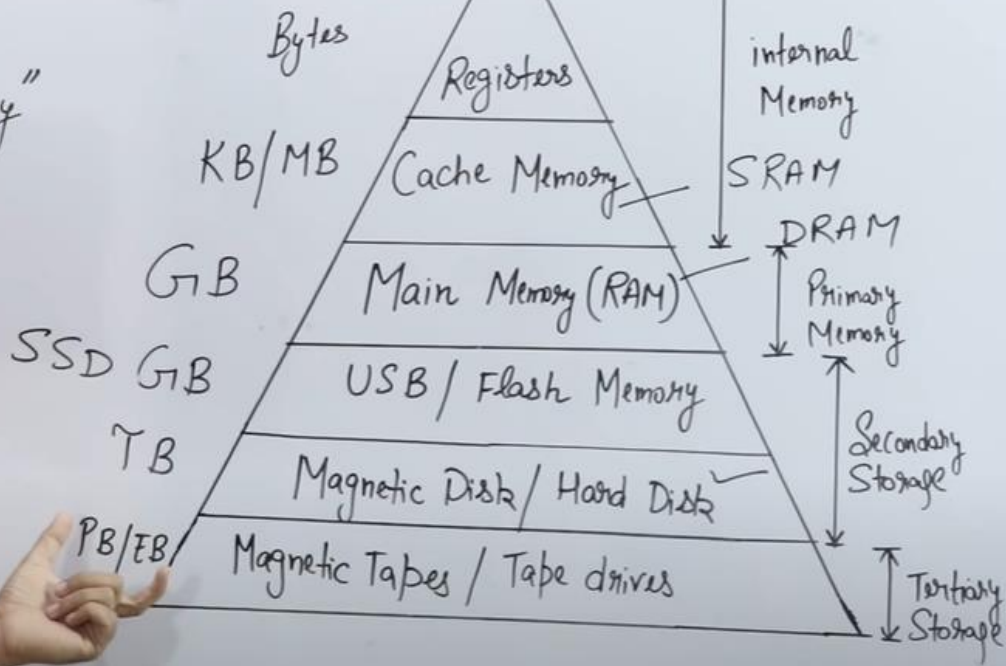
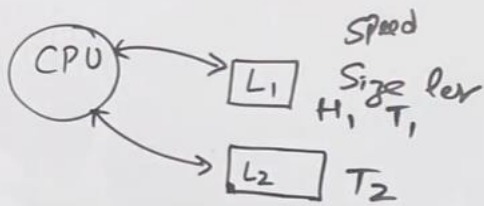


# Memory Hierarchy



## 2-level Memory Organization

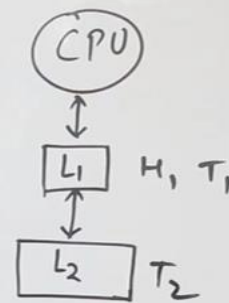
Independent Organization



$$AT(L_1) < AT(L_2)$$

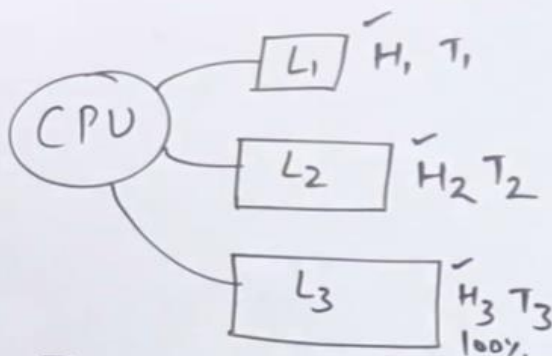
$$T_{avg} = H_1 T_1 + (1 - H_1) T_2$$

Hierarchical



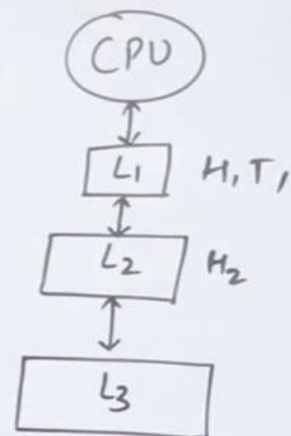
$$T_{avg} = H_1 T_1 + (1 - H_1) (T_1 + T_2)$$

## 3-Level Memory Organisation



$$T_{avg} = H_1 T_1 + (1 - H_1) H_2 T_2 + (1 - H_1) (1 - H_2) T_3$$

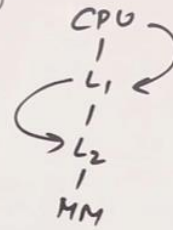
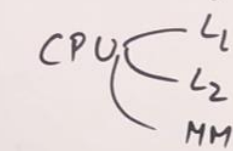
$$T_{avg} = H_1 T_1 + (1 - H_1) H_2 (T_1 + T_2) + (1 - H_1) (1 - H_2) (T_1 + T_2 + T_3)$$



Consider a system with 2-level caches. Access times of  $L_1$  and  $L_2$  and main memory are 1 ns, 10 ns and 500 ns. Hit ratio of  $L_1$  and  $L_2$  are .8 and .9. What is Avg. access time of system ignoring search time within cache?

- A) 13.0 ns  
B) 12.8 ns  
C) 12.6 ns  
D) 12.4 ns

(GATE 2004)



$$T_{avg} = H_1 T_1 + (1 - H_1) H_2 T_2 + (1 - H_1)(1 - H_2) T_3$$

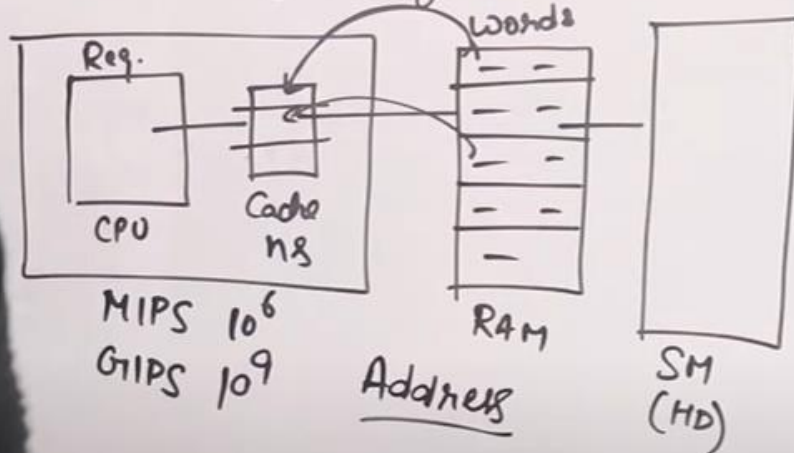
$$= .8 \times 1 + .2 \times .9 \times 10 + .2 \times .1 \times 500$$

$$= .8 + 1.8 + 10$$

$$= 12.6 \text{ ns}$$

## 'Cache Mapping and its types'

Direct Mapping      Fully associative      K-way Set associative



## 'Direct Mapping'

$$K \bmod n$$

Block No      No. of lines

$B_0$

$0 \bmod 4$

$1 \bmod 4$

$2 \bmod 4$

$4 \bmod 4$

Cache	
$L_0$	$B_0 B_4 B_8 B_{12} B_{28}$
$L_1$	$B_1 B_5 B_9 B_{13} B_{29}$
$L_2$	$B_2 B_6 B_{10} B_{14} B_{30}$
$L_3$	$B_3 B_7 B_{11} B_{15} B_{31}$

16 words  
Lines size  
 $\frac{16}{4} = 4$

$10$

$0001010$

2

$0 - 127 = 7 \quad 2^7 = 128$

$0 - 3 = 2^2 = 4$

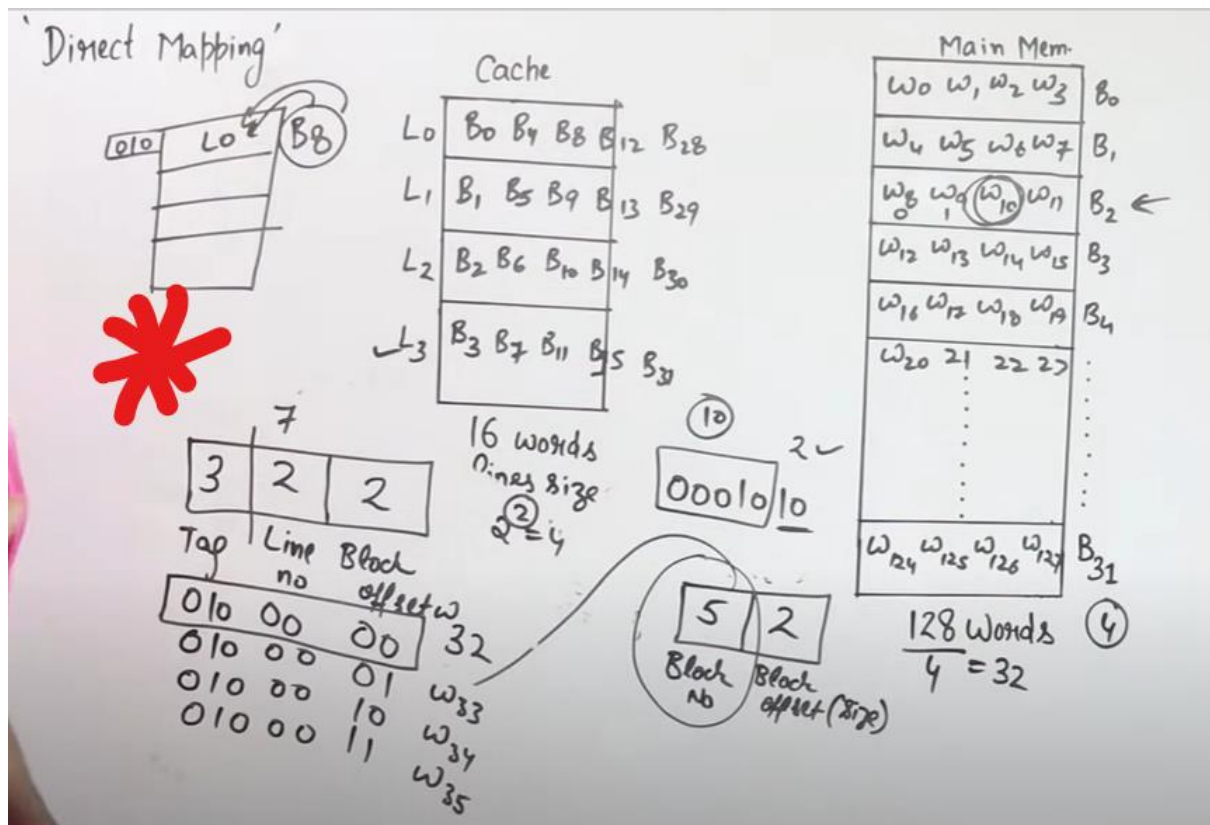
$5 \quad 2$

Block No      Block offset (size)

Main Mem	
$W_0 W_1 W_2 W_3$	$B_0$
$W_4 W_5 W_6 W_7$	$B_1$
$W_8 W_9 W_{10} W_{11}$	$B_2 \leftarrow$
$W_{12} W_{13} W_{14} W_{15}$	$B_3$
$W_{16} W_{17} W_{18} W_{19}$	$B_4$
$\vdots$	$\vdots$
$W_{124} W_{125} W_{126} W_{127}$	$B_{31}$

128 Words  
 $\frac{128}{4} = 32$

4

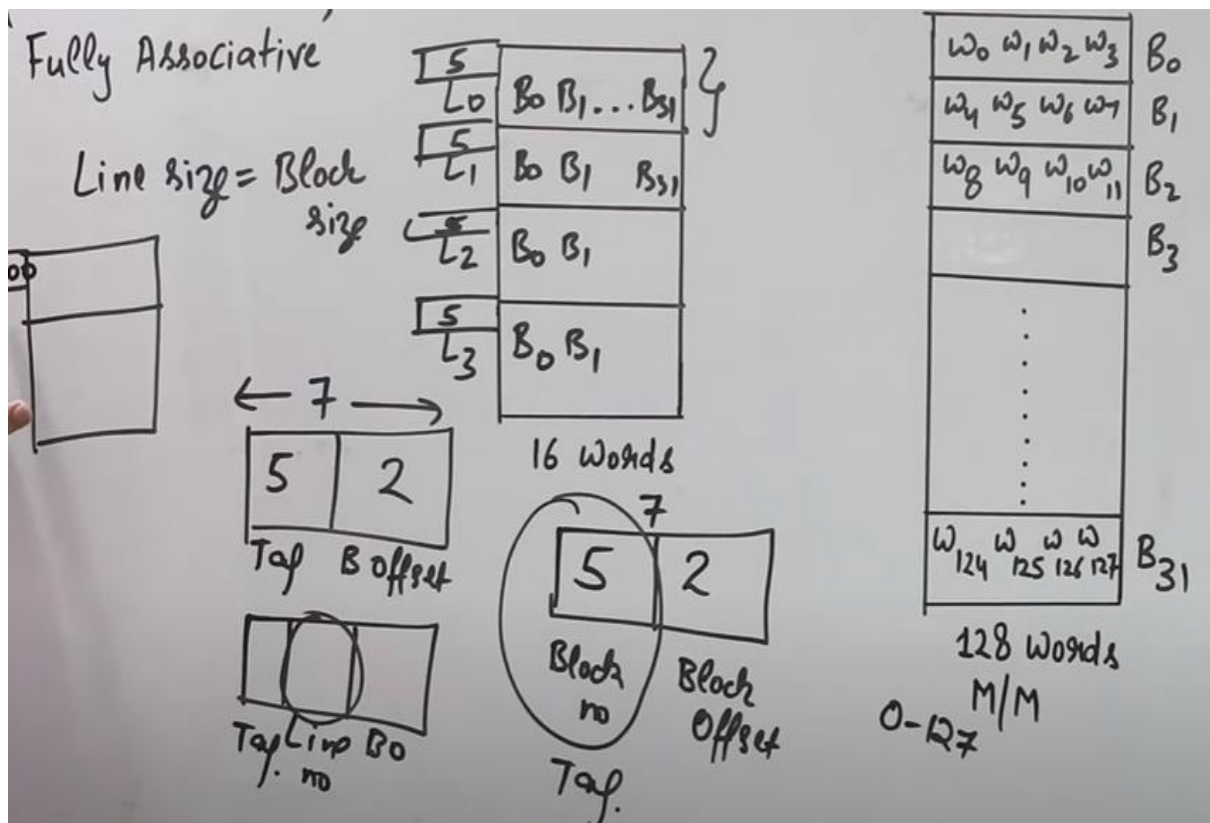


### Advantages:

1. Line number is fixed hence we do not need to search whole lines.
2. Hits are easy and fast (Searching time is low).

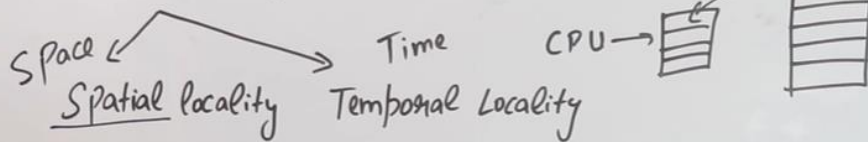
### Disadvantages:

1. Due to fixed place of block, even if we have space in other lines, we cannot place the block there (**Conflict Miss**).

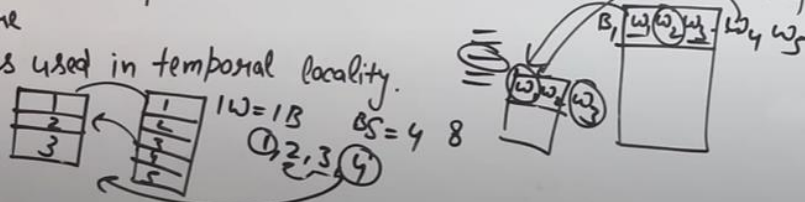




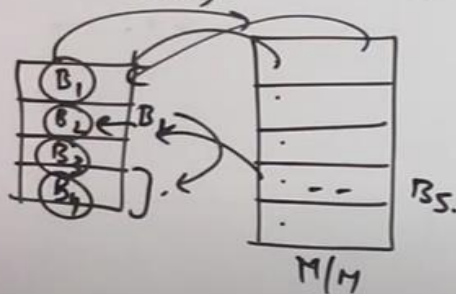
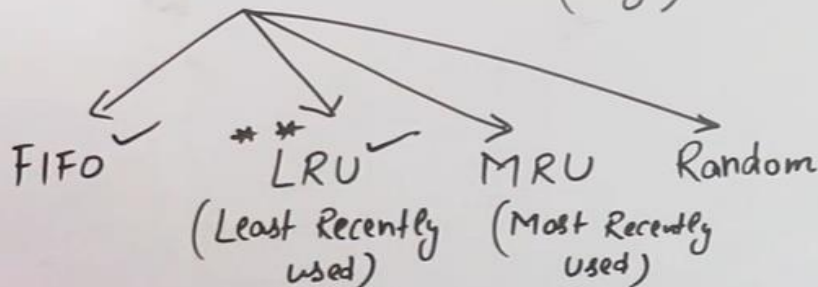
## Locality of Reference



- $\rightarrow$  if a word is accessed now then the word adjacent to it (close proximity) will be accessed next.
- $\rightarrow$  Keeping more words in a block affects spatial locality (block size)
- $\rightarrow$  if a word is referenced now then the same word will be referenced again in future
- $\rightarrow$  LRU is used in temporal locality.



## Cache Replacement Policies (Algo)



Miss Penalty Reduce  
Hit increase

Consider a fully associative cache with 8 cache blocks (0-7) and the following sequence of memory block requests:

$\overset{M}{4}, 3, 25, 8, 19, \textcircled{6}, 25, 8, 16, 35, 45, 22, 8, 3, 16, 25, 7$

if LRU replacement policy is used, which cache block will have memory block 7 (GATE 2004)

0	<del>4</del> 45	} 45 22
1	<del>3</del> 22	
2	25	
3	8	
4	<del>19</del> 3	
5	<del>6</del> 7	}
6	16	
7	35	
cache		