BCN1043

COMPUTER ARCHITECTURE & ORGANIZATION

By
Dr. Mritha Ramalingam

Faculty of Computer Systems & Software Engineering
mritha@ump.edu.my

AUTHORS

- **Dr. Mohd Nizam Mohmad Kahar** (mnizam@ump.edu.my)
- **Jamaludin Sallim** (jamal@ump.edu.my)
- **Dr. Syafiq Fauzi Kamarulzaman** (syafiq29@ump.edu.my)
- **Dr. Mritha Ramalingam** (mritha@ump.edu.my)

Faculty of Computer Systems & Software Engineering

BCN1043

COMPUTER ARCHITECTURE & ORGANIZATION

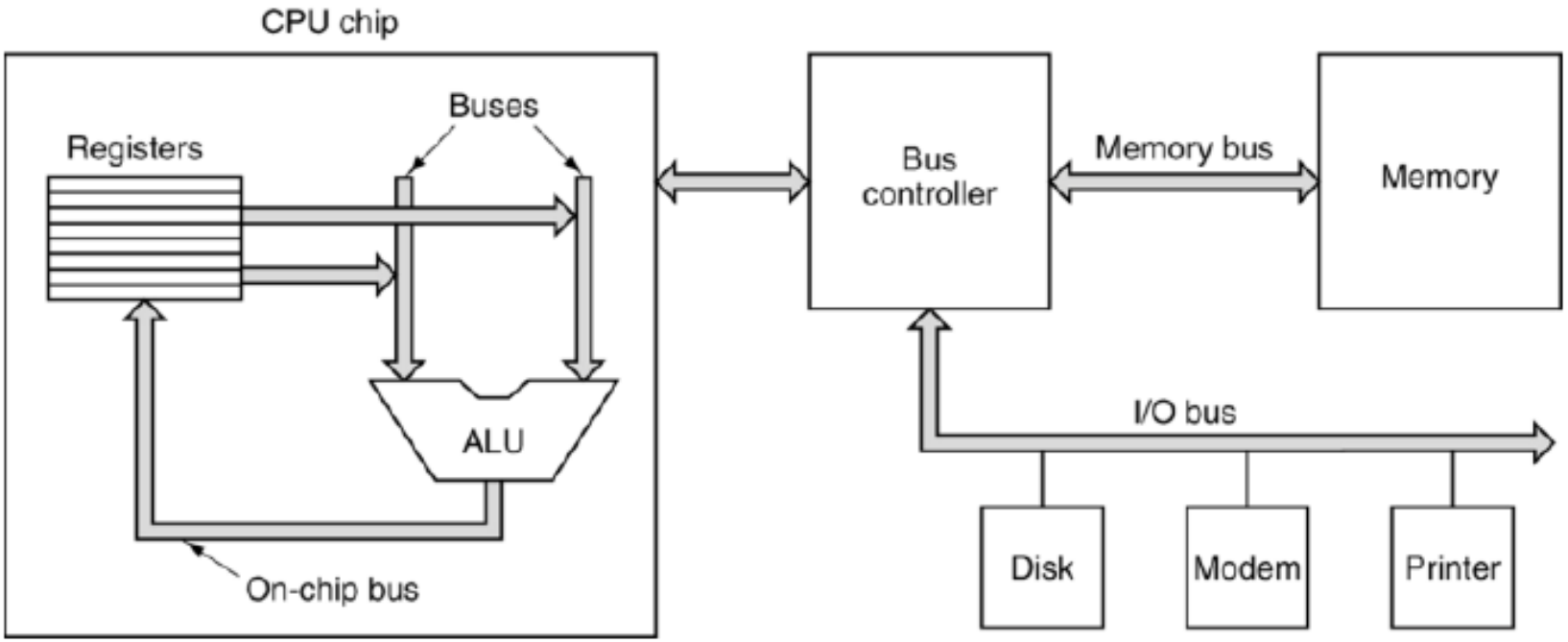
Chapter 6 continues…

BUSES

- A computer consists of a set of components (CPU, memory, I/O) that communicate with each other.
- The collection of paths connecting the various modules is call the *buses/interconnection structure*.
- The design of this structure will depend on the exchange that must be made between modules/devices.

Buses

- A bus is a common electrical pathway between multiple devices



Source: William Stallings, Computer Organization and Architecture,9th Edn

- This so called buses is used as a mechanism to let these devices communicate. So, what are their feature?
  - Is there a dedicated bus/line for each device to communicate with one another; CPU to memory, CPU to printer, memory to scanner or CPU to speaker etc? if it not dedicated, then how?
  - or is it being shared by all of these devices? Wait, if it is shared, who determine which devices will used the buses?
  - How many type of buses? Is it one, two or three?

- It is a *shared transmission* medium (allows connection to multiple devices). However, only ONE device can transmit data (or instruction) at a time.
- Each line is capable of transmitting signals represent binary values (i.e. 1 or 0).
- Hence, several lines of a bus can be used to transmit binary digits simultaneously (in parallel). For example, an 8-bit unit of data can be transmitted over eight bus lines.

Buses

- The major parts of a computer model are the central processing unit (CPU), memory and the Input and output circuitry (I/O).
- Connecting these part are 3 sets of parallel lines called buses.
- These buses consist of 50 to 100 separate lines and each line is assigned particular function.
- The 3 buses are the **address** bus, **data** bus and **control** bus.

Data bus (or data line)

- It provide a path for moving data between devices.
- Consist of from 32 to hundreds of separate lines (= width)
- Each line carry only 1 bit at a time. The number of lines (width) determine how many bits can be transferred at a time - determine the overall system performance.
- For example: if the data bus is 8-bit wide and each instruction is 16 bits long, the processor must access the memory module twice (2) during each instruction cycle.
- The data bus lines are bi-directional.
- Note: Many devices in a system will have their outputs connected to the data bus, but only one device at a time will have its output enabled.

Address Bus (or address line)

- Used to designate the source or destination of the data on the data bus (or address)
- The width of the address line determine the maximum possible memory capacity of the system.
- On these line the CPU sends out the address of the memory location that is to be written to or read from.
- For example, if the processor want to read a word (8, 16 or 32 bits) of data from memory, it puts the address of the desired word on the address line.

Address bus (cont…)

- The number of memory location that the CPU can address is determined by the number of address lines. This is calculate using 2^n for example CPU with 16 address lines can address 65536 memory (20 address line = _____ location?)
- Note: When the CPU reads data from or writes data to a port, it sends the port address out on the address bus.

Control Bus (or control lines)

- Used to control the access to and the use of the data and address lines.
  - Because data and address line are shared by all components, hence the control line act as mechanism to control their use.
- The CPU sends out signal on the control bus to enable the outputs of addressed memory devices or port devices.
- Typical control bus signals are: Memory Read, Memory Write, I/O Read, I/O Write, Bus request, Interrupt request etc.

Example:

- The control bus is an eclectic collection of signals that control how the processor communicates with the rest of the system.
- **Example**, consider data bus. The CPU sends data to memory and receives data from memory on the data bus. This prompts the question, "Is it sending or receiving?" There are two lines on the control bus, *read* and *write*, which specify the direction of data flow.

Example: READ instruction process

- Example: Read a byte of data from a memory location,
  1. the CPU sends out a Memory Read signal on the control bus. The memory Read signal enables the address memory device to output a data word on the data bus.
  2. It then sends out the memory address of the desired byte on the address bus
  3. Finally, the data word from memory travels along the data bus to the CPU.

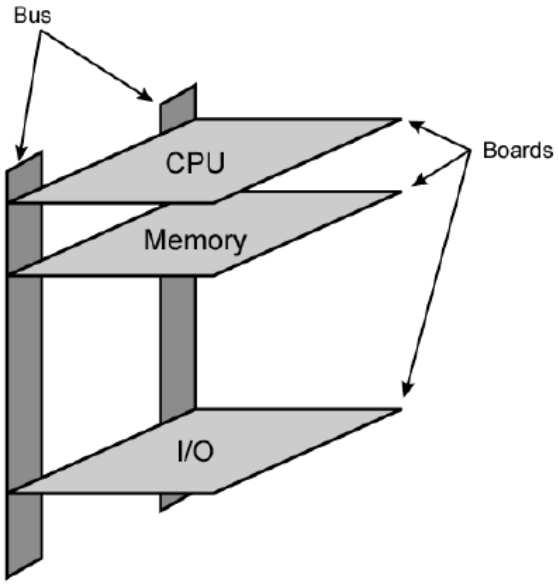
Q: Explain, how would the Write instruction process work?

summarize

- Buses - data, address and control bus → buses are bidirectional
- **Data bus** - the width determine overall system performance
- **Address bus** - the width determine the maximum possible memory capacity of the system.
- **Control bus** - determine the direction of the data/address bus flow
- These buses play an important role during reading or writing instruction process
  - If one module wishes to send data (write)
    - obtain the use of the bus
    - transfer data via the bus
  - If one module wishes to request data (read)
    - obtain the use of the bus
    - transfer request to the other module over the control and address lines, then wait for that second module to send the data.

Physical realization of Bus Architecture

- System bus is a number of parallel electrical conductors.
- The conductors are metal lines etched in a card or printed circuit board (PCB).
- The bus extends across all of the components that taps into the bus lines.
- Analogy: highway interchange



Source: William Stallings, Computer Organization and Architecture,9th Edn

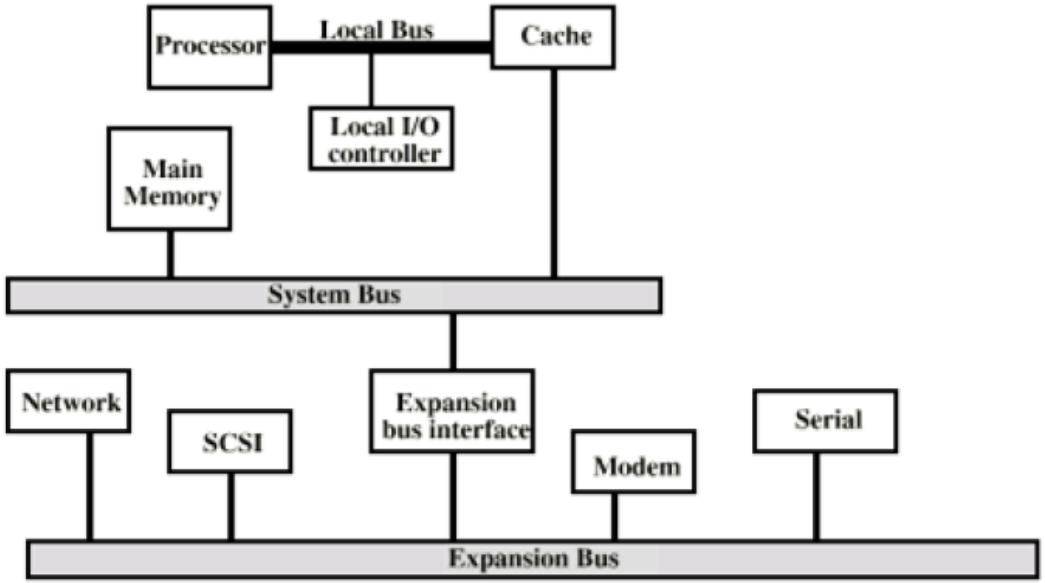
Single Bus problems

- The more devices attached to the bus
  - The greater the bus length - Therefore propagation time is more. This propagation delay can affect performance. When control of the bus passes from one device to another frequently
  - The bus may become bottleneck as aggregate data transfer demand approaches the capacity of bus. Because data rate generated by attached devices like graphics and video controller are growing rapidly
  - Only one master bus can operate at a time, other waits.
- Most system use multiple buses to overcome these problems.

Traditional bus architecture

- The traditional bus connection uses three buses local bus, system bus and expansion bus
  - **Local bus** connects the processor to cache memory
  - The cache memory controller connects the cache to local bus and to the system bus.
  - **System bus** also connects main memory module
  - Input/output transfer to and from the main memory across the system bus, and do not interface with the processor activity because processor accesses cache memory.

    - It is possible to connect I/O controllers directly on to the system bus?



Source: William Stallings, Computer Organization and Architecture,9th Edn

Summarize

- Single bus vs multiple buses
  - Bus length = propagation delay
  - Bottleneck = aggregated data
- Traditional bus vs high performance architecture

Bus Design

1. Type - Dedicated or Multiplexed
2. Bus Width - Address or Data
3. Timing - Synchronous or Asynchronous
4. Method of Arbitration - Centralized or Distributed
5. Data Transfer Type
   - Read
   - Write
   - Read-modify-write
   - Read-after-write
   - Block

bus Type

- **Dedicated bus** - When a bus is permanently assigned only 1 function, it is called dedicated bus.
  e.g. separate address and data lines, with separate bus for memory and I/O modules
  Advantages: It gives high performance and less bus contention
  Disadvantages: Increased size and cost.

- **Multiplexed bus -** When the bus is used for more than 1 function.
  e.g. 8085 microprocessor outputs A7- A0 in first clock cycles on pins. AD7 – AD0.
  Advantages: few pins lines are required. less cost and save space
  Disadvantages: slow in speed

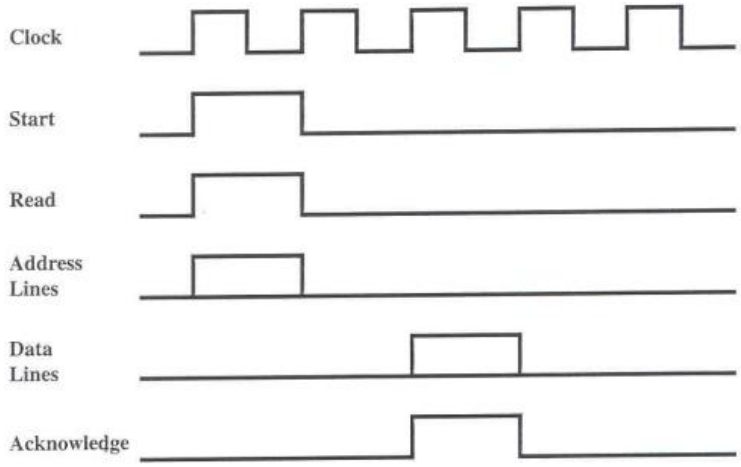
Bus Width

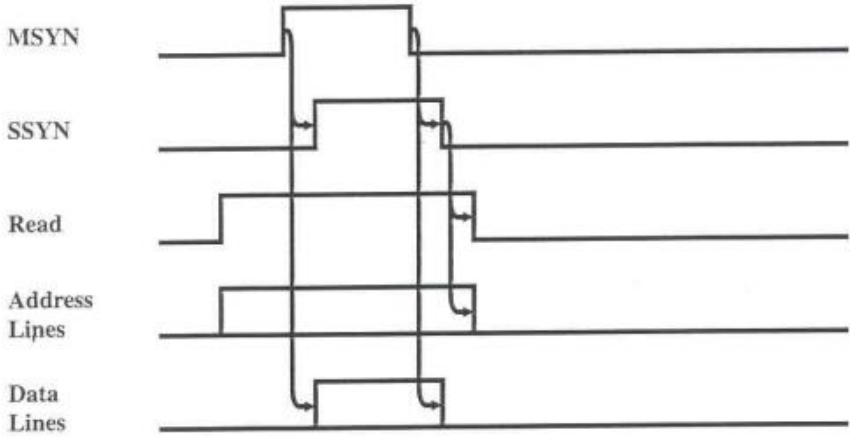
- It decides the number of lines to be used for address and data.
- Data the wider of data bus has an impact on the number of bits transferred at one time. Therefore speed increases.
- Address the wider of address bus has an impact on range of locations that can be referenced/accessed; e.g. 16 line address make 2^{16} = 64 kb , 20 address line makes 2^{20} = 1 Mb memory access .

Timing

- **Synchronous**
  occurrence of events on the bus is determined by a clock (Clock Cycle or Bus Cycle)

- **Asynchronous**
  occurrence of one event follows and depends on the previous event of bus.





Source: William Stallings, Computer Organization and Architecture,9th Edn

Data Transfer Type

- **Read Dedicated** address is put on bus and remain there while data are put on the data bus
- **Write Dedicated** data put on data bus as soon as the address has stabilized
- **Read Multiplexed** bus is used to specifying address and then for transferring data after a wait while data is being fetched
- **Write Multiplexed** bus is used to specifying address and then transferring data (same as read operation)
- **Read-modify-write** address is broadcast once at beginning a simply read is followed immediately by a write to the same address
- **Read-after-write** a write followed immediately by a read from the same address,performed for checking purposes
- Block one address cycle is followed by n data cycles.
- The first data item is transferred to or from the specified address; remainder data items are transferred to or from subsequent addresses

Chapter 6 ends!