I Evolution of Microprocessors

We divide the years of development of microprocessors as 5 generations

First generation (1971 – 73)
Intel Corporation introduced 4004, the first microprocessor in 1971. It is evolved from the development effort while designing a calculator chip.

There were three other microprocessors in the market during the same period:

- Rockwell International's PPS-4 (4 bits)
- Intel's 8008 (8 bits)
- National Semiconductor's IMP-16 (16 bits)

They were fabricated using PMOS technology which provided low cost, slow speed and low output currents
They were not compatible with TTL.

Second Generation (1974 – 1978)
Marked the beginning of very efficient 8 – bit microprocessors.
Some of the popular processors were:

- Motorola's 6800 and 6809
- Intel's 8085
- Zilog's Z80

They were manufactured using NMOS technology.
This technology offered faster speed and higher density than PMOS
It is TTL compatible

Third generation microprocessors (1979 – 80)
This age is dominated by 16 – bits microprocessors
Some of them were:

- Intel's 8086/80186/80286
- Motorolla's 68000/68010

They were designed using HMOS technology
HMOS provides some advantages over NMOS as
Speed-power-product of HMOS is four times better than that of NMOS
HMOS can accommodate twice the circuit density compared to NMOS

Intel used HMOS technology to recreate 8085A and named it as 8085AH with a higher price tag.

Fourth Generation (1981 – 1995)

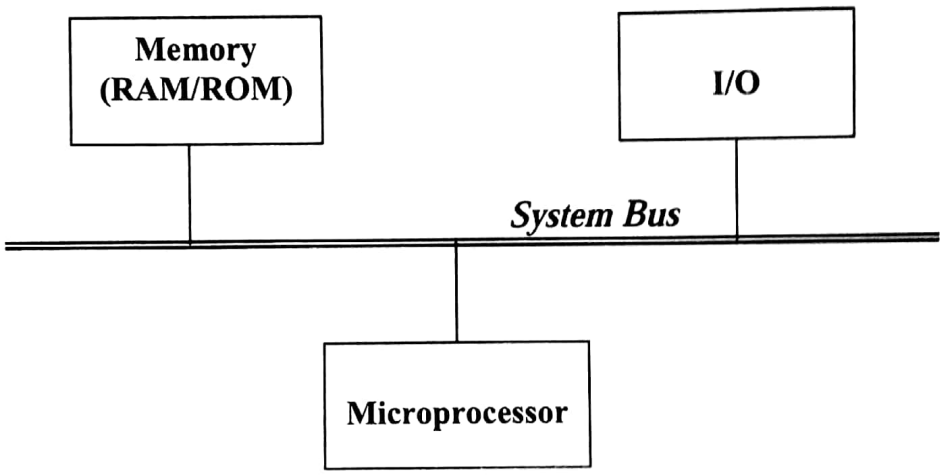
- This era marked the beginning of 32 bits microprocessors
- Intel introduced 432, which was bit problematic
- Then a clean 80386 in launched.
- Motorola introduced 68020/68030.

They were fabricated using low-power version of the HMOS technology called HCMOS. Motorola introduced 32-bit RISC processors called MC88100

Fifth Generation (1995 – till date)

This age the emphasis is on introducing chips that carry on-chip functionalities and improvements in the speed of memory and I/O devices along with introduction of 64-bit microprocessors.
Intel leads the show here with Pentium, Celeron and very recently dual and quad core processors working with up to 3.5GHz speed.

II Microcomputer Hardware:



The micro computer consists of

1. System Bus
2. Microprocessor
3 Memory unit
4 Input/Output unit

1. The System Bus:

This is further divided into

- Address Bus,
- Data Bus, and
- Control Bus

They together connect microprocessor to each of memory and I/O elements which facilitates the information transformation between them.

Address Bus:

- Unidirectional – from microprocessor to memory or I/O elements
- Usually 8 to 32 bits wide
- The number of unique addresses a microprocessor can generate depends on the width of this bus
  For example, 8085 has 16 – bits address bus. So, it can generate $2^{16} = 65,536$ different addresses.
  A different memory location or an I/O element can be represented by these addresses

Data Bus:

- Bidirectional – data is brought on these lines prior to an operation and results are sent back to selected memory location or I/O using these lines only.
- The width determines the amount of information that can be brought/sent at once, more precisely in one machine cycle into or out of processor.
  For Example, 8085 microprocessor has 8-bit data bus. Only one byte of information can be fetched in or sent out of processor using this data bus.

Control Bus:

- Some signals are unidirectional and some are bidirectional
- Transmits signals that are used to synchronize the operation of the individual microcomputer elements.
- Typical control signals include READ, WRITE, and RESET

2. The Microprocessor

The Microprocessor is fabricated on a single chip using MOS technology.
It comprises of

i) Register section
ii) One or more ALU, and
iii) A control unit

i) Register Section

Classification of processors based on register section:

Accumulator based microprocessors – Intel's 8085, Motorola's 6809

- Here, one of the operands is assumed to be held a special register called "accumulator".
- All arithmetic and logic operations are performed using this register as one of the data source and result is stored back in the accumulator.
- One-operand instructions are predominant in this organization

The general-purpose register based microprocessors – Intel's 8086/386, Motorola's 68000/020

- These processors have a set of registers which can be used to hold data, memory addresses or the results of an arithmetic or logic operations for indefinite amount of time.

- The number and size of these registers vary from processor to processor
- Some registers are general purpose while others are earmarked with some functions.
- General purpose registers are used to store addresses or data for an indefinite time and are capable of manipulating data by shift or rotate operations.
- Typical dedicated registers include,
  a. *Program Counter (PC)*
  b. *Instruction Register (IR)*
  c. *Status Register or Flag Register*
  d. *Stack Pointer (SP)*
  e. *Barrel Shifter*

The *PC* always contains the address of the next instruction to be executed. Its contents are automatically updated by ALU. The microcomputer executes a program sequentially unless it encounters a jump/branch/call instruction. At that time, PC will be loaded with the address present in the instruction. The size of the PC itself varies from one processor to another. For example, the 8085 has a 16-bit PC, while 68029 has 32-bits PC.

The *Instruction register* contains the instruction to be executed. After fetching the instruction from memory, microprocessor places it in IR for translation.

The *Status Register* contains individual bits each having a special meaning. The bits are termed as flags. Each flag is set or reset by an ALU operation. These flags or used by Conditional branch instructions. Typical flags include carry, sign, zero and overflow.

- The carry (C) flag is used to reflect whether or not an arithmetic operation such as ADD generates a carry. If carry is generated then CF = 1 else CF = 0. The carry is generated out of 8^{th} bit for byte operations, 16^{th} bit for word operations etc. Carry is used as Borrow flag for subtraction.

- The Zero (Z) flag is used to indicate whether the result of an arithmetic or logic operation is zero. ZF = 1 for zero result and ZF = 0 for a non-zero result.

- The Sign(S) flag indicates whether the result is positive or negative. SF = 1 indicates negative result means the most significant bit of the result is 1. If SF = 0, the result is a positive number. It is observed only for signed operations. This flag can be ignored for the result on an unsigned arithmetic or logic operations.

- The Overflow (O) flag is set if the result of an arithmetic and logical operation on signed numbers is too large for the microprocessor's maximum word size. OF can be shown as OF = C7 $\oplus$ C8 where C7 is the final carry and C6 is the previous carry. Once again, this applies to signed numbers only.

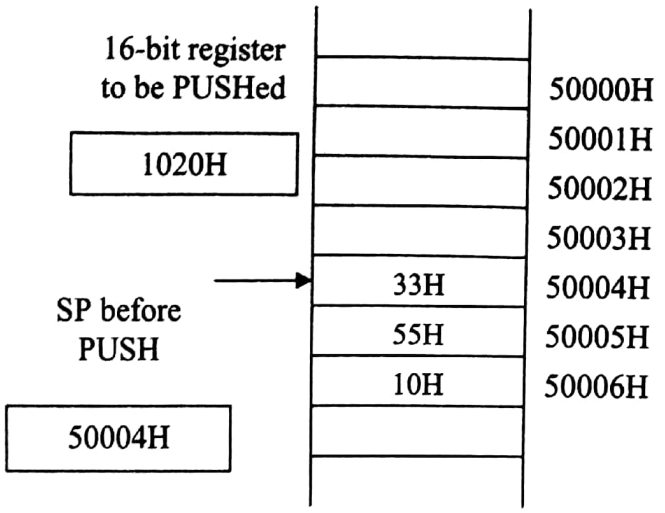
The *Stack Pointer (SP)* register addresses the stack.

- A stack is a Last-In-First-Out read/write memory. The items that go in last will come out first. This is because all read (POP) and write (PUSH) operations will take place from one end called top of the stack(tos).
- Stack is implemented using hardware or software.
- The hardware stack is designed by using a set of high speed registers to provide a fast response. The disadvantage is that stack size is limited. But push and pop operations are
- very fast. Intel's 4040, an 8 – bit processor used hardware stack.
- The software stack on the other hand, is implemented using a portion of memory. Some RAM locations are earmarked as stack. The advantage is that they provide unlimited space for stack, depends on the amount of memory we interface to microprocessor, though. But it is slower than hardware stack.

- The SP always contains the memory address of the last byte of the currently pushed item on tos i.e. it always points to the tos. Stack is normally used by subroutines or interrupts for saving certain registers such as the program counter and status register.

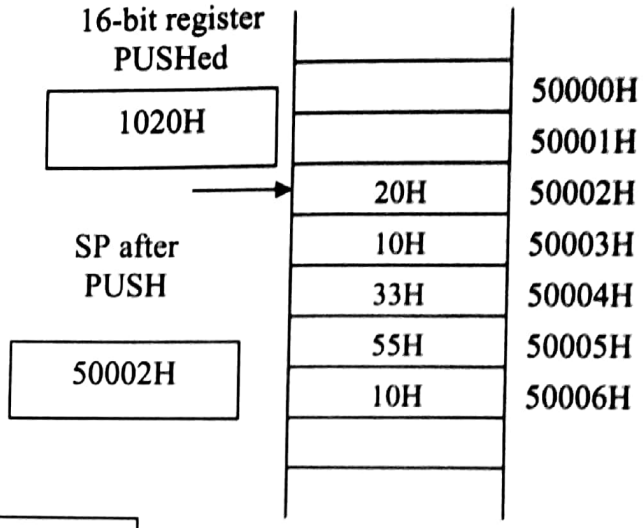
PUSH and POP operations:

- If the stack is accessed from the top, the stack pointer is decremented after a PUSH operation and incremented before POP.
- On the other hand, if the stack is accessed from the bottom, SP is incremented after a PUSH and decremented after a POP.
- Typical microprocessors access stack from top.
- Depending upon the microprocessor, 8-, 16- or 32-bits can be pushed onto or popped from the stack.
- The value by which the SP is incremented or decremented after PUSH or POP operations depends on the register size.
- For example, in 8086 microprocessor, PUSH and POP operations can be done only on 16-bit data. Hence, SP is incremented or decremented by a value of 2 always.

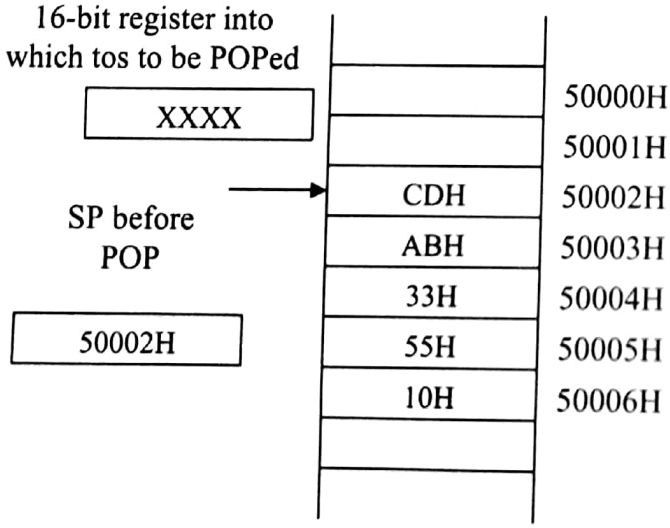


Before PUSH

After PUSH

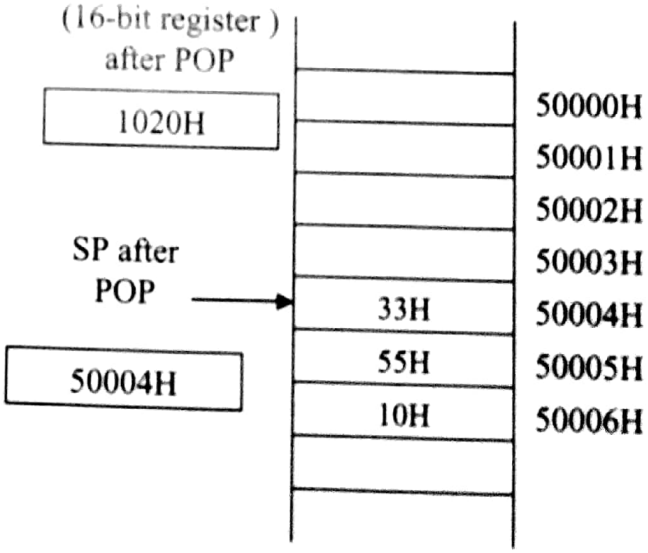


(SP) = (SP) – 2



Before POP

After POP



(SP) = (SP) + 2

Consider the PUSH operation as shown in the fig. when the stack is accessed from the top. SP is decremented by 2 after PUSH.
Similarly, after POP operation SP is incremented by 2, since we are accessing stack from top.

Index Register:
Index register is useful with instructions where tables or arrays of date are accessed. Here, Index Register can be used to manipulate the address portion of the instruction.. Thus appropriate data in the table can be accessed. The actual address called physical address of the data is calculated by adding address portion in the instruction with contents of the index register.

In 8086: MOV AL, 200[SI] means one byte present at an address (DS) + 200 + (SI) will be moved into AL register. (DS – Data Segment register).

Barrel Shifter:
32-bit processors include a special type of register called Barrel Shifter. This register provides faster shift operation. For example, Intel's 80386 barrel shifter can shift a number from 0 through 64 positions in one clock period.

ALU (Arithmetic and Logical Unit)

- ALU performs all arithmetic and logic operations on data.
- The size of ALU defines the size of the microprocessor.
- For example Intel 8086 is a 16-bit microprocessor since its ALU 16-bits wide. Intel 8088 is also a 16-bit microprocessor even though its data bus is 8-bits wide. That is because of its 16-bit ALU
- Some 32-bit microprocessors like Motorola 68030 include multiple ALUs for parallel operations to achieve faster speed.

The Control Unit:
The CU performs basically two tasks:
Instruction interpretation:

i) CU reads instruction from memory using PC
ii) It then recognizes the instruction type, gets necessary operands, and routes then to appropriate functional units of execution unit
iii) Necessary signals are issued to perform desired operation
iv) Results are routed to the specified destination.

Instruction Sequencing:
The CU determines the address of the next instruction to be executed and loads it into PC.

The CU is designed using one of the three techniques:

i) Hardwired Control
Designed by physically connecting typical components such as gated and flip-flops. For example, Zilog's 16-bit Z8000

ii) *Microprogramming*
This type of CUs include a control ROM for translating the instructions.
Intel's 8086 is a microprogrammed microprocessor

iii) *Nanoprogramming*
It includes two ROMs inside CU. The first ROM, which is called microROM stores all the addresses of the second ROM, which is called nanoROM. If the microinstructions repeat many times in a microprogram, use of two level ROMs provides tremendous memory savings.
Motorola's 68000, 68020 and 68030 are nanoprogrammed.

3. Memory Organization:

i) Memory unit is the integral part of any microcomputer system and its primary purpose is to hold program and data

ii) The major design goal of memory unit is to allow it to operate at a speed close to that of the processor.

iii) The cost factor inhibits the design of entire memory unit with single technology that guarantees high speed.

iv) In order to seek a trade-off between the cost and operating speed, a memory system is usually designed with different technologies such as solid state, magnetic and optical.

In a broad sense, a microcomputer memory can be logically divided into three groups:

i) *Processor Memory*
ii) *Primary or Main Memory*
iii) *Secondary memory*

Processor Memory refers to a set of CPU registers. These registers are useful to hold temporary results when a computation is in progress. Also, there is no speed disparity between the registers and the microprocessor because they are fabricated using the same technology.
The main disadvantage is the cost involved which forces the architect to include very few registers (usually 8 to 16 only) in the microprocessor.

Primary memory –

- is the storage area in which all the programs are executed.
- The processor can directly access only those items that are stored in the primary memory.
- All the programs and corresponding data must be within primary memory prior to execution. MOS technology is normally used in primary memory design.
- The size of primary memory is much larger compared to processor memory but its operating speed is slower than processor registers by a factor of 25.

Secondary memory refers to the storage medium for huge files such as program source codes, compilers, operating systems, RDBMSs etc. These are not needed very frequently. They comprises of slow devices such as magnetic tapes and optical disks.

Sometimes they are referred to as auxiliary or backup store.

Classification of Primary Memory:

Primary memory normally includes ROM (Read Only Memory) and RAM(Random Access Memory).

As the name implies, a **ROM** permits only a read access. There are many kinds of this category. For example,

- Some ROMS are custom made, their contents are programmed by the manufacturer. They are called mask programmable ROMs. Since they are mass produced, they are inexpensive.
- Sometimes a user has to program the ROM in field. Such types of ROMs which allow this operation are called PROMs (Programmable ROMs). The main disadvantage is that they cannot be reprogrammed.
- In practice, it is necessary to alter the programs before they are put in market. ROMs that allow reprogramming are called Erasable Programmable Read-Only Memories (EPROMs). In an EPROM, programs are entered using electrical impulses and the stored information is erased using UV rays.
- With advances in IC technology, it is possible to achieve an electrical means of erasure. These new ROMs are called Electrically Alterable ROMs (EAROMs) or Electrically Erasable PROMs (EEPROMs).

⇒ These memories are usually called Read Mostly Memories (RMMs), since they have much slower writing times than read times.

Information stored in semiconductor random access memories will be lost if the power is turned off. This property is known as volatility and hence, **RAMs** are usually called volatile memories. Stored information in a magnetic tape or magnetic disk is not lost when the power is turned off. Therefore these storage devices are called nonvolatile memories. ROM is a nonvolatile memory.

- In a semiconductor memory constructed using bipolar transistors, the information is stored in the form of voltage levels in flip-flops. These voltage levels do not usually get drifted away. Such memories are called *static RAMs* because stored information remains constant for some period of time.
- On the other hand, semiconductor memories designed using MOS transistors, the information is held in the form of electrical charges in capacitors. Here the stored charge has the tendency o get leaked away. These memories are referred to as *dynamic RAMs*. In order to prevent any information loss, dynamic RAMS have to be refreshed at regular intervals. Refreshing means boosting the signal level and writing it back. This activity is performed using a hardware unit called "refresh logic".

- Since the static RAM maintains information in active circuits, power is required even when the chip is inactive or standby mode. Hence, static RAMs require large power supplies. Also each static RAM cell is about four times larger in area than an equivalent dynamic cell.

Differences between static and dynamic RAMs:

Static RAM	Dynamic RAM
1. This semiconductor memory is constructed using bipolar transistors	1. This semiconductor memory is constructed using MOS transistors
2. Information is stored in the form of voltage levels in flip-flops	2. Information is stored in the form of electrical charges in capacitors
3. These voltage levels do not get drifted away	3. Has tendency of leakage
4. No refresh logic is needed	4. Refresh logic is necessary since leakage of electrical charges
5. Power is required even when the chip is in standby mode	5. Refresh login is inbuilt, so draws less power comparatively.
6. Four time larger in size compared to an equivalent dynamic cell	6. Four times as many bits as a static RAM chip.

Classification of Primary Memories: