OBJETIVOS DEL TEMA

Objetivo general

- Comprender cómo se organiza y cuáles son los principios que rigen el funcionamiento de la jerarquía de memoria de un computador

Objetivos específicos

- Enunciar el objetivo de la jerarquía de memoria del ordenador
- Explicar los principios en los que se basa la jerarquía de memoria para conseguir su objetivo
- Mostrar ejemplos de localidad espacial y temporal en un programa dado
- Definir memoria cache
- Explicar la relación de la memoria cache con memoria principal
- Enunciar las diferencias entre memoria cache y principal en cuanto a tecnología de fabricación y rendimiento
- Enumerar políticas de emplazamiento, sustitución y escritura de memoria cache
- Razonar sobre los pros y contras de tener un tamaño de bloque grande o pequeño
- Mostrar el esquema de organización de una memoria cache, sea cual sea la política de emplazamiento utilizada
- Calcular qué marco de memoria cache se corresponde con cada bloque de memoria principal
- Interpretar una dirección de memoria según el tipo de cache utilizada
- Simular el uso de memoria que hace un determinado programa durante su ejecución (calcular direcciones de acceso a memoria y determinar qué accesos producen fallo)

BIBLIOGRAFÍA

- Capítulos 4 (jerarquía de memoria y memoria cache) del libro "Organización y Arquitectura de Computadores", de W. Stallings, recomendado como parte de la bibliografía básica

CUESTIONES

1. ¿Qué tipo de tecnología se utiliza para implementar memorias cache? ¿Y para memorias RAM? ¿Qué diferencias hay entre ellas?

2. ¿En qué consiste el principio de localidad espacial? ¿Por qué puede asegurarse que se cumple en todos los programas?

3. ¿En qué consiste el principio de localidad temporal? ¿Por qué puede asegurarse que se cumple en todos los programas?

4. Muestra en el siguiente código un ejemplo de localidad espacial y otro de localidad temporal:

```
for (i=0; i<20; i++)
  for (j=0; j<20; j++)
      a[i] = a[i]*j;
```

5. Independientemente de la política de emplazamiento utilizada, si se tiene una cache de 1k líneas, de 8 palabras cada una, ¿cómo se divide la dirección de memoria entre bloque y palabra?
6. Con una cache de emplazamiento directo, de *m* marcos, con *k* palabras por marco, ¿qué bloques serán asignados al marco *i*?
7. ¿En qué se diferencian la post-escritura y la escritura directa?
8. ¿Cómo pueden clasificarse los tipos de fallos en los accesos a una cache?
9. ¿Qué relación existe entre el tamaño de un bloque y la tasa de fallos forzosos? ¿Y si consideramos los fallos de conflicto?
10. Se tiene una memoria cache asociativa por conjuntos de 4 vías. En total, consta de 32 conjuntos. El tamaño de bloque es de 8 palabras, cada una de ellas de 1 byte. Y la memoria principal consta de 1024 bloques.
    a) ¿Cuál es el tamaño de la cache (en KB)? ¿Y el de la memoria principal?
    b) ¿Con qué bloques puede entrar en conflicto el bloque 57?
    c) ¿Cuántos bits deben usarse para codificar la dirección de memoria? ¿Cómo se reparten esos bits entre palabra, conjunto y etiqueta?
    d) ¿Cómo se codificará la dirección $(2567)_{10}$? ¿En qué conjunto quedará situado la palabra de memoria asociada a esa dirección?

EJERCICIOS

11. Considera el siguiente programa en lenguaje ensamblador del QuasiZ80-16.

```
        ORG $0000

        di
ini:    ld ix, var1
        ld a,(ix+0)
        ld b,(ix+1)

mien:   cp b
        jr c, sino
        jr z, fmien
        sub b
        jr mien
sino:   ld e,b
        ld b,a
        ld a,e
        jr mien
fmien:  ld (mcd), a

fin:    halt

var1:   db 75
var2:   db 40
mcd:    db 0
```

a) ¿Cuál es la secuencia de direcciones de acceso a memoria durante la ejecución de las dos primeras iteraciones del bucle?

b) Muestra, a la vista de la anterior secuencia, un ejemplo de localidad espacial y otro de localidad temporal

12. Supongamos que se incorpora al QuasiZ80-16 una memoria cache de emplazamiento directo de 8 marcos, y que cada bloque está compuesto por 4 palabras.

    a) Dibuja un esquema de la estructura de dicha cache

    b) Comenta en cuántos campos se divide la dirección para hallar una palabra en cache y cuántos bits ocupa cada campo

    c) Para la secuencia de direcciones del ejercicio 12, determina si el primer acceso a la instrucción "jr z, fmien" produce acierto o fallo.

13. Supongamos que se incorpora al QuasiZ80-16 una memoria cache asociativa por conjuntos de 2 vías, con un total de 4 conjuntos. Cada bloque está compuesto por 4 palabras. Responde a las mismas cuestiones que en el ejercicio 12 para este tipo de cache.

14. Considera un microprocesador de 32 bits que tiene una cache interna (on-chip) de 16 Kbytes asociativa por conjuntos de 4 vías. Supón que la cache tiene un tamaño de línea de 4 palabras de 32 bits. Dibuja un diagrama de bloques de esta cache, mostrando su organización y cómo se utilizan los diferentes campos de dirección. ¿Dónde se asigna, dentro de la cache, la palabra de la posición de memoria ABCDE8F8?

15. Considera una máquina con una memoria principal de 64 KBytes, direccionable por bytes, y un tamaño de bloque de 8 bytes. Supón que se utiliza una cache de 32 líneas y correspondencia directa.

    a) ¿Cómo se divide la dirección de memoria de 16 bits entre etiqueta, número de línea y número de byte?

    b) ¿En qué líneas se almacenarían los bytes que se encuentran en las siguientes direcciones?

    0001 0001 0001 1011 1100 0011 0011 0100
    1101 0000 0001 1101 1010 1010 1010 1010

    c) Suponga que se almacena en la cache el byte de dirección 0001 1010 0001 1010. ¿Cuáles son las direcciones de los bytes que se almacena junto a él?

    d) ¿Cuántos bytes de memoria pueden almacenarse en total en la cache?

16. El procesador Intel 80486 tiene una cache unificada on-chip, de 8 Kbytes, con una organización asociativa por conjuntos de 4 vías y una longitud de bloque de 4 palabras de 32 bits. La cache está estructurada en 128 conjuntos. En un fallo de cache, el 80486 lee una línea de 16 bytes de memoria principal en una ráfaga de lectura de memoria a través del bus. Dibuja un diagrama simplificado de la cache y muestra como son interpretados los diferentes campos de la dirección.