Arquitectura e Ingeniería de Computadores - 3er Curso. Sol. de 2º parcial 19/1/2018

Responde a cada pregunta en una hoja de respuesta distinga. Tiempo disponible: 2h

1. **(6 puntos)** Se dispone de un procesador MIPS superescalar de **dos vías** y que posee gestión dinámica de instrucciones con especulación hardware basado en el algoritmo de Tomasulo. El lanzamiento de las instrucciones se realiza en orden y alineado. Las instrucciones atraviesan las siguientes etapas: IF (búsqueda de instrucciones), I (decodificación y lanzamiento de las instrucciones), En (ejecución en el operador multiciclo correspondiente), WB (escritura en los buses comunes de datos; duración de la transferencia 1 ciclo) y C (confirmación de las instrucciones). El ROB tiene 64 entradas, identificándose la primera de ellas como entrada #0.

   Las características de las unidades funcionales **no segmentadas** son los siguientes:

	Nº Operadores	Latencia	Características
Carga/Almacenamiento	2	2	4 buffers de lectura y 2 de escritura
Suma/Resta CF	1	2	4 estaciones de reserva
Multiplicador CF.	1	5	4 estaciones de reserva
Enteros/Saltos	2	1	4 estaciones de reserva

   Se pretende evaluar el comportamiento del procesador ante el siguiente fragmento de código:

```
      .data
zero: .double 0.0
prod: .double 0.0
x:    .double ...
y:    .double ...

      .text 0x00010000   ; Este fragmento de código
                         ; comienza en 0x00010000
      l.d f0, zero(r0)
loop: l.d f1, x(r1)
      l.d f2, y(r1)
      mult.d f1, f1, f2
      add.d f0, f0, f1
      dsub r1, r1, #8
      bnez r1, loop
      s.d f0, prod(r0)
```

   Supóngase que al comenzar la última iteración del bucle todas las estructuras de datos se encuentran libres (*reorder buffer*, estaciones de reserva, ...).

   Se solicita:

   a) Dibujar el diagrama instrucciones-tiempo de **la última iteración**. Refleja en el diagrama **sólo las instrucciones de esa iteración y de la siguiente que se busca erróneamente** al fallar el predictor.

   b) Indica en qué ciclos de reloj, desde el comienzo de la última iteración, se actualizan los valores de los registros f0, f1, f2 y r1. Si el valor de un registro es actualizado más de una vez deberás indicar todos los ciclos de reloj en los que se produzcan dichas actualizaciones.

   c) Indica las marcas que contenían los registros f0, f1, f2 y r1 en el ciclo de reloj en el que la primera instrucción `l.d f2, y(r1)` realiza la fase C.

   d) ¿Cuál será el número de ciclos consumido por una iteración cuando el predictor acierta? ¿Y cuando falla?.

   e) Si asumimos que X e Y son vectores de 512 elementos y que el procesador considerado funciona a una frecuencia de 900 MHz, ¿Cuántos MFLOPS ofrecerá el procesador ejecutando el código bajo estudio? Asume que el predictor falla en las predicciones de la primera y la última iteración.

   f) ¿Cuál será el CPI efectivo que obtendremos en el procesamiento de vectores de tamaño muy grande?

Solución:

a) () Diagrama instrucciones-tiempo de la última iteración.

```
                          1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
      l.d f0, zero(r0)   IF  X
loop: l.d f1, x(r1)      IF  I AC L1 L2 WB  C
      l.d f2, y(r1)         IF  I AC L1 L2 WB  C
      mult.d f1, f1, f2     IF  I  -  -  -  - M1 M2 M3 M4 M5 WB C
      add.d f0, f0, f1         IF  I  -  -  -  -  -  -  -  -  - A1 A2 WB C
      dsub r1, r1, #8          IF  I EX WB                             C
      bnez r1, loop               IF  I  - EX WB                          C
      s.d f0, prod(r0)            IF  X
      l.d f0, zero(r0)               IF  X
loop: l.d f1, x(r1)                  IF  I AC L1 L2 WB                    X
      l.d f2, y(r1)                     IF  I AC L1 L2 WB                X
      mult.d f1, f1, f2                 IF  I  -  -  -  - M1 M2 M3 M4 M5  X
      add.d f0, f0, f1                     IF  I  -  -  -  -  -  -  -  -  X
      dsub r1, r1, #8                      IF  I EX WB                    X
      bnez r1, loop                           IF  I  - EX WB              X
      s.d f0, prod(r0)                        IF  X
```

b) Indica en qué ciclos de reloj se actualizan los valores de los registros f0, f1, f2 y r1.

- F0 se actualiza en el ciclo 17.
- F1 se actualiza en los ciclos 7 y 14.
- F2 se actualiza en el ciclo 8 .
- R1 se actualiza en el ciclo 17.

c) Indica las marcas que contenían los registros f0, f1, f2 y r1 en el ciclo de reloj en el que la primera instrucción `l.d f2, y(r1)` realiza la fase C.
Para el marcado que realizamos asumimos que la primera entrada del ROB se etiqueta como #0.

- F0 contiene la marca #9.
- F1 contiene la marca #8.
- F2 contiene la marca #7.
- R1 contiene la marca #10.

d) Indica el número de ciclos consumido por una iteración cuando el predictor acierta y también cuando falla.
Una iteración normal se ejecuta en 4 ciclos. La penalización por fallo en la predicción es de 14 ciclos, es decir, que una iteración con fallo require 18 ciclos para ejecutarse.

e) Si asumimos que X e Y son vectores de 512 elementos y que el procesador considerado funciona a una frecuencia de 900 MHz, ¿Cuantos MFLOPS ofrecerá el procesador cuando ejecute el código bajo estudio? Asume que el predictor sólo falla en la primera y la última de las predicciones que efectúa.

$$T(n) = 2 * 18 + 510 * 4 = 2076 ciclos$$

$$MFLOPS = \frac{2 \times 512}{2076} = 0,4932 \frac{op.}{ciclo} \text{ @ } 0{,}9\, GHZ \approx 443,93\, MFLOPS$$

f) ¿Cuál será el CPI efectivo que obtendremos en el procesamiento de vectores de tamaño muy grande?
Si el número de elementos de X e Y es muy elevado entonces:

$$CPI_{efectivo} = \frac{4}{6} = 0,67$$

Nota: este valor sólo es alcanzable cuando el número de elementos a procesar en los vectores considerados (X e Y) es muy grande ($\lim_{n \to \infty}$).

□

2. **(1 punto)** Responde a las siguientes preguntas:

   a) En el contexto de las memorias caché no bloqueantes, ¿a qué nos referimos cuando decimos que pueden servir peticiones siguiendo una aproximación de "acierto ante fallo" ("*hit under miss*" en inglés)?

   b) Un procesador multihilo de grano grueso conmuta de hilo cada vez que falla la búsqueda de una instrucción en la caché de instrucciones L1 o la búsqueda de un dato por parte de una instrucción de carga en la caché de datos L1. En este contexto, ¿Qué podría ocurrir, y cómo se verían afectadas las prestaciones, si la caché L1 no ofreciera la posibilidad de "acierto ante fallo"?

Solución:

   a) Non-blocking caches allow the CPU to continue executing instructions while a miss is being handled. In this context, the expression "hit under miss"means that the cache can only handle one miss at a time, but meanwhile it is able to service hits.

   b) If the L1 cache does not support hit under miss, it will not be able to service hits until it completes the last access that produced a miss. Since context is switched to another task on a L1 instruction miss, the new task could not be executed (because instruction fetches cannot be completed) until the previous miss is serviced. Thus, mutithreading support becomes useless.

□

3. **(3 puntos)** Un diseño de procesador incluye una caché L1 y una L2. Para reducir el tiempo medio de acceso a la memoria, la próxima generación de este procesador incluirá otro nivel de caché (L2) entre los dos ya existentes, de modo que el L2 antiguo se convertirá en el L3, sin modificar sus características. Las cachés tienen un diseño inclusivo.

   En el diseño original, la caché L1 tenía las siguientes características: Tiempo de acierto, $TA_{L1} = 1$ ciclo; tasa de fallos local, $TF_{L1} = 0{,}2$. Y para L2: Tiempo de acierto, $TA_{L2} = 15$ ciclos; tasa de fallos local, $TF_{L2} = 0{,}25$; penalización por fallo $PF_{L2} = 100$ ciclos (a la frecuencia del reloj del procesador).

   La nueva caché L2 es tal que si fuera el único nivel de caché en el sistema, la tasa de fallos global (y local) sería de $0{,}1$.

   Se solicita calcular:

   a) Tasa de fallos global para el diseño original.

   b) Tiempo medio de acceso a memoria para el diseño original.

   c) Tasa de fallos local para la nueva L2 cuando se incluye en el nuevo diseño.
   Hipótesis: la tasa de fallos global sólo depende de la geometría del último nivel de caché y, por lo tanto, no cambia al insertar la nueva caché L2. Además, la tasa de fallos local para L1 no cambia cuando se agregan, eliminan o modifican los niveles de caché restantes.

   d) Tiempo medio de acceso a memoria del nuevo diseño si el tiempo de acierto para la nueva caché L2 es de 10 ciclos. Supóngase que la tasa de fallos local de la nueva L2 cuando se incluye en el diseño es de $0{,}4$, y que la tasa de fallos global para el nuevo diseño es de $0{,}04$.

Solución:

   a) Global miss rate for the original design.
   $MR\ global_{L2} = MR_{L1} \times MR_{L2} = 0{,}2 \times 0{,}25 = 0{,}05$

   b) Average memory access time for the original design.

$$T_{\text{access}} = Ht_{L1} + MR_{L1} \times (Ht_{L2} + MR_{L2} \times MP_{L2})$$

$$T_{\text{access}} = 1 + 0{,}2 \times (15 + 0{,}25 \times 100) = 9 \text{ cycles}$$

c) Local miss rate for the new L2 when included in the new design.
If the new L2 was the only cache, $MR\ global_{New\ L2\ alone}$ = 0,1
If we include only the old L1 and the new L2, the global miss rate would remain the same, since it only depends on the LLC:
$MR\ global_{L1\ and\ new\ L2}$ = $MR_{L1} \times MR_{L2}$ = 0,2 × MR_{L2} = 0,1
Thus, MR_{L2} = 0.5

d) Average memory access time for the new design if the hit time for the new L2 is 10 cycles.
First, we have to compute the local miss rate of the L3 when incorporating the new L2.
$MR\ global_{L3}$ = $MR_{L1} \times MR_{L2} \times MR_{L3}$ = 0,2 × 0,4 × MR_{L3} = 0,04
Thus, MR_{L3} = 0.5

$$T_{\text{access}} = Ht_{L1} + MR_{L1} \times (Ht_{L2} + MR_{L2} \times (Ht_{L3} + MR_{L3} \times MP_{L3}))$$

$$T_{\text{access}} = 1 + 0{,}2 \times (10 + 0{,}4 \times (15 + 0{,}5 \times 100)) = 8{,}2 \text{ cycles}$$

□