Chapter 3

Memory Interface, I/O Interface

LEARNING OBJECTIVES

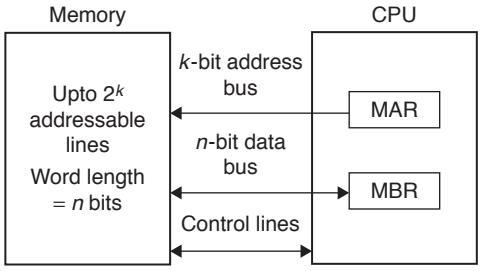
- *Memory interface*
- *RAM*
- *ROM*
- *Memory interfacing*
- *Input–output interfacing*
- *Handshaking*
- *Data transfer mode*
- *Design techniques for interrupts*
- *Direct memory access*
- *Input–output processor*

Memory Interface

Basic Concepts

Computer memory is used to store programs and data. The maximum size of a memory that can be used in any computer is determined by the addressing scheme.

Example: If the memory address has 16-bits, then the size of memory will be 2^{16} Bytes.



If MAR is k-bits long and MDR is n-bits long, then the memory may contain up to 2^k addressable locations and the n-bits of data are transferred between the processor and memory. This transfer takes place over processor bus. The processor bus has

1. Address line
2. Data line
3. Control line

Control line is used for coordinating data transfer.

Processor reads the data from the memory by loading the address of the required memory location into MAR and setting the $R/\overline{W}$ line to 1.

The memory responds by placing the data from the addressed location onto the data lines and confirms the actions. Upon confirmation, the processor loads the data onto the data lines, into MDR register. The processor writes the data into the memory location by loading the address of this location into MAR and loading the data into MDR sets the $R/\overline{W}$ line to 0.

- **Memory Access Time:** It is the time that elapses between the initiation of an operation and the completion of that operation.
- **Memory Cycle Time:** It is the minimum time delay that required between the initiations of two successive memory operations.

RAM (Random Access Memory)

In RAM, if any location that can be accessed for a read/write operation in fixed amount of time, it is independent of the location's address:

- Memory cells are usually organized in the form of array, in which each cell is capable of storing one bit of information.
- Each row of cells constitutes a memory word and all cells of a row are connected to a common line called as word line.
- The cells in each column are connected to sense/write circuit by two bit lines.

The data input and data output of each sense/write circuit are connected to a single bidirectional data line that can be connected to a data bus.

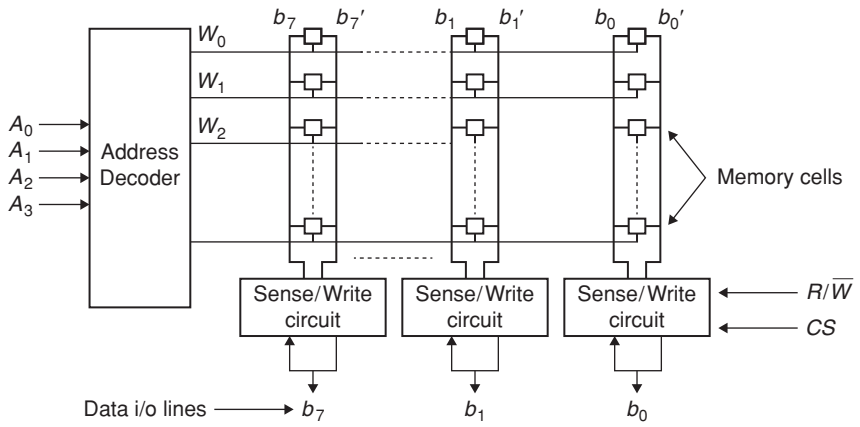


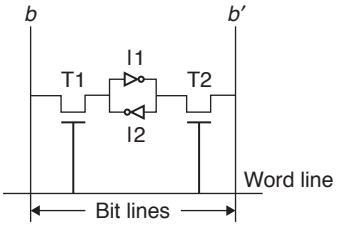
Figure 1 Organization of bit cells in a memory chip

- $R/\overline{W}$: Specifies the required operation.
- CS: Chip select input selects a given chip in the multi-chip memory system.

Static memories

Memories that consist of circuits capable of retaining their state as long as power is applied are known as static memories.

SRAM (static RAM) SRAM consists of two inverters, two transistors. In order to read the state of the SRAM cell, the word line is activated to close switches T1 and T2.



Advantages of SRAM:

1. It has low power consumption, because the current flows in the cell only when the cell is being activated or accessed.
2. SRAM can be accessed quickly.

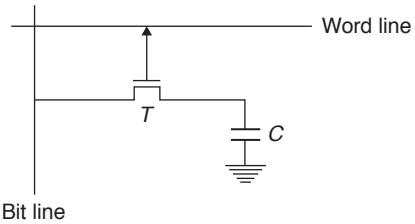
Disadvantages of SRAM: SRAMs are said to be volatile memories, because their contents are lost when the power is interrupted.

DRAM (Dynamic RAM) Less expensive RAMs can be implemented if simplex cells are used, such cells cannot retain their state indefinitely. Hence they are called dynamic RAMs.

The information stored in a dynamic memory cell in the form of a charge on a capacitor and this charge can be maintained only for tens of milliseconds.

The contents must be periodically refreshed by restoring the capacitor charge to its full value.

Example: Single-transistor dynamic memory cell:



If charge on capacitor > threshold value, then bit line will have '1'. If charge on capacitor < threshold value, then bit line will have '0'.

DRAM	SRAM
1. Volatile	1. Volatile
2. Simple to build and slower than SRAM	2. Faster than DRAM
3. Need refresh circuitry	3. More expensive to build
4. Favoured for large memory units	4. Favoured for cache memory

Latency It is the amount of time it takes to transfer a word of data to or from the memory.

- For the transfer of a single word, the latency provides the complete indication of memory performance.
- For a block transfer, the latency denotes the time it takes to transfer the first word of data.

Bandwidth It is defined as the number of bits or bytes that can be transferred in one second.

Note: All dynamic memories have to be refreshed.

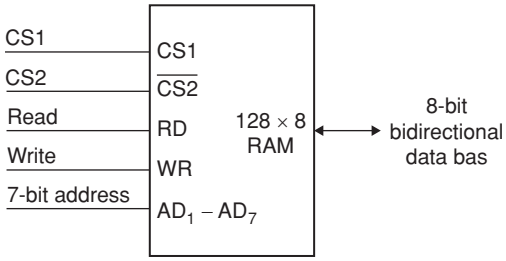


Figure 2 RAM chip block diagram

Read-only Memory (ROM)

Both SRAM and DRAM chips are volatile, which means that they lose the stored information if power is turned off. If the normal operation involves only reading of stored data, use ROM memory.

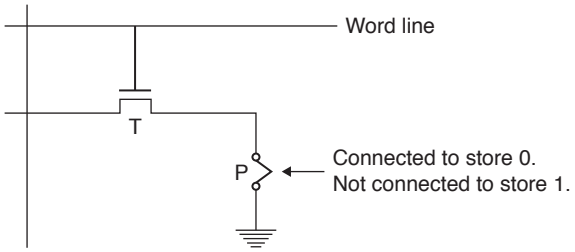


Figure 3 ROM cell

Types of ROM

Different types of non-volatile ROM are:

1. **PROM (Programmable ROM):**
   - Allows the data to be loaded by the user.
   - Less expensive, faster, flexible.
2. **EPROM (Erasable PROM):**
   - Allows the stored data to be erased and new data to be loaded.
   - Flexible, retain information for a long time.
   - Contents erased by UV light.
3. **EEPROM (Electrically Erasable PROM):**
   - Programmed and erased electrically.
   - Allows the erasing of all cell contents selectively.
   - Requires different voltage for erasing, writing and reading of stored data.
4. **Flash memory:** Allows to read the contents of a single cell but it is only possible to write the entire contents of a block.

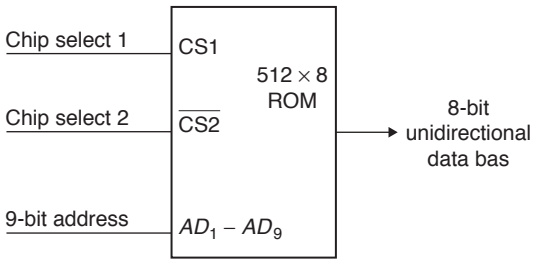


Figure 4 Block diagram of ROM chip

Memory Interfacing

The interfacing circuit enables the access of processor to memory. The function of memory interfacing is that the processor should be able to read from and write into a given register of a memory chip. To perform this, the microprocessor should be

1. able to select the chip.
2. identify the register.
3. enable the appropriate buffer.

Input–output Interfacing

Basic Concepts of I/O Module

I/O module contains logic for performing a communication function between the peripherals and the bus. The peripherals are not connected to the system bus directly. The reasons for this are

1. Peripherals are electromechanical and electromagnetic devices and their manner of operation is different from the operation of the CPU and memory, which are electronicdevices. So a conversion of signal values may be required.
2. The data transfer rate of peripherals is usually slower than the transfer rate of the CPU and hence a synchronization mechanism may be needed.
3. Data codes and formats in peripherals differ from the word format in the CPU and memory.
4. The operating modes of peripherals are different from each other and each must be controlled so as not to disturb the operation of other peripherals connected to the CPU.

To resolve these differences, computer systems include special hardware components between the CPU and peripherals to supervise and synchronize all input and output transfers. These components are called 'interface' units.

By using this interfacing,

1. interface to the processor and memory via the system bus or central switch.
2. interface to one or more peripheral devices by tailored data links.

Input–output devices

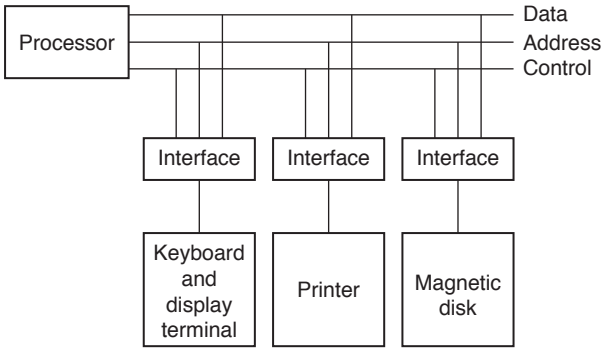
- Input and Output devices provide a means for people to make use of a computer.
- Some I/O devices function as an interface between a computer system and other physical system.

Input–output interface

Input/output Interface provides a method for transferring information between internal storage (such as memory and CPU Register) and external I/O devices. It resolves the difference between the computer and peripheral devices.

Input–output bus and interface modules

Each peripheral has an interface module associated with it. The interface module decodes the device address (device code), decodes signals for the peripheral controller, synchronizes the data flow and supervises the transfer rate between peripheral and CPU or memory.



Function of buses

1. **Memory bus:** It is used for information transfer between CPU and main memory.
2. **I/O bus:** It is used for information transfers between CPU and I/O devices through their I/O interface.

Isolated versus memory mapped I/O

1. **Isolated I/O:**
   - Separate I/O read/write control lines in addition to memory read/write control lines.
   - Separate (isolated) memory and I/O address space
   - Distinct input and output instructions.
2. **Memory-mapped I/O:**
   - A single set of Read/write control lines (i.e., no distinction between memory and I/O transfer).
   - Memory and I/O address share the common address space (reduces memory address range available).
   - No specific input or output instruction.
   - The same memory reference instructions can be used for I/O transfer.
   - Considerable flexibility in handling I/O operations.

Asynchronous serial transfer

In serial data transmission, each bit in the message is sent in sequence one at a time. Serial transmission can be synchronous or asynchronous.

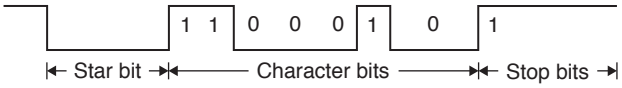
In synchronous transmission, the two units share a common clock frequency and bits are transmitted continuously at the rate dictated by the clock pulses.

In asynchronous transmission, binary information is sent only when it is available and the line remains idle when there is no information to be transmitted.

In serial asynchronous transmission technique, each character consists of three parts:

1. start bits
2. character bits
3. stop bits

Example:



A transmitted character can be detected by the receiver from the knowledge of the transmission rules:

1. When a character is not being sent, the line is kept in the 1-state.
2. The initiation of a character transmission is detected from the start bit, which is always 0.
3. The character bits always follow the start bit.
4. After the last bit of the character is transmitted, a stop bit is detected when the line returns to the 1-state for at least one bit time.
   - The baud rate is defined as the rate at which serial information is transmitted and is equivalent to the data transfer in bits per second.

Strobe control Employs a single control line to time each transfer. Strobe may be activated by either the source or the destination units.

(a) Source initiated transfer:

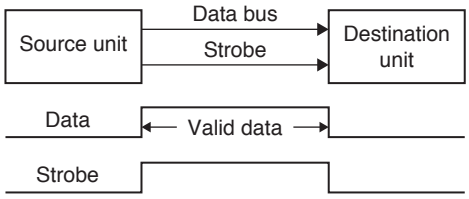
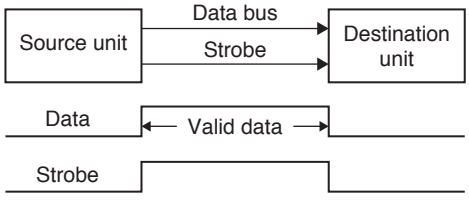


Figure 5 Source initiated strobe for data transfer

- The data bus carries the binary information from source unit to the destination unit.
- The strobe is a single line that informs the destination unit when a valid data word is available in the bus.

(b) Destination initiated strobe for data transfer:



- The destination unit activates the strobe pulse, informing the source to provide the data. The source unit responds by placing the requested binary information on the data bus.
- The data must be valid and remain in the bus long enough for the destination unit to accept it.
- The falling edge of the strobe pulse can be used again to trigger a destination register. The destination unit then disables the strobe.

Handshaking

Disadvantage of strobe method: Source unit which initiated the transfer has no way of knowing whether the destination unit has actually received the data item that was placed in the bus.

The handshake method solves this problem by introducing a second control signal that provides a reply to the unit that initiates the transfer.

Principle of two-wire handshaking: One control line is in the same direction as the data flow in the bus from the source to the destination. It is used by the source unit to inform the destination unit whether there are valid data in the bus.

The other control line is in the other direction from the destination to the source. It is used by the destination unit to inform the source whether it can accept data.

The sequence of control during the transfer depends on the unit that initiates the transfer.

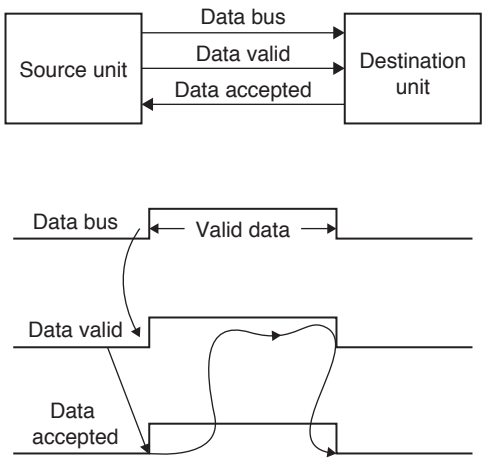


Figure 6 Source initiated transfer using hand shaking

Similarly a destination unit may also initiate the transfer.

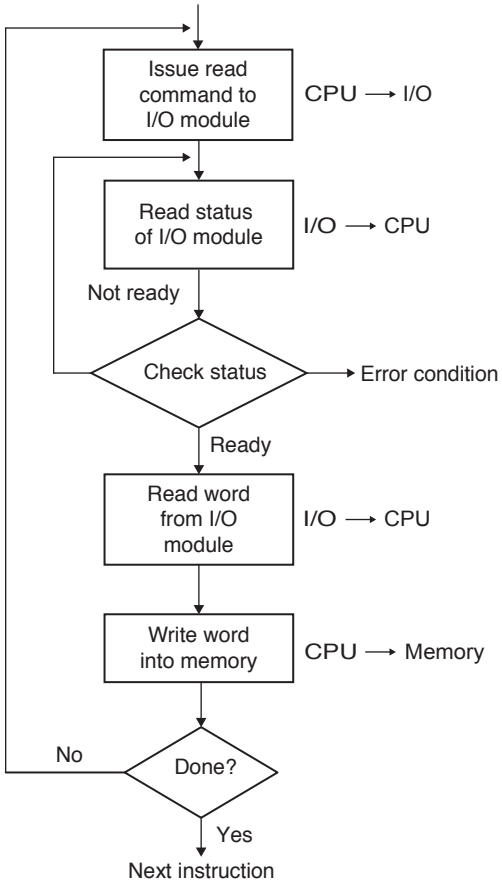
Advantage: Handshaking scheme provides a high degree of flexibility and reliability because the successful completion of a data transfer relies on active participation by both units.

Modes of transfer

There are three different data transfer modes between the central computer (CPU or Memory) and peripherals:

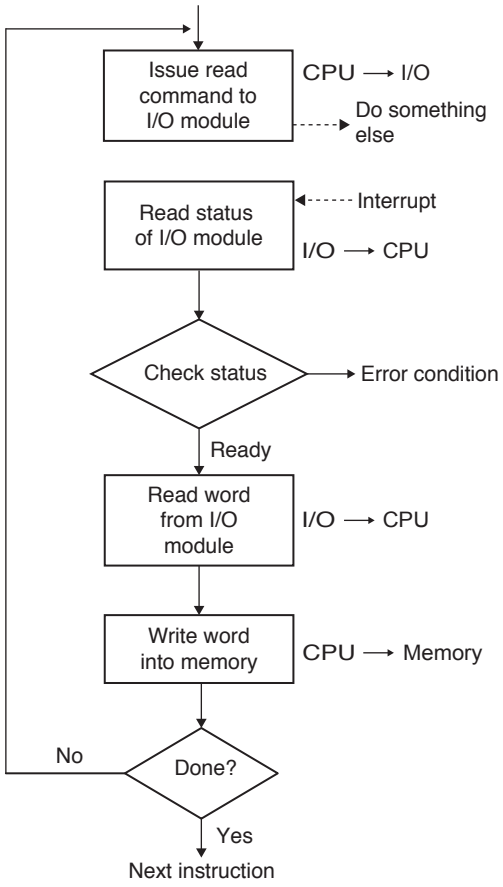
1. Program-controlled I/O
2. Interrupt-initiated I/O
3. Direct memory access

Program-controlled input–output With programmed I/O, the I/O module will perform the requested action and then set the appropriate bits in the I/O status register. The I/O module takes no further action to alert the CPU. In particular it does not interrupt the CPU. Thus, it is the responsibility of the CPU to periodically check the status of the I/O module until it finds that the operation is complete.



Interrupt initiated input–output The problem with programmed I/O is that the CPU has to wait a long time for the I/O module of concern to be ready for either reception or transmission of data. The CPU, while waiting must repeatedly interrogate the status of the I/O module. As a result, the level of the performance of the entire system is severely degraded.

An alternation is for the CPU to issue an I/O command to a module and then go on to do some other useful work. The I/O module will then interrupt the CPU to request service when it is ready to exchange data with the CPU. The CPU then executes the data transfer, as before and then resumes its former processing.



Interrupt Processing: In all computers, there is a mechanism by which the normal processing of the processor is interrupted by other modules like I/O, memory. The interrupts may be of the following class:

1. Program: Generated by some condition that occurs as a result of an instruction execution.

   Examples: Arithmetic overflow, division by zero, etc.
2. Timer: Generated by timer within the processor.
3. I/O: Generated by an I/O controller, to signal normal completion of an operation or to signal a variety of error conditions.
4. Hardware failure: Generated by a failure such as power failure or memory parity error.

Interrupts are provided primarily as a way to improve processing efficiency.

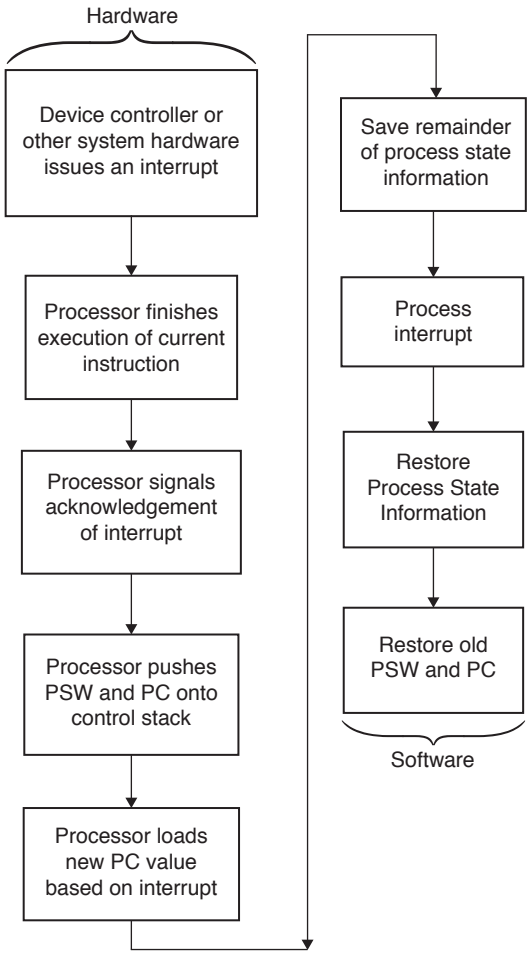


Figure 7 Interrupt Processing Flowchart

Consider the following figures, which show the contents of memory and registers before and after interrupt instruction processing.

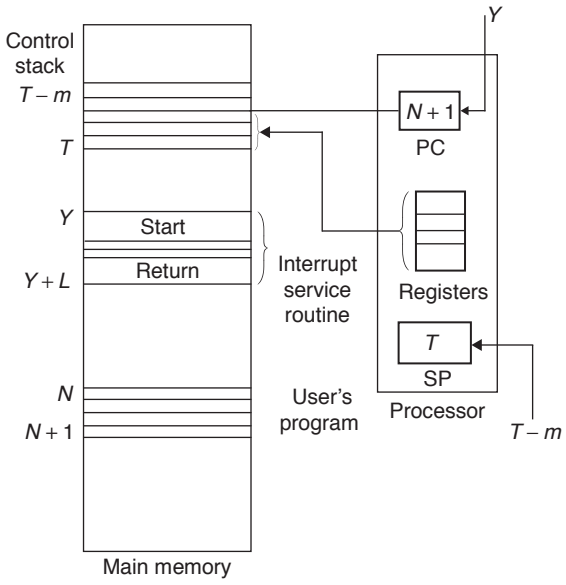


Figure 8 Interrupt occurs after instruction at location *N*

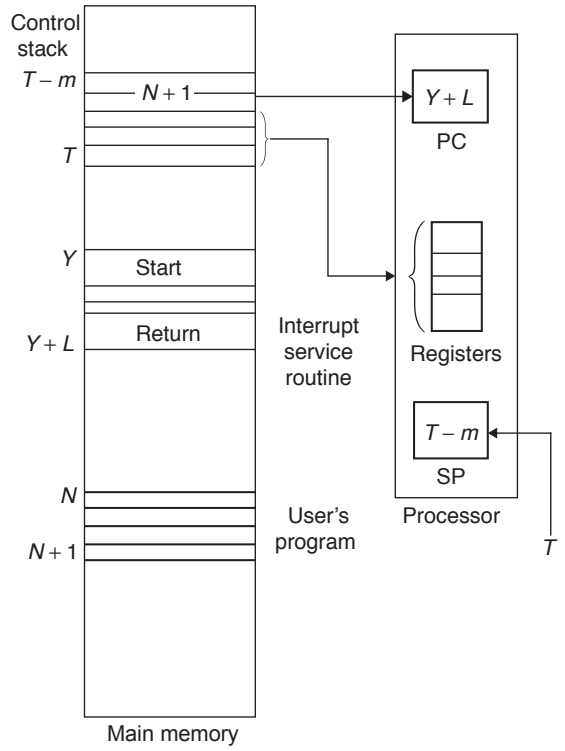


Figure 9 Return from interrupt

Interrupt priority: Priority determines which interrupt is to be served first when two or more requests are made simultaneously. Priority also determines which interrupts are permitted to interrupt the computer while another is being serviced. Higher priority interrupts can make requests while servicing a lower priority interrupt.

Design techniques for interrupts: Two design issues arise in implementing interrupt I/O:

1. Since there will almost invariably be multiple I/O module, how does the CPU determine which device issued the interrupt.
2. If multiple interrupts have occurred, how does the CPU decide which one to process.

Four general categories of techniques are there which are common in use:

(a) Multiple interrupt lines: In this technique, multiple interrupt lines are provided between the CPU and the I/O modules. However, it is impractical to dedicate more than a few bus lines or CPU pins to interrupt lines. Consequently, even if multiple lines are used, it is likely that each line will have multiple I/O modules attached to it. Thus, one of the other three techniques must be used one each line.

(b) Software poll: When the CPU detects an interrupt, it braches to an interrupt-service routine whose job is to poll each I/O module to determine which module generated the interrupt. The poll could be in the form of a separate command line. The CPU receives the command and places the address of a particular I/O module on the address lines. The I/O module responds positively if it set the interrupt. Alternatively, each I/O module could contain an addressable status register. The CPU then read the status register of each I/O module to identify the interrupting module. Once the correct module is identified, the CPU branches to a device service routine specified to that device. It is time consuming.

(c) Daisy chain: Daisy chain in effect provides a hardware poll. For interrupts all I/O modules share a common interrupt request line. The interrupt acknowledge line is daisy chained through the modules. When the CPU is interrupted, it sends out an interrupt acknowledgement. This signal propagates through a series of I/O modules until it gets to a requesting module. The requesting module typically responds by placing a word on the data lines. This word is referred to as a vector and is either the address of the I/O module or some other unique identifier. In either case, the CPU uses the vector as a pointer to the appropriate device-service routine. This avoids the need to execute a general interrupt-service routine first. This technique is referred to as a vectored Interrupt.

(d) Bus arbitration: Bus arbitration is also another technique which makes use of vectored Interrupts. With bus arbitration, an I/O module must first gain control of the bus before it can raise the interrupt request line. Thus only one module can raise the line at a time. When the CPU detects the interrupt, it responds on the interrupt acknowledge line. The requesting module then places its vector on the data lines.

Direct Memory Access (DMA)

Drawbacks of programmed and interrupt-driven I/O:

1. The I/O transfer rate is limited by the speed with which the processor can test and service a device.
2. The processor is tied up in managing I/O transfer; a number of instructions must be executed for each I/O transfer.

Both methods have an adverse impact on both processor activity and I/O transfer rate.

When large volume of data is to be moved, a more efficient technique is required: Direct Memory Access (DMA).

DMA function: DMA involves an additional module on the system bus.

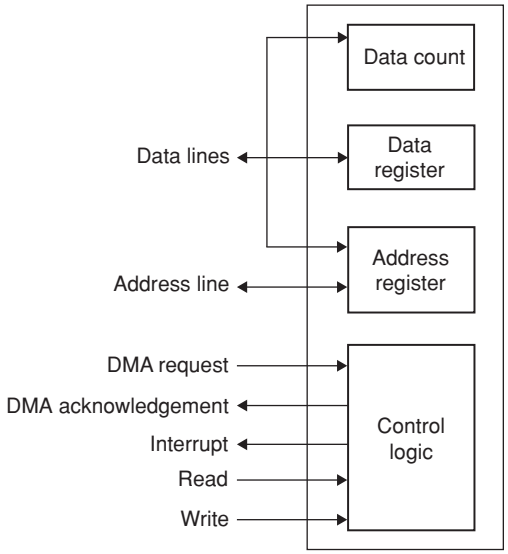
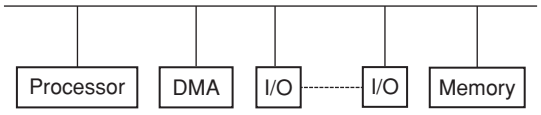


Figure 10 DMA Block Diagram

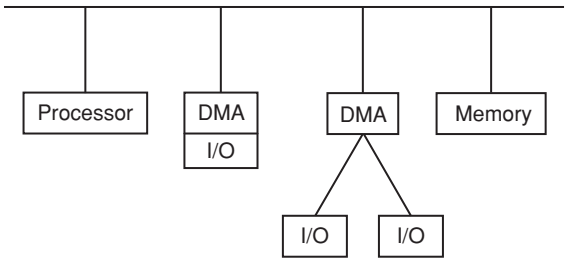
The DMA module is capable of mimicking the processor and indeed, of taking over control of the system from the processor. It needs to do this to transfer data to and from memory over the system bus. For this purpose, the DMA module must use the bus only when the processor does not need it or it must force the processor to suspend operation temporarily. The latter technique is more common and is referred as cycle stealing.

DMA configurations

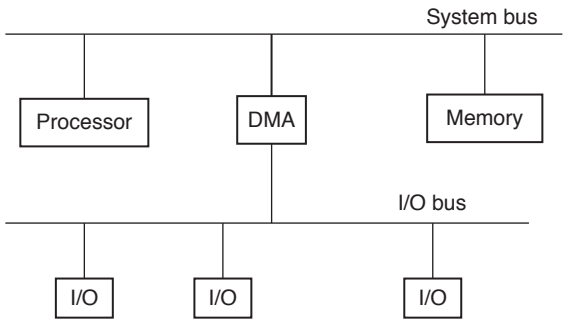
1. **Single bus, detached DMA**
   - Inexpensive, inefficient
   - Each transfer of a word consumes two bus cycles.



2. **Single bus, integrated DMA I/O** There is a path between the DMA module and one or more I/O modules that does not include system bus.



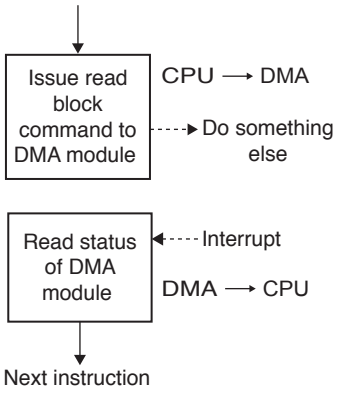
3. **I/O bus**
   - Reduces the number of I/O interfaces in the DMA module to one.
   - Easily expandable configuration.



With DMA, when the CPU wishes to read or write a block of data, it issues a command to the DMA module, by sending the following information to the DMA module.

1. Whether a read or write is requested.
2. The address of the I/O device involved.
3. The starting location in memory to read from or write to.
4. The number of words to be read or written.

The CPU then continues with other work. It has delegates this I/O operation to the DMA module, and that module will take care of it. The DMA module transfers the entire block of data, one word at a time, directly to or from memory, without going through the CPU. When the transfer is complete, the DMA module sends an interrupt signal to the CPU. Thus, the CPU is involved only at the beginning and end of the transfer.



DMA transfer can either happen as:

1. **Burst transfer**
   - A block sequence consisting of a number of memory words is transferred in continuous burst.
   - DMA controller is master of memory Buses.
   - This mode of transfer is needed for fast devices such as Magnetic Disks, where transmission cannot be stopped or slowed down.
2. **Cycle stealing**
   - CPU is usually much faster than I/O (DMA), thus CPU uses the most of the memory cycles.
   - DMA controller steals the memory cycles from CPU.
   - For those stolen cycles, CPU remains idle.

- For those slow CPU, DMA Controller may steal most of the memory.
- Cycle stealing, which may cause CPU remain idle long time.

Input–Output Processor (IOP)

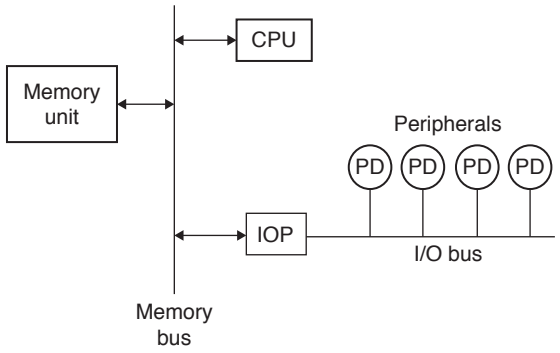
An IOP is a processor, having a direct memory access capability, used to communicate with I/O devices.

In this configuration, the computer system can be divided into a memory unit and a number of processors comprised of the CPU and one or more IOPs.

Each IOP takes care of input and output tasks, relieving the CPU from the house keeping chores involved in I/O transfers.

- IOP is similar to a CPU except that it is designed to handle the details of I/O processing.
- Unlike DMA, the IOP can fetch and execute its own instructions.

The following figure shows a computer with two processors:



Exercises

Practice Problems I

Directions for questions 1 to 20: Select the correct alternative from the given choices.

1. Consider a DRAM that must be given a refresh cycle 64 times per ms. Each refresh operation requires 150 ns, a memory cycle requires 250 ns. What is the approximate percentage of the memory's total operating time must be given to refreshes?
(A) 1% (B) 2%
(C) 9% (D) 60%

2. A DMA controller transfers 16-bit words to memory using cycle stealing. The words are assembled from a device that transmits characters at a rate of 2400 characters per second. The CPU is fetching and executing instructions at an average rate of 1 million instructions per second. By how much time will the CPU be slowed down because of the DMA transfer?
(A) 0.6% (B) 0.1%
(C) 0.12% (D) 0.24%

3. A system is based on a 16-bit microprocessor and has two I/O devices. The I/O controllers for this system use separate control and status registers. Both devices handle data on a one-byte-at-a time basis. The first device has two status lines and three control lines. The second device has three status lines and four control lines. How many 16-bit I/O control module registers do we need for status reading and control of each device?
(A) 1, 2 (B) 2, 1
(C) 2, 2 (D) 1, 1

4. In a programmed I/O technique, the processor is stuck in a wait loop doing status checking of an I/O device. To increase efficiency, the I/O software could be written so that the processor periodically checks the status of the device. If the device is not ready, the processor can jump to other tasks. After some timed interval, the processor comes back to check status again. Let us assume that above scheme is used for outputting data one character at a time to a printer that operates at 10 characters per second (CPS). Which of the following statement is true if its status is scanned every 200 ms?
(A) The printing speed is increased by 5 CPS
(B) The printing rate is slowed to 5 CPS
(C) The printing rate is at 10 CPS only
(D) The printing rate is at 20 CPS

5. Consider a system employing interrupt-driven I/O for a particular device that transfers data at an average of 8 KB/s on a continuous basis. The interrupt processing takes about 100 μs and the I/O device interrupts processor for every byte. Let assume that the device has two 16-byte buffers and interrupts the processor when one of the buffer is full. While executing the ISR, the processor takes about 8 μs for the transfer of each byte. Then what is the fraction of processor time is consumed by this I/O device?
(A) 8% (B) 11%
(C) 50% (D) 65%

6. A 32-bit computer has two selector channels one multiplexor channel. Each selector channel supports two magnetic disks and three magnetic tape units. The multiplexor channel has two line printers, two card readers and 10 VDT terminals connected to it. Assume the following transfer rates:
Disk drive: 1000 KB/sec
Magnetic tape drive: 300 KB/sec
Line printer: 6.2 KB/sec
Card reader: 2.4 KB/sec
VDT: 1 KB/sec

What is the maximum aggregate I/O transfer rate of this system?
(A) 1625.6 KB/s (B) 1327.2 KB/s
(C) 2027.2 KN/s (D) 2327.2 KB/s

7. Consider a disk drive with 16 surfaces, 512 tracks per surface and 512 sectors per track, 1 kilo bytes per sector and a Rotation speed of 3000 RPM. The disk is operated in cycle stealing mode where by whenever one 4 byte word is ready it is sent to memory; similarly, for writing, the disk interface read a 4 byte word from the memory in each DMA cycle. The memory cycle time is 40 nsec. Find the maximum percentage of time that the CPU gets blocked during DMA operation?
(A) 2.62% (B) 26.21%
(C) 0.26% (D) 0.52%

8. How many RAM chips of size (256 K × 1-bit) are needed to build a 1 M Byte memory?
(A) 16 (B) 8
(C) 32 (D) 24

9. Four memory chips of 16 × 4 size have their address bases connected together. The whole system will have a size of
(A) 16 × 8 (B) 64 × 64
(C) 16 × 16 (D) 256 × 1

10. In which of following I/O techniques, there will be no interrupt?
(A) Programmed I/O (B) Interrupt-driven I/O
(C) DMA (D) Both (B) and (C)

11. The capacity of a memory unit is defined by the number of words multiplied by the number of bits/word. How many separate address and data lines are needed for a memory 16K × 16?
(A) 10 address, 4 data lines
(B) 14, 4
(C) 14, 16
(D) 14, 14

12. The main problem of strobe asynchronous data transfer is
(A) it employs a single control line
(B) it is controlled by clock pulses in the CPU.
(C) the falling edge again used to trigger
(D) no way of knowing whether the destination has received the data item.

13. Which of the following DMA transfer modes and interrupt handling mechanisms will enable the highest I/O bandwidth?
(A) Block transfer and polling interrupt
(B) Cycle stealing and polling interrupt
(C) Block transfer and vectored interrupt
(D) Transparent DMA and vectored interrupt

14. Which of the following enables peripherals to pass a signal down the bus to the next device on the bus during polling of the device?
(A) Interrupt vectoring (B) Cycle stealing
(C) DMA (D) Daisy chain

15. What will be the response of the CPU, on receiving an interrupt from an input/output device?
(A) It hands over the control of address bus and data bus to the interrupting device.
(B) It branches off to the interrupt service routine after completion of the current instruction.
(C) It halts for a predetermined time.
(D) It branches off to the interrupt service routine immediately.

16. What is the bandwidth of memory system that has a latency of 50ns, a pre charge time of 10ns and transfers 2 bytes of data per access?
(A) 60 B/sec (B) 1.67 B/sec
(C) 1.67×10^7 B/sec (D) 3.33×10^7 B/sec

17. A hard disk is connected to a 50MHz processor through a DMA controller. Assume that the initial set-up of a DMA transfer takes 2000 clock cycles for the processor and also assume that the handling of the interrupt at DMA completion requires 1000 clock cycles for the processor. The hard disk has a transfer rate of 4000 K bytes/sec and average block size transferred is 8 K bytes. What fraction of the processor time is consumed by the disk, if the disk is actively transferring 100% of the time?
(A) 1% (B) 1.5%
(C) 2% (D) 3%

18. A device with transfer rate of 20KB/sec is connected to a CPU. Data is transferred byte wise. Let the interrupt overhead is 6 micro seconds. The byte transfer time between the device interface register and CPU or memory is negligible. What is minimum performance gain of operating the device under interrupt mode over operating it under program-controlled mode?
(A) 6 (B) 8
(C) 10 (D) 12

19. A DMA module is transferring characters to main memory from an external device at 76800 bits per second. The processor can fetch instructions at a rate of 2 million instructions per second. How much will the processor be slowed down due to DMA activity? (Express this as a percent of the time from when there is a conflict between DMA and the CPU)
(A) 0.24% (B) 0.48%
(C) 0.96% (D) 0.50%

20. Let us suppose that we want to read 2048 bytes in programmed I/O mode of CPU. The bus width is 32-bits. Each time an interrupt occurs from Hard disk drive and it taken 4 μsec to service it. How much CPU time is required to read 2048 bytes?
(A) 512 msec (B) 768 msec
(C) 1024 msec (D) 2048 msec

Practice Problems 2

Directions for questions 1 to 21: Select the correct alternative from the given choices.

1. Memory which is ultraviolet erasable is
(A) RAM (B) EPROM
(C) PROM (D) EEPROM

2. Memory which is electrically erasable is
(A) EPROM (B) EEPROM
(C) ROM (D) PROM

3. The minimum time delay that is required between the initiation of two successive memory operations is called
(A) Memory access time (B) Transmission time
(C) Seek Time (D) Memory cycle

4. The memory that is programmed at the time of manufacture is
(A) RAM (B) PROM
(C) ROM (D) EEPROM

5. The disadvantage of dynamic RAM over static RAM is
(A) High power consumption
(B) Higher bit density
(C) Need to refresh the capacitor charge every once in two milliseconds.
(D) Variable speed

6. If an error is detected, a part of the memory can be erased in
(A) PROM (B) EPROM
(C) EAROM (D) EROM

7. What are sequences of events in source initiated hand shaking transfer?
(A) Source enable data valid, destination enable data accepted, source disable data valid, destination disable data accepted
(B) Source disable data valid, destination enable data accepted, source disable data valid, destination enable data accepted
(C) Source disable data valid, destination Disable data valid, source enable data valid, destination enable data accepted.
(D) Source disable data valid, destination enable data valid, source enable data valid, destination disable data accepted.

8. Processor needs software interrupt to
(A) return from subroutine
(B) implement co-routines
(C) test the interrupt system of the processor
(D) obtain system services which need execution of privileged instructions

9. A microcomputer has primary memory of 512 KB. what is the exact number of bytes contained in this memory?
(A) 512 × 1000 (B) 512 × 100
(C) 512 × 1024 (D) 512 × 1028

10. The number of address lines required in a microprocessor which has to access 1 K bytes of memory is
(A) 6 (B) 4
(C) 10 (D) 8

11. Software interrupt is
(A) used to stimulate an external device
(B) generated by an external device
(C) Both (A) and (B)
(D) None of these

12. The bus that is used to transfer data from main memory to peripheral devices and vice-versa is
(A) Control bus (B) input bus
(C) output bus (D) DMA bus

13. The bus which is connected between the CPU and the main memory that permits transfer of information between the CPU and main memory is called
(A) memory bus (B) address bus
(C) control bus (D) DMA bus

14. An interrupt in which the external device supplies the interrupt requests as well as its address is called
(A) maskable interrupt
(B) vectored interrupt
(C) designated interrupt
(D) non-maskable interrupt

15. A temporarily ignored interrupt is called
(A) designated interrupt
(B) maskable interrupt
(C) non-maskable interrupt
(D) low priority interrupt

16. Which of the following device is used to connect a peripheral to a bus?
(A) control register
(B) interface
(C) communication protocol
(D) None of these

17. Which of the following is true for the daisy scheme of connecting input/output devices?
(A) It gives non-uniform priority to various devices.
(B) It gives uniform priority to all devices.
(C) It is only useful for connecting slow devices to a processor device.
(D) It requires a separate interrupt pin on the processor for each device.

18. In direct memory access data are directly transferred
(A) from CPU to input/output device and memory
(B) from an input/output device to memory only.
(C) from memory to an input/output device only.
(D) from an input/output device to the memory or vice versa

19. Which one of the following is true for a CPU having a single interrupt request line and a single interrupt grant line?
(A) Vectored interrupt multiple interrupting devices are always possible.
(B) Vectored interrupts are not possible but multiple interrupting devices are possible
(C) Vectored interrupts and multiple interrupting devices are sometimes possible
(D) Vectored interrupt is possible but multiple interrupting devices are not possible

20. In which of the following I/O, there is a single address space for memory locations and I/O devices
(A) Isolated I/O
(B) Memory mapped I/O
(C) DMA
(D) Both (A) and (B)

21. ____ signal used to interrupt processor and to execute service routine that takes an error recovery action.
(A) Strobe (B) Handshaking
(C) Polling (D) Time out

Previous Years' Questions

1. A device with data transfer rate 10 KB/sec is connected to a CPU. Data is transferred byte-wise. Let the interrupt overhead be 4 μsec. The byte transfer time between the device interface register and CPU or memory is negligible. What is the minimum performance gain of operating the device under interrupt mode over operating it under program-controlled mode? **[2005]**
(A) 15 (B) 25
(C) 35 (D) 45

2. A computer handles several interrupt sources of which the following are relevant for this question.
- Interrupt from CPU temperature sensor (raises interrupt if CPU temperature is too high)
- Interrupt from Mouse (raises interrupt if the mouse is moved or a button is pressed)
- Interrupt from Keyboard (raises interrupt when a key is pressed or released)
- Interrupt from Hard Disk (raises interrupt when a disk read is completed)

Which one of these will be handled at the HIGHEST priority? **[2011]**
(A) Interrupt from Hard Disk
(B) Interrupt from Mouse
(C) Interrupt from Keyboard
(D) Interrupt from CPU temperature sensor

The following information pertains to 3 and 4: Consider the following program segment for a hypothetical CPU having three user registers R_1, R_2 and R_3.

Instruction	Operation	Instruction Size (in words)
MOV R_1, 5000;	$R_1 \leftarrow$ Memory [5000]	2
MOV R_2 (R1);	$R_2 \leftarrow$ Memory [(R_1)]	1
ADD R_2, R3;	$R_2 \leftarrow R_2 + R_3$	1
MOV 6000, R_2;	Memory [6000] $\leftarrow R_2$	2
HALT;	Machine halts	1

3. Consider that the memory is byte addressable with size 32 bits, and the program has been loaded starting from memory location 1000 (decimal). If an interrupt occurs while the CPU has been halted after executing the HALT instruction, the return address (in decimal) saved in the stack will be **[2004]**
(A) 1007 (B) 1020
(C) 1024 (D) 1028

4. Let the clock cycles required for various operations be as follows:
Register to/from memory transfer: 3 clock cycles
ADD with both operands in register: 1 clock cycle
Instruction fetch and decode: 2 clock cycles per word
The total number of clock cycles required to execute the program is **[2004]**
(A) 29 (B) 24
(C) 23 (D) 20

5. A CPU generally handles an interrupt by executing an interrupt service routine **[2009]**
(A) As soon as an interrupt is raised.
(B) By checking the interrupt register at the end of fetch cycle.
(C) By checking the interrupt register after finishing the execution of the current instruction.
(D) By checking the interrupt register at fixed time intervals.

6. A hard disk with a transfer rate of 10 Mbytes/second is constantly transferring data to memory using DMA. The processor runs at 6000 MHz, and takes 300 and 900 clock cycles to initiate and complete DMA transfer respectively. If the size of the transfer is 20 Kbytes, what is the percentage of processor time consumed for the transfer operation?
(A) 5.0% (B) 1.0%
(C) 0.5% (D) 0.1%

7. On a non-pipelined sequential processor, a program segment, which is a part of the interrupt service routine, is given to transfer 500 bytes from an I/O device to memory.

Initialize the address register.

Initialize the count to 500

LOOP: Load a byte from device

Store in memory at address given by address register.

Increment the address register

Decrement the count

If count! = 0 go to LOOP

Assume that each statement in this program is equivalent to a machine instruction which takes one clock cycle to execute if it is a non-load/store instruction. The load-store instructions take two clock cycles to execute.

The designer of the system also has an alternate approach of using the DMA controller to implement the same transfer. The DMA controller requires 20 clock cycles for initialization and other overheads. Each DMA transfer cycle takes two clock cycles to transfer one byte of data from the device to the memory.

What is the approximate speedup when the DMA controller based design is used in place of the interrupt driven program based input-output? **[2011]**

(A) 3.4 (B) 4.4
(C) 5.1 (D) 6.7

8. Which of the following statements about synchronous and asynchronous I/O is NOT true? **[2008]**

(A) An ISR is invoked on completion of I/O in synchronous I/O but not in asynchronous I/O
(B) In both synchronous and asynchronous I/O, an ISR (Interrupt Service Routine) is invoked after completion of the I/O
(C) A process making a synchronous I/O call waits until I/O is complete, but a process making an asynchronous I/O call does not wait for completion of the I/O
(D) In the case of synchronous I/O, the process waiting for the completion of I/O is woken up by the ISR that is invoked after the completion of I/O

9. A main memory unit with a capacity of 4 megabytes is built using 1M × 1-bit DRAM chips. Each DRAM chip has 1K rows of cells with 1K cells in each row. The time taken for a single refresh operation is 100 nanoseconds. The time required to perform one refresh operation on all the cells in the memory unit is **[2010]**

(A) 100 nanoseconds
(B) $100 * 2^{10}$ nanoseconds
(C) $100 * 2^{20}$ nanoseconds
(D) $3200 * 2^{20}$ nanoseconds

10. A processor can support a maximum memory of 4GB, where the memory is word - addressable (a word consists of two bytes). The size of the address bus of the processor is atleast ___ bits. **[2016]**

11. The size of the data count register of a DMA controller is 16 bits. The processor needs to transfer a file of 29,154 kilobytes from disk to main memory. The memory is byte addressable. The minimum number of times the DMA controller needs to get the control of the system bus from the processor to transfer the file from the disk to main memory is _____. **[2016]**

12. The following are some events that occur after a device controller issues an interrupt while process *L* is under execution.

(P) The processor pushes the process status of *L* onto the control stack.
(Q) The processor finishes the execution of the current instruction.
(R) The processor executes the interrupt service routine.
(S) The processor pops the process status of *L* from the control stack.
(T) The processor loads the new PC value based on the interrupt.

Which one of the following is the correct order in which the events above occur? **[2018]**

(A) QPTRS (B) PTRSQ
(C) TRPQS (D) QTPRS

13. A 32-bit wide main memory unit with a capacity of 1 GB is built using 256 M × 4-bit DRAM chips. The number of rows of memory cells in the DRAM chip is 2^{14}. The time taken to perform one refresh operation is 50 nanoseconds. The refresh period is 2 milliseconds. The percentage (rounded to the closest integer) of the time available for performing the memory read write operations in the main memory unit is _____. **[2018]**

Answer Keys

Exercises

Practice Problems 1

1. A	**2.** C	**3.** D	**4.** B	**5.** B	**6.** C	**7.** B	**8.** C	**9.** C	**10.** A
11. C	**12.** D	**13.** A	**14.** D	**15.** B	**16.** D	**17.** D	**18.** B	**19.** B	**20.** D

Practice Problems 2

1. B	**2.** B	**3.** A	**4.** C	**5.** A	**6.** C	**7.** B	**8.** D	**9.** C	**10.** C
11. A	**12.** D	**13.** A	**14.** B	**15.** B	**16.** B	**17.** A	**18.** D	**19.** B	**20.** B
21. D									

Previous Years' Questions

1. B	**2.** D	**3.** D	**4.** B	**5.** C	**6.** D	**7.** A	**8.** A	**9.** D	**10.** 31
11. 456	**12.** A	**13.** 59 to 60							