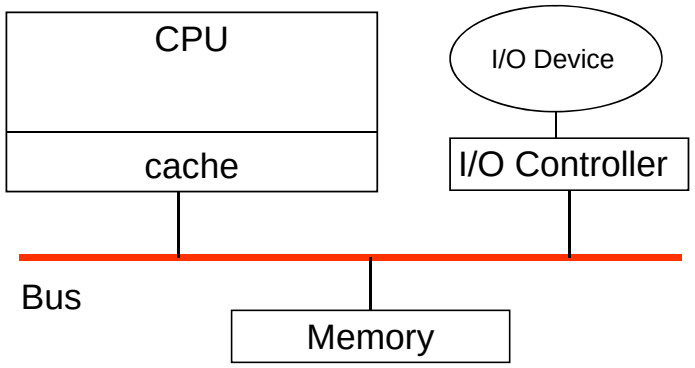
Buses

COMP375 Computer Architecture and Organization

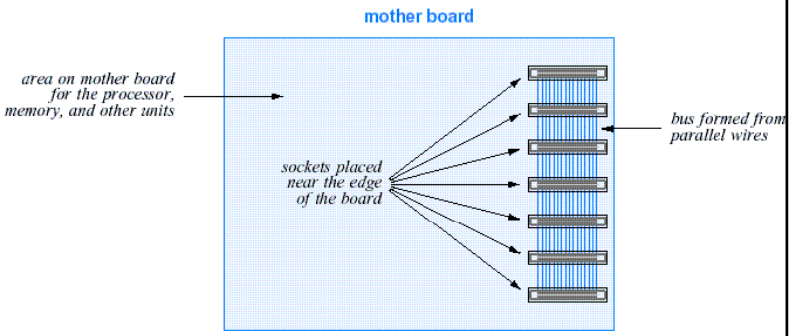
Definition Of A Bus

- Digital interconnection mechanism
- A set of parallel wires with rules for putting and retrieving information on the wires.
- A digital communication mechanism that allows two or more functional units to transfer control signals or data.
- The connection medium allowing the CPU, memory and I/O controllers to communicate

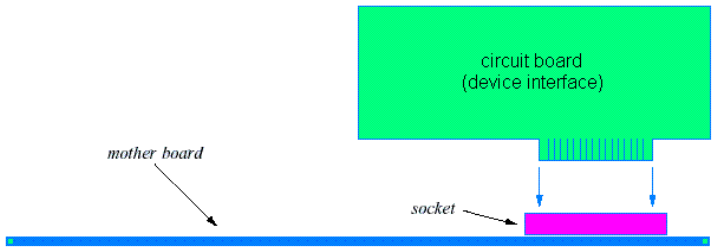
Basic Computer Components



Physical Connections



Physical Bus Interface



External and Internal

- Internal buses
  - connect the primary computer components
  - transfer data at up to 20 GB/sec
  - about 30 – 40 cm maximum length
- External buses
  - USB, IEEE-1394 and Firewire
  - USB 2.0 runs at 60 MB/sec
  - USB 2.0 cables can be 5m in length
  - Allow hot swapping of devices

Bus Design Issues

- Dedicated or Multiplexed
- Width
- Access Protocol
- Arbitration
- Timing
- Operation

Bus Lines

- Transfer of data
- Address information
- Control of the bus
  - Memory fetch or store
  - Ready
  - Bus Request and Bus Grant
  - Interrupt and Interrupt Acknowledge
  - Clock

Bus Width

- The width of a bus is the number of lines.
- The more data lines, the more data that can be transferred simultaneously.
- A “32 bit bus” has 32 data lines.
- The more address lines, the larger the maximum amount of memory that can be accessed.
- The greater the width, the more hardware required to implement the bus.

Example Bus Width

- Pentium is a 32-bit processor with a 64-bit data bus
- Itanium is a 64-bit processor with a 128-bit data bus
- Address bus width
  - Determines the system addressing capacity
  - *N* address lines directly address 2^N memory locations
    - 8086: 20 address lines
      - Can address 1 MB of memory
    - Pentium: 32 address lines
      - Can address 4 GB of memory
    - Itanium: 64 address lines
      - Can address 2^{64} bytes of memory
    - AMD Athlon™ 64: 40 address lines
      - Can address 1 TB of memory

Dedicated or Multiplexed

- With a dedicated bus there are separate wires for data and addresses.
- With a multiplexed bus, the same lines are used at different times to hold either data or addresses.

Dedicated Bus

- A store operation can put both the address and the data on the bus at the same time.
- Having separate data and address lines simplifies the bus protocol.

Multiplexed Bus

- Multiplexed buses require fewer lines.
- Chips can be limited in the number of pins that can be physically attached.
- For a given number of pins, it is usually advantageous to transfer more data.
- Data and addresses may appear on the bus at different times.

Fetch-Store Paradigm

- A processor can fetch a value from memory or store a value to memory.
- Fetch and store are also used to transfer data to an I/O device.
- Only one device at a time can put a value on the bus data or address lines.

Fetch

1. Use control lines to obtain access to bus
2. Place an address on the address lines
3. Use control lines to request a fetch operation
4. Wait until operation complete
5. Read the value from the data lines
6. Set controls line to allow another device to use the bus

Store

1. Use control lines to obtain access to bus
2. Place an address on the address lines
3. Place value on the data lines
4. Use control lines to specify a store function
5. Wait until operation complete
6. Set controls line to allow another device to use the bus

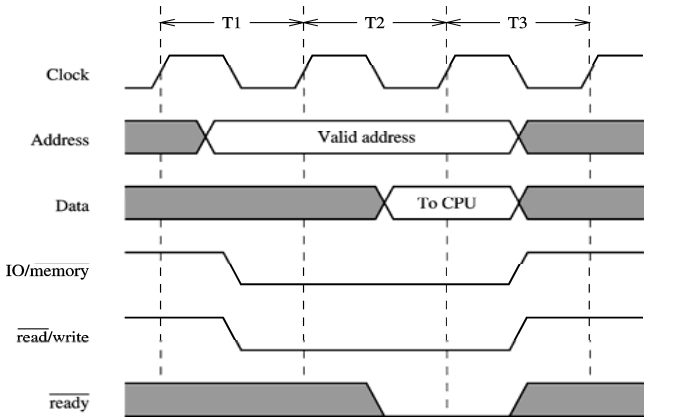
Block Transfers

- With cache systems, memory requests to the RAM are for a whole line of data.
- The CPU requests an address and the RAM provides a series of data values.
- I/O controllers may still communicate with the CPU or the memory with arbitrarily sized data.

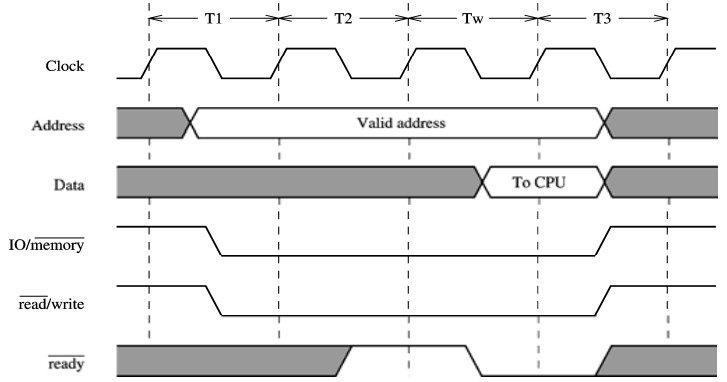
Wait States

- Some devices are not as fast as the CPU.
- When the CPU requests data from RAM or an I/O device, it may not be able to get it the next clock cycle.
- A wait state is created when the CPU must wait for a device to be ready.
- The CPU needs to wait until a device signals it is ready to provide the data.

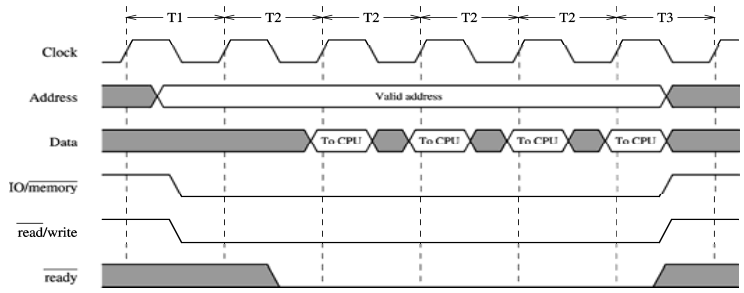
Memory read with no wait states



Memory read with a wait state



Block transfer of data



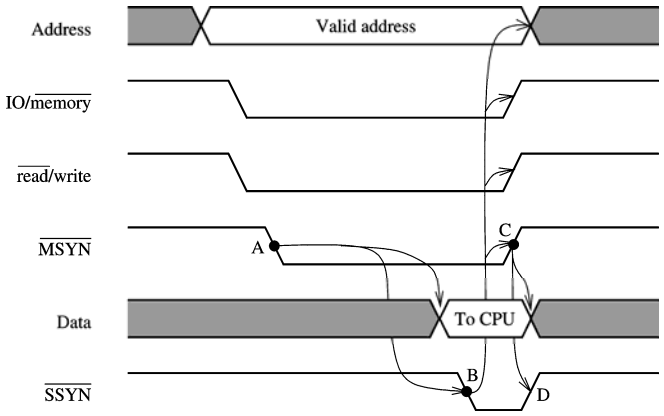
Cycles and Bus Width

- If a bus is **n** bits wide, it can transfer **n** bits (or **n**/8 bytes) every cycle.
- If you need to transfer more than **n** bits, it will take multiple cycles.
- To transfer **x** bytes over an **n** bit wide bus, it takes **8x / n** cycles plus any cycles previous to the data transfer.

Synchronous or Asynchronous

- In a synchronous bus, a clock signal provides timing for all operations.
- A device presents the address on a given clock pulse and expects the data during another predefined clock pulse.
- In an asynchronous bus, a device waits for a ready signal to know when data is available.

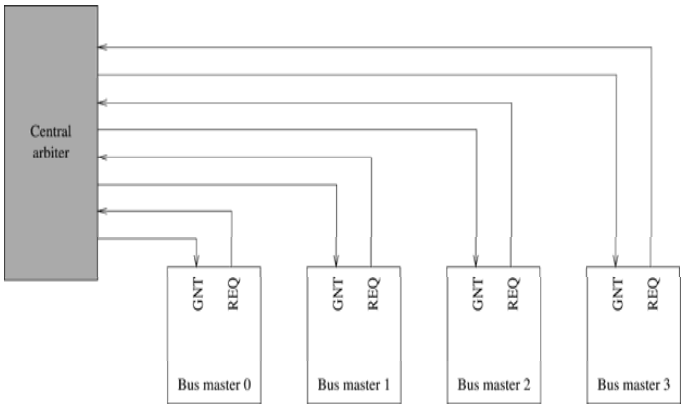
Asynchronous Bus



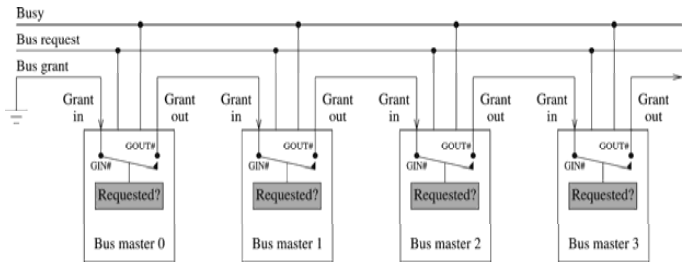
Arbitration

- Only one device can put data on the bus at a time. Many devices can sense the data, but only one can assert it.
- The bus arbitration protocol determines which device gets to use the bus at any given time.
- Bus arbitration can be centralized or distributed.

Centralized Arbitration



Daisy Chain Bus



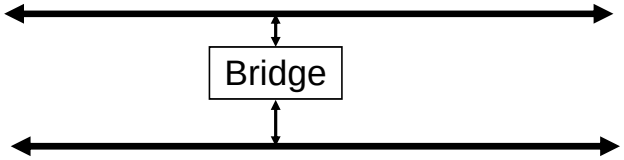
The devices determine who gets to use the bus.

Multiple Buses

- A single computer usually has several buses.
- Different devices have different requirements
- A 56K modem only needs about 7 KB/sec bandwidth while a graphics device may need 70 MB/sec or more.
- Multiple buses allow devices using different technologies to connect to the same computer.

Bridging Buses

- A bridge is a device that connects two buses.
- A bridge converts the addresses and protocols of one bus to another.



Chipsets

- The chipset controls the bus.
- Intel Pentium® 4 processors are available with system bus speeds of 400, 533, 800 and 1066 MHz
- Intel's latest chipset is the Intel X38 Express chipset which operates at 1333 MHz and can transfer data at up to 21.2 GB/s

Frontside Bus

- The Frontside bus is the primary bus connecting the processor to the RAM.
- It is the fastest bus in the system.
- All other buses connect to the frontside bus, either directly or indirectly.
- The computers chipset controls the frontside bus and bus bridges.

Bus Standards

- To allow different equipment to connect together, devices need to follow standards
- Bus designs are often developed by individual companies and then standardized by industry organizations.

ISA Bus

- The Industry Standard Architecture (ISA) bus was used in the first 8088 PCs.
- It was originally a 8 bit data bus with 20 dedicated address lines.
- Bus design similar to 8088 local bus.
- The bus was updated to have 16 data lines and operate at 8.33 MHz providing 8 MB/second bandwidth.
- Updated again to the Extended ISA (EISA)

PCI Bus

- The Peripheral Component Interconnection (PCI) bus was developed by Intel in the early 1990s.
- PCI has 64 data lines running at 66 MHz providing up to 528 MB/sec bandwidth.
- Data and address lines are multiplexed.
- Centralized arbitration.

SCSI Bus

- Pronounced “scuzzy”
- Small Computer System Interface
  - Supports both internal and external connection
- Comes in multiple bus widths
- Allows the connection of up to 16 devices. Each device has a bus ID.
- Popular for hard drive

Types of SCSI Buses

SCSI type	Bus width (bits)	Transfer rate MB/s
SCSI 1	8	5
Fast SCSI	8	10
Ultra SCSI	8	20
Ultra 2 SCSI	8	40
Wide Ultra SCSI	16	40
Wide Ultra 2 SCSI	16	80
Ultra 3 (Ultra 160) SCSI	16	160
Ultra 4 (Ultra 320) SCSI	16	320

Accelerated Graphics Port

- Designed to provide a high speed path to the monitor.
- Provides connection for only one device.
- **AGP 8x**, double-pumped at 266 MHz to give a maximum of 2.133 MB / second.
- *Double-pumping* means transferring data on both the rising and falling edges of the clock waveform

Beyond Buses

- Although there are several devices that communicate over the bus, only one device can send data at a time.
- Other interconnection schemes allow multiple simultaneous connections.
- There are many designs of *switching fabrics* to interconnect devices.
- Most switching fabrics can be expensive to implement.

Crossbar Switch