Computer Architecture

Lecture 04

Interconnection Structure

Reference Books

- Computer Organization and Architecture: Designing for Performance-William Stallings
  (8th Edition)
  – Any later edition is fine

Overview

- The collection of paths connecting the various modules is called the *interconnection structure*
- The interconnection structure must support the following types of transfers:
  - Memory to processor: The processor reads an instruction or a unit of data from memory
  - Processorto memory: The processor writes a unit of data to memory
  - I/O to processor:The processor reads data from an I/O device via an I/O module
  - Processor to I/O: The processor sends data to the I/O device
  - I/O to or from memory: For these two cases, an I/O module is allowed to exchange data directly with memory, without

BUS Interconnection

- A bus is a communication pathway connecting two or more devices
  - It is a shared transmission medium
  - Consists of multiple communication pathways, or lines, each line is capable of transmitting signals representing binary 1 and binary 0
- Multiple devices connect to the bus, and a signal transmitted by any one device is available for reception by all other devices attached to the bus
- If two devices transmit during the same time period, their signals will overlap and become garbled
  - Only one device at a time can successfully transmit
- A bus that connects major computer components (processor, memory, I/O) is called a system bus

Bus Structure

- A system bus consists, typically, of from about 50 to hundreds of separate lines
- Each line is assigned a particular meaning or function
- Any bus the lines can be classified into three functional groups: data, address, and control lines
- There may be power distribution lines that supply power to the attached modules

DATA BUS

- The data lines provide a path for moving data among system modules
  - These lines, collectively, are called the *data bus*
- The data bus may consist of 32, 64, 128, or even more separate lines, the number of lines being referred to as the *width* of the data bus
- Because each line can carry only 1 bit at a time, the number of lines determines how many bits can be transferred at a time
- Width: If the data bus is 32 bits wide and each instruction is 64 bits long, then the processor must access the memory module twice during each instruction cycle

ADDRESS BUS

- The address lines are used to designate the source or destination of
  the data on the data bus
- For example, if the processor wishes to read a word (8, 16, or 32 bits) of data from memory, it puts the address of the desired word on the address lines
- The width of the address bus determines the maximum possible
  memory capacity of the system.
- Typically, the higher-order bits are used to select a particular module on the bus, and the lower-order bits select a memory location or I/O port within the module
- For example, on an 8-bit address bus, address 01111111 and below might reference locations in a memory module (module 0) with 128 words of memory, and address 10000000 and above refer to devices attached to an I/O module (module 1)

CONTROL BUS

- The control lines are used to control the access to and the use of the data and address lines
- Control signals transmit both command and timing information among system modules
  - Timing signals indicate the validity of data and address information
  - Command signals specify operations to be performed
- Typical control lines include:
  - Memory write, Memory read, I/O write, I/O read, Transfer ACK, Bus request, Bus grant, Interrupt request, Interrupt ACK, Clock, Reset

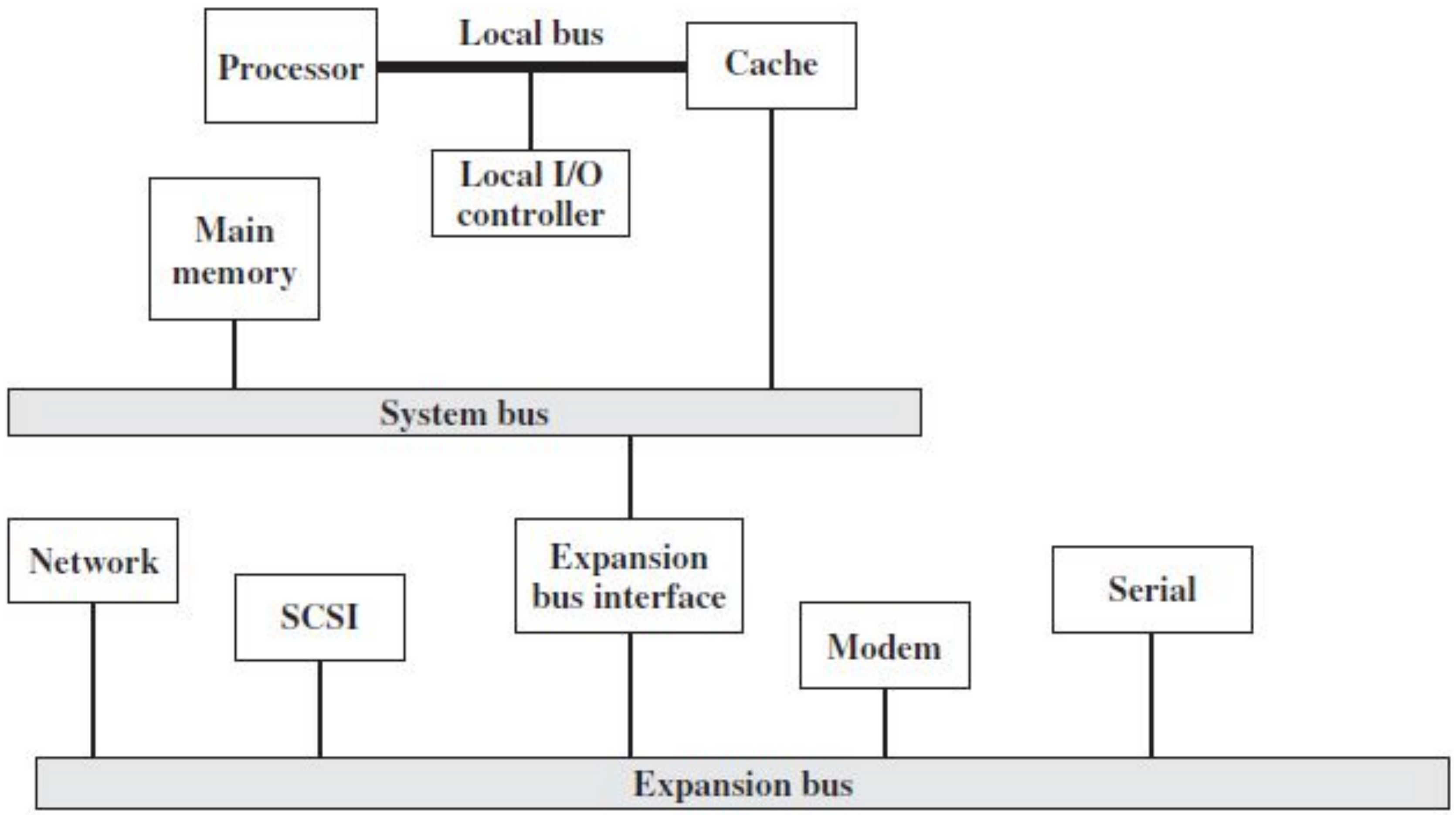
Multiple-Bus Hierarchies

- Motivated by two major concerns:

1) In general, the more devices attached to the bus, the greater the bus length and hence the greater the propagation delay
2) The bus may become a bottleneck as the aggregate data transfer demand approaches the capacity of the bus

- Most computer systems use multiplebuses, generally laid out in a hierarchy

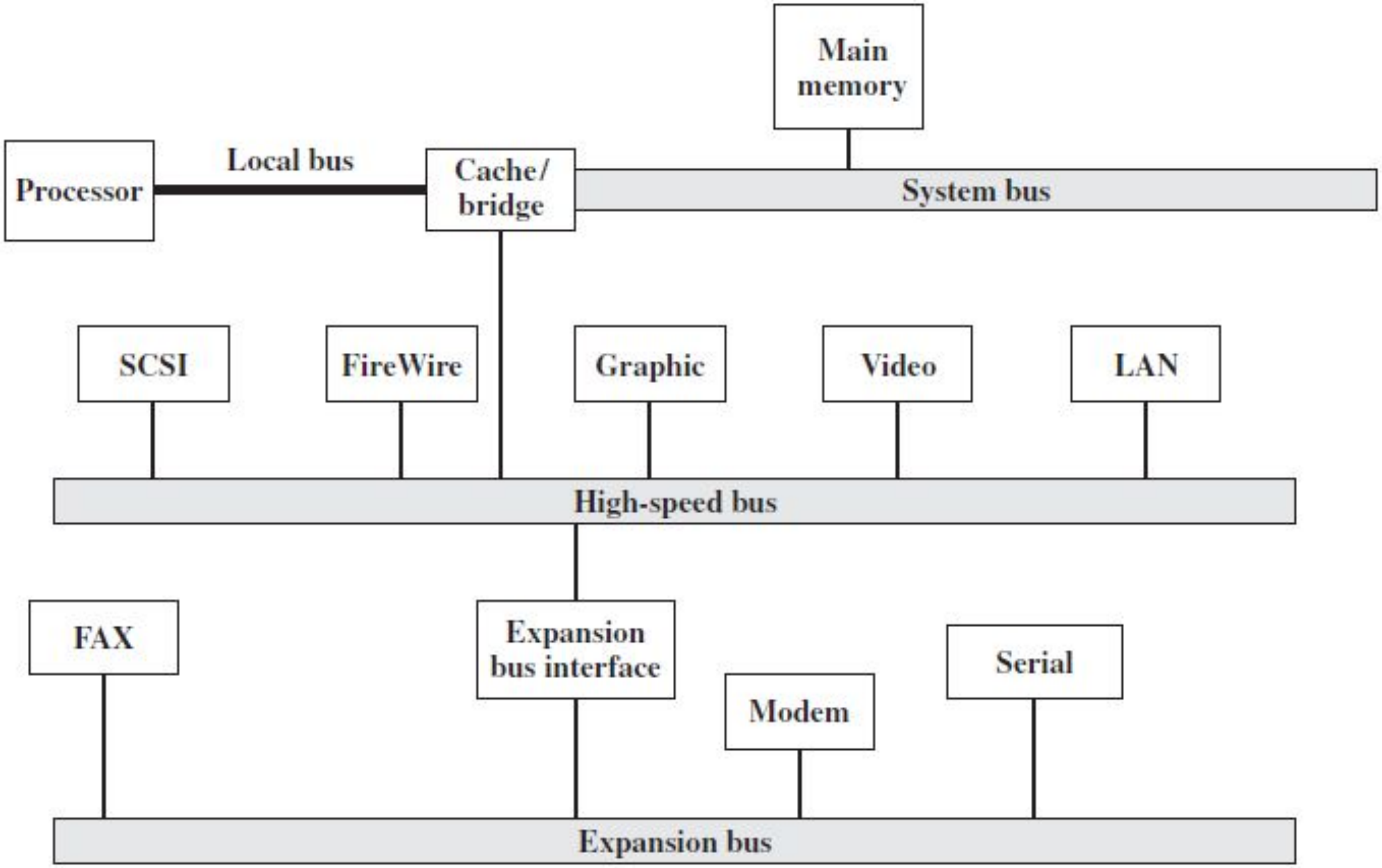
Traditional BUS Architecture



Traditional BUS Architecture

- Network connections include local area networks (LANs) such as a 10-Mbps Ethernet and connections to wide area networks (WANs) such as a packet-switching network
- SCSI (small computer system interface) is itself a type of bus used to support local disk drives and other peripherals
- A serialport couldbe used to support a printer or scanner

High-Performance



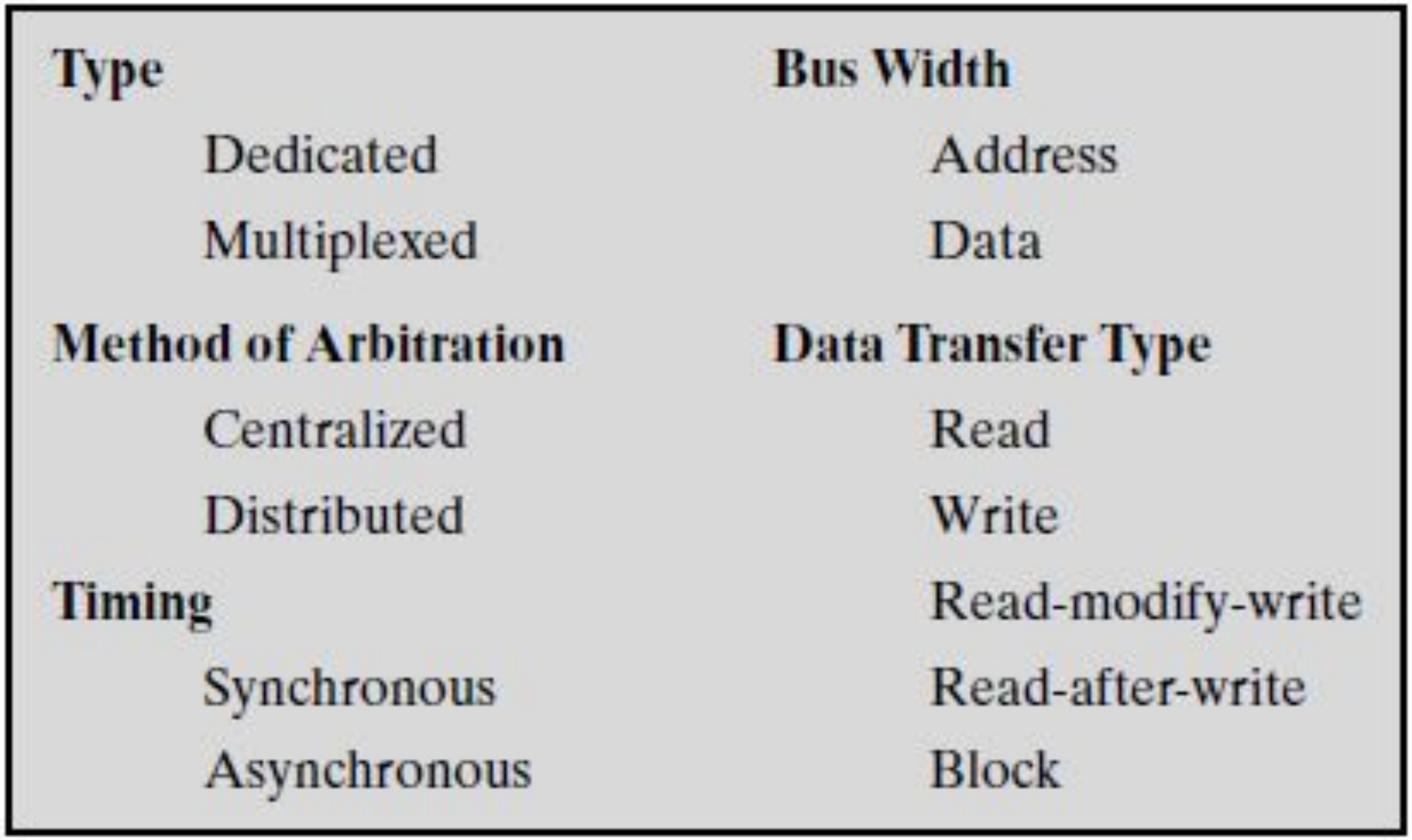
High-Performance Architecture

- There is a local bus that connects the processor to a cache controller, which is in turn connected to a system bus that supports main memory
- The cache controller is integrated into a bridge, or buffering device, that connects to the high-speed bus
- This bus supports connections to high-speed LANs, such as Fast Ethernet at 100 Mbps, video and graphics workstation controllers, as well as interface controllers to local peripheral buses, including SCSI and FireWire
- The latter is a high-speed bus arrangement specifically designed to support high-capacity I/O devices
- Lower-speed devices are still supported off an expansion bus, with an interface buffering traffic between the expansion bus and the high-speed bus

High-Performance Architecture

- The advantage of this arrangement is that the high-speed bus brings high demand devices into closer integration with the processor and at the same time is independent of the processor
- Thus, differences in processor and high-speed bus
  speeds and signal line definitions are tolerated
- Changes in processor architecture do not affect
  the high-speed bus, and vice versa

Elements of Bus Design



Type

- Dedicated
- Multiplexed

Method of Arbitration

- Centralized
- Distributed

Timing

- Synchronous
- Asynchronous

Bus Width

- Address
- Data

Data Transfer Type

- Read
- Write
- Read-modify-write
- Read-after-write
- Block

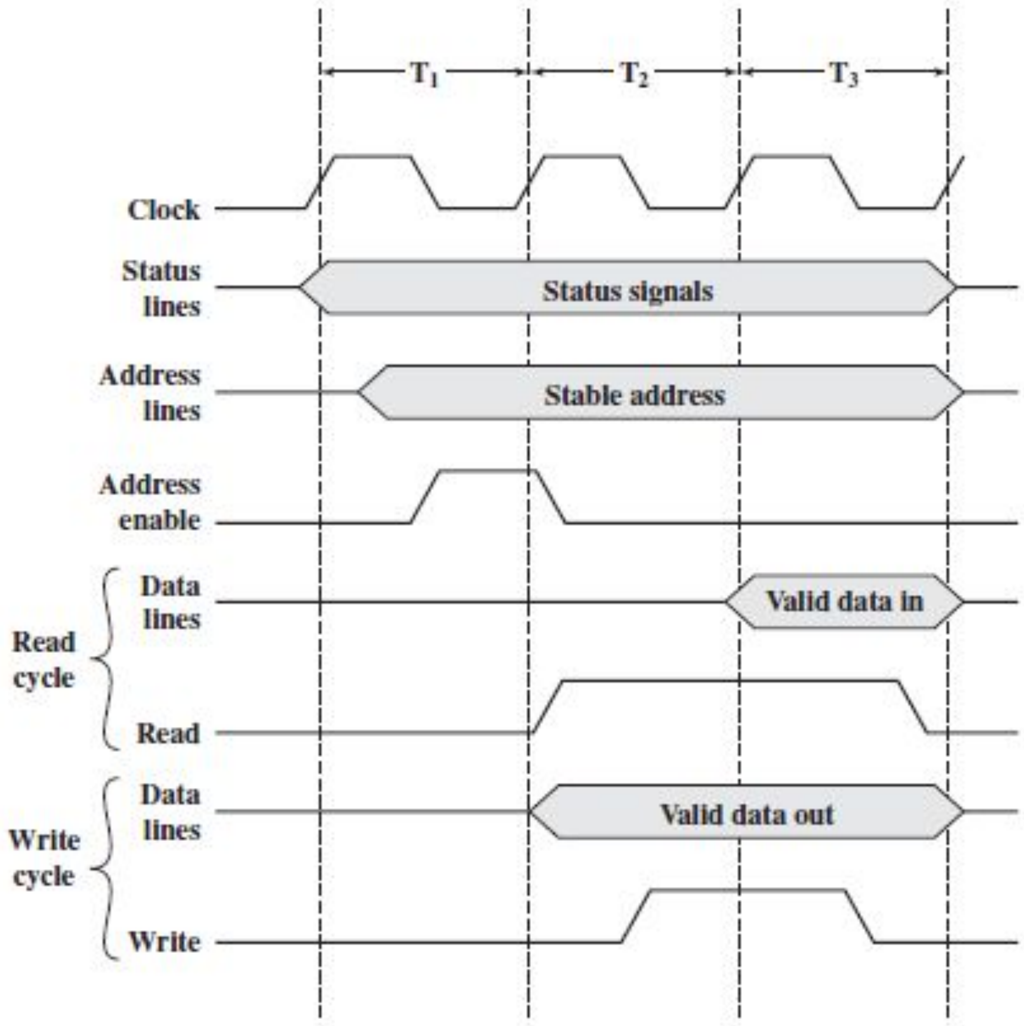
Method of Arbitration

- Roughly Two categories: Centralized and Distributed
- In a centralized scheme, a single hardware device, referred to as a *bus controller* or *arbiter,* is responsible for allocating time on the bus
- The device may be a separate module or part of the

  processor
- In a distributed scheme, there is no central controller
- Rather, each module contains access control logic and the modules act together to share the bus
- With both methods of arbitration, the purpose is to designate one device, either the processor or an I/O module as master

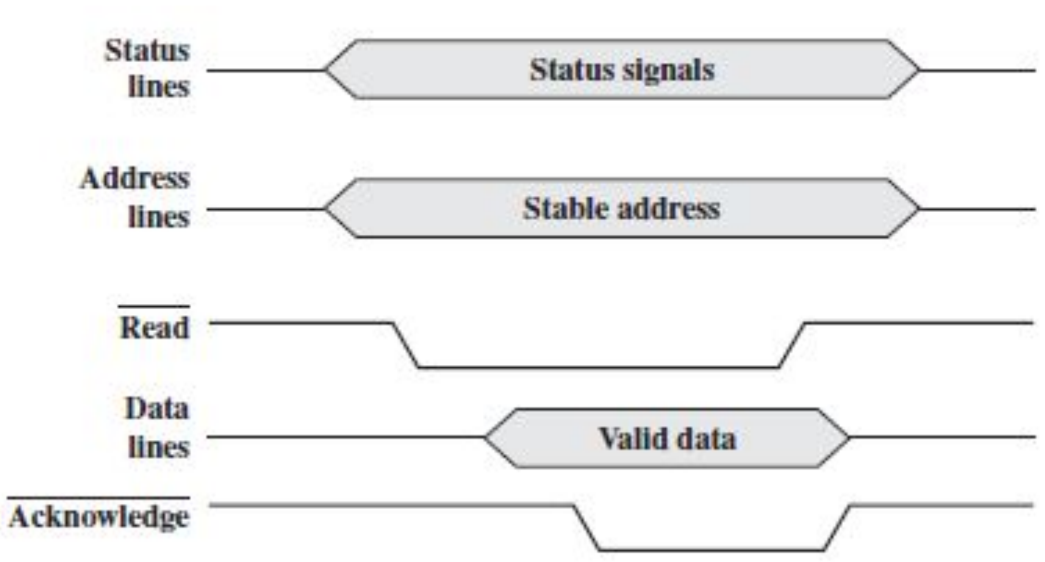
Timing

- With **synchronous timing**, the occurrence of events on the bus is determined by a clock
- The bus includes a clock line upon which a clock transmits a regular sequence of alternating 1s and 0s of equal duration



Timing

- With **asynchronous timing,** the occurrence of one event on a bus follows and depends on the occurrence of a previous event


Status lines
Status signals
Address lines
Stable address
Read
Data lines
Valid data
Acknowledge