

# Memoria cache

Leonardo Bizzoni

May 20, 2023

La cache è un livello della gerarchia di memoria presente tra il processore e la memoria centrale, è quindi una memoria di piccole dimensioni ma molto veloce.

Viste le dimensioni ridotte rispetto alla dimensione della memoria centrale ci sono 3 modi distinti per poter associare un dato presente in memoria centrale alla sua copia nella cache.

## 1 Direct mapped cache - le migliori

È il metodo più semplice per stabilire l'indirizzo nella cache di una word in memoria centrale.

Ogni word della memoria centrale viene associata ad una sola locazione della cache definendo una corrispondenza tra l'indirizzo della word e la locazione nella cache.

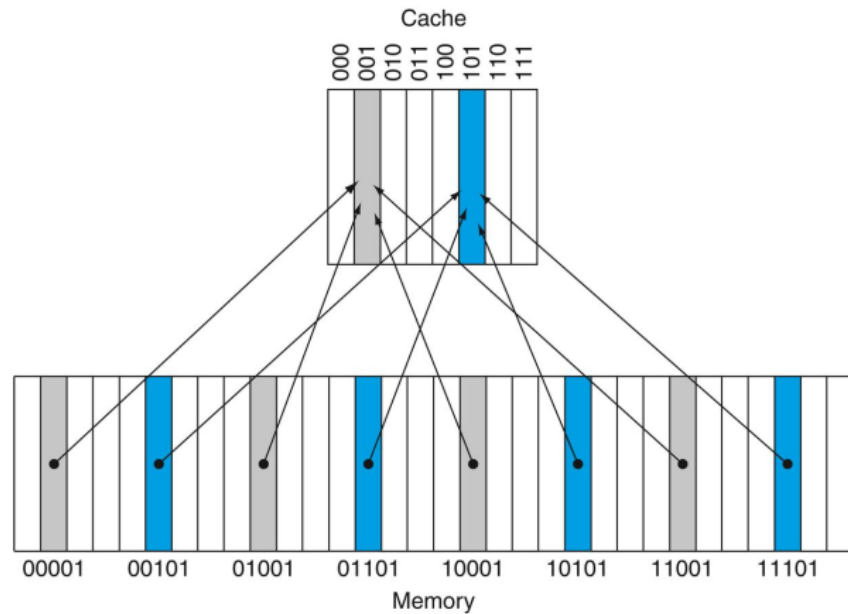
Per trovare il blocco corrispondente ad una indirizzo della memoria centrale si utilizza la formula:

$$\text{numero del blocco} = \frac{\text{Indirizzo del dato}}{\text{byte per blocco}} \pmod{\text{(numero di blocchi nella cache)}}$$

Se però il numero di blocchi nella cache è una potenza di 2 allora la posizione della word nella cache è data dai  $\log_2(\text{numero di blocchi nella cache})$  bit meno significativi.

### 1.1 Esempio

Una cache ad 8 blocchi utilizzerà per indirizzare una word gli ultimi 3bit.



## 1.2 Contenuto di una linea di cache

Dato che ogni cella della cache viene mappata a più indirizzi di memoria è necessario aggiungere un **tag** contenente i bit più significativi per poter avere la certezza che il dato richiesto è quello effettivamente mappato nella cache.

Si ha anche la necessità di riconoscere se una blocco della cache contiene informazioni valide. Per esempio, quando il compilatore viene avviato, la cache non contiene alcuni dato e quindi le informazioni fornite dai tag sono inutili. Questo problema viene risolto aggiungendo un **bit di validità**.

In una cache ad indirizzamento diretto ogni linea di cache include:

- bit di validità
- tag
- blocco di dati formato da una o più word

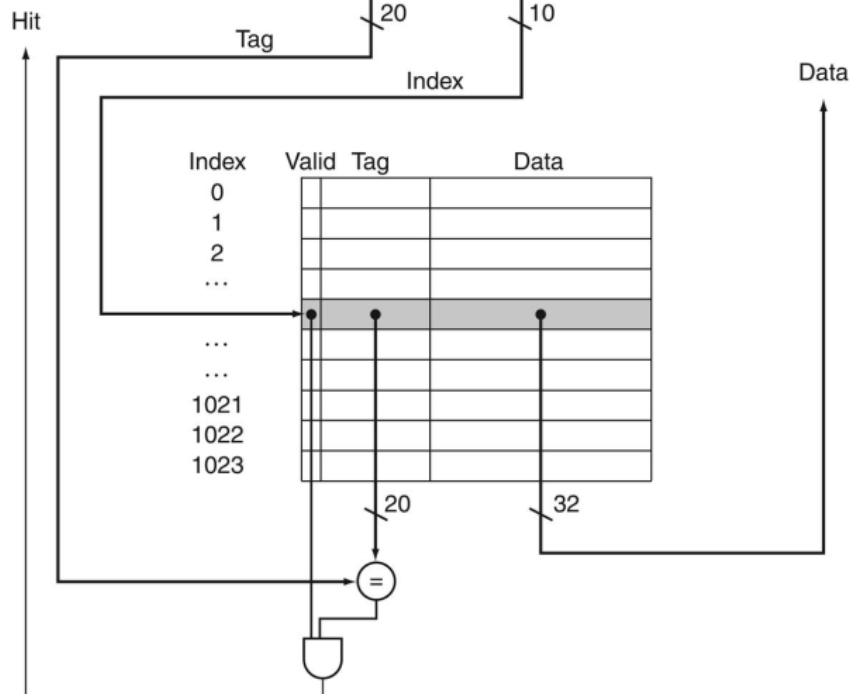
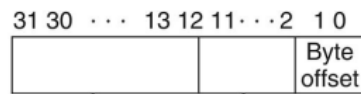
Memory (16 parole)

0 (0000)	
1 (0001)	
2 (0010)	
3 (0011)	ccc
4 (0100)	
5 (0101)	
6 (0110)	
7 (0111)	
8 (1000)	xyz
9 (1001)	
A (1010)	bbb
B (1011)	
C (1100)	
D (1101)	aaa
E (1110)	
F (1111)	

Direct Mapped Cache (4 parole)

	Valid	Tag	Data
0 (00)	1	10	xyz
1 (01)	1	11	aaa
2 (10)	1	10	bbb
3 (11)	1	00	ccc

Address (showing bit positions)



## 2 Fully associative cache - la merda

A differenza delle direct mapped cache, nelle **fully associative cache** un blocco può essere posizionato dovunque all'interno della cache.

Per trovare un determinato blocco nella cache è necessario cercarlo iterando su tutti i blocchi.

### 2.1 Contenuto di una linea di cache

In una cache ad indirizzamento diretto ogni linea di cache include:

- bit di validità
- blocco di dati formato da una o più word

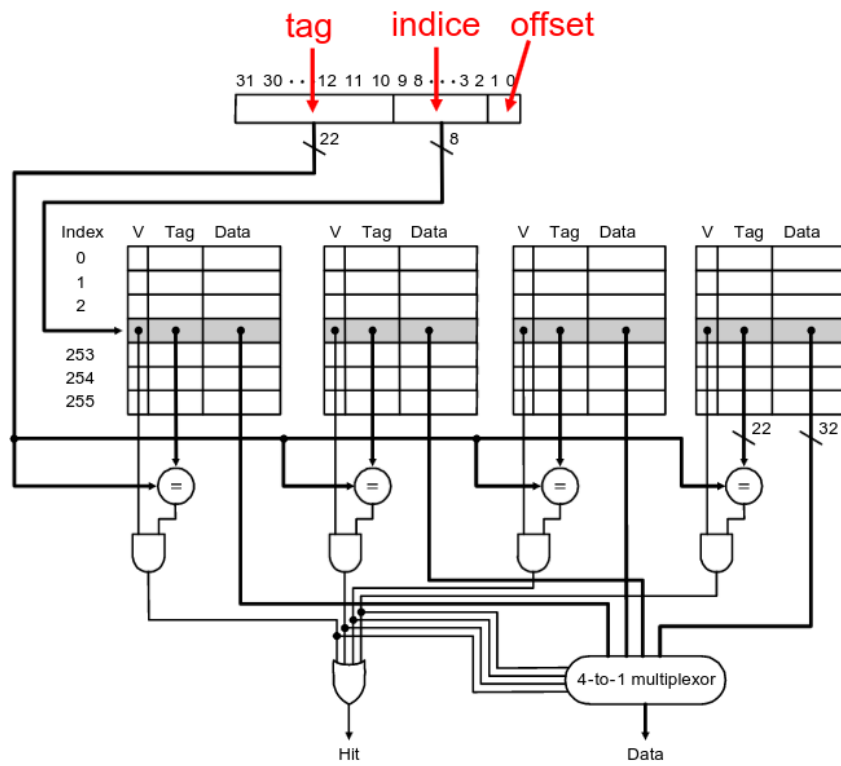
## 3 N-way set associative cache - a metà

Set associative cache è una via di mezzo tra fully associative e direct mapped cache.

Una set associative cache è divisa in set ognuno composto da  $n$  blocchi. Ogni blocco di memoria viene mappato ad un unico set nella cache in qualsiasi posizione. Stabilito il set, per determinare se un certo indirizzo è presente in un blocco del set è necessario confrontare i tag di tutti i blocchi del set.

Per trovare il set corrispondente ad un indirizzo della memoria centrale si utilizza la formula:

$$(\text{numero del blocco}) \% (\text{numero di set nella cache})$$



## 4 Confronto tra le tecniche di indirizzamento su cache

