

Práctico 11 – Pipeline, Jerarquía de memoria

Objetivo

El objetivo de este práctico es trabajar sobre algunos de los conceptos de pipeline y jerarquía de memoria, y entender los diferentes esquemas de memoria caché.

Notas

- La escritura de una variable en memoria es un acceso a memoria.
-

Preguntas teóricas:

- ¿Qué es un Hazard? ¿Cómo se clasifican?
 - Explique la técnica de forwarding utilizada en procesadores con pipeline para mitigar el impacto de las dependencias de datos.
 - Explique una técnica utilizada en los procesadores con pipeline para reducir el impacto de los hazards de control.
 - Describa la organización de una caché con correspondencia directa. ¿Qué políticas de reemplazo son aplicables en este caso?
 - Describa la organización de una caché con correspondencia asociativa por conjuntos. ¿Qué políticas de reemplazo son aplicables en este caso?
 - Explique el funcionamiento de las políticas de escritura write-through y write-back.
 - ¿A qué se le llama “problema de coherencia de caché”? ¿Cómo se resuelve?
-

Ejercicio 1 ★

Considere el siguiente código:

```
for (i = 0; i < 20; i++)  
    for (j = 0; j < 10 ; j++)  
        a[i]=a[i]*a[j]
```

- Dé un ejemplo de localidad espacial en el código.
- Dé un ejemplo de localidad temporal en el código.

Ejercicio 2 ★

Describa las características generales de un programa que muestra escasa localidad temporal pero mucha localidad espacial, con respecto al acceso a datos. Escriba un pseudocódigo con estas características.

Ejercicio 3 ★★

Dada una arquitectura con las siguientes características:

- Direcciones de 20 bits.
- Memoria caché de 4 Kb
- 256 líneas de caché.

Especifique la estructura de la caché y el esquema de las direcciones para:

- (a) Correspondencia directa.
- (b) Correspondencia totalmente asociativa.
- (c) Correspondencia asociativa por conjuntos de 2 vías.

Ejercicio 4 ★★

Considere un microprocesador con direcciones de 32 bits que tiene un caché on-chip de 32 KBytes asociativa por conjuntos de 4 vías. Suponga que la caché tiene un tamaño de línea de 4 palabras de 32 bits. Dibuje un diagrama de bloques de esta caché, mostrando su organización y cómo se utilizan los diferentes campos de dirección para determinar un acierto/fallo de caché.

¿Dónde se asigna, dentro de la caché, la palabra de la posición de memoria 0xABCDE8F8?

Ejercicio 5 ★★★ (en OpenFing)

Sea un computador de 32 bits y memoria direccionable por bytes que dispone de una memoria caché de 128 KB y líneas de 32 bytes, con una organización de correspondencia directa. Considere el siguiente fragmento de código:

```
for (i = 0; i < 512; i++)  
    for (j = 0; j < 512; j++)  
        c[i][j] = a[i][j] + b[i][j]
```

Donde a, b y c representan matrices cuadradas de 512 x 512 números enteros, que se almacenan en memoria de forma consecutiva. Cada matriz, a su vez, se almacena por filas. La matriz a comienza en la dirección 0x00100000.

- (a) Especifique cómo se divide la dirección de memoria entre etiqueta, línea, desplazamiento. Justifique.
- (b) Calcule la tasa de aciertos que se produce en la memoria caché para el fragmento de código anterior, suponiendo que la caché está inicialmente vacía.
- (c) Se propone un cambio en el algoritmo que recorra primero las columnas y luego las filas de las matrices ¿Esto mejoraría la tasa de aciertos? ¿Y almacenando las matrices por columnas? Justifique.
- (d) Proponga un diseño de memoria caché que mejore la tasa de aciertos para el algoritmo sin que ello suponga un incremento excesivo en el tiempo de búsqueda en la memoria caché. Calcule la tasa de aciertos para este diseño.

Ejercicio 6 ★★★ (con contenido complementario)

Considere un CPU de 16 bits, que emite direcciones de 24 bits, operando con una memoria RAM y un Cache L1 de 1 KB, con 16 líneas y una función de correspondencia totalmente asociativa.

- (a) Indique cómo se interpreta la dirección de memoria para el cache indicado. Halle el valor de todos los parámetros relevantes
- (b) En caso de producirse un miss en el Cache L1, ¿Cuántos bytes se transfieren entre memoria principal y el cache?

Suponga que entre el Cache L1 y la memoria principal se agrega un Cache L2 de 32 KB, con 128 líneas y una función de correspondencia asociativa en conjuntos de 4 vías. En este tipo de jerarquías, en los caches de niveles inferiores (L2, L3), los pedidos desde los niveles superiores son a nivel de bloque. Suponiendo buses de datos de 16 bits entre el CPU y el cache L1, entre el cache L1 y el cache L2, y entre el cache L2 y la memoria principal, todos operando a la misma frecuencia:

- (c) ¿Cómo se interpreta la dirección de memoria para el Cache L2? Halle el valor de todos los parámetros relevantes.
- (d) ¿Cuántos bytes se transfieren entre la CPU y el cache L1 en caso de ocurrir hit en la cache L1? ¿Cuántos ciclos de bus se demora en hacer la transferencia?
- (e) ¿Cuántos bytes se transfieren entre el cache L1 y el cache L2 en caso de ocurrir miss en el cache L1 pero hit en el cache L2? ¿Cuántos ciclos de bus se demora en hacer la transferencia?
- (f) ¿Cuántos bytes se transfieren entre el cache L2 y la memoria principal en caso de ocurrir miss en el cache L1 y también en el L2? ¿Cuántos ciclos de bus se demora en hacer la transferencia?