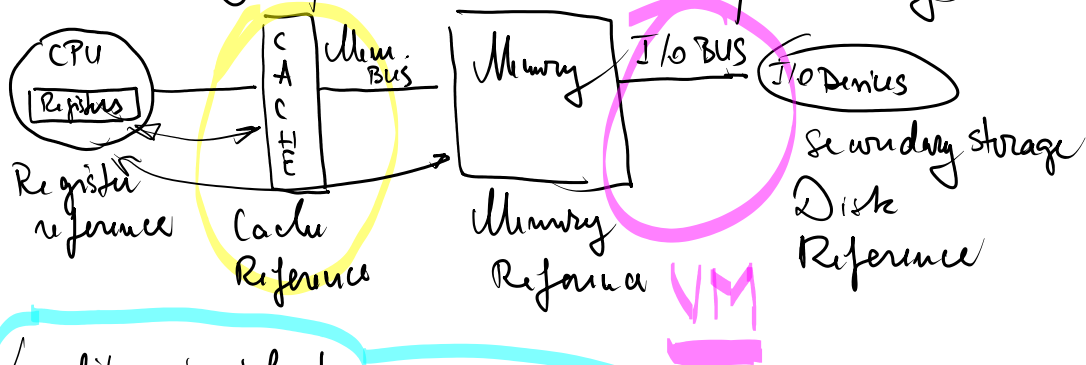


Chapter 3 Memory Hierarchy Performance

- In CS we do not have a monolithic memory
- hierarchy of memories (levels of memory)



Locality principle!

CPU reads instructions

- Cache first!

Hit

Miss

Hit rate, dominantly bigger than Miss Rate (~90%)

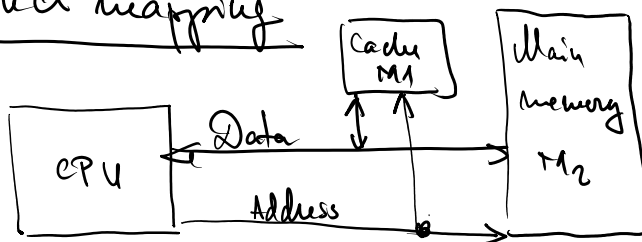
- 1 word → data / instruction

1 word = 2^8 bytes
1 byte = 8 bits = 1B

- 1 block = 2^P words

3.1 Address mapping for caches

3.1.1. Direct mapping



- direct mapping
- set-associativity
- full associativity

M_2 is divided in blocks and pages

M_1 is divided in blocks of the same size

CPU operates with words

We admit that M_2 is divided in 2^u blocks ($u > m$)

M_1 is divided in 2^m blocks

Each block has 2^P words

Each word has 2^2 bytes

M_2 : $M_2(0), M_2(1), M_2(2), \dots, M_2(j), \dots, M_2(2^u - 1)$

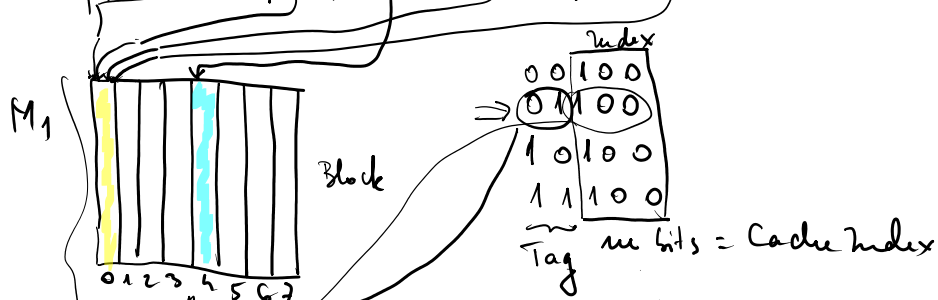
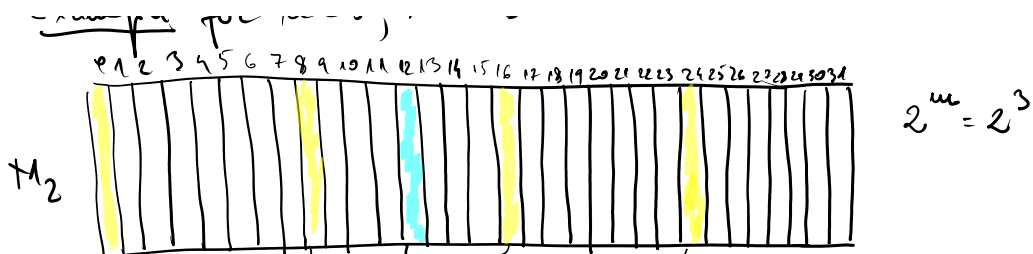
$$i \equiv j \pmod{2^m}$$

M_1 : $M_1(0), M_1(1), M_1(2), \dots, M_1(i), \dots, M_1(2^m - 1)$

Example for $u=5, m=3$

$$12 \pmod{8} = 4$$





Address word

