

# Guía de Trabajos Prácticos N° 4

## Organización de la memoria

### Teoria

Suponga que la memoria es direccionable por bytes y las palabras son de 64 bits, a menos que se especifique lo contrario.

1. El siguiente código está escrito en C, donde los elementos dentro de la misma fila se almacenan de forma contigua. Suponga que cada palabra es un entero de 64 bits.

```
for (I=0; I<8; I++)
    for (J=0; J<8000; J++)
        A[I][J]=B[I][0]+A[J][I];
```

- a. ¿Cuántos enteros de 64 bits se pueden almacenar en un bloque de cache de 16 bytes?
- b. ¿Qué variables muestran **localidad temporal**?
- c. ¿Qué variables muestran **localidad espacial**?

La localidad se ve afectada tanto por el orden de referencia como por el diseño de los datos. El mismo cálculo también se puede escribir en Matlab, que se diferencia de C en que almacena elementos de matriz dentro de la misma columna de forma contigua en la memoria.

```
for I=1:8
    for J=1:8000
        A(I,J)=B(I,0)+A(J,I);
    end
end
```

- d. ¿Qué variables muestran **localidad temporal**?
- e. ¿Qué variables muestran **localidad espacial**?
- f. ¿Cuántos bloques de cache de 16 bytes se necesitan para almacenar todos los elementos de la matriz de 64 bits a los que se hace referencia utilizando el almacenamiento de la matriz de Matlab? ¿Cuántos usa el almacenamiento matricial de C? (Suponga que cada fila contiene más de un elemento).

2. Los cache son importantes para proporcionar una jerarquía de memoria de alto rendimiento a los procesadores. A continuación se muestra una lista de referencias de direcciones de memoria de 64 bits, expresadas como direcciones

0x03, 0xb4, 0x2b, 0x02, 0xbf, 0x58, 0xbe, 0x0e, 0xb5, 0x2c, 0xba, 0xfd

- a. Para cada una de estas referencias, identifique la dirección de la palabra binaria, la etiqueta y el índice con un cache mapeado directamente con 16 bloques de una palabra. Enumere si cada referencia es un acierto o un error, asumiendo que la cache está inicialmente vacía.
- b. Para cada una de estas referencias, identifique la dirección de la palabra binaria, la etiqueta, el índice y el desplazamiento dado un cache mapeado directamente con bloques de dos palabras y un tamaño total de ocho bloques. Enumere si cada referencia es un acierto o un error, asumiendo que la cache está inicialmente vacía.

3. Por convención, un cache se nombra de acuerdo con la cantidad de datos que contiene (es decir, un cache de 4 KiB puede contener 4 KiB de datos), sin embargo las caches también requieren SRAM para almacenar metadatos como etiquetas y bits válidos. Examine cómo la configuración de una cache afecta la cantidad total de SRAM necesaria para implementarla, así como el rendimiento de la cache. Para todas los Incisos suponga que las caches son direccionables por bytes y que las direcciones y palabras son de 64 bits.
  - a. Calcule el número total de bits necesarios para implementar una cache de 32 KiB con bloques de dos palabras.
  - b. Calcule el número total de bits necesarios para implementar una cache de 64 KiB con bloques de 16 palabras. ¿Cuánto más grande es esta cache que la de 32 KiB descrita en el Inciso a. (Observe que, al cambiar el tamaño del bloque, duplicamos la cantidad de datos sin duplicar el tamaño total de la cache)
  - c. Explique por qué esta cache de 64 KiB, a pesar de su tamaño de datos más grande, puede proporcionar un rendimiento más lento que la primera cache.
4. Los cache son importantes para proporcionar una jerarquía de memoria de alto rendimiento a los procesadores, incluyendo varios niveles para aumentar el desempeño del sistema.
  - a. Se emplean búferes entre diferentes niveles de jerarquía de memoria para reducir la latencia de acceso. Enumere los posibles búferes necesarios entre los caches L1 y L2, así como el cache y la memoria L2.
  - b. Describa el procedimiento de manejo de un error de escritura en la cache L1, considerando los componentes involucrados y la posibilidad de reemplazar un bloque sucio.
5. Para un diseño de cache mapeado directamente con una dirección de 64 bits, los siguientes bits de la dirección se utilizan para acceder a la cache.

Tag	Index	Offset
63-10	9-5	4-0

- a. ¿Cuál es el tamaño (en palabras) del bloque de cache?
- b. ¿Cuántos bloques tiene la cache?
- c. ¿Cuál es la relación entre el total de bits necesarios para una implementación de cache de este tipo sobre los bits de almacenamiento de datos?

A partir del encendido, se registran las siguientes referencias de cache con direcciones de bytes

Address												
Hex	00	04	10	84	E8	A0	400	1E	8C	C1C	B4	884
Dec	0	4	16	132	232	160	1024	30	140	3100	180	2180

- d. Para cada referencia, enumere (1) su etiqueta, índice y desplazamiento, (2) si es un acierto o un error, y (3) qué bytes fueron reemplazados (si corresponde).
- e. ¿Cuál es la proporción de aciertos?

6. En general, el tiempo de acceso a la cache es proporcional a la capacidad. Suponga que los accesos a la memoria principal toman 70 ns y que el 36% de todas las instrucciones acceden a la memoria de datos. La siguiente tabla muestra los datos de las memorias cache L1 adjuntas a cada uno de los dos procesadores, P1 y P2.

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

- Suponiendo que el Hit-time de L1 determina los tiempos de ciclo para P1 y P2, ¿cuáles son sus respectivas frecuencias de reloj?
- ¿Cuál es el tiempo medio de acceso a la memoria, dado por  

$$AMAT = Hit\ Time + Miss\ rate \times Miss\ penalty$$
para P1 y P2 (en ciclos)?
- Suponiendo un CPI base de 1.0 sin paradas de memoria, ¿cuál es el total de Ciclos por Instrucción (CPI) para P1 y P2? ¿Qué procesador es más rápido? (Cuando decimos "CPI base de 1.0", nos referimos a que las instrucciones se completan en un ciclo, a menos que el acceso a las instrucciones o el acceso a los datos provoque una falla de cache).

Para los siguientes Incisos consideraremos la inclusión de una cache L2 a P1 (para supuestamente compensar su capacidad limitada de cache L1). Utilice las capacidades de la cache L1 y los Hit-time de la tabla anterior para resolver estos problemas. La tasa de fallos L2 indicada es su tasa de fallos local.

L2 Size	L2 Miss Rate	L2 Hit Time
1 MiB	95%	5.62 ns

- ¿Cuál es el tiempo medio de acceso a la memoria para P1 con la adición de una cache L2? ¿El AMAT es mejor o peor con la cache L2?
  - Suponiendo un CPI base de 1.0 sin paradas de memoria, ¿cuál es el CPI total para P1 con la adición de una cache L2?
7. Este ejercicio compara las caches asociativas con las caches mapeadas directamente. Para resolver los Incisos consulte la siguiente secuencia de direcciones

0x03, 0xb4, 0x2b, 0x02, 0xbe, 0x58, 0xbf, 0x0e, 0x1f, 0xb5, 0xbf, 0xba, 0x2e, 0xce

- Dibuje la organización de un cache asociativo de tres vías con bloques de dos palabras y un tamaño total de 48 palabras. El esquema debe tener un estilo similar al presentado en clase pero mostrando el ancho de la etiqueta y los campos de datos.
- Analice el comportamiento de la cache del Inciso a. Asuma una política de reemplazo de LRU. Para cada referencia, identifique la dirección de la palabra binaria, la etiqueta, el índice, el desplazamiento si la referencia es un acierto o un error, y qué etiquetas están en cada vía de la cache después de que se haya manejado la referencia.
- Dibuje la organización de un cache completamente asociativo con bloques de una palabra y un tamaño total de ocho palabras. Su esquema debe tener un estilo similar al presentado en clase pero mostrando el ancho de la etiqueta y los campos de datos.

- d. Analice el comportamiento de la cache del ejercicio Inciso c. Asuma una política de reemplazo de LRU. Para cada referencia, identifique la dirección de la palabra binaria, la etiqueta, el índice, el desplazamiento, si la referencia es un acierto o un error, y el contenido de la cache después de que se haya manejado cada referencia.
- e. Repita el Inciso b utilizando una política MRU.
8. La memoria virtual usa una tabla de páginas para rastrear la asignación de direcciones virtuales a direcciones físicas. Este ejercicio muestra cómo se debe actualizar esta tabla a medida que se accede a las direcciones. Los siguientes datos constituyen un flujo de direcciones de bytes virtuales como se ve en un sistema.

Decimal	4669	2227	13916	34587	48870	12608	49225
hex	0x123d	0x08b3	0x365c	0x871b	0xbec6	0x3140	0xc049

Suponga páginas de 4 KB, un TLB completamente asociativo de cuatro entradas y una política de reemplazo de LRU. Si las páginas se deben traer desde el disco, incremente el siguiente número de página más grande.

Bufere de Traducción (TLB)

Valid	Tag	Physical Page Number	Time Since Last Access
1	0xb	12	4
1	0x7	4	1
1	0x3	6	3
0	0x4	9	7

Tabla de Página

Index	Valid	Physical Page or in Disk
0	1	5
1	0	Disk
2	0	Disk
3	1	6
4	1	9
5	1	11
6	0	Disk
7	1	4
8	0	Disk
9	0	Disk
a	1	3
b	1	12

- a. Para cada acceso que se muestra arriba, indique si el acceso es un acierto o un error en la TLB, si el acceso es un acierto o no en la tabla de páginas, si el acceso es un error de página, el estado actualizado de la TLB.
- b. Repita el Inciso a, pero esta vez use páginas de 16 KB en lugar de páginas de 4KB. ¿Cuáles serían algunas de las ventajas de tener un tamaño de página más grande? ¿Cuáles son algunas de las desventajas?
- c. Analice por qué una CPU debe tener un TLB para un alto rendimiento. ¿Cómo se manejarían los accesos a la memoria virtual si no hubiera TLB?
9. En este ejercicio, examinaremos las optimizaciones de espacio / tiempo para tablas de páginas. La siguiente lista proporciona los parámetros de un sistema de memoria virtual.

Virtual Address (bits)	Physical DRAM Installed	Page Size	PTE Size (byte)
43	16 GiB	4 KiB	4

- a. Para una tabla de página de un solo nivel, ¿cuántas entradas de tabla de página (PTE) se necesitan? ¿Cuánta memoria física se necesita para almacenar la tabla de páginas?
- b. El uso de una tabla de páginas de varios niveles puede reducir el consumo de memoria física de las tablas de páginas manteniendo solo los PTE activos en la memoria física. ¿Cuántos niveles de tablas de página se necesitarán si se permite que las tablas de segmento (las tablas de página de nivel superior) tengan un tamaño ilimitado? ¿Cuántas referencias de memoria se necesitan para la traducción de direcciones si faltan en TLB?
- c. Suponga que los segmentos están limitados al tamaño de página de 4 KiB (para que se puedan paginar). ¿Son 4 bytes lo suficientemente grandes para todas las entradas de la tabla de páginas (incluidas las de las tablas de segmentos)?
- d. ¿Cuántos niveles de tablas de páginas se necesitan si los segmentos están limitados al tamaño de página de 4 KiB?