

Progettazione di una cache

Dimensione della Cache

- è influenzata da
 - costo
 - prestazioni
 - al crescere delle dimensioni
 - la cache diventa lenta
 - ma h cresce
- Dimensioni frequenti sono tra **KB-MB**

Dimensione del Blocco

- al crescere della dimensione del blocco avvine
 - raddoppia hit radio
 - località di rif
 - poi decresce
 - diminuscie il numero di blocchi
- Dimensioni frequenti sono da **4 a 32 byte**

Funzione di traduzione (mapping)

- permette di stabilire in quale linea si trova un certo blocco
- possibili soluzioni
 - memoria associatibva
 - costosa
 - memeoria sequenziale
 - lenta
 - meccanisomo di mapping
 - molto usata

Meccnaismo di mapping: Direct Mapping

ogni **blocco i** della **memoria principale** è messo in corrispondenza fissa con una **linea k** della cache

$k = i \bmod N$ **N** il numero di linee della cache

se $N = 2^n$ allora per stabilire il valore di **k**,basta prendere i primi **n** bit meno significativi del blocco

Direct Mapping: esempio

Si assuma che

- la cache sia composta da 4 linee
- la memoria sia composta da 16 blocchi.

esempio 19

Vantaggi facimlente impementabile

Svantaggi se il programma fa **accesso frequentemente a 2 blocchi diversi corrispondenti alla stessa linea** della cache

ad ogni accesso avvine sempre un miss

Etichetta Blocco Parola

Identifica il blocco di memoria. È il campo che viene confrontato con il tag.

Identifica la parola all'interno del blocco

Attraverso la funzione di mapping identifica la linea di cache

21

La memoria cache

La memoria cpu

dato un indirizzo di memoria

Meccnaismo di mapping: Associative Mapping

- ogni blocco può essere memorizzato in un qualsiasi blocco della cache
 - vantaggi massima flessibilità
 - svantaggi alta complessita HW
 - Struttura dell' indirizzo
- Etichetta Parola

Identifica il blocco di memoria. È il campo sul quale la cache esegue il controllo.

Identifica la parola all'interno del blocco

Aggiornamento della memoria principale

- Bit di validità**
 - ad ogni linea di cache si introduce questo bit
 - il blocco presente in quella linea ha un valore diverso della memoria
 - qundo disattivato
 - ogni accesso in questo caso comporta ad un miss
 - Cache di Livello**
 - alche la cache è organizzata in livelli
 - write-back**
 - La CPU ha un canale di connessione diretto con la memoria pricipale
 - esistono 2 possibili metodi per l'aggiornamento
 - dirty-bit**
 - per linea della cache esite un bit
 - che ricorda se il blocco è stato modificato o meno
 - il dirty bit
 - a 0
 - qundo un nuovo blocco è scritto in una linea
 - a 1
 - quando la linea viene eseguità una scrittura
 - quando bisogna cambiare blocco
 - il blocco viene copiato nella meoria principale
 - Svantaggi**
 - la gestione del miss è più lenta
 - nei sistemi a multi processori si può avere concorrenza sui dati
 - il ripristono può non avvenire
 - in pratica fai tutto sulla cache
 - poi lo scrivi in memoria
- write-through**
 - ogni volta che la CPU esegue un operazione di **scrittura**
 - la esegue sia sul dato nella cache sia in quello nella memoria principale
 - la predata di efficienza è limitata dal numoero di scritture

Algorirmo di rimpiazzo

- Definisce **quale** linea di cache deve esse utilizzata per memorizzare il nuovo blocco
- tra quelle associare al blocco
 - più usato Least Recently Used LRU
 - più economico First-In First-Out FIFO
 - teoricamente più efficiente ma troppo costosa Least Frequently Used LFU
 - esay pisy random

Meccnaismo di mapping: Set Associative Mapping

- caratteristiche
 - le linee della cache sono suddivise in **S** insiemi,ciascuno composto da **W** linne
 - $k = i \bmod S$ un blocco i e associato all' insime k
 - il blocco i può essere messo in una qualunque delle **W** linee dell' insieme k
 - valori comuni di **W** sono 2 e 4
- in caso di
 - direct-mapping **S = N**
 - associative mapping **S = 1**
- Struttura dell'indirizzo

Etichetta Blocco Parola

Identifica il blocco di memoria. È il campo che viene confrontato con il tag.

Identifica la parola all'interno del blocco

Attraverso la funzione di mapping identifica l'insieme di linee di cache

25