Arquitectura de Computadores
Problemas (hoja 4). Curso 2006-07

1. Sea un computador superescalar similar a la versión Tomasulo del DLX capaz de lanzar a ejecución dos instrucciones independientes por ciclo de reloj (incluyendo operaciones enteras). El siguiente código implementa la operación SAXPY (Y = a * X + Y) sobre un vector de 100 elementos:

```
foo:  LD    F2, 0(R1)   ; leer X(i)
      MULTD F4, F2, F0  ; a*X(i)
      LD    F6, 0(R2)   ; leer Y(i)
      ADDD  F6, F4, F6  ; a*X(i)+Y(i)
      SD    0(R2), F6   ; almacenar Y(i)
      ADDI  R1, R1, #8  ; incrementar el índice de X
      ADDI  R2, R2, #8  ; incrementar el índice de Y
      SGTI  R3, R1, done
      BEQZ  R3, foo
```

Las unidades funcionales están segmentadas y se asume que existen suficientes estaciones de reserva.

a) Desenrollar el bucle SAXPY para obtener 4 copias del cuerpo y planificarlo de tal forma que se alcance el mayor IPC posible.
b) ¿Cuántos ciclos emplea cada iteración del bucle original?
c) ¿Cuál es la ganancia producida por el desenrollado del bucle?

2. Sea un computador superescalar similar al del ejercicio anterior pero además con ejecución especulativa. Las unidades funcionales segmentadas tienen las siguientes características:

Unidad funcional	**Cantidad**	**Latencia**
FP mult	1	7 ciclos
FP add	1	5 ciclos
ALU Int	1	1 ciclos

a) Considerando la misma versión del código SAXPY que en el ejercicio anterior mostrar el estado del procesador cuando se lanza a ejecución la instrucción de salto por segunda vez. Se supone que el salto es correctamente predicho y tomado en la primera iteración.
b) ¿Cuantos ciclos emplea cada iteración del bucle?

3. Sea un procesador segmentado con planificación dinámica mediante el algoritmo de Tomasulo

- Los datos que se escriben en la etapa de escritura se pueden usar en la etapa de ejecución de una instrucción en el mismo ciclo.
- Las instrucciones SGTI, la BNEZ y NOP tienen tratamiento de instrucciones enteras.
- Los LOAD Y STORE tienen una latencia de dos ciclos, utilizan su propia unidad funcional no segmentada
- Hay un solo bus de datos común (CDB)
- La estructura del procesador tiene las siguientes características:

UF	CANTIDAD	LATENCIA	SEGMENTADA
FP ADD	1	2	SI
FP DIV	1	8	SI
FP MUL	1	4	SI

ESTACIONES RESERVA	CANTIDAD
FP ADD	2
FP DIV	2
FP MUL	2

INT ALU	1	1	SI		INT ALU	3
MEMORIA	1	2	NO		LOAD	2
					STORE	2

Dado el siguiente fragmento de programa:

```
      ADDI R1,R0,#DIR
      LD    F0,0(R7)
LOOP: LD    F4,0(R1)
      DIVD  F8,F4,F0
      SUBI  R1,R1, #4
      LD    F2,0(R1)
      MULD  F8,F2,F8
      SD    0(R1),F8
      SUBI  R3,R3,,#8
      SGTI  R5,R3,#1000
      BEQZ  R5,LOOP
      NOP
```

a) Representar el diagrama instrucción – tiempo para todo el fragmento de programa considerando sólo la primera iteración, indicando en cada caso el tipo de parada que se produce

b) Indicar el diagrama instrucción – tiempo para todo el fragmento de programa considerando sólo la primera iteración, suponiendo un superescalar con las mismas unidades funcionales ya vistas, que lanza un par de instrucciones de cualquier tipo en un ciclo de reloj. En caso de dependencia de datos entre dos instrucciones de un par se congela el lanzamiento de la segunda. Suponer que existen dos buses comunes de datos, que el banco de registros puede realizar dos escrituras en cada ciclo de reloj, y que tiene suficientes estaciones de reserva para que no se produzcan paradas.

4. Si el CPI de un computador con un sistema perfecto de memoria es 1.5.

Suponiendo que:

- La latencia de memoria son 40 ciclos.
- Las transferencias de bloques se realizan a 4 bytes/ciclo.
- El 50% de los bloques de datos son modificados.
- El tamaño de bloque es de 32 bytes.
- El 20% de las instrucciones son transferencias de datos.
- No hay buffer de escritura.

a) ¿Cuál sería el CPI para una cache directa unificada con post-escritura de 16KB y tasa de fallos de 0.029?

Si ahora suponemos además la existencia de un TLB tal que:

- El fallo de TLB produce una penalización de 20 ciclos.
- El acceso al TLB no afecta al tiempo de acierto de cache.
- El 0.2% de las referencias no se encuentran en el TLB.

b) Calcular de nuevo el CPI efectivo teniendo en cuenta el TLB.
c) Determinar el impacto del TLB en el rendimiento si se accede a la cache con dirección virtual o con dirección física.

5. Un computador tiene memorias cache separadas de datos (8KB) e instrucciones (16KB) y las siguientes características:

- La tasa de aciertos de la cache de instrucciones es del 95%.
- La tasa de aciertos de la cache de datos es del 85%.
- El tamaño de bloque de la cache de instrucciones es de 8 bytes y el de la cache de datos es de 4 bytes. En ambos casos, cuando se falla se lee el bloque completo.

La CPU genera 10^8 referencias a instrucciones y $4*10^7$ a datos por unidad de tiempo. Cada una de estas referencias se suponen de un byte. El 20 % de las referencias a datos son escrituras. El 25% de los bloques en la memoria cache de datos son modificados (escritos) durante su permanencia en ella.
Calcular la anchura de banda media necesaria entre memoria principal y memoria cache en los siguientes casos:

a) Con política de escritura directa (sin asignación en escritura).
b) Con política de post-escritura (con asignación en escritura).

6. Calcular la formula del tiempo medio de acceso para una cache de tres niveles.

7. Una cache de instrucciones de tipo directo puede en algún caso proporcionar mejor rendimiento que una cache totalmente asociativa con reemplazamiento LRU. Explicar como es posible.

8. Un computador con memoria cache separada de datos e instrucciones tiene las siguientes características:
- El 90% de todas las referencias a instrucciones generadas por la CPU son encontradas en la cache.
- El 80% de todas las referencias a datos generadas por la CPU son encontradas en la cache.
- El bloque de la cache de instrucciones es de cuatro palabras y en cada fallo se lee el bloque completo.

El bloque de la cache de datos es de dos palabras y en cada fallo se lee el bloque completo.
La CPU genera 10^7 referencias a instrucciones y $2*10^6$ referencias a datos. El 20% de estas referencias son escrituras. El 25% de los bloques en la memoria cache de datos son modificados (escritos) durante su permanencia en ella.
Calcular la anchura de banda media necesaria entre memoria principal y memoria cache en los siguientes casos:

(a) Con política escritura directa (write-through) sin asignación en escritura
(b) Con política post-escritura (copy-back) con asignación en escritura

9. Considérese un procesador con instrucciones de 32 bits y una cache de instrucciones de 32 bytes con líneas de 8 bytes. Para los dos fragmentos de código DLX siguientes, indicar la organización (y sus parámetros) que da lugar a una ejecución con el mínimo número de fallos (si hay varias que produzcan el mismo rendimiento, deberá escogerse la de menor coste hardware).

```
0     lw    r1,100(r2)        0     lw    r1,100(r2)
1     add   r1,r1,r3          1     add   r1,r1,r3
2     sub   r4,r5,r1          2     sub   r4,r5,r1
3     j     8                 3     j     8
...                           ...
6     mul   r2,r2,r8          6     mul   r2,r2,r8
7     sub   r2,r2,r9          7     sub   r2,r2,r3
8     div   r8,r1,r4          8     add   r8,r1,r0
9     j     0                 9     j     6
```

10. Un procesador a 300 Mhz solicita un acceso a memoria cache para recoger una palabra cada ciclo de reloj. Ejecutando un determinado programa la tasa de fallos de memoria cache es 0.1, en caso de fallo se procede a reemplazar el bloque de 4 palabras. La cache utiliza la política de “copy-back” y cada bloque dispone de un bit que indica si se ha modificado desde que llegó a la cache. El porcentaje de bloques que se modifican es del 35%. Se construye un computador paralelo de memoria compartida con este procesador donde los procesadores (con sus cache locales) se conectan a memoria principal por medio de un bus con un ancho de

banda de 5,2 Gpalabras/seg. ¿Cual es el número máximo de procesadores que soportará esta configuración sin agotar el ancho de banda del bus?

11. Un computador de 32-bits (palabra de 32bits) y con una memoria direccionable en bytes tiene una cache única de 512 bytes asociativa por conjuntos y con 2 bloques por conjunto. Cada bloque es de 4 palabras (16bytes) y una reemplazamiento LRU.

a) Para la siguiente secuencia de referencias indicar si el acceso es un acierto o un fallo. En caso de fallo indicar el tipo (inicial, capacidad, conflicto). Suponer la cache vacía al inicio.

Dirección	Acierto/fallo	Tipo
0x300		
0x1BC		
0x206		
0x109		
0x308		
0x1A1		
0x1B1		
0x2AE		
0x3B2		
0x10C		
0x205		
0x301		
0x3AE		
0x1A8		
0x3A1		
0x1BA		

b) Calcular la tasa de aciertos y la tasa de fallos.

12. Tenemos un procesador a 500Mhz con 2 niveles de cache, memoria principal y un disco para memoria virtual. El primer nivel de cache es separado para datos e instrucciones. Los parámetros del sistema de almacenamiento son los siguientes:

	Tiempo de acceso	Tasa de fallos	Tamaño del bloque
Cache nivel 1	1 ciclo	4% datos 1% instrucciones	64 bytes
Cache nivel 2	20 ciclos + 1ciclo/64 bits	2%	128 bytes
DRAM	100 ns + 25ns/ 8 bytes	1%	16K bytes
Disco	50ms + 20ns/byte	0%	16K bytes

El TLB produce 0.1% de fallos para los datos con una penalización de 40 ciclos (Las instrucciones no producen fallos en el TLB). Calcular el tiempo medio de acceso (TAMA) para instrucciones y para datos asumiendo que solo se hacen lecturas.

13 [Se adjunta solución]. Disponemos de una memoria principal de 64 Kb direccionable en bytes. Consideremos una memoria caché de 4Kb con organización directa, líneas de 256 bytes y que realiza precarga en todos los casos (tanto que haya fallo como acierto se trae el bloque siguiente al que se referencia).
Suponiendo que tenemos la siguiente secuencia de accesos a memoria principal:
04F5h, 11E0h, 1500h, 2000h, 241Fh, 16FFh, 1233h, mostrar el contenido del directorio caché, según el modelo, así como el número de fallos y de precargas que se producen. Indicar con el símbolo "*" un fallo, con "+" un acierto y con "-" una precarga.

	04F5h	11E0h	1500h	2000h	241Fh	16FFh	1233h
0							
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							

Solución

Como el tamaño de la memoria principal son 64 Kb necesitamos 16 bits para construir una dirección. La caché tiene 4 Kb de tamaño, y como cada línea tiene 256 bytes tenemos un total de 16 bloques. Por tanto necesitamos 8 bits para especificar el byte dentro del bloque y 4 bits para especificar la línea de cache. Por tanto el campo etiqueta tendrá 16 - (4 + 8) = 4 bits.

	04F5h	11E0h	1500h	2000h	241Fh	16FFh	1233h
0				2*	2	2	2
1		1*	1	2-	2	2	2
2		1-	1	1	1	1	1+
3							1-
4	0*	0	0	0	2*	2	2
5	0-	0	1*	1	2-	2	2
6			1-	1	1	1+	1
7						1-	1
8							
9							
10							

11							
12							
13							
14							
15							