

Figure : Internal structure of CACHE Memory

- **CACHE HIT:**
- Number of times CPU get the data from CACHE.
- In CACHE HIT CPU initially start it's search with CACHE only.
- **HIT Ratio:**
- It is the ratio of number of successful hits to total number of hits.
- HIT ratio should be as high as possible.
- Usually it should be 98 percent out of 100 percent.

- **CACHE MISS:**
- **The number of times CPU do not get data from CACHE.**
- **CACHE MISS is also known as MISS penalty time.**
- **MISS penalty time is measured in terms of time taken to replace the new block in CACHE from main memory.**
- **Replacement of new block by old block in CACHE by dropping the old block in CACHE.**
- **SWAP IN:**
- **SWAP OUT:**

- **Direct Mapped CACHE:**
- **CPU places main memory address and cache memory address has to be derived.**

CACHE	0	1	2	3	4	5	6	7	8	9
MAIN	0	1	2	3	4	5	6	7	8	9
MEMORY	10	11	12	13	14	15	16	17	18	19
	20	21	22	23	24	25	26	27	28	29
	30	31	32	33	34	35	36	37	38	39

Figure : Direct mapped CACHE block

- **Algorithm:** To find out the cache address.
- **CACHE ADD.= MODULO(No. of locations of CACHE)(location of main memory).**
- **Example:**
CACHE ADD.= mod(10)(23)
CACHE ADD.= 3

where 23 is the main memory location , 10 are total locations of cache and 3 is remainder.

- **CACHE ADD.= mod(8)(21)**
- **CACHE ADD.=5**

CACHE	0	1	2	3	4	5	6	7
MAIN	0	1	2	3	4	5	6	7
MEMORY	8	9	10	11	12	13	14	15
	16	17	18	19	20	21	22	23
	24	25	26	27	28	29	30	31

Figure : Direct mapped CACHE block

- **CACHE Contents:**

- **Example:**

(32)=Main memory location

Where:

CACHE ADD.=2

VALID=1

TAG=3

DATA=(MM content at location 32)

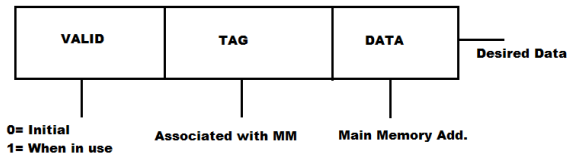


Figure : CACHE contents in terms of field values

- Let us assume CPU generating the followings MM address continuously.
- 22, 23, 17, 22, 17, 7, 17, 22,.....

MM ADD.	CACHE	HIT/MISS	VALID	TAG	DATA
22	2	MISS	0 (int.) & 1	2	(22)
23	3	MISS	0 (int.) & 1	2	(23)
17	7	MISS	0 (int.) & 1	1	(17)
22	2	HIT	1	2	(22)
7	7	MISS	1	0	(7)
22	2	HIT	1	2	(22)

Figure : Different examples of HIT/MISS

- **Block Size of CACHE Memory:**
- **It is unit of the information passes between two level i.e. CPU and MM.**
- **Example:**
- **Block size=Word/Block**
- **Block size of 32 bit= 4 Word/Block and each Word belongs to 32 bit.**
- **CPU get the data from CACHE in the form of bits while on the other hand CACHE receive the data in the form of Block.**
- **Block size may vary between CACHE and MM.**

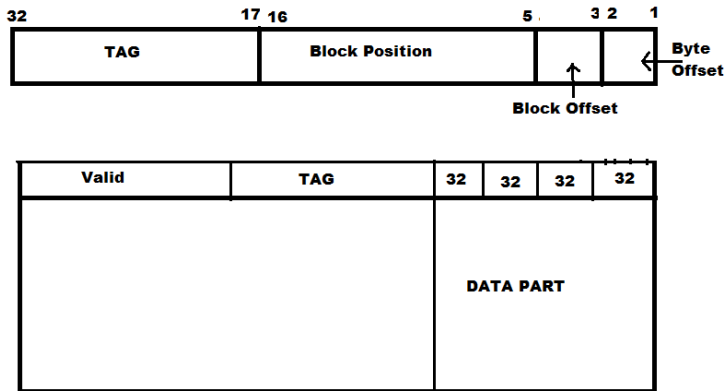


Figure : Example to calculate the size of CACHE

- **Byte Offset:**
 - It is the 2-bit representation of byte address field.
 - It indicate the which of the Byte out 4 Bytes/Block.
- **Block Offset:**
 - It is the 2-bit representation of Block address field.
 - It indicate the within the Block which of the 32 bit data.

- **Block Position:**
 - It is the 12-bit representation of the field.
 - It indicate the which of the Block out of number of Blocks.
- **TAG Filed:**
 - It indicates within the block whether the required information is available or not.

- **Example:**
- **For 32 bit data/4 blocks, 32 bit address is given for the CACHE of 64 KB(data part only). Calculate the size of the CACHE?**
- **Solution:**
- **Size of CACHE= 64KB**
- **Byte size of 4 blocks=4x4=16B**
- **Total no. of Blocks= 64KB/16B=4K=4000**
- **12 bits are required to represent the block position field.**
- **16 bits are required to represent the TAG filed.**
- **Size of the CACHE=(1+16+12+32x4)x4K**

- **Example:**
- **For a given CACHE the total no. of Blocks are 64 and 16 Bytes/Blocks. Calculate the position of Block whose Byte address is 1600.**
- **Solution:**
- **Size of CACHE= 64KB**
- **Byte size of Block=16B**
- **Total no. Blocks= Byte Add./Block size=1600/16=100**
- **Actual Block no.=100 MOD(64)=Q=1 and R=36**
- **Block no. 36 belongs to Byte address 1600.**

- **Read and Write Policies:**
- **Read Policy:**
- **In Read cycle CPU communicates with CACHE in terms of CACHE HIT or MISS.**
- **Write Policy:**
- **In Write cycle CPU may get following two difficulties.**
- **1:CPU can modify the data in CACHE but the content of MM and CACHE will be different.**
- **2:CPU can parallel write the data in CACHE and MM but speed will be issue with MM.**

- Write through the CACHE !! writing.
- Copy back and swap out
- I/O may required that data before updated data.
- Solution:
- Dirty Bit (DB):
- Like valid filed dirty bit also belongs to either 0 or 1 at a time.
- DB=0 (Data not altered)
- DB=1 (Data altered)

- Memory access time: Miss Penalty as low as possible.

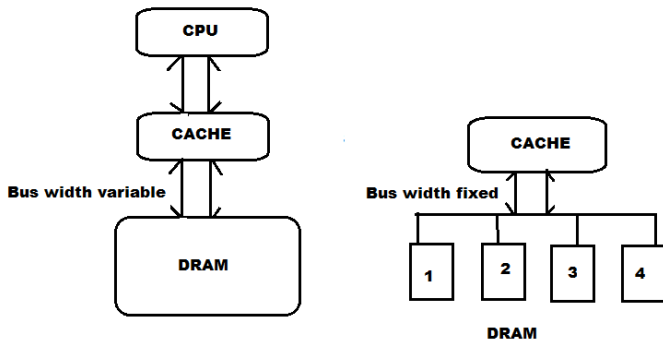


Figure : Examples of CACHE bus width

- **Example:**
- **One clock to send the address.**
- **10 clocks for each DRAM access.**
- **1 clock for send the memory word to CACHE from DRAM.**

C (CACHE Width)	D (DRAM Width)	MISS Penalty
4W	1W	$1 + (4W \times 10) + 1 \times 4 = 45$ (CLK)
4W	4W	$1 + (1W \times 10) + 1 = 12$ (CLK)
4W	2W	$1 + (2W \times 10) + 1 \times 2 = 23$ (CLK)
4W	4W	$1 + (1W \times 10) + 1 \times 4 = 15$ (CLK)

Figure : Examples of CACHE MISS penalty