Memory Segmentation in 8086 Microprocessor

Segmentation is the process in which the main memory of the computer is logically divided into different segments and each segment has its own base address. It is basically used to enhance the speed of execution of the computer system, so that the processor is able to fetch and execute the data from the memory easily and fast.

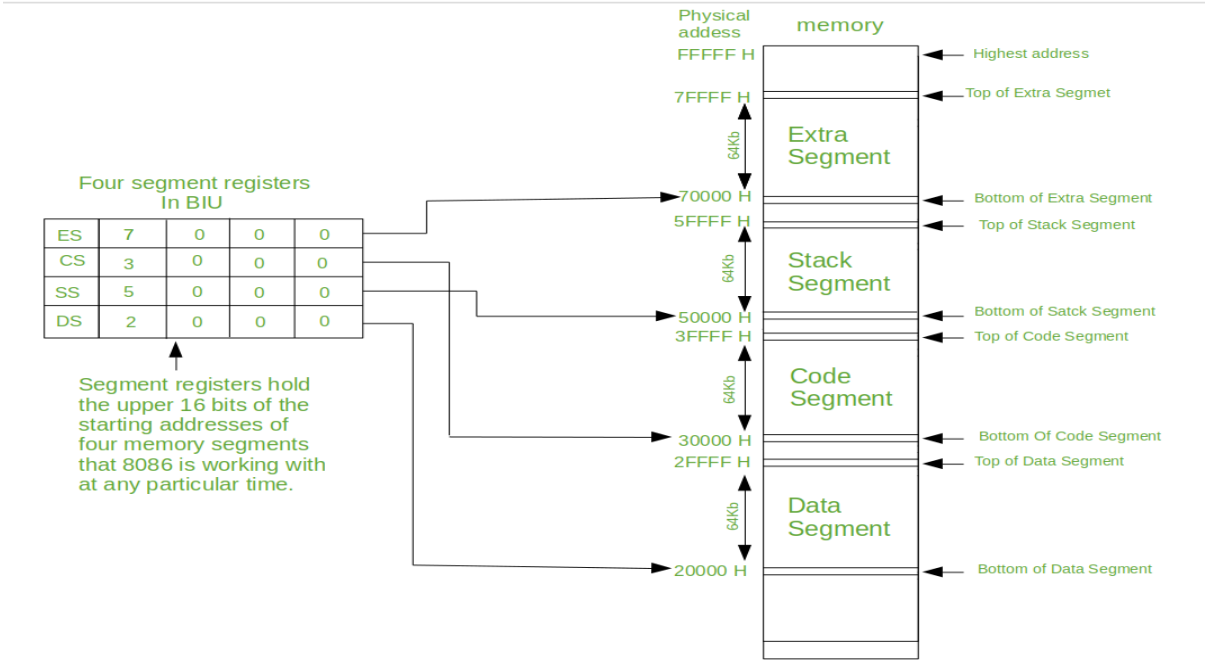
Need for Segmentation –
The Bus Interface Unit (BIU) contains four 16 bit special purpose registers (mentioned below) called as Segment Registers.

- **Code segment register (CS):** is used for addressing memory location in the code segment of the memory, where the executable program is stored.
- **Data segment register (DS):** points to the data segment of the memory where the data is stored.
- **Extra Segment Register (ES):** also refers to a segment in the memory which is another data segment in the memory.
- **Stack Segment Register (SS):** is used for addressing stack segment of the memory. The stack segment is that segment of memory which is used to store stack data.

The number of address lines in 8086 is 20, 8086 BIU will send 20bit address, so as to access one of the 1MB memory locations. The four segment registers actually contain the upper 16 bits of the starting addresses of the four memory segments of 64 KB each with which the 8086 is working at that instant of time. A segment is a logical unit of memory that may be up to 64 kilobytes long. Each segment is made up of contiguous memory locations. It is an independent, separately addressable unit. Starting address will always be changing. It will not be fixed.

Note that the 8086 does not work the whole 1MB memory at any given time. However, it works only with four 64KB segments within the whole 1MB memory.

Below is the one way of positioning four 64 kilobyte segments within the 1M byte memory space of an 8086.



Types of Segmentation –

1. **Overlapping Segment –** A segment starts at a particular address and its maximum size can go up to 64kilobytes. But if another segment starts along with this 64kilobytes location of the first segment, then the two are said to be *Overlapping Segment*.
2. **Non-Overlapped Segment –** A segment starts at a particular address and its maximum size can go up to 64kilobytes. But if another segment starts before this 64kilobytes location of the first segment, then the two segments are said to be *Non-Overlapped Segment*.

Rules of Segmentation

Segmentation process follows some rules as follows:

- The starting address of a segment should be such that it can be evenly divided by 16.
- Minimum size of a segment can be 16 bytes and the maximum can be 64 kB.

Segment	Offset Registers	Function
CS	IP	Address of the next instruction
DS	BX, DI, SI	Address of data
SS	SP, BP	Address in the stack
ES	BX, DI, SI	Address of destination data (for string operations)

Advantages of the Segmentation The main advantages of segmentation are as follows:

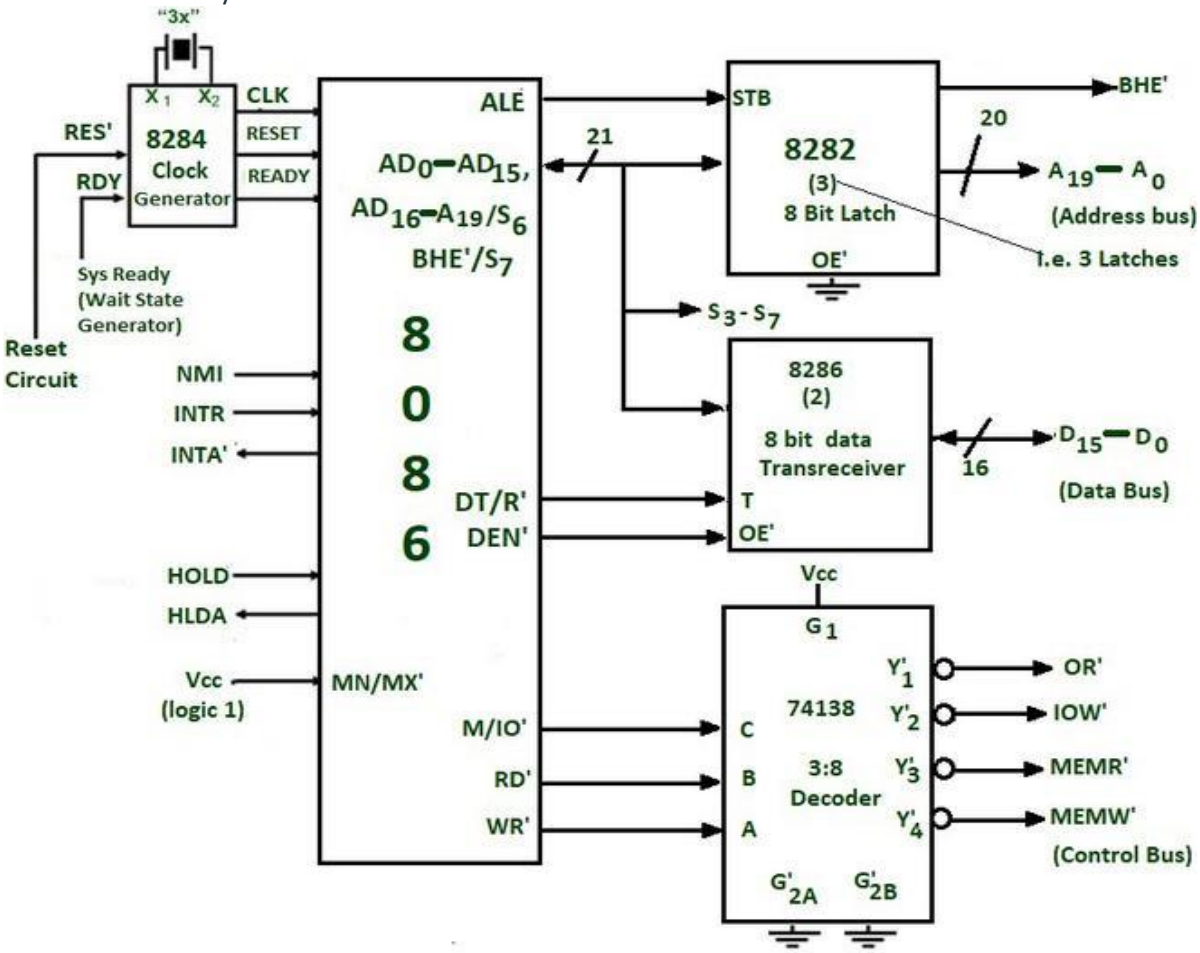
- It provides a powerful memory management mechanism.
- Data related or stack related operations can be performed in different segments.
- Code related operation can be done in separate code segments.
- It allows to processes to easily share data.
- It allows to extend the address ability of the processor, i.e. segmentation allows the use of 16 bit registers to give an addressing capability of 1 Megabytes. Without segmentation, it would require 20 bit registers.
- It is possible to enhance the memory size of code data or stack segments beyond 64 KB by allotting more than one segment for each area.

Minimum mode configuration of 8086 microprocessor (Min mode)

Overview:

- The 8086 microprocessor operates in minimum mode when MN/MX' = 1.
- In minimum mode,8086 is the only processor in the system which provides all the control signals which are needed for memory operations and I/O interfacing.
- Here the circuit is simple but it does not support multiprocessing.
- The other components which are transceivers, latches, 8284 clock generator, 74138 decoder, memory and i/o devices are also present in the system.
- The address bus of 8086 is 20 bits long. By this we can access 2^{20} byte memory i.e. 1MB . Out of 20 bits, 16 bits A_0 to A_{15}(or 16 lines) are multiplexed with a data bus. By multiplexing, it means they will act as address lines during the first T state of the machine cycle and in the rest, they act as data lines. A_{16} to A_{19} are multiplexed S_3 to S_6 and BHE' is multiplexed with S_7.

Control signals provided by 8086 for memory operations and i/o interfacing :
They are used to identifying whether the bus is carrying a valid address or not , in which direction data is needed to be transferred over the bus, when there is valid write data on the data bus and when to put read data on the system bus. Therefore, their sequence pattern makes all the operations successful in a particular machine cycle.



Min mode circuit

8282 (8 bits) latch :
The latches are buffered D FF. They are used to separate the valid address from the multiplexed Address/data bus by using the control signal ALE, which is connected to strobe(STB) of 8282. The ALE is active high signal. Here three such latches are required because the address is 20 bits.
8286 (8 bits) transceivers :
They are bidirectional buffers and also known as data amplifiers. They are used to separate the valid data from multiplexed add/data bus. Two such transceivers are needed because the data bus is 16 bits long. 8286 is connected to DT/R' and DEN' signals. They are enabled through the DEN signal .The direction of data on the data bus is controlled by the DT/R' signal. DT/R' is connected to T and DEN' is connected to OE'.

DEN'	DT/R'	Action
1	X (don't care)	Transreceiver is disabled
0	0	Receiver data
0	1	Transmit data

- 8284 clock generator is used to provide the clock.
- M/IO'= 1,then I/O transfer is performed over the bus. and when M/IO' = 0, then I/O operation is performed.
- The signals RD' and write WR' are used to identify whether a read bus cycle or a write bus cycle is performing. When WR' = 0 ,then it indicates that valid output data on the data bus.
- RD' indicates that the 8086 is performing a read data or instruction fetch process is occurring .During read operations, one other control signal is also used, which is DEN (data enable) and it indicates the external devices when they should put data on the bus.
- Control signals for all operations are generated by decoding M/IO', RD', WR'. They are decoded by 74138 3:8 decoder.

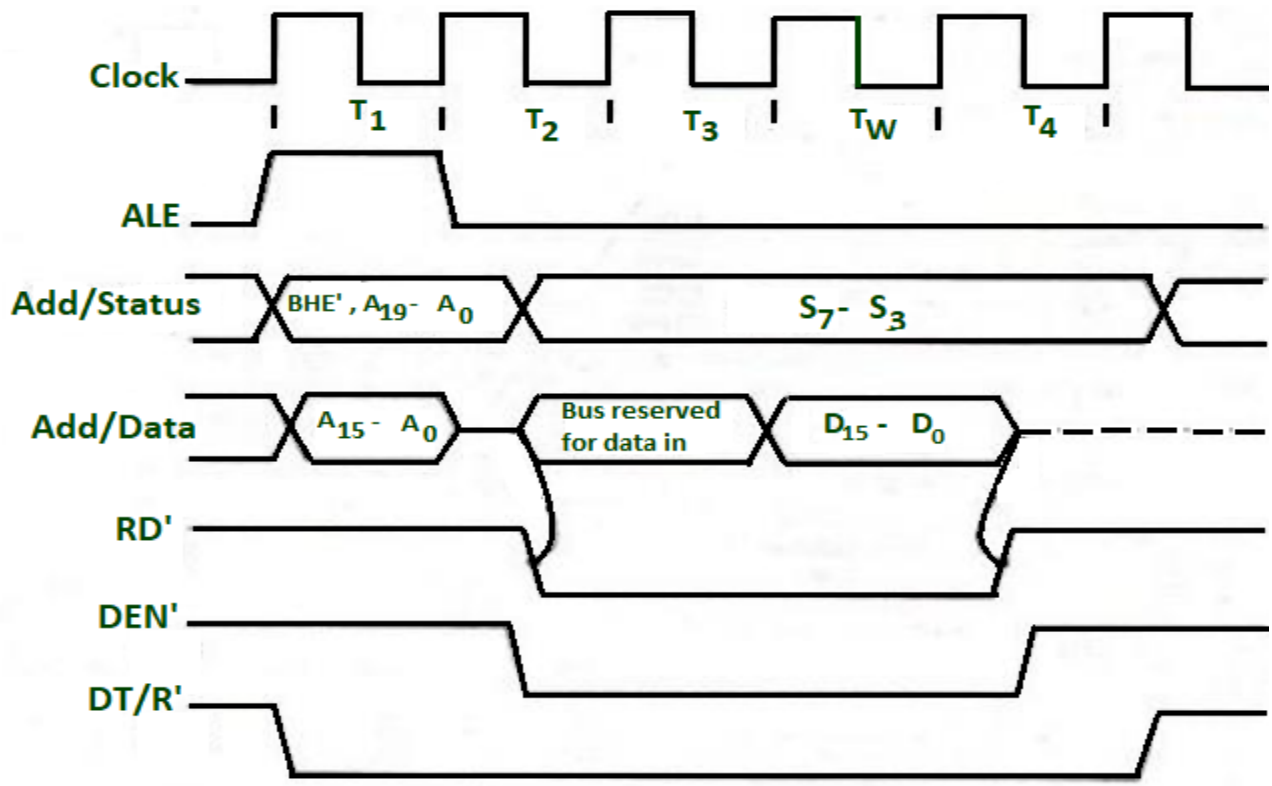
M/IO'	RD'	WR'	Action
1	0	1	Memory Read
1	1	0	Memory Write
0	0	1	I/O Read
0	1	0	I/O Write

- **INTR and INTA :**
  When INTR = 1,then there is an interrupt to 8086 by other devices for their service. When INTA'= 0,then it indicates that the processor is ready to service them.
- The bus request is made by other devices using the HOLD signal and the processor acknowledges them using the HLDA output signal.
- For more details about the 8086 minimum mode pins please refer(this article).

Timing diagram :

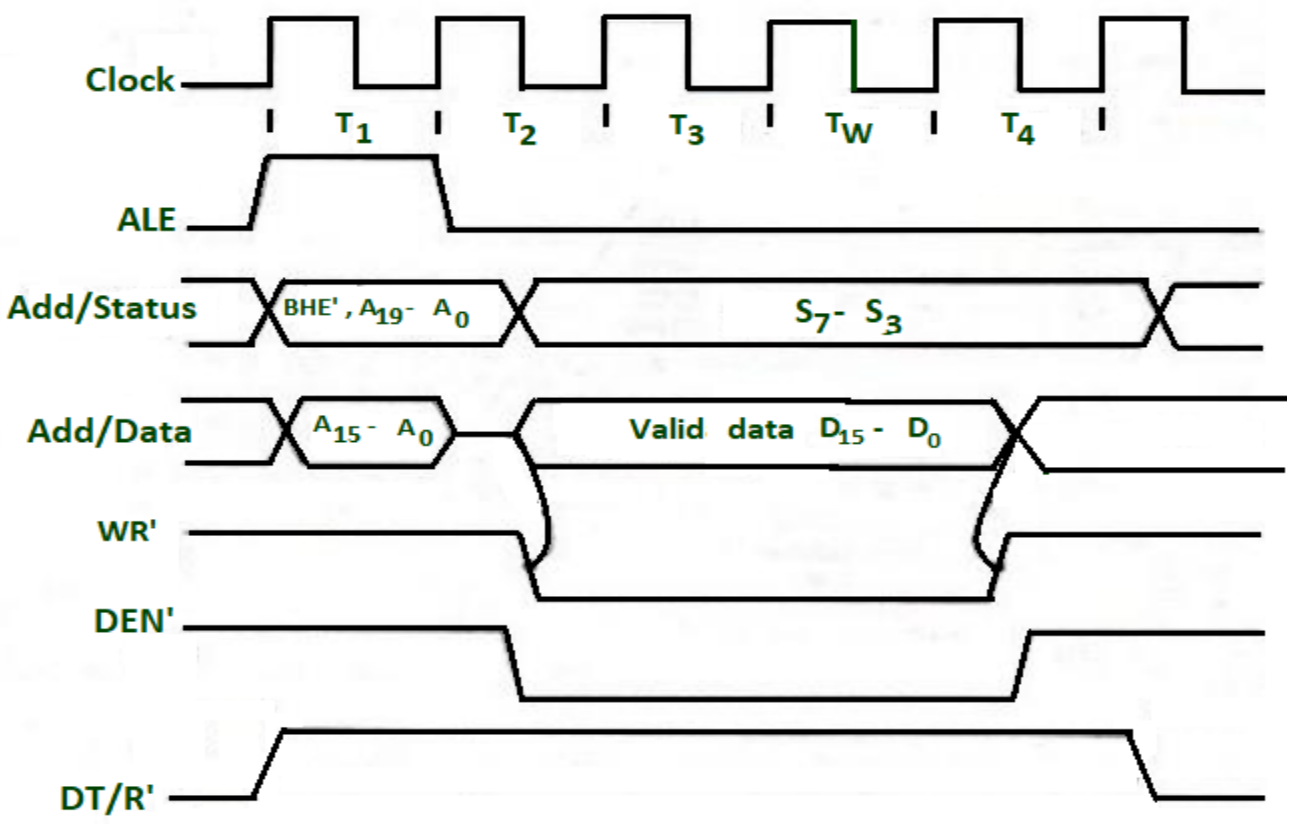
The working of min mode can be easily understood by timing diagrams.

- All processors bus cycle is of at least 4 T-states(T_1,T_2,T_3,T_4) .The address is given by processor in the T1 state. It is available on the bus for **one T-state.**
- In T_2, the bus is tristated for changing the direction of the bus(in the case of a data read cycle.)
- The data transfer takes place between T_3 and T_4.
- If the addressed device is slower, then the wait state is inserted between T_3 and T_4.



Opcode fetch or read timing diagram

- At T_1 state ALE =1 ,this indicates that a valid address is latched on the address bus and also M / IO'= 1, which indicates the memory operation is in progress.
- In T_2, the address is removed from the local bus and is sent to the addressed device. Then the bus is tristated.
- When RD' = 0 , the valid data is present on the data bus.
- During T_2 DEN' =0, which enables transceivers and DT/R' = 0 ,which indicates that the data is received.
- During T_3, data is put on the data bus and the processor reads it.
- The output device makes the READY line high. This means the output device has performed the data transfer process. When the processor makes the read signal to 1, then the output device will again tristate its bus drivers.



Write memory cycle

- At T_1 state ALE =1 ,this indicates that a valid address is latched on the address bus and also M / IO'= 1, which indicates the memory operation is in progress.
- In T_2, the processor sends the data to be written to the addressed location.
- The data is buffered on the bus until the middle of T_4 state.
- The WR'=0 becomes at the beginning of T_2.
- The BHE' and A0 signals are used to select the byte or bytes of memory or I/O word.
- During T_2 DEN' =0, which enables, transceivers and DT/R' = 1 ,which indicates that the data is transferred by the processor to the addressed device.

All kinds of memory and i/o operations are performed using the decoding of M/IO'and RD' WR' as shown in the table above.

Maximum mode configuration of 8086 microprocessor (Max mode)

8086 microprocessor characteristics:

- It contains 20 bit address bus.
- It contains 16-bit data bus, therefore 8086 is called as **16-bit microprocessor.**
- It is 2-stage pipelined processor. It can prefetch 6 bytes from memory and store into queue to increase the speed of the execution.
- It's control bus carries signals for executing operations such as read ,write etc.

- It has Memory Banks. 2 banks of 512KB each. These banks are called as lower Bank (even) and higher Bank (odd).
- In 8086 the entire memory is divided into four memory segments which are code ,stack, data and extra segment.
- 8086 has 16 bit IO address.
- It has 256 interrupts.

8086 has two operating Modes:

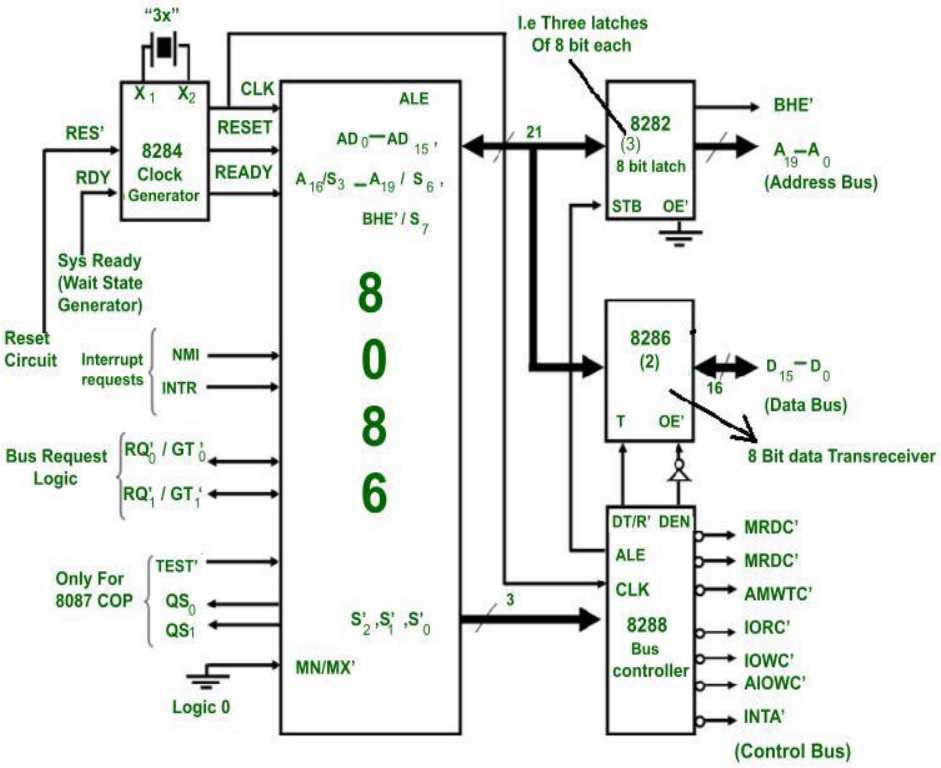
1. Minimum mode
2. Maximum mode

Minimum mode:

- In this 8086 is the only processor in the system . In a minimum mode 8086 system.
- 8086 is operated in minimum mode when MN/MX' pin to logic 1.
- In this mode, all the control signals are given out by the 8086 itself.

Maximum mode*:*

- In this we can connect more processors to 8086 (8087/8089).
- 8086 max mode is basically for implementation of allocation of global resources and passing bus control to other coprocessor(i.e. second processor in the system), because two processors can not access system bus at same instant.
- All processors execute their own program.
- The resources which are common to all processors are known as global resources.
- The resources which are allocated to a particular processor are known as local or private resources.



Maximum mode circuit

Circuit explanation:

- When MN/ MX' = 0 , 8086 works in max mode.
- Clock is provided by **8284 clock generator.**
- **8288 bus controller-** Address form the address bus is latched into 8282 8-bit latch. Three such latches are required because **address bus is 20 bit**. The ALE(Address latch enable) is connected to STB(Strobe) of the latch. **The ALE for latch is given by 8288 bus controller.**
- The data bus is operated through 8286 8-bit transceiver. Two such transceivers are required, because **data bus is 16-bit**. The transceivers are enabled the DEN signal, while the direction of data is controlled by the DT/R signal. DEN is connected to **OE'** and **DT/ R'** is connected to T. **Both DEN and DT/ R' are given by 8288 bus controller.**

DEN (Of 8288)	DT/ R'	Action
0	X	Transreceiver is disabled
1	0	Receive data
1	1	Transmit data

- **Control signals for all operations are generated by decoding S'_2, S'_1 and S'_0 using 8288 bus controller.**

S'_2	S'_1	S'_0	Processor State (What the µP wants to do)	8288 Active Output (What Control signal should 8288 generate)
0	0	0	Interrupt Acknowledge	INTA'
0	0	1	Read I/O Port	IORC'
0	1	0	Write I/O Port	IOWC' and AIOWC'
0	1	1	Halt	None
1	0	0	Instruction Fetch	MRDC'
1	0	1	Memory Read	MRDC'
1	1	0	Memory Write	MWTC' and AMWTC'
1	1	1	Inactive	None

- Bus request is done using RQ' / GT' lines interfaced with 8086. RQ_0/GT_0 has more priority than RQ_1/GT_1.
- INTA' is given by 8288, in response to an interrupt on INTR line of 8086.
- In max mode, the advanced write signals get enabled one T-state in advance as compared to normal write signals. This gives slower devices more time to get ready to accept the data, therefore it reduces the number of cycles.

Advantages of max mode of 8086:

- It helps to interface more devices like 8087.This interface is also called a **closely coupled co-Processor configuration**. In this 8086 is called as the host and 8087 as Co-processor.
- It supports multiprocessing, Therefore it helps to increase the efficiency.
- The 8087 was the first floating-point coprocessor for the 8086 series of microprocessors. The purpose of the 8087 was to increase calculations for floating point operations, such as add, sub, multiply, divide, and square root.

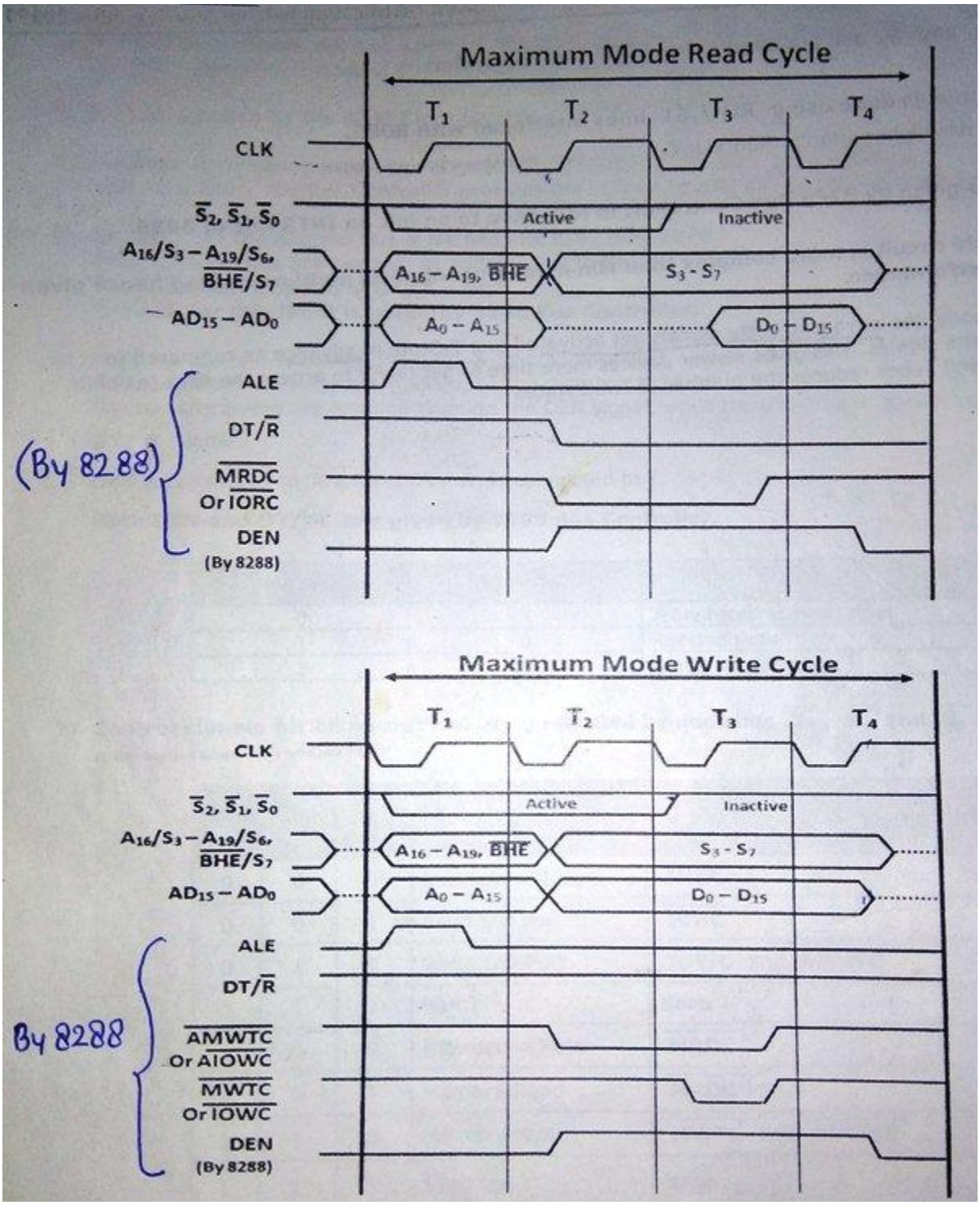
Disadvantages of max mode over min mode :

- It has more complex circuit than min mode.

Applications of 8086:

- Microcomputer are built using 8086. **For example** : IBM PC, used the Intel 8088, a version of the 8086 with 8-bit data bus.
- It is used in calculators.
- It is used for control purposes like in traffic signals(uses micro controllers which are nothing but contains **one or more CPUs along with memory and programmable i/o peripherals**).

Timing diagram :



Differences between 8086 and 8088 microprocessors

Though the architecture and instruction set of both 8086 and 8088 processors are same, still there are differences between them.
Following is the table listing the differences between the 2 microprocessors:

S. No.	8086 microprocessor	8088 microprocessor
1	The data bus is of 16 bits.	The data bus is of 8 bits.
2	It has 3 available clock speeds (5 MHz, 8 MHz (8086-2) and 10 MHz (8086-1)).	It has 3 available clock speeds (5 MHz, 8 MHz)
3	The memory capacity is 512 kB.	The memory capacity is implemented as a single 1 MX 8 memory banks.
4	It has memory control pin (M/IO) signal.	It has complemented memory control pin (IO/M) signal of 8086.
5	It has Bank High Enable (BHE) signal.	It has Status Signal (SSO).
6	It can read or write either 8-bit or 16-bit word at the same time.	It can read only 8-bit word at the same time.
7	Input/Output voltage level is measured at 2.5 mA.	Input/Output voltage level is measured at 2.0 mA
8	It has 6 byte instruction queue.	It has 4 byte instruction queue as it can fetch only 1 byte at a time.
9	It draws a maximum supply current of 360 mA.	It draws a maximum supply current of 340 mA.