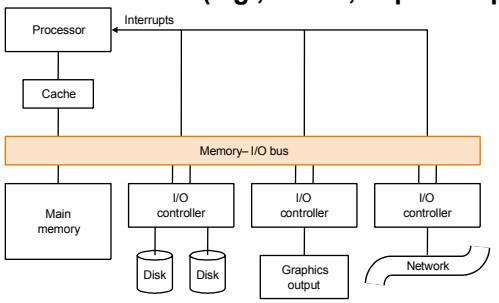
Chapters 8

Interfacing Processors and Peripherals

- **I/O Design affected by many factors (expandability, resilience)**
- **Performance:**
  - **access latency**
  - **throughput**
  - **connection between devices and the system**
  - **the memory hierarchy**
  - **the operating system**
- **A variety of different users (e.g., banks, supercomputers, engineers)**

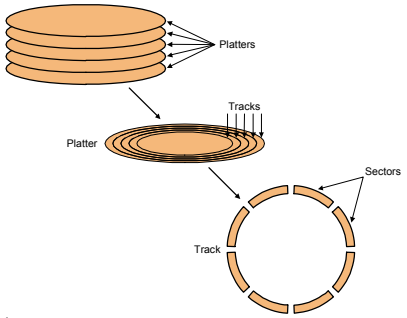


I/O

- **Important but neglected**
  - **"The difficulties in assessing and designing I/O systems have often relegated I/O to second class status"**
- **Performance measures for I/O systems: see pp.639**
- **Very diverse devices (Section 8.3)**
  - **behavior (i.e., input vs. output), partner (who is at the other end?)**

Device	Behavior	Partner	Data rate (KB/sec)
Keyboard	input	human	0.01
Mouse	input	human	0.02
Voice input	input	human	0.02
Scanner	input	human	400.00
Voice output	output	human	0.60
Line printer	output	human	1.00
Laser printer	output	human	200.00
Graphics display	output	human	60,000.00
Modem	input or output	machine	2.00-8.00
Network/LAN	input or output	machine	500.00-6000.00
Floppy disk	storage	machine	100.00
Optical disk	storage	machine	1000.00
Magnetic tape	storage	machine	2000.00
Magnetic disk	storage	machine	2000.00-10,000.00

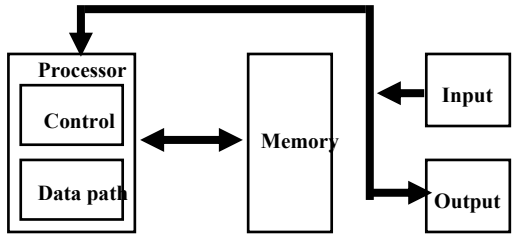
I/O Example: Disk Drives



- **To access data:**
  - **seek: position head over the proper track (8 to 20 ms. avg.)**
  - **rotational latency: wait for desired sector (.5 / RPM), see pp.647**
  - **transfer: grab the data (one or more sectors) 2 to 15 MB/sec**
- **Disk Read time: average seek time + average rotational delay+transfer time +controller overhead; see pp.648**

8.4 Buses

- What is a bus? A bunch of wires;Shared communication link (one or more wires); Used to connect multiple subsystems: A Bus is also a fundamental tool for composing large, complex systems

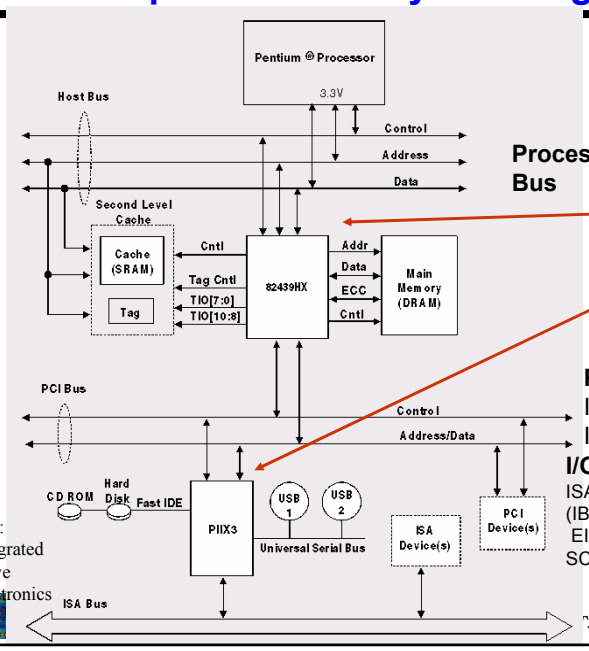


- Difficult design:
  - may be bottleneck
  - length of the bus
  - number of devices
  - cost
  - tradeoffs (buffers for higher bandwidth increases latency)
  - support for many different devices
- Advantages of bus: versatility
  - New devices can be added easily
  - Peripherals can be moved between computer systems that use the same bus standard

Types of Buses

- Processor-Memory Bus (design specific)
  - Short and high speed
  - Only need to match the memory system
    - Maximize memory-to-processor bandwidth
  - Connects directly to the processor
  - Optimized for cache block transfers, custom design
- I/O Bus (industry standard, e.g., SCSI:small computer system interface)
  - Usually is lengthy and slower
  - Need to match a wide range of I/O devices
  - Connects to the processor-memory bus or backplane bus
- Backplane Bus (standard or proprietary, e.g., PCI: Peripheral component interconnect bus)
  - Backplane: an interconnection structure within the chassis
  - Allow processors, memory, and I/O devices to coexist
  - Cost advantage: one bus for all components

Example: Pentium System Organization



Processor/Memory Bus

Northern Bridge

Southern Bridge

PCI Bus

I: 4byte per cycle@33 MHz

II:8byte per cycle @66 MHz

I/O Busses

ISA:Industry standard architecture (IBM PC/AT2 bytes per cycle@8.33 MHz; EISA:4 bytes per cycle@8.33 MHz)

SCSI: ultra SCSI; 2 bytes @20 Mhz

IDE: Integrated Drive Electronics

The General Organization of a Bus

- **Control lines:**
  - **Signal requests and acknowledgments**
  - **Indicate what type of information is on the data lines**
- **Data lines carry information between the source and the destination:**
  - **Data and Addresses**
  - **Complex commands**
- **What defines a bus ?**

Transaction Protocol

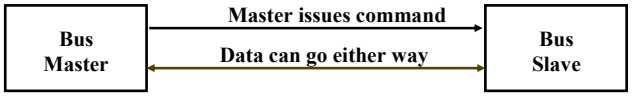
Timing and Signaling Specification

Bunch of Wires

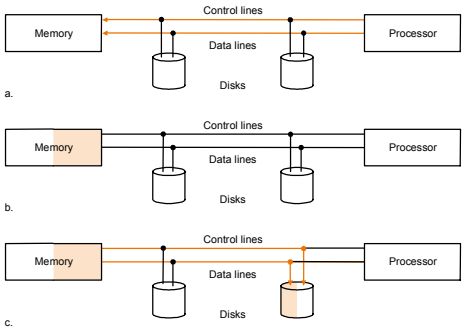
Electrical Specification

Physical / Mechanical Characterisics ?the connectors

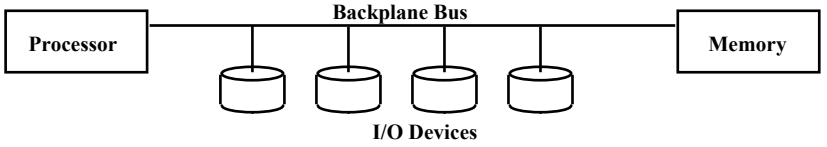
Master versus Slave



- A bus transaction includes three parts:
  - Arbitration:Which bus master gets to use the bus next
  - Request: Issuing the command (and address)
  - Action:Transferring the data
- Master is the one who starts the bus transaction by issuing the command (and address)
- Slave is the one who responds to the request by sending data to the master if the master ask for data or receiving data from the master if the master wants to send data

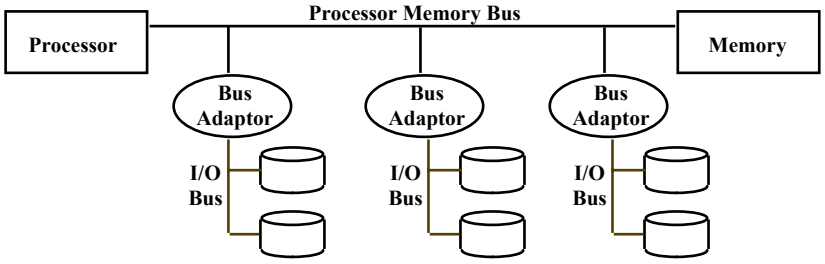


A Computer System with One Bus: Backplane Bus



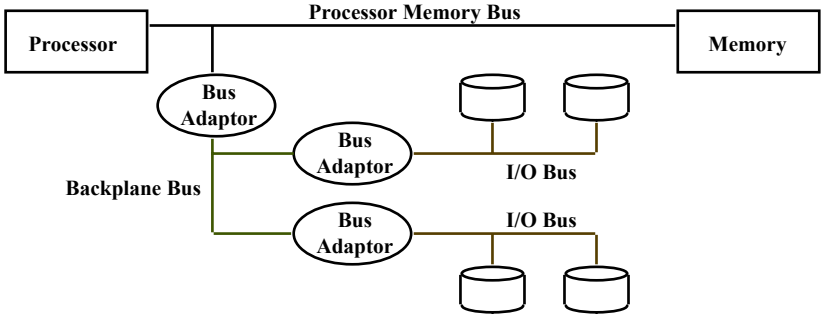
- A single bus (the backplane bus) is used for:
  - Processor to memory communication
  - Communication between I/O devices and memory
- Advantages: Simple and low cost
- Disadvantages: slow and the bus can become a major bottleneck
- Example: IBM PC - AT

A Two-Bus System



- I/O buses tap into the processor-memory bus via bus adaptors:
  - Processor-memory bus: mainly for processor-memory traffic
  - I/O buses: provide expansion slots for I/O devices
- Apple Macintosh-II
  - NuBus: Processor, memory, and a few selected I/O devices
  - SCCI Bus: the rest of the I/O devices

A Three-Bus System

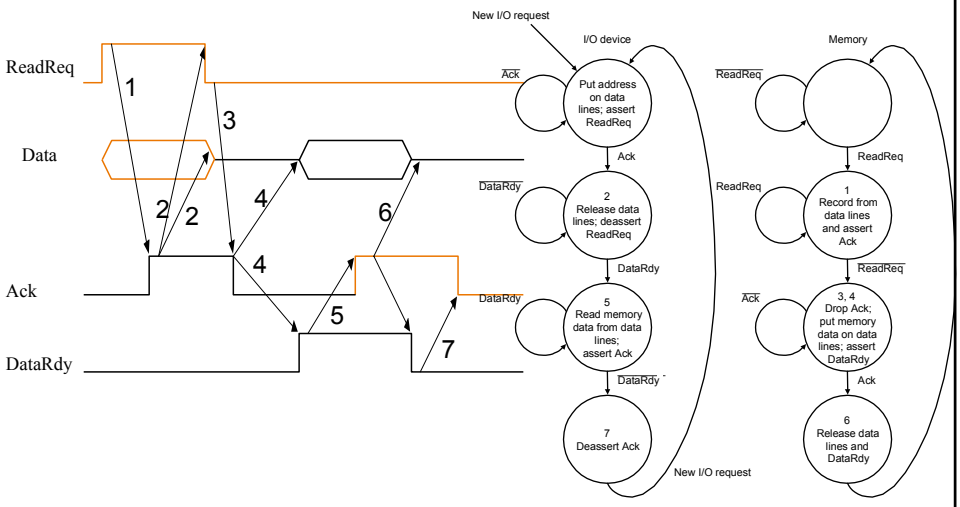


- A small number of backplane buses tap into the processor-memory bus
  - Processor-memory bus is used for processor memory traffic
  - I/O buses are connected to the backplane bus
- Advantage: loading on the processor bus is greatly reduced

Synchronous and Asynchronous Bus

- **Synchronous Bus:**
  - **Includes a clock in the control lines**
  - **A fixed protocol for communication that is relative to the clock**
  - **Advantage: involves very little logic and can run very fast**
  - **Disadvantages:**
    - Every device on the bus must run at the same clock rate
    - To avoid clock skew, they cannot be long if they are fast
- **Asynchronous Bus:**
  - **It is not clocked**
  - **It can accommodate a wide range of devices**
  - **It can be lengthened without worrying about clock skew**
  - **It requires a handshaking protocol**

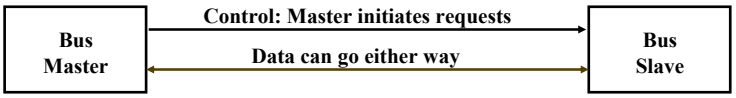
Fig.8.10/11: Asynchronous Handshaking Protocol



Increasing the Bus Bandwidth

- **Separate versus multiplexed address and data lines:**
  - **Address and data can be transmitted in one bus cycle if separate address and data lines are available**
  - **Cost: (a) more bus lines, (b) increased complexity**
- **Data bus width:**
  - **By increasing the width of the data bus, transfers of multiple words require fewer bus cycles**
  - **Example: SPARCstation 20 memory bus is 128 bit wide**
  - **Cost: more bus lines**
- **Block transfers:**
  - **Allow the bus to transfer multiple words in back-to-back bus cycles**
  - **Only one address needs to be sent at the beginning**
  - **The bus is not released until the last word is transferred**
  - **Cost: (a) increased complexity**
    (b) decreased response time for request

Arbitration: Obtaining Access to the Bus

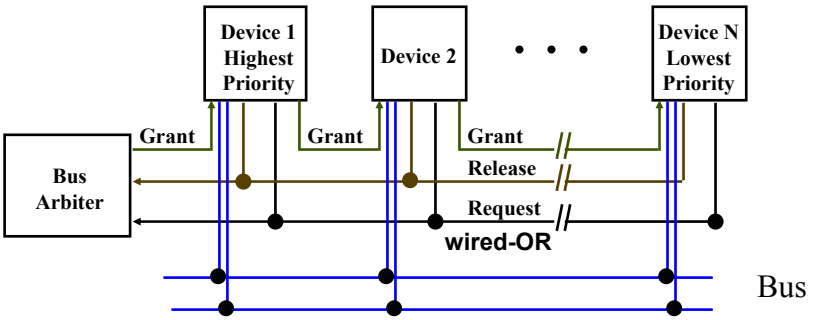


- **One of the most important issues in bus design:**
  - **How is the bus reserved by a devices that wishes to use it?**
- **Chaos is avoided by a master-slave arrangement:**
  - **Only the bus master can control access to the bus:**
    It initiates and controls all bus requests
  - **A slave responds to read and write requests**
- **The simplest system:**
  - **Processor is the only bus master**
  - **All bus requests must be controlled by the processor**
  - **Major drawback: the processor is involved in every transaction**

Multiple Potential Bus Masters: the Need for Arbitration

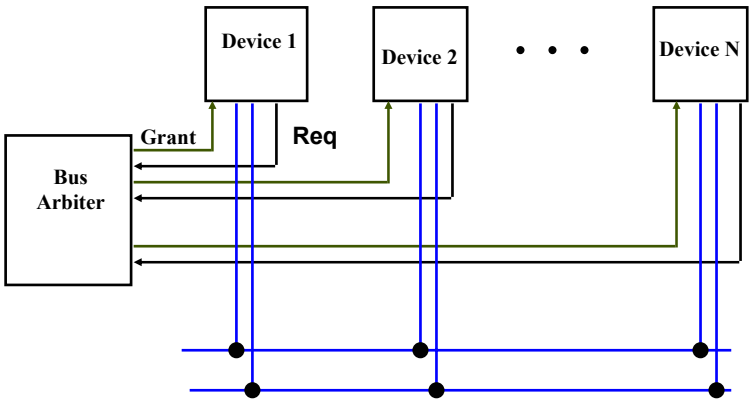
- **Bus arbitration scheme:**
  - **A bus master wanting to use the bus asserts the bus request**
  - **A bus master cannot use the bus until its request is granted**
  - **A bus master must signal to the arbiter after finish using the bus**
- **Bus arbitration schemes usually try to balance two factors:**
  - **Bus priority: the highest priority device should be serviced first**
  - **Fairness: Even the lowest priority device should never be completely locked out from the bus**
- **Bus arbitration schemes can be divided into four broad classes:**
  - **Daisy chain arbitration: single device with all request lines.**
  - **Centralized, parallel arbitration: see next-next slide**
  - **Distributed arbitration by self-selection: each device wanting the bus places a code indicating its identity on the bus.**
  - **Distributed arbitration by collision detection: Ethernet uses this.**

The Daisy Chain Bus Arbitrations Scheme



- **Advantage: simple**
- **Disadvantages:**
  - **Cannot assure fairness:**
    A low-priority device may be locked out indefinitely
  - **The use of the daisy chain grant signal also limits the bus speed**

Centralized Parallel Arbitration



- Used in essentially all processor-memory busses and in high-speed I/O busses

1993 MP Server Memory Bus Survey: GTL revolution

Bus	MBus	Summit	Challenge	XDBus
Originator	Sun	HP	SGI	Sun
Clock Rate (MHz)	40	60	48	66
Address lines	36	48	40	muxed
Data lines	64	128	256	144 (parity)
Data Sizes (bits)	256	512	1024	512
Clocks/transfer		4	5	4?
Peak (MB/s)	320(80)	960	1200	1056
Master	Multi	Multi	Multi	Multi
Arbitration	Central	Central	Central	Central
Slots		16	9	10
Busses/system	1	1	1	2
Length		13 inches	12? inches	17 inches

1993 Backplane/IO Bus Survey

Bus	SBus	TurboChannel	MicroChannel	PCI
Originator	Sun	DEC	IBM	Intel
Clock Rate (MHz)	16-25	12.5-25	async	33
Addressing	Virtual	Physical	Physical	Physical
Data Sizes (bits)	8,16,32	8,16,24,32	8,16,24,32,64	8,16,24,32,64
Master	Multi	Single	Multi	Multi
Arbitration	Central	Central	Central	Central
32 bit read (MB/s)	33	25	20	33
Peak (MB/s)	89	84	75	111 (222)
Max Power (W)	16	26	13	25

- High speed I/O bus
  - Examples: graphics, fast networks
  - Limited number of devices
  - Data transfer bursts at full rate
  - DMA transfers important:small controller spools stream of bytes to or from memory

Summary of Bus Options

Option	*High performance*	*Low cost*
Bus width	Separate address & data lines	Multiplex address & data lines
Data width	Wider is faster (e.g., 32 bits)	Narrower is cheaper (e.g., 8 bits)
Transfer size	Multiple words has less bus overhead	Single-word transfer is simpler
Bus masters	Multiple (requires arbitration)	Single master (no arbitration)
Clocking	Synchronous	Asynchronous
Protocol	pipelined	Serial

Multimedia Bandwidth Requirements

- **High Quality Video**
  - **Digital Data = (30 frames / second) (640 x 480 pels) (24-bit color / pel) = 221 Mbps (27.6 MB/s)**
- **Reduced Quality Video**
  - **Digital Data = (15 frames / second) (320 x 240 pels) (16-bit color / pel) = 18 Mbps (2.2 MB/s)**
- **High Quality Audio**
  - **Digital Data = (44,100 audio samples / sec) (16-bit audio samples)**
  - **(2 audio channels for stereo) = 1.4 Mbps**
- **Reduced Quality Audio**
  - **Digital Data = (11,050 audio samples / sec) (8-bit audio samples) (1 audio channel for monaural) = 0.1 Mbps**
- **compression changes the whole story!**

Other important issues

- **Interfacing I/O devices to the memory, processor: Operating system**
- **Devices communicate with processor; Transferring the data between a device and memory by**
  - **polling:simplest way;processor is totally in control;can waste a lot of processor time; example:p676**
  - **interrupt-driven:like exception; control unit check for pending I/O interrupt at the time it starts a new instruction;example:p679**
  - **DMA: specialized controller is used to transfers data between an I/O device and memory independent of the processor; example: p681**