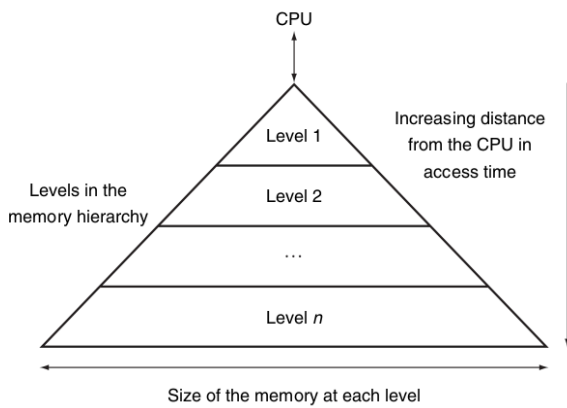


Práctico N° 5: Jerarquía de Memorias - CACHE



Structure of a memory hierarchy.

One-way set associative (direct mapped)

Block	Tag	Data
0		
1		
2		
3		
4		
5		
6		
7		

Two-way set associative

Set	Tag	Data	Tag	Data
0				
1				
2				
3				

Four-way set associative

Set	Tag	Data	Tag	Data	Tag	Data	Tag	Data
0								
1								

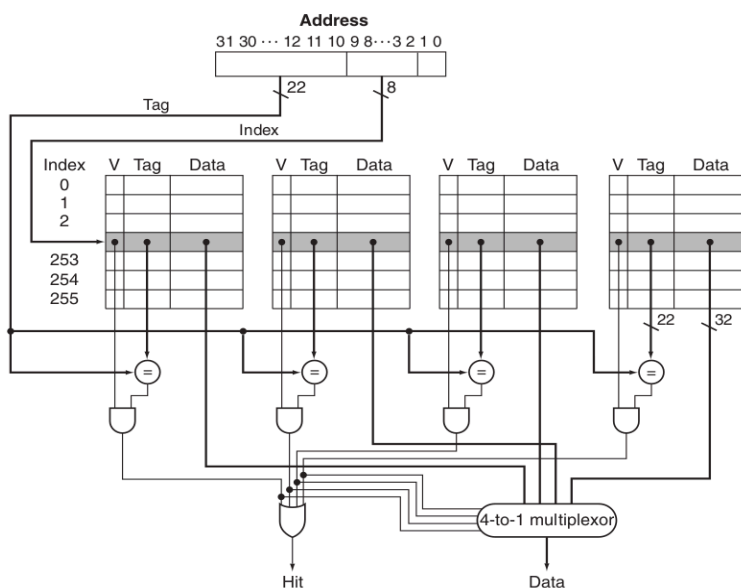
Eight-way set associative (fully associative)

Tag	Data	Tag	Data	Tag	Data	Tag	Data	Tag	Data	Tag	Data	Tag	Data	Tag	Data	Tag	Data

An eight-block cache configured as direct-mapped, two-way set associative, four-way set associative, and fully associative.



The location of a memory block whose address is 12 in a cache with eight blocks.



Four-way set-associative cache.

Ejercicio 1:

Indicar si son correctas o no las siguientes afirmaciones, si se intenta realizar una comparación entre distintos sistemas de memoria tecnológicamente diferentes. Justificar su respuesta.

- A menor tiempo de acceso, mayor coste por bit.
- A menor capacidad, menor coste por bit.
- A mayor capacidad, menor tiempo de acceso.
- A mayor tiempo de acceso, mayor capacidad.
- A mayor frecuencia de acceso, menor capacidad y mayor coste por bit.

Ejercicio 2:

Considerando un procesador que trabaja a 1.7GHz y los siguientes tiempos de acceso a memoria:

Memory technology	Typical access time	\$ per GiB in 2012
SRAM semiconductor memory	0.5–2.5ns	\$500–\$1000
DRAM semiconductor memory	50–70ns	\$10–\$20
Flash semiconductor memory	5,000–50,000 ns	\$0.75–\$1.00
Magnetic disk	5,000,000–20,000,000ns	\$0.05–\$0.10

¿Cuántos ciclos de clock implica leer un dato de caché (SRAM) y de memoria principal (DRAM)?

Ejercicio 3:

Calcular el tamaño total (en bits) de una caché de mapeo directo de 16KiB (16K x 8 bits) de datos y tamaño de bloque de 4 palabras de 32 bits c/u. Asuma direcciones de 64 bits.

Ejercicio 4:

Las memorias caché son fundamentales para elevar el rendimiento de un sistema de memoria jerárquico respecto del procesador. A continuación se da una lista de referencias de acceso a memoria (direcciones de 64 bits) las cuales deben ser consideradas como accesos secuenciales en ese mismo orden. El formato que se utiliza para cada dirección está reducido a sólo 16 bits, solo con fines prácticos:

0x000C, 0x02D0, 0x00AC, 0x0008, 0x02FC, 0x0160,
0x02F8, 0x0038, 0x02D4, 0x00AC, 0x00B0, 0x0160

Se debe:

- Para cada una de estas referencias a memoria, determinar el binario de la dirección de cada palabra (cada palabra de 32 bits), la etiqueta (tag), el número de línea (index) asignado en una caché de mapeo directo, con un tamaño de 16 bloques de 1 palabra c/u. Además liste qué referencias produjeron un acierto (hit) o un fallo (miss) de caché, suponiendo que la caché se inicializa vacía.
- Para cada una de estas referencias a memoria, determinar el binario de la dirección de cada palabra (cada palabra de 32 bits), la etiqueta (tag), el número de línea (index) asignado en una caché de mapeo directo, con un tamaño de 8 bloques de 2 palabras c/u. Además liste qué referencias produjeron un acierto (hit) o un fallo (miss) de caché, suponiendo que la caché se inicializa vacía.

Ejercicio 5:

Un computador tiene una memoria principal de 64K bytes y una memoria caché de 1K bytes para datos. La memoria caché utiliza correspondencia (mapeo) directa, con un tamaño de línea de 4 palabras de 16 bits c/u. Obtener:

- ¿Cuántos bits hay en los diferentes campos del formato de dirección de memoria principal?
- ¿Cuántas líneas contiene la memoria caché?
- ¿Cuántos bits hay en cada línea de la memoria caché y cómo se dividen según su función?

Ejercicio 6:

Una caché asociativa por conjuntos consta de 64 líneas, dividida en 4 vías. La memoria principal contiene 4K bloques de 128 palabras cada uno. Muestre el formato de dirección de memoria principal suponiendo que cada palabra es direccionable directamente en memoria.

Ejercicio 7:

Considere una CACHE con los siguientes parámetros:

- Criterio de correspondencia: Asociativa por conjuntos
- Asociatividad (N-vías): 2
- Tamaño de bloque: 2 words
- Tamaño de palabra (word): 32 bits
- Tamaño de la cache: 32K words
- Tamaño de dirección: 32 bits
- Cada palabra es directamente direccionable en memoria

Responder:

- a. Muestre el formato de dirección de memoria principal.
- b. ¿Cuál es el tamaño de toda el área de Tag de la cache, expresada en bits?
- c. Suponga que cada LINEA de la cache contiene además un bit de validación (V) y un tiempo de vida de 8 bits. Cual es el tamaño completo (expresado en bits) de un CONJUNTO de la cache, considerando datos, tags y los bits de status antes mencionados?

Ejercicio 8:

Sea un sistema con una memoria principal de 1M palabras divididas en 4K bloques, donde cada palabra es direccionable directamente en memoria. Definir el formato de la dirección de memoria principal en los siguientes casos, sabiendo que la memoria caché posee 64 líneas:

- a. Memoria caché con función de correspondencia directa.
- b. Memoria caché con función de correspondencia full-asociativa.
- c. Memoria caché con función de correspondencia asociativa de 8 vías.

Ejercicio 9:

Considere que un sistema tiene una CACHE para DATOS de correspondencia ASOCIATIVA POR CONJUNTOS de 2 VÍAS de 256Kbyte y 4 palabras por línea, sobre un procesador de 32bits, CPI = 1, que resuelve todos los data y control hazard sin necesidad de stalls, tiene una memoria principal de 4G palabras de 1 byte cada una.

- a. Muestre el formato de memoria principal.
- b. Considere ahora la ejecución en este sistema de los siguientes segmentos de código LEGv8 equivalentes para $N > 0$, donde $N \rightarrow X3$, start = 0x000203C y X6 contiene la dirección base del arreglo A[] del tipo uint32_t. Considerar que el resto de los registros están inicializados en 0 y la CACHE vacía al inicio de la ejecución de cada segmento.

Segmento #1:

```
start: ADD X9,X3,XZR
loop: CBZ X9, end
      LDURH X10, [X6,#0]
      LDURH X11, [X6,#8]
      ADD X12, X10, X11
      LSR X12, X12, #1
      ADDI X6, X6, #4
      SUBI X9, X9, #1
      B loop
end: ...
```

Segmento #2:

```
start: LDURH X10, [X6,#0]
      LDURH X11, [X6,#8]
      ADD X12, X10, X11
      LSR X12, X12, #1
      ADDI X6, X6, #4
      ADDI X9, X9, #1
      SUBS XZR, X9, X3
      B.NE start
end: ...
```

¿Cuántos MISS de CACHE se produjeron en 5 iteraciones del lazo para el segmento #1 si la dirección base del arreglo A es 0x00001008?

- c. Ahora considere que se incorpora una CACHE exclusiva para INSTRUCCIONES de correspondencia FULL ASOCIATIVA de una sola línea de 8 palabras de 32 bits c/u. Considerando que cada MISS de caché supone una penalidad de 5 ciclos de clk. Determine qué segmento de código se ejecuta en menor tiempo para $N = 4$. Argumente su respuesta indicando los ciclos totales en cada caso. Suponer que el pipeline ya se encuentra en régimen y que la CACHE de DATOS solo produce aciertos en los accesos, por lo que no se tiene penalidades por acceso a datos.

Ejercicio 10:

Encuentre el AMAT (Average Memory Access Time) para un procesador con un tiempo de ciclo de 1 ns, un miss penalty de 20 ciclos de reloj, un miss rate de 5% fallos por instrucción y un tiempo de acceso a caché (incluida la detección de acierto) de 1 ciclo de reloj. Suponga que las penalizaciones por falla de lectura y escritura son las mismas e ignore otros stalls en la escritura.

Ejercicio 11:

Una computadora posee un sistema de memoria jerárquica que consiste en dos cachés separadas (una de instrucciones y una de datos), seguidas por la memoria principal y un procesador ARM que trabaja a 1 GHz, con hit time = 1 ciclo de clock y CPI = 1.

- La caché de instrucciones es perfecta (siempre acierta), pero la caché de datos tiene un 15% de miss rate. En una falla de caché, el procesador se detiene durante 200 ns para acceder a la memoria principal y luego reanuda el funcionamiento normal. Teniendo en cuenta los errores de caché, ¿cuál es el tiempo promedio de acceso a la memoria?
- ¿Cuántos ciclos de reloj por instrucción (CPI) se requieren en promedio para instrucciones load y store, considerando las condiciones del punto a)?
- Asumiendo que la distribución de instrucciones se divide en las siguientes categorías: 25% loads, 10% stores, 13% branches y 52% instrucciones tipo R, ¿cuál es el CPI promedio?
- Suponiendo ahora que la caché de instrucciones tampoco es ideal y tiene un miss rate del 10%, ¿cuál es el CPI promedio considerando los fallos de ambas cachés?. Asuma la misma distribución de instrucciones de programa del punto c).

Ejercicio 12:

La siguiente tabla muestra los datos de la caché de un solo nivel (L1) para dos procesadores distintos (P1 y P2). En ambos procesadores el tiempo de acceso a memoria principal es de 70ns y el 36% de las instrucciones acceden a la memoria de datos.

	L1 Size	L1 Miss Rate	L1 Hit Time
P1	2 KiB	8.0%	0.66ns
P2	4 KiB	6.0%	0.90ns

- Asumiendo que el *hit time* de la caché L1 determina el tiempo de ciclo de ambos procesadores, calcule sus respectivas frecuencias de CLK.
- ¿Cuál es el AMAT para ambos procesadores?
- Asumiendo un CPI de 1 (sin memory stalls) y considerando las penalidades derivadas de instrucciones y datos (mismo Miss Rate), ¿cuál es el CPI total para ambos procesadores? ¿Cuál de los dos procesadores es más rápido?