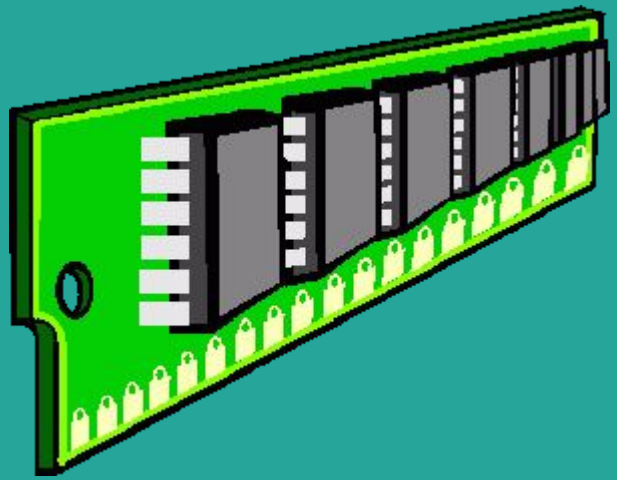
Different Mapping Techniques in Cache Memory

—

CPE355 - Team Ziggiotti
ESPINA, MACASO, & TAGAYON

What is a Cache Memory?



This is a computer memory that helps in reading data faster since data stored within this memory are the ones that are frequently used.

Static random access memory (SRAM) is used on the cache since it gives faster access to data rather than the dynamic random access memory (DRAM), in this case it is more expensive and there is less storage, but the use of SRAM provides better speed.

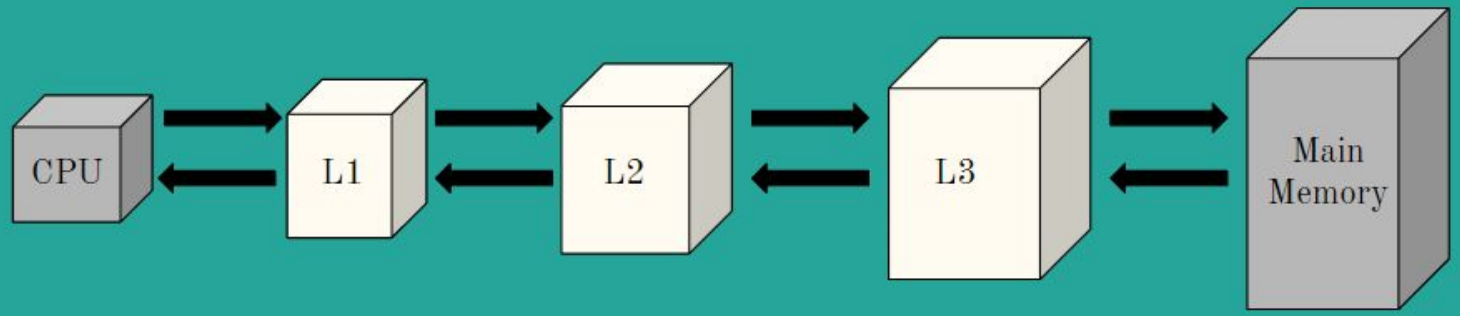
The introduction of multi level cache helps to lessen the cost of a processor and also improving its performance.

Levels of Cache Memory

Level 1 Cache
- The fastest among the levels.
- Located in the CPU.
- Has 2 kinds of cache: the instruction cache and data cache.

Level 2 Cache
- Slower than the L1 cache.
- Shared among cores in a CPU or can be set up individually for each core.

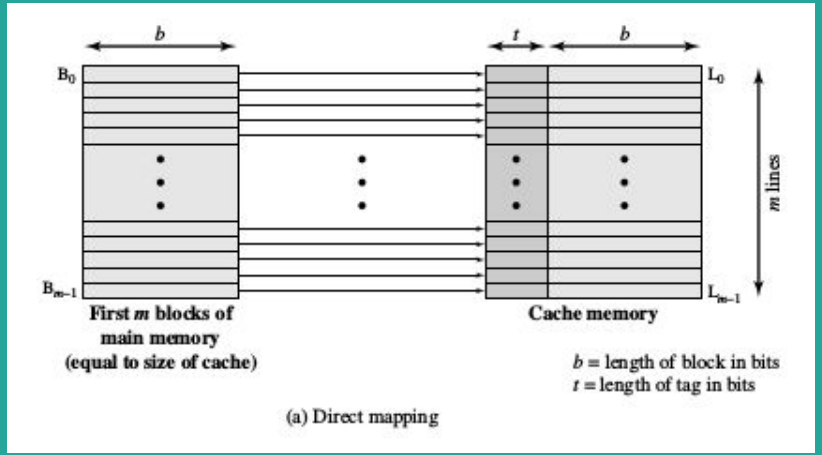
Level 3 Cache
- The slowest among the levels.
- Enhances the performance of the L1 and L2 caches.



Mapping Techniques

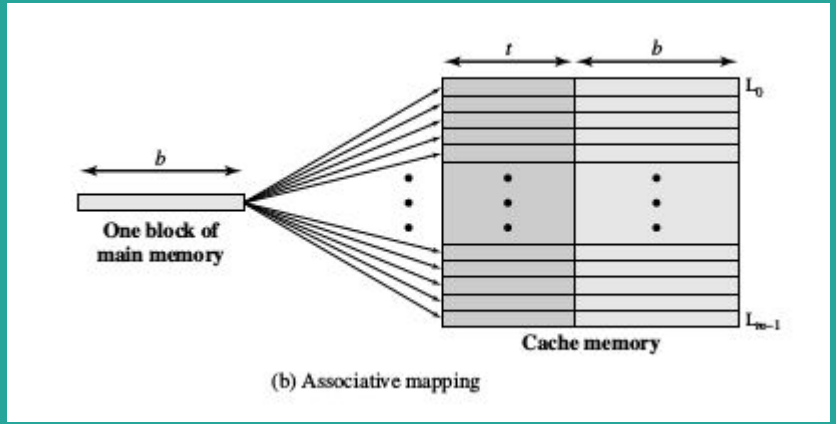
Direct Mapping

- It is the simplest out of all the techniques.
- Maps out each block of main memory into only one possible cache line.
- “Thrashing of lines” is a possibility.



Associative Mapping

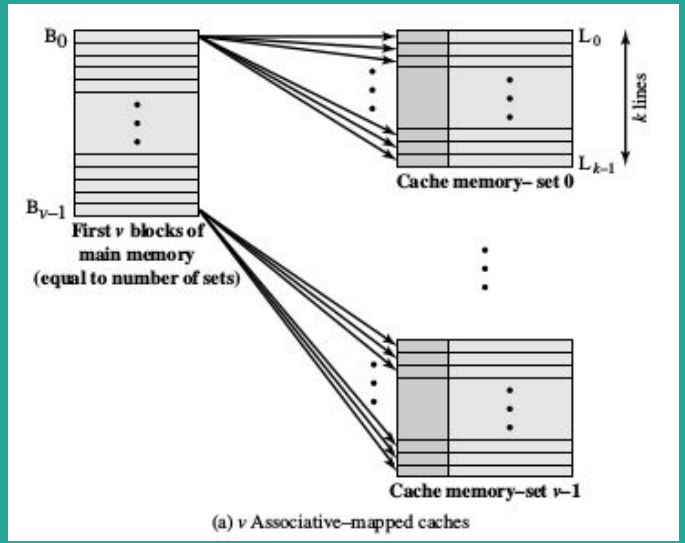
- More flexible compared to Direct Mapping.
- Any block of the Main memory can reside in any cache block position.
- Enables the placement of any word at any place in the cache memory.



(b) Associative mapping

Set-Associative Mapping

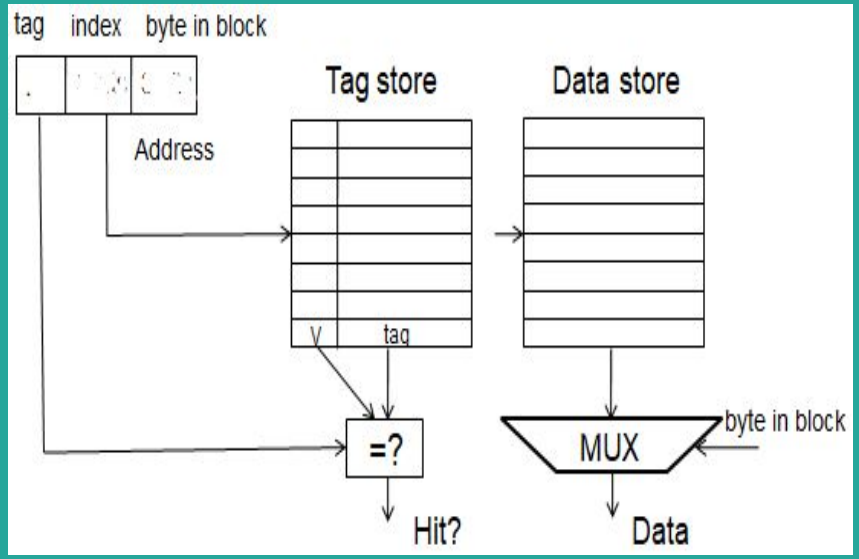
- Blocks of cache are grouped together into sets.
- This mapping technique allows a block of main memory to reside in any block of the specific set.
- Addresses the drawbacks of the possible thrashing in Direct Method.



Difference of Mapping Techniques

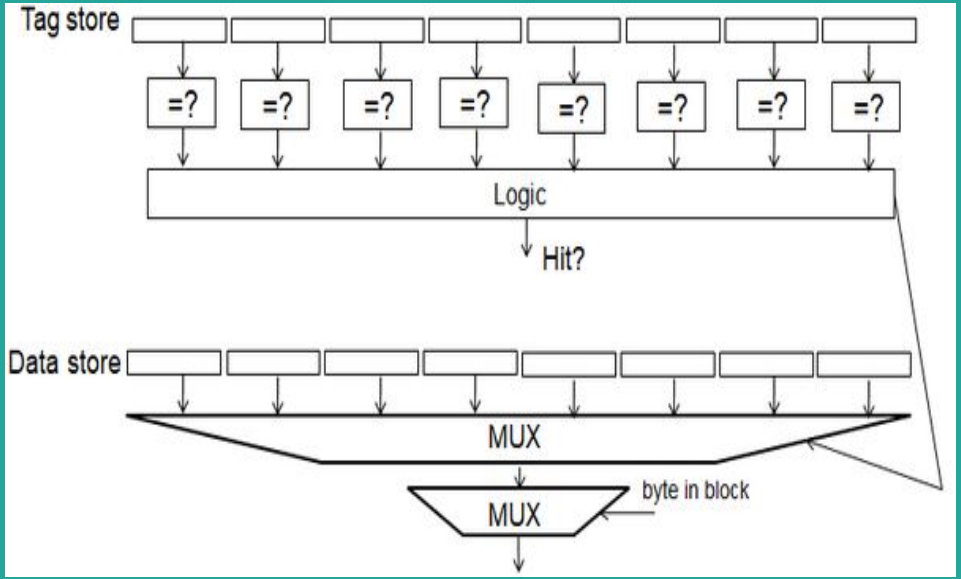
Direct Mapping

- A block can be placed only in one position in a cache.
- Not flexible enough.
- Hit ratio is poor as there is only one fixed location for each main memory location.



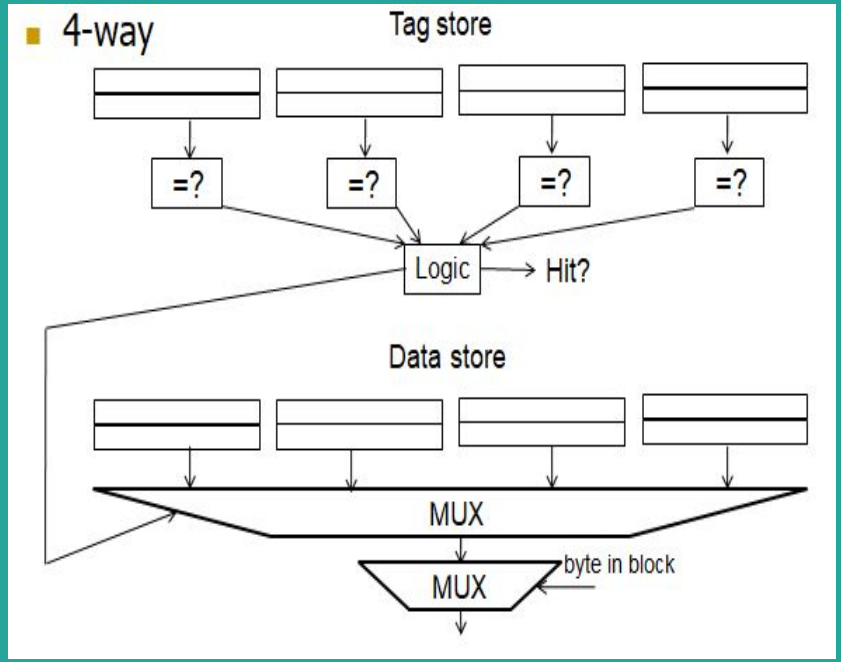
Associative Mapping

- Enables each main memory block to be loaded by any line of the cache.
- Has less potential for collisions between blocks trying to occupy a cache.
- Larger hardware needed to store large data.



Set-Associative Mapping

- The caches is divided into numerous sets of cache.
- Each main memory block can be placed in number of ways within a set.
- More variety of blocks could be accommodated.



Relevance of Using Different Mapping Techniques

We must remember that caches are use to make processors faster by reducing the memory of the performance wall and it is made of smallest storage elements known as block.

Thus, the different mapping techniques will determine the arrangement of blocks and will then, tells us how fast can we access the cache and how the latency could be reduced.