

EJERCICIOS 1

Ejercicio 1.1

Considere la ejecución de los siguientes cuatro programas en cada una de las tres máquinas que se indican

Programa	Tiempo de ejecución (segundos)		
	computador A	computador B	computador C
programa 1	1	10	20
programa 2	1000	100	20
programa 3	500	1000	50
programa 4	100	800	100

Suponga que se han ejecutado 100.000.000 de instrucciones en cada uno de los 4 programas.

Pregunta 1: Calcule los MIPS de cada programa en cada una de las 3 máquinas.

Considere que los 4 programas representan una carga de trabajo. Además, suponga que los programas se ponderan con el mismo peso al considerar la carga de trabajo.

Pregunta 2: Calcule el tiempo medio de ejecución por instrucción para cada máquina.


Pregunta 3: Calcule los MIPS para cada máquina. Efectúe el calculo a partir de los MIPS de cada programa.

Ejercicio 1.2

Disponemos de 3 versiones de un procesador: A, B y C y de un conjunto de programas de prueba que se utiliza para evaluarlos. El procesador B es una versión mejorada del procesador C y el procesador A es una versión mejorada del procesador B. En los procesadores A y B la cache es no bloqueante y los load son bloqueantes. En el procesador C la cache es bloqueante. Todos los procesadores funcionan a la misma frecuencia.

El CPI medio medido en la versión A es 1.25. Por otro lado conocemos que el CPI medio de A es un 30% menor que el CPI medio de B.

Pregunta 1: Calcule el CPI de B. Así mismo calcule la ganancia en rendimiento de A respecto de B.



El lenguaje máquina de los procesadores A y B se ha ampliado con instrucciones de prebúsqueda, las cuales no están disponibles en el procesador C. Conocemos que en ambos procesadores el número de instrucciones ejecutadas es un 3% mayor que en el procesador C.

En los procesadores A y B el CPI de las instrucciones de prebúsqueda es uno.

Pregunta 2: Calcule el CPI medio del resto de las instrucciones (que no son de prebúsqueda) en el procesador B.

Sea G la ganancia en rendimiento de A respecto de B y sea RR la reducción en rendimiento de B respecto de A. Ambos procesadores funcionan a la misma frecuencia.

Pregunta 3: Expresar G y RR en función del tiempo de ejecución de A y B. Muestre una expresión que permita calcular G en función de RR y otra expresión que permita el cálculo recíproco.

El mayor IPC de A respecto de B está determinado por una mejora arquitectónica que reduce la frecuencia de fallos por instrucción (f), sin incrementar el número de instrucciones ejecutadas. Todos los otros parámetros arquitectónicos permanecen invariables. En concreto, el CPI de la unidad de proceso (acierto siempre en cache, CPI_{UP}) es el mismo.

En las siguientes preguntas suponga que el CPI de la unidad de proceso (CPI_{UP}) es uno y la penalización por fallo son 10 ciclos.

Pregunta 4: Calcule los fallos por instrucción en los procesadores A y B.

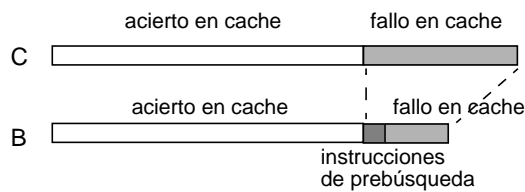
Pregunta 5: Calcule la reducción en fallos por instrucción del procesador A respecto del procesador B.

El cálculo de la ganancia mediante la ley de Amdahl utiliza los parámetros g y m , siendo m la fracción de tiempo, del caso sin mejora, donde se utiliza la mejora y g la ganancia en esta fracción de tiempo.

Pregunta 6: Para los procesadores A y B determine expresiones para los parámetros g y m en función de factores tales como: CPI medio (con jerarquía de memoria), CPI de la unidad de proceso

(acierto siempre en cache, CPI_{UP}), la frecuencia de fallo por instrucción, la penalización por fallo o el número de instrucciones. Corrobore el valor de la ganancia de la pregunta a).

Al comparar el procesador B y el procesador C hay que considerar que las instrucciones de prebúsqueda se incluyen en el tramo de tiempo donde se observa la mejora.



Suponga que el CPI del procesador C es 3.

Pregunta 7: Para los procesadores B y C determine expresiones para los parámetros g y m en función de factores tales como: CPI medio (con jerarquía de memoria), CPI de la unidad de proceso (acierto siempre en cache, CPI_{UP}), la frecuencia de fallo por instrucción, la penalización por fallo o el número de instrucciones.


Ejercicio 1.3

En un procesador A el CPI de las instrucciones de tipo multimedia es 2.5 y en una nueva generación del procesador (B), que funciona a la misma frecuencia, se consigue reducir el CPI de estas instrucciones a 1.5, ya que internamente la instrucción de lenguaje máquina multimedia se interpreta mediante una sola microinstrucción en lugar de las 2 microinstrucciones utilizadas en la versión A del procesador. En un conjunto de programas de prueba (P) las instrucciones multimedia representan un 20% de las ejecutadas. Al ejecutar este conjunto de programas de prueba en el procesador A se mide un CPI medio de 1.9. El CPI mínimo o ideal en los dos procesadores es 1.

Nota: los valores de CPI están calculados utilizando las instrucciones de lenguaje máquina.

Pregunta 1: Calcule la ganancia de reducir el CPI en las instrucciones multimedia.

Partiendo del procesador B se evalúa introducir la técnica multihilo para mejorar la productividad. Supondremos que la técnica multihilo es capaz de utilizar todos los ciclos perdidos, por



riesgos de datos y de secuenciamiento, al ejecutar un programa para ejecutar otro programa de forma concurrente. Esto es, se aprovechan todos los ciclos perdidos por riesgos de datos y de secuenciamiento. Además, no se incrementan los riesgos estructurales al utilizar la técnica multihilo.

En el procesador B, suponga que el tiempo de ejecución de otro conjunto de programas de prueba (P1) sin utilizar la técnica multihilo es 10 segundos, el CPI medido es 2.5 y los ciclos perdidos por instrucción por riesgos estructurales representan un 6% de los ciclos perdidos totales.

Pregunta 2: En el procesador B calcule el número de ciclos por instrucción perdidos por riesgos de datos y de secuenciamiento.

Pregunta 3: Calcule la latencia media de inicio al utilizar la técnica multihilo en el procesador B. ¿Cuál es la ganancia potencial en productividad al utilizar la técnica multihilo en el procesador B?

Pregunta 4: En 10 segundos cuantas veces se puede ejecutar el conjunto de programas de prueba P1 cuando se utiliza la técnica multihilo en el procesador B.

Suponga que, cuando se utiliza la técnica multihilo en el procesador B, el conjunto de programa de prueba tarda en ejecutarse 4.37 segundos.

Pregunta 5: Introducir la técnica multihilo representa incrementar la potencia consumida en un 10%. Calcule la ganancia energética potencial de introducir la técnica multihilo en el procesador B al ejecutar el conjunto de programas de prueba P1.

Partiendo del procesador B, sin la técnica multihilo, también se evalúan dos posibilidades para incrementar la productividad en una nueva versión tecnológica: a) utilizar los transistores disponibles para reducir el CPI en un 10% (versión C) o b) utilizar los transistores disponibles para crear un chip con 2 procesadores idénticos (versión D). En los dos casos la frecuencia se mejora un 40%.

Note que en la versión D cada procesador puede estar ejecutando uno de los programas del conjunto de programas de prueba. Suponga que todos los procesadores inician y finalizan el trabajo al mismo tiempo y que el CPI de los programas es el mismo que en la versión B (uniprosesador).

En el procesador B, sin utilizar la técnica multihilo, suponga que el tiempo de ejecución de otro conjunto de programas de prueba (P2) es 10 segundos y el CPI medido es 2.5.

Pregunta 6: Al ejecutar el conjunto de programas de prueba P2, calcule el tiempo de ejecución del conjunto de programas de prueba en la versión C.

Pregunta 7: Al ejecutar el conjunto de programas de prueba P2, calcule el tiempo de ejecución del conjunto de programas de prueba en la versión D.

En la nueva versión tecnológica del procesador el consumo de potencia de la versión C es 10/16 el consumo de potencia de la versión B. En cambio en la versión D el consumo de potencia se mantiene igual al de la versión B.


Pregunta 8: Al ejecutar el conjunto de programas de prueba P2, indique cuál de las versiones C o D es energéticamente mejor.

Ejercicio 1.4

Un procesador utiliza una jerarquía de memoria con un nivel de cache. En este nivel se utiliza una cache para instrucciones y otra cache para datos. Las dos cache utilizan mapeo 4 asociativo, el tamaño de bloque es de 16 bytes y el tamaño de cada cache es 32Kbytes. La penalización por fallo de cache son 10 ciclos. Las caches son bloqueantes.

En este procesador se ejecuta el siguiente código.

L.A.N	dirección	instrucción	direc.	instrucción
	1\$: 200	load R1, 0(R2)	232	add R6, R6, #8
do I = 1,100	204	load R3, 0(R4)	236	sub R7, R7, #1
A(I) = B(I) + C(I)	208	add R5, R1, R3	240	bne R7, 1\$
enddo	220	store R5, 0(R6)		
	224	add R2, R2, #8		
	228	add R4, R4, #8		



El 1º elemento de los vectores A, B y C está alineado a tamaño de bloque. El tamaño de un elemento de los vectores es de 8 bytes.

La 1ª instrucción del bucle está alineada a tamaño de bloque. El tamaño de una instrucción son 4 bytes.

Cuando empieza a ejecutarse el bucle las caches están vacías. Esto es, la información de estado de todos los contenedores indica bloque inválido.

Pregunta 1: Calcule los fallos que se producen en la cache de instrucciones al ejecutar el bucle.

Cuando se ejecuta el bucle y se accede en cache, tanto al acceder al código como a los datos se mide un CPI de 4 / 3.

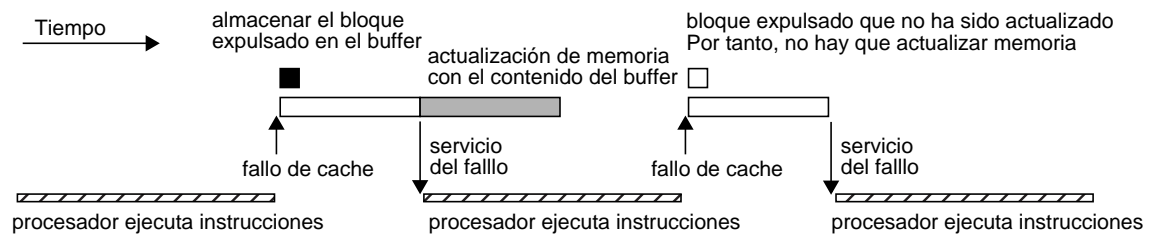
Pregunta 2: Calcule el número de ciclos que tarda en ejecutarse la 1ª iteración del bucle.

Pregunta 3: Calcule el número de ciclos que tarda en ejecutarse la 2ª iteración del bucle.

Pregunta 4: Calcule el número de ciclos que tarda en ejecutarse el bucle.

En la cache de datos se utiliza escritura retardada para mantener la coherencia en la jerarquía de memoria. Además, en caso de fallo en una operación de escritura se trae el bloque al 1º nivel de la jerarquía. Esto es, en una operación de escritura se actualiza el bloque en el 1º nivel de cache (si era fallo se ha traído antes) y el bloque se marca como actualizado. Posteriormente cuando el bloque debe expulsarse del contenedor para almacenar otro bloque (acción de reemplazo) se actualiza memoria.

El tiempo de actualizar memoria son 10 ciclos. Para no retardar el servicio del fallo de cache, el cual ha provocado la expulsión de un bloque que ha sido actualizado previamente, el bloque expulsado se almacena en un buffer y concurrentemente se inicia el servicio del fallo de cache (siguiente figura). Posteriormente cuando finaliza el servicio del fallo de cache se actualiza memoria con el contenido del buffer mientras concurrentemente el procesador sigue ejecutando instrucciones.



En las siguientes preguntas supondremos que todos los contenedores de cache almacenan bloques válidos y que los vectores A y B están almacenados en cache cuando se inicia la ejecución del bucle. Sólo se falla al acceder al vector C, lo cual en ocasiones requiere actualizar memoria.

Pregunta 5: Justifique si en el bucle previo el tiempo entre dos fallos consecutivos de cache es mayor que el tiempo necesario para actualizar memoria.

La potencia consumida por el procesador y el nivel de cache cuando no hay fallos es P julios/ciclo y cuando está bloqueado esperando el servicio de un fallo de cache es $0.2 \times P$. La potencia consumida cuando se accede a memoria (fallo o actualización de memoria) es $0.1 \times P$.

Pregunta 6: Suponga que un 30% de los bloques reemplazados al ejecutar el bucle han sido actualizados previamente y por tanto hay que actualizar memoria. Calcule la energía consumida al ejecutar el bucle.

Ejercicio 1.5

Sea un procesador de 10MIPS que efectúa 1.3 referencias a memoria por instrucción. La frecuencia de fallo de cache por referencia es 0.05 y la probabilidad de que el fallo se produzca en una línea que debe actualizarse en memoria principal es 0.5. Cada vez que se accede a memoria se ocupa el bus durante 300 ns que es el tiempo que se tarda en acceder a una línea de cache; durante este tiempo el procesador está parado.

Pregunta 1: Calcule la frecuencia de fallo por instrucción ejecutada (considerar el efecto del reemplazo como un fallo de cache).

Pregunta 2: Calcule el tiempo de procesador transcurrido entre accesos a memoria que producen fallo.

Pregunta 3: Calcule el número de transacciones por segundo en el bus.

Pregunta 4: Calcule los MIPS efectivos del procesador.

Pregunta 5: Calcule la utilización (ocupación) del bus por parte del procesador. Es decir, la probabilidad de que una instrucción utilice el bus.

Ejercicio 1.6

El tamaño del primer nivel de la cache de datos (L1D) de un procesador es 32Kbytes y la asociatividad es 4, siendo el tamaño de bloque 64 bytes. Respecto a la cache de instrucciones de primer nivel (L1I) supondremos que siempre se acierta. El segundo nivel de la jerarquía es una cache compartida (datos e instrucciones, L2). La penalización en fallo de L1D son 5 ciclos de procesador. Supondremos que siempre se acierta en L2 y que la cache L1D es bloqueante. Al computador con las características descritas lo denominamos A.

El tiempo de interpretación de cualquier instrucción es 1 ciclo a menos que se produzca fallo en cache.

Como programa de prueba utilizaremos el siguiente código. Al iniciarse la ejecución del programa de prueba ninguno de los elementos de los vectores A y B está almacenado en la cache L1D.

```
do K = 1 , M
    S = S + A(I) * B(I)
enddo
```

El tamaño de un elemento de los vectores A y B es 8 bytes.

R1 y R3 almacenan la dirección del primer elemento de los vectores A y B respectivamente. R8 almacena el contenido de la variable S. El valor de M es múltiplo de 8. Los vectores A y B están alineados a tamaño de bloque.

```
1$: load R2, 0(R1)
    load R4, 0(R3)
    add R1, R1, #8
    add R3, R3, #8
    add R5, R5, # -1
    mul R6, R2, R4
    add R8, R6, R8
    bne R5, 1$
```

Pregunta 1: Calcule los fallos totales que se producen al ejecutar el bucle, los fallos por instrucción, los accesos a memoria por instrucción y los fallos por acceso a memoria.

Pregunta 2: Muestre en un diagrama temporal las dos primeras iteraciones del programa de prueba. Para indicar la ejecución de varias instancias de una misma instrucción utilice la misma fila, indicando la ejecución en los ciclos correspondientes. Calcule los ciclos totales de penalización debido a fallos en L1D al ejecutar el bucle. Así mismo, calcule la fracción de tiempo que representa esta penalización en el tiempo de ejecución del programa.

El consumo de potencia cuando se acierta en la cache L1D es P vatios y cero cuando el procesador está bloqueado. El consumo de potencia de la cache L2 es un 15% del consumo del procesador.


Pregunta 3: Calcule la energía consumida al ejecutar el programa de prueba en el computador A.

En un nuevo diseño del computador, al que denominamos B, la cache es no bloqueante y la instrucción load tampoco es bloqueante. Así mismo, el procesador soporta varios fallos concurrentes. Además, el segundo nivel de la jerarquía de memoria (L2) tiene dos bancos que pueden servir accesos concurrentes. Los bloques de memoria se almacenan de forma entrelazada en los bancos. Para ello se utiliza el bit menos significativo de los bits de la dirección que identifican el bloque. Esto es, bloques consecutivos en memoria se almacenan en bancos distintos.

Suponga que el primer bloque de cada vector del programa de prueba se almacena en bancos distintos de L2.

Pregunta 4: Muestre en un diagrama temporal las dos primeras iteraciones del programa de prueba. Para indicar la ejecución de varias instancias de una misma instrucción utilice la misma fila, indicando la ejecución en los ciclos correspondientes. Calcule el tiempo de ejecución del programa de prueba en ciclos y la ganancia respecto al computador A.

Pregunta 5: El consumo de potencia de un banco de la cache L2 es un 15% del consumo del procesador. Muestre en un diagrama temporal la potencia consumida en cada ciclo durante las dos primeras iteraciones del programa de prueba. Calcule la energía



consumida al ejecutar el programa en el diseño B. ¿Cuál de los computadores A y B tiene la mejor métrica $MIPS^2$ /Potencia media?