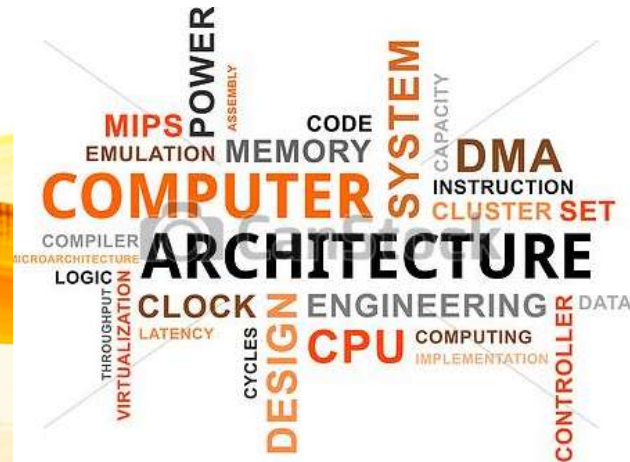


UNIT IV

MEMORY SYSTEM

Basic concepts of Semiconductor RAMs - ROMs – Speed, Size and Cost – **Cache memories – Performance consideration** – Virtual memory – Memory Management requirements – Secondary storage.

Case Study: Memory Organization in Multiprocessors



Recap the previous Class



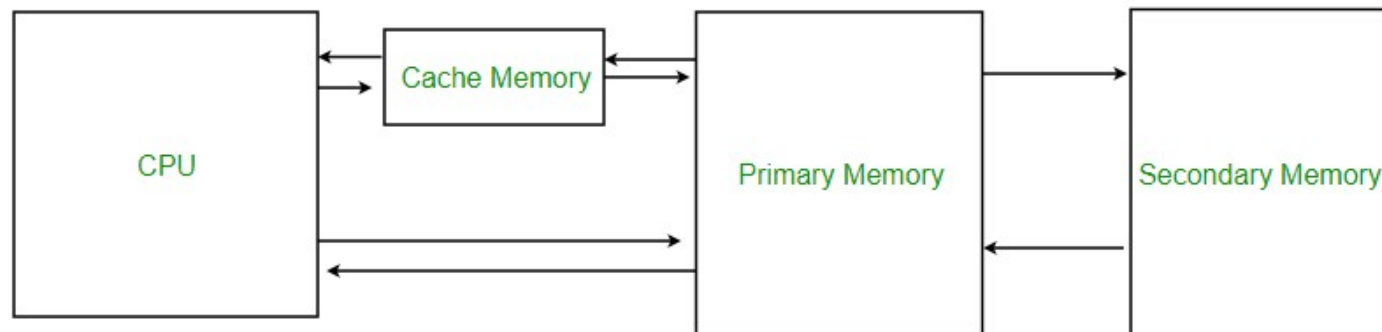


Cache Memories

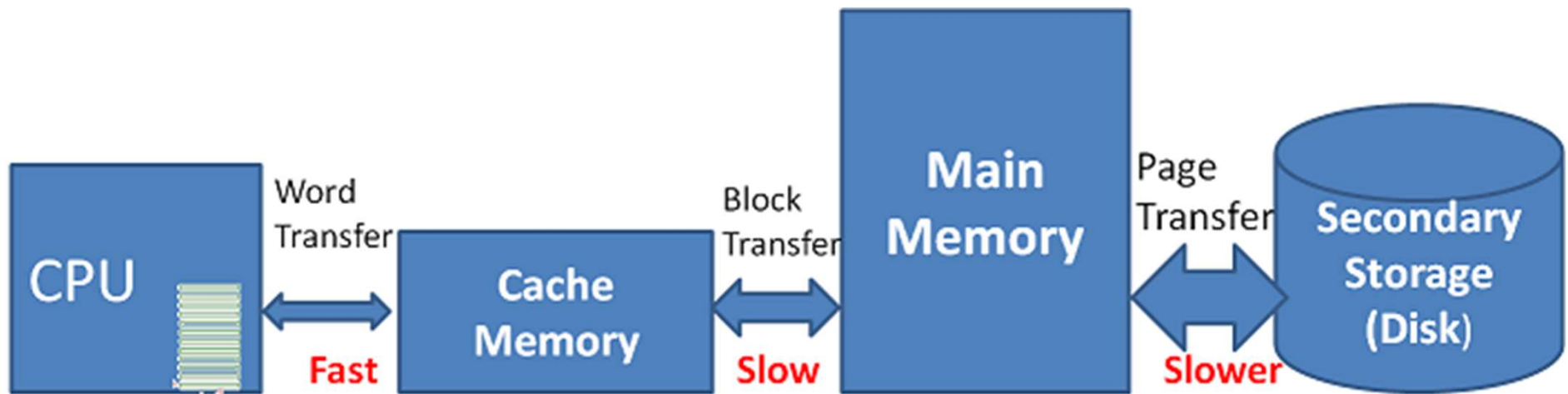
- special very **high-speed memory**
- used to **speed up and synchronizing** with high-speed CPU
- **costlier** than main memory
- acts as a **buffer between RAM and the CPU**
- holds **frequently** requested data and instructions
- used to **reduce the average time** to access data from the Main memory.

Cache Memories

- smaller and faster memory which stores copies of the data from frequently used main memory locations.
- Different independent caches in a CPU, which store instructions and data.



Cache Memories



Levels of memory

Level 1 or Register

- It is a type of memory in which data is stored and accepted that are immediately stored in CPU.
- Commonly used register is accumulator, Program counter, address register etc.



Levels of memory

Level 2 or Cache memory

- It is the fastest memory which has faster access time where data is temporarily stored for faster access.

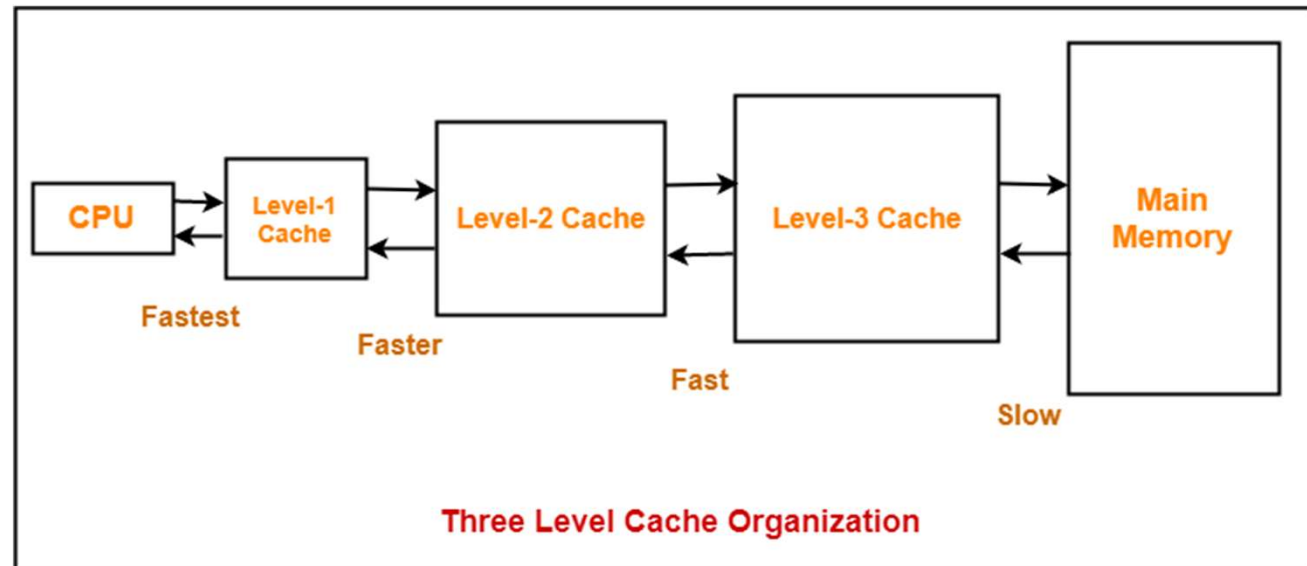
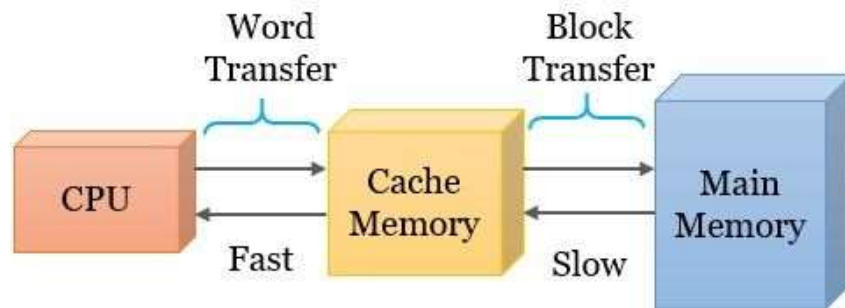
Level 3 or Main Memory

- It is memory on which computer works currently. It is small in size and once power is off data no longer stays in this memory.

Level 4 or Secondary Memory

- It is external memory which is not as fast as main memory but data stays permanently in this memory.

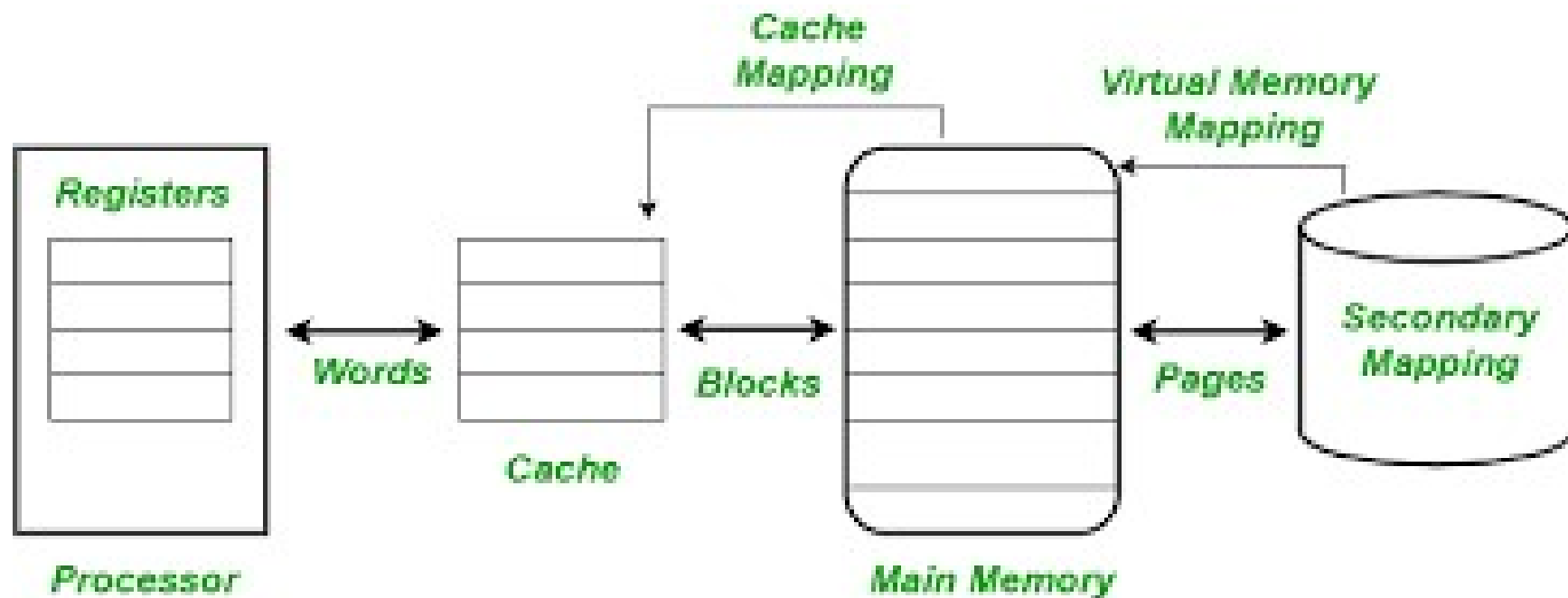
Cache Performance



Cache Performance

- When the processor needs to read or write a location in main memory, it first checks for a corresponding entry in the cache.
- **Cache Hit** - processor finds that the memory location in the cache and data is read from cache
- **Cache Miss** - processor does not find that the memory location in the cache. It allocates the new location in the cache to be fulfilled from the contents of the cache.

Cache Performance

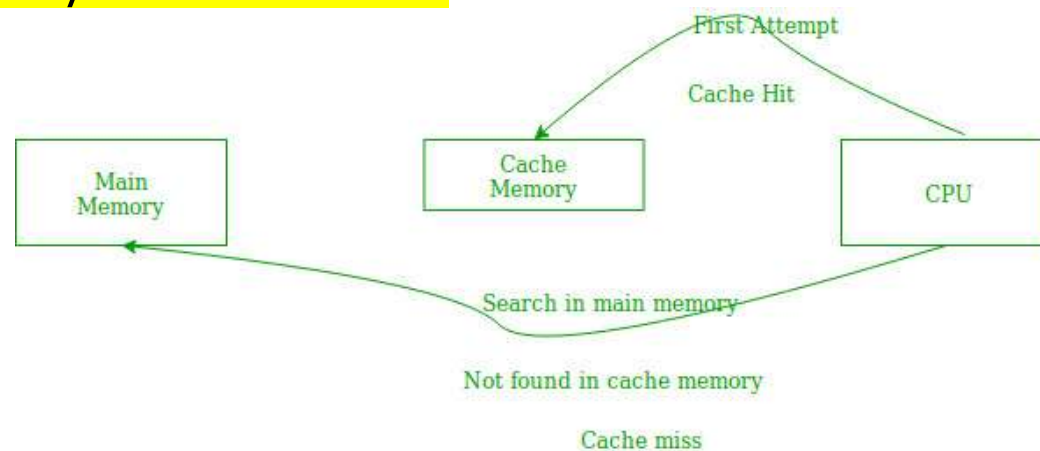


Hit ratio = hit / (hit + miss) = no. of hits/total accesses
Hit ratio = hit / (hit + miss) = no. of hits/total accesses
Hit ratio = hit / (hit + miss) = no. of hits/total accesses



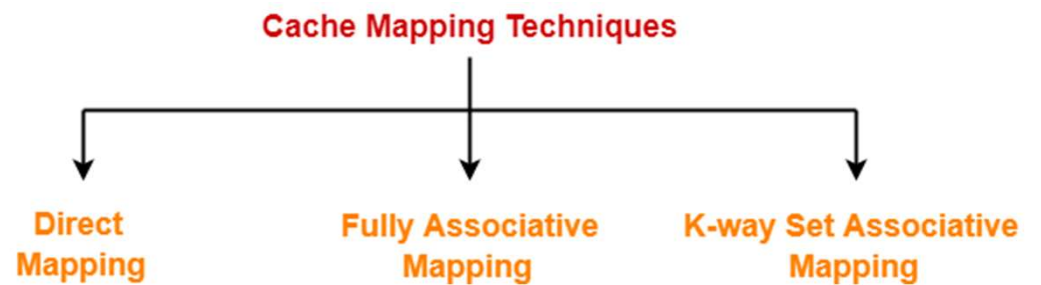
Cache Performance

- The performance of cache memory is frequently measured in terms of a quantity called **Hit ratio**.
- **Hit ratio = hit / (hit + miss) = no. of hits/total accesses**
- Improve Cache performance
 - higher cache block size,
 - reduce miss rate,
 - reduce miss penalty, and
 - reduce the time to hit in the cache.



Cache Mapping

- Direct mapping
- Associative mapping,
- Set-Associative mapping



Direct Mapping

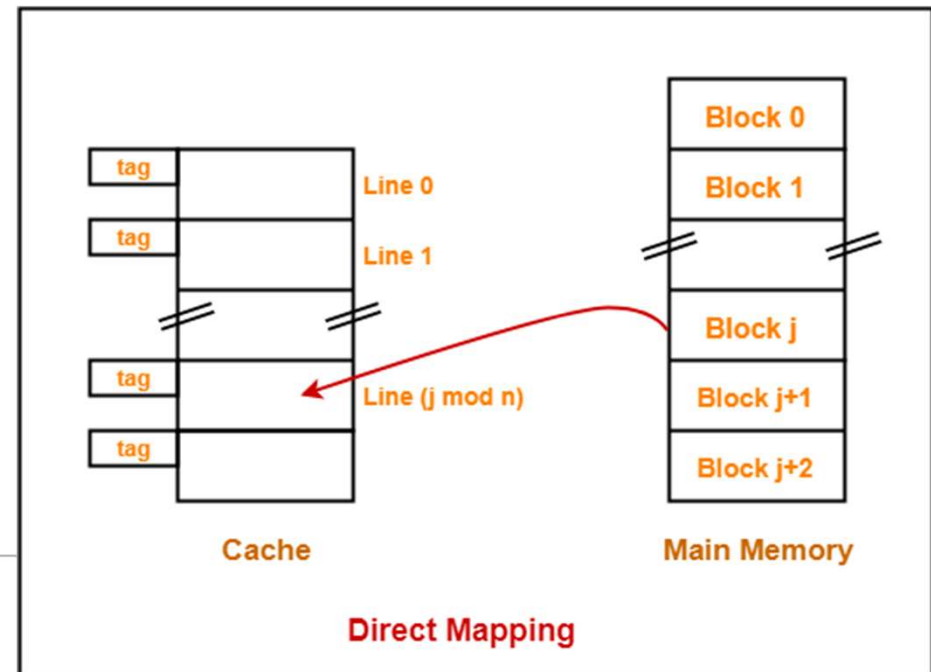
A particular block of main memory can map only to a particular line of the cache.

$i = j \text{ modulo } m$ where

i = cache line number

j = main memory block number

m = number of lines in the cache



Cache line number

= (Main Memory Block Address) Modulo (Number of lines in Cache)



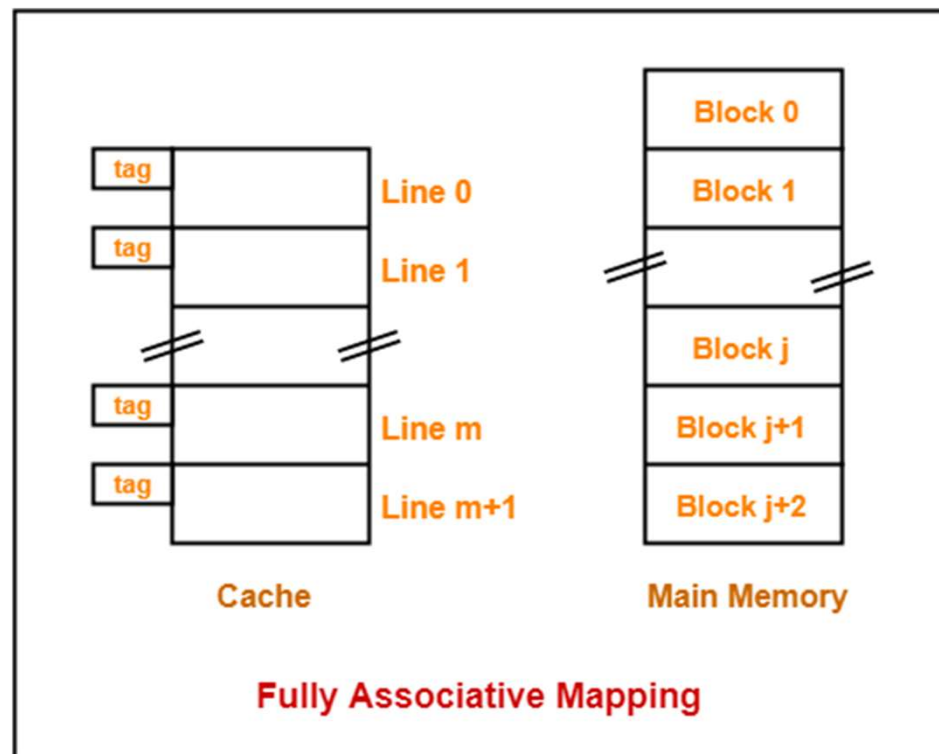
Direct Mapping

Replacement Algorithm

- no need of any replacement algorithm.
- A main memory block can map only to a particular line of the cache.
- new incoming block will always replace the existing block (if any) in that particular line.

Associative Mapping

- used to store content and addresses of the memory word.
- A block of main memory can map to any line of the cache that is freely available at that moment.





Associative Mapping

Replacement Algorithm

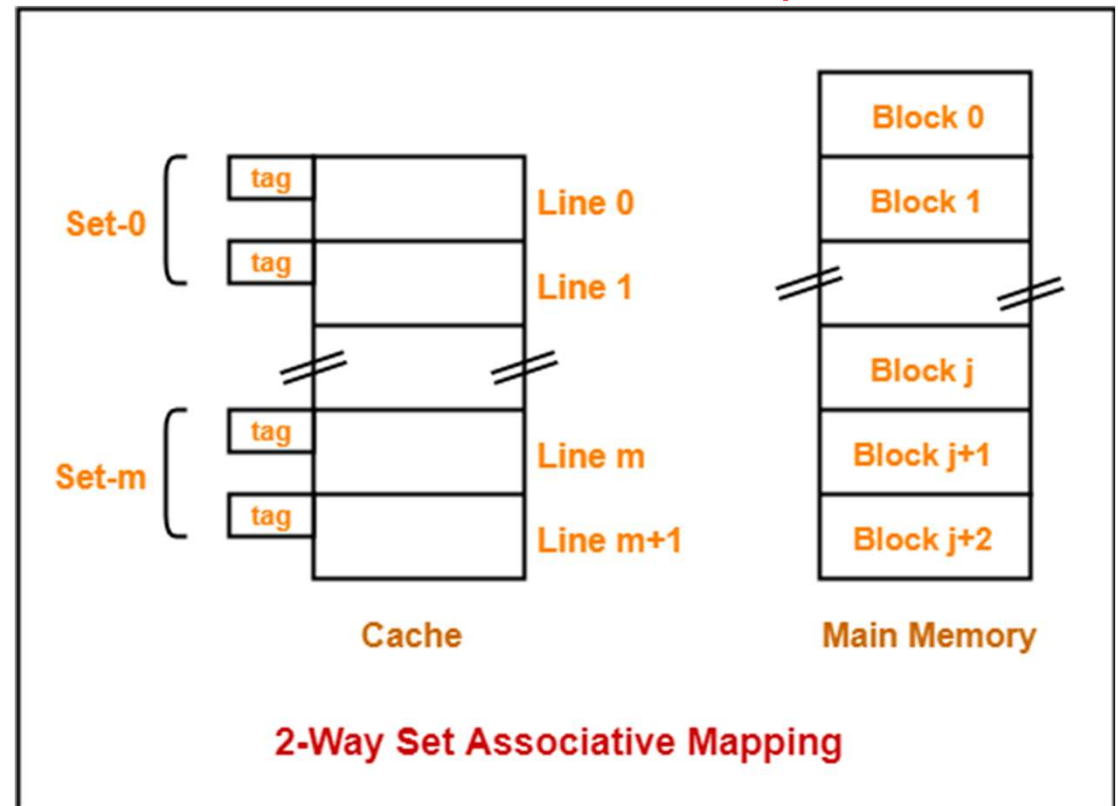
- A replacement algorithm is required.
- Replacement algorithm suggests the **block to be replaced if all the cache lines are occupied.**
- Thus, replacement algorithm like **FCFS Algorithm, LRU Algorithm** etc is employed.

Set-associative Mapping

$K = 2$, cache contains 6 lines, so number of sets in the cache = $6 / 2 = 3$ sets.

Combination of direct mapping and fully associative mapping

- block of main memory can map to only one particular set of the cache.
- within that set, the memory block can map any cache line that is freely available.





Set Associative Mapping

Replacement Algorithm

- A replacement algorithm is required.



Types of Cache

Primary Cache

- located on the processor chip.
- small and its access time is comparable to that of processor registers.

Secondary Cache

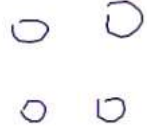
- placed between the primary cache and the rest of the memory.
- It is referred to as the level 2 (L2) cache. Often, the Level 2 cache is also housed on the processor chip.



Types of Cache Memory

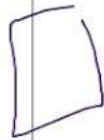
L1: Level 1 cache

- First level of cache
- Memory is present inside the CPU itself.
- It can work at the same speed as of the CPU.
- Each core of CPU have its own level 1 cache.



L2: Level 2 cache

- Second level of cache
- It can be present inside or outside the CPU.
- They are slower than L1.
- Each core of CPU can have its own L2 cache or they can share single L2 cache.



L3: Level 3 cache

- Third level of cache
- It is located outside the CPU.
- It is slower than L1 & L2.
- It is shared by all the cores of a CPU.

Locality of reference

- data or instruction is fetched from main memory and get stored in cache memory

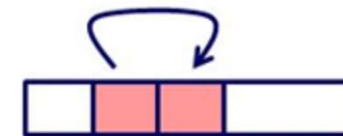
Temporal locality:

- Recently referenced items are likely to be referenced again in the near future



Spatial locality:

- Items with nearby addresses tend to be referenced close together in time





sns
INSTITUTIONS



Thank You