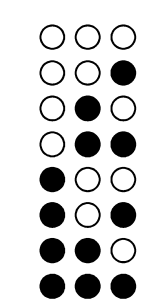
- Dual channel introduces a second memory controller and memory
- This effectively doubles the bus width to 128 bits
- With Dual channel DDR 2 RAM
  - 128 bits are transferred four times each clock cycle
  - 128 x 133 x 4 = 68096/8 = 8512 MB/s

Increasing the bus speed



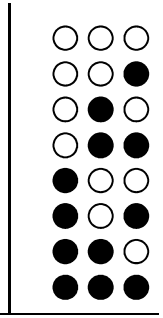
- If the bus speed is changed to 400 MHz
- With Dual channel DDR 2 RAM
  - 128 bits are transferred four times each clock cycle
  - 128 x 400 x 4 = 204800/8 = 25600 MB/s

What speeds can DDR3 reach?



- An activity for you

Task



- You have been asked to design a performance PC system
- You must make choices of
  - The bit size of the CPU (32 or 64)
  - The memory type
  - The bus speed
  - The address and data bus widths
  - Any techniques you will use (eg DMA or memory mapping)
- Describe the choices you make and how they impact the overall performance and complexity of the system