

Responde cada pregunta en una hoja distinta. Tiempo disponible: 1h30m

1. (6 puntos) Se dispone de un procesador MIPS con ejecución fuera de orden y especulación hardware basada en el algoritmo de Tomasulo. Las instrucciones atraviesan las siguientes etapas: IF (búsqueda de instrucciones), I (decodificación y lanzamiento de las instrucciones), En (ejecución en el operador multi-ciclo correspondiente), WB (escritura en el bus común de datos) y C (confirmación de las instrucciones). El ROB tiene 32 entradas, identificándose la primera de ellas como entrada #0. El procesador dispone de un predictor de saltos del tipo *Branch Target Buffer* (BTB) de 1 bit que ofrece la predicción al final de la etapa IF.

Las características de las unidades funcionales son las siguientes:

	Nº Operadores	Latencia	Características
Carga/Almacenamiento	1	2	No segmentada; 4 buffers de lectura y 4 de escritura
Suma/Resta CF	1	2	Segmentada; 4 estaciones de reserva
Multiplicador CF.	1	4	Segmentada; 4 estaciones de reserva
Enteros/Saltos	1	1	8 estaciones de reserva

Se pretende evaluar el comportamiento del procesador ante el siguiente bucle, el cual calcula el resultado de la operación $\vec{z} = a\vec{x} + b\vec{y} + c$.

```
.data
x:      .double 1, 2, 3, 4, 5, 6, 7, 8
y:      .double 100, 100, 100, 100, 100, 100, 100, 100
```

```
.text 4096
loop:
    l.d f1, 0(r1)      ; X
    l.d f2, 0(r2)      ; Y
    mul.d f3, f1, f10   ; aX
    mul.d f4, f2, f11   ; bY
    add.d f3, f3, f4    ; aX+bY
    add.d f3, f3, f12   ; aX+bY+c
    s.d f3, 0(r3)      ; Z
    dadd r1, r1, 8
    dadd r2, r2, 8
    dadd r3, r3, 8
    bne r1,r4, loop
    trap #0
```

Antes de la ejecución del bucle, el estado del procesador es el siguiente:

r1	r2	r3	r4	f10	f11	f12
x	y	z	x+512	a	b	c

El diagrama instrucciones–tiempo que se muestra a continuación refleja la secuencia de instrucciones ejecutada durante la primera iteración del bucle:

PC	Instruccion	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
loop	l.d f1,0(r1)	IF	I	AC	L1	L2	WB	C																			
4100	l.d f2,0(r2)		IF	I	AC	-	L1	L2	WB	C																	
4104	mul.d f3,f1,f10			IF	I	-	-	M1	M2	M3	M4	WB	C														
4108	mul.d f4,f2,f11				IF	I	-	-	-	M1	M2	M3	M4	WB	C												
4112	add.d f3,f3,f4					IF	I	-	-	-	-	-	-	A1	A2	WB	C										
4116	add.d f3,f3,f12						IF	I	-	-	-	-	-	-	-	-	A1	A2	WB	C							
4120	s.d f3,0(r3)							IF	I	AC	-	-	-	-	-	-	-	-	-	-	C	L1	L2				
4124	dadd r1,r1,8								IF	I	E1	-	WB	-	-	-	-	-	-	-	-	-	-	-	C		
4128	dadd r2,r2,8									IF	I	E1	-	-	WB	-	-	-	-	-	-	-	-	-	-	C	
4132	dadd r3,r3,8										IF	I	E1	-	-	WB	-	-	-	-	-	-	-	-	-	C	
salto	bne r1,r4,loop											IF	I	E1	-	-	-	WB	-	-	-	-	-	-	-	-	C
4140	trap 0												IF	I	-	-	-	-	-	-	-	-	-	-	-	-	x
_etext	<nop>														if	if	if	if	if	if	if	if	if	if	if	if	X
loop	l.d f1,0(r1)																										IF

Se solicita:

- Considera la instrucción `mul.d f3, f1, f10` e indica el ciclo de reloj en el cual:
 - Se copia el valor del primer operando fuente a su estación de reserva.
 - Se copia el valor del segundo operando fuente a su estación de reserva.
 - Se libera el operador físico utilizado.
 - Se difunde el resultado de la operación por el bus común de datos.
 - Se almacena el resultado en el registro destino `f3`.
- Suponiendo, para simplificar, que el ROB, las estaciones de reserva, los operadores y la tabla BTB estaban vacíos al comenzar la ejecución del bucle, considera el ciclo de reloj en el que la instrucción `mul.d f3, f1, f10` realiza la fase C:
 - Indica los valores y las marcas de los registros `f3`, `f4`, `r1` y `r2` al final del ciclo. Los registros no inicializados almacenan el valor 0 inicialmente.
 - Muestra el estado del ROB al final del ciclo.
- ¿Cuál será el número de ciclos consumido por una iteración cuando el predictor acierta? ¿Y cuando falla?
- Si asumimos que `X` e `Y` son vectores de 100 elementos y que el procesador considerado funciona a una frecuencia de 1 GHz, ¿Cuántos MFLOPS ofrecerá el procesador ejecutando el código bajo estudio? Asume que la tabla BTB está vacía al empezar la ejecución del bucle.
- Con el objeto de mejorar las prestaciones, se diseña la siguiente generación del procesador para que sea superescalar de **2 vías**, replicando asimismo el número de unidades funcionales disponibles. La fase IF proporciona dos instrucciones alineadas en una dirección múltiplo de 8. Dibuja el diagrama instrucciones-tiempo de **una** iteración intermedia (en la que el predictor acierta). Refleja en el diagrama **las instrucciones de esa iteración, y las dos primeras de la siguiente**. Considera que al inicio de la iteración el ROB y las estaciones de reserva están vacías, y que los registros no tienen marcas.

Solución:

- Considera la instrucción `mul.d f3, f1, f10` e indica el ciclo de reloj en el cual:

- 1) Se copia el valor del primer operando fuente a su estación de reserva. Ciclo 6
 - 2) Se copia el valor del segundo operando fuente a su estación de reserva. Ciclo 4
 - 3) Se libera el operador físico utilizado. Ciclo 10
 - 4) Se difunde el resultado de la operación por el bus común de datos. Ciclo 11
 - 5) Se almacena el resultado en el registro destino f3. Ciclo 12
- b) Suponiendo, para simplificar, que el ROB, las estaciones de reserva y los operadores estaban vacíos al comenzar la ejecución del bucle, considera el ciclo de reloj en el que la instrucción `mul.d f3, f1, f10` realiza la fase C:

- 1) indica los valores y las marcas de los registros f3, f4, r1 y r2 al final del ciclo. Los registros no inicializados almacenan el valor 0 inicialmente.

	f3	f4	r1	r2
Valor	a		x	y
Marca	#5	#3	#7	#8

- 2) Muestra el estado del ROB al final del ciclo.

ROB						
#	Busy	Instrucción	Destino	Valor	Completado	Predicción
0	NO					
1	NO					
2	NO					
3	SI	<code>mul.d f4,f2,f11</code>	f4		NO	
4	SI	<code>add.d f3,f3,f4</code>	f3		NO	
5	SI	<code>add.d f3,f3,f12</code>	f3		NO	
6	SI	<code>s.d f3,0(r3)</code>	s1		NO	
7	SI	<code>dadd r1,r1,8</code>	r1	x+8	SI	
8	SI	<code>dadd r2,r2,8</code>	r2		NO	
9	SI	<code>dadd r3,r3,8</code>	r3		NO	
10	SI	<code>bne r1,r4,loop</code>	loop		NO	No Salta

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

Acierto: 11 ciclos

Fallo: 25 ciclos

- d) Si asumimos que X e Y son vectores de 100 elementos y que el procesador considerado funciona a una frecuencia de 1 GHz, ¿Cuántos MFLOPS ofrecerá el procesador ejecutando el código bajo estudio? Asume que la tabla BTB está vacía al empezar la ejecución del bucle.

Operaciones FP: 4 por iteración * 100 = 400 op FP

Tiempo por iteración. El predictor falla dos veces (la primera iteración por no estar el salto en la tabla y la última por cambiar el comportamiento). Por tanto: $98 \cdot 11 + 2 \cdot 25 = 1128$ ciclos

$MFLOPS = \frac{400}{1128} = 0,355 \text{ op/ciclo} = 0,355 \cdot 1000 \text{ MHz} = 355 \text{ MFLOPS}$

- e) Con el objeto de mejorar las prestaciones, se diseña la siguiente generación del procesador para que sea superescalar de **2 vías**, duplicando asimismo el número de unidades funcionales disponibles. La fase IF proporciona dos instrucciones alineadas en una dirección de byte múltiplo de 8. Dibuja el diagrama instrucciones-tiempo de una iteración intermedia (en la que el predictor acierta). Refleja en el diagrama **sólo las instrucciones de esa iteración, y las dos primeras de la siguiente**.

PC	Instruccion	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
loop	<code>l.d f1,0(r1)</code>	IF	I	AC	L1	L2	WB	C														
4100	<code>l.d f2,0(r2)</code>	IF	I	AC	L1	L2	WB	C														
4104	<code>mul.d f3,f1,f10</code>		IF	I	-	-	-	M1	M2	M3	M4	WB	C									
4108	<code>mul.d f4,f2,f11</code>		IF	I	-	-	-	M1	M2	M3	M4	WB	C									
4112	<code>add.d f3,f3,f4</code>			IF	I	-	-	-	-	-	-	-	A1	A2	WB	C						

4116	add.d f3,f3,f12	IF	I	-	-	-	-	-	-	-	-	-	-	A1	A2	WB	C		
4120	s.d f3,0(r3)	IF	I	AC	-	-	-	-	-	-	-	-	-	-	-	-	C	L1	L2
4124	dadd r1,r1,8	IF	I	E1	WB	-	-	-	-	-	-	-	-	-	-	-	-	C	
4128	dadd r2,r2,8	IF	I	E1	WB	-	-	-	-	-	-	-	-	-	-	-	-	C	
4132	dadd r3,r3,8	IF	I	E1	WB	-	-	-	-	-	-	-	-	-	-	-	-	C	
salto	bne r1,r4,loop	IF	I	E1	WB	-	-	-	-	-	-	-	-	-	-	-	-	C	
4140	trap 0	IF	X																
loop	l.d f1,0(r1)	IF	I	AC	L1	L2	WB	-	-	-	-	-	-	-	-	-	-	C	
4100	l.d f2,0(r2)	IF	I	AC	L1	L2	WB	-	-	-	-	-	-	-	-	-	-	C	



2. (4 puntos) Durante el proceso de diseño del nuevo modelo M1 Pro, los ingenieros de Apple han estado valorando diversas configuraciones de la jerarquía de memoria que se encuentra integrada en el mismo encapsulado. En la tabla siguiente se pueden observar algunas características de la memoria principal unificada de los modelos del procesador M1.

Características	Apple M1	M1 Pro
Tipo	LPDDR4X	LPDDR5
Ancho del bus (bits)	128	256
Ancho de banda (GB/s)	68.27	204.8

Los procesadores M1 siguen una arquitectura big.LITTLE con dos tipos de núcleos: los *IceStorm* y los *Firestorm* con ejecución fuera de orden. Los núcleos *Firestorm* funcionan a 3.2 GHz y tiene caches L1 de instrucciones y de datos separadas de uso privado y comparten la cache L2. Los tiempos en caso de acierto, TA , son de 1 ciclo para el nivel L1 y de 6 ciclos de CPU para L2. El tamaño de bloque en ambos niveles es de 128 bytes.

La temporización de la memoria principal es 16-16-16 para la memoria LPDDR4X y de 24-24-24 para la memoria LPDDR5.

En las pruebas de evaluación que se han realizado sobre los núcleos *Firestorm* se ha obtenido que la cache L1 sufre una tasa de fallos TF_{L1} de un 2 % y 5 % para instrucciones y datos respectivamente, mientras que la tasa de fallos de la cache L2, TF_{L2} , es de un 50 %.

Se solicita, justificando claramente la respuesta:

- En base al ancho de bus mostrado en la tabla, determinar el ancho de banda, Bw , de ambos modelos, Apple M1 y M1 Pro, en bytes por ciclo de memoria.
- A partir del dato anterior y la información mostrada en la tabla, determinar la frecuencia del bus de memoria, F_{bus} , de ambos modelos.
- Determinar la latencia normal, L , y reducida, L_r , para la memoria del Apple M1 y del M1 Pro en ciclos de bus.
- Calcular la penalización por fallo de L2, PF_{L2} , en ciclos de CPU para el modelo Apple M1 suponiendo una tasa de aciertos en el buffer de fila, TA_{bf} , del 80 %.
- Calcular la penalización por fallo de L1, PF_{L1} , en ciclos de CPU para el modelo Apple M1.
- Calcular el tiempo de ejecución en segundos del programa P, que ejecuta 20 millones de instrucciones con un CPI de 1.2, para el modelo Apple M1 sobre un solo núcleo. Supóngase una fracción no solapada del tiempo extra de acceso a datos del 75 % y un porcentaje de instrucciones de carga del 20 %. Supóngase a su vez que las instrucciones de almacenamiento no producen ninguna penalización apreciable debido al uso de buffers de escritura.
- Dado el incremento del ancho de banda de memoria en el modelo M1 Pro, se está considerando la posibilidad de **duplicar** el tamaño de bloque de la cache L2. Esto permite aumentar la tasa de **aciertos** de la cache L2 a un 60 %. Suponiendo que se mantiene la tasa de aciertos en el buffer de fila, TA_{bf} , del 80 %, calcular cuál sería la penalización por fallo de L1, PF_{L1} , sobre el M1 Pro en ciclos de CPU.

Solución:

- En base al ancho de bus mostrado en la tabla, determinar el ancho de banda, Bw , de ambos modelos, Apple M1 y M1 Pro, en bytes por ciclo de memoria.

$$Bw_{M1} = 2 \text{ transf/ciclo} \times \frac{128 \text{ bits/transf}}{8 \text{ bits/byte}} = 32 \text{ bytes/ciclo}$$

$$Bw_{Pro} = 2 \times \frac{256}{8} = 64$$

- b) A partir del dato anterior y la información mostrada en la tabla, determinar la frecuencia del bus de memoria, F_{bus} , de ambos modelos.

$$F_{bus-M1} = \frac{68,27 \text{ GB/s}}{32 \text{ bytes/ciclo}} = 2,1333 \text{ GHz}$$

$$F_{bus-pro} = \frac{204,8}{64} = 3,2 \text{ GHz}$$

- c) Determinar la latencia normal, L , y reducida, L_r , para la memoria del Apple M1 y del M1 Pro en ciclos de bus.

$$L_{DDR4} = CL + t_{RCD} + t_{RP} = 16 + 16 + 16 = 48 \text{ ciclos}$$

$$L_{DDR4-r} = CL = 16 \text{ ciclos}$$

$$L_{DDR5} = 24 + 24 + 24 = 72 \text{ ciclos}$$

$$L_{DDR5-r} = 24 \text{ ciclos}$$

- d) Calcular la penalización por fallo de L2, PF_{L2} , en ciclos de CPU para el modelo Apple M1 suponiendo una tasa de aciertos en el buffer de fila, TA_{bf} , del 80 %.

$$PF_{L2} = (1 - TA_{bf}) \times L + TA_{bf} \times L_r + \frac{B}{Bw}$$

$$PF_{L2-M1} = (1 - 0,8) \times 48 + 0,8 \times 16 + \frac{128}{32}$$

$$PF_{L2-M1} = 0,2 \times 48 + 0,8 \times 16 + 4$$

$$PF_{L2-M1} = 9,6 + 12,8 + 4 = 26,4 \text{ ciclos bus}$$

$$PF_{L2-M1} = 26,4 \times \frac{3,2}{2,1333} = 26,4 \times 1,5 = 39,6 \text{ ciclos}$$

- e) Calcular la penalización por fallo de L1, PF_{L1} , en ciclos de CPU para el modelo Apple M1.

$$PF_{L1} = T_{acc-L2} = TA_{L2} + TF_{L2} \times PF_{L2}$$

$$PF_{L1-M1} = 6 + 0,5 \times 39,6 = 25,8 \text{ ciclos}$$

- f) Calcular el tiempo de ejecución en segundos del programa P, que ejecuta 20 millones de instrucciones con un CPI de 1.2, para el modelo Apple M1 sobre un solo núcleo. Supóngase una fracción no solapada del tiempo extra de acceso a datos del 75 % y un porcentaje de instrucciones de carga del 20 %. Supóngase a su vez que que las instrucciones de almacenamiento no producen ninguna penalización apreciable debido al uso de buffers de escritura.

$$T_{ej} = I \times CPI \times T + I \times TF_I \times PF_{L1} \times T + I \times API_L \times TF_D \times PF_{L1} \times FNS \times T$$

$$T_{ej} = I \times (CPI + TF_I \times PF_{L1} + API_L \times TF_D \times PF_{L1} \times FNS) \times T$$

$$T_{ej-M1} = 20 \times 10^6 \times (1,2 + 0,02 \times 25,8 + 0,2 \times 0,05 \times 25,8 \times 0,75) \times 0,3125 \times 10^{-9}$$

$$T_{ej-M1} = 20 \times (1,2 + 0,516 + 0,1935) \times 0,3125 \text{ ms}$$

$$T_{ej-M1} = 20 \times 1,9095 \times 0,3125 \text{ ms} = 11,934375 \text{ ms}$$

- g) Dado el incremento del ancho de banda de memoria en el modelo M1 Pro, se está considerando la posibilidad de **duplicar** el tamaño de bloque de la cache L2. Esto permite aumentar la tasa de **aciertos** de la cache L2 a un 60 %. Suponiendo que se mantiene la tasa de aciertos en el buffer de fila, $TAbf$, del 80 %, calcular cuál sería la penalización por fallo de L1, PF_{L1} , sobre el M1 Pro en ciclos de CPU.

$$PF_{L2-pro} = 0,2 \times 72 + 0,8 \times 24 + \frac{256}{64}$$

$$PF_{L2-pro} = 14,4 + 19,2 + 4 = 37,6 \text{ ciclos bus}$$

$$PF_{L2-pro} = 37,6 \times \frac{3,2}{3,2} = 37,6 \times 1 = 37,6 \text{ ciclos}$$

$$PF_{L1-pro} = 6 + 0,4 \times 37,6 = 21,04 \text{ ciclos}$$

□