

Arquitectura de Computadores



Los Del DGIIM, losdeldgiim.github.io

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0).

Eres libre de compartir y redistribuir el contenido de esta obra en cualquier medio o formato, siempre y cuando des el crédito adecuado a los autores originales y no persigas fines comerciales.

Arquitectura de Computadores

Los Del DGIIM, losdeldgiim.github.io

Arturo Olivares Martos

Granada, 2023-2024

Índice general

1. Relaciones de Problemas	5
1.1. Arquitecturas Paralelas: Clasificación y Prestaciones	5
1.2. Programación paralela	9

1. Relaciones de Problemas

1.1. Arquitecturas Paralelas: Clasificación y Prestaciones

Ejercicio 1.1.1. En el código de prueba (benchmark) que ejecuta un procesador no segmentado que funciona a 300 MHz, hay un 20 % de instrucciones **LOAD** que necesitan 4 ciclos, un 10 % de instrucciones **STORE** que necesitan 3 ciclos, un 25 % de instrucciones con operaciones de enteros que necesitan 6 ciclos, un 15 % de instrucciones con operandos en coma flotante que necesitan 8 ciclos por instrucción, y un 30 % de instrucciones de salto que necesitan 3 ciclos.

1. ¿Cuál es la ganancia que se puede obtener por reducción a 3 ciclos de las instrucciones con enteros?
2. ¿Cuál es la ganancia que se puede obtener por reducción a 3 ciclos de las instrucciones en coma flotante?

Ejercicio 1.1.2. Un circuito que implementaba una operación en un tiempo de $T_{op} = 450$ ns se ha segmentado mediante un cauce lineal con cuatro etapas de duración $T_1 = 100$ ns, $T_2 = 125$ ns, $T_3 = 125$ ns y $T_4 = 100$ ns respectivamente, separadas por un registro de acoplo que introduce un retardo de 25 ns.

1. ¿Cuál es la máxima ganancia de velocidad posible? ¿Cuál es la productividad máxima del cauce?

Tenemos que el ciclo de reloj es de $T_c = 125ns + 25ns = 150ns$. La ganancia de velocidad es:

$$S(N) = \frac{T^b(N)}{T^s(N)} = \frac{N \cdot T_R}{TLI + (N - 1)T_c} = \frac{N \cdot T_R}{4 \cdot T_c + (N - 1)T_c}$$

donde el 4 se debe a que es el número de etapas del cauce.

La ganancia máxima es:

$$S(N \gg) = \lim_{N \rightarrow \infty} S(N) = \lim_{N \rightarrow \infty} \frac{N \cdot T_R}{4 \cdot T_c + (N - 1)T_c} = \frac{T_R}{T_c}$$

2. ¿A partir de qué número de operaciones ejecutadas se consigue una productividad igual al 90 % de la productividad máxima?

Ejercicio 1.1.3. En un procesador sin segmentación de cauce, determine cuál de estas dos alternativas para realizar un salto condicional es mejor:

- ALT1: Una instrucción **COMPARE** actualiza un código de condición y es seguida por una instrucción **BRANCH** que comprueba esa condición. Se usan dos instrucciones.
- ALT2: Una sola instrucción incluye la funcionalidad de las instrucciones **COMPARE** y **BRANCH**. Se usa una única instrucción.

Hay que tener en cuenta que hay un 20 % de instrucciones **BRANCH** para ALT1 en el conjunto de programas de prueba; que las instrucciones **BRANCH** en ALT1 y **COMPARE+BRANCH** en ALT2 necesitan 4 ciclos mientras que todas las demás necesitan sólo 3; y que el ciclo de reloj de la ALT1 es un 25 % menor que el de la ALT2, dado que en este caso la mayor funcionalidad de la instrucción **COMPARE+BRANCH** ocasiona una mayor complejidad en el procesador.

Ejercicio 1.1.4. ¿Qué ocurriría en el problema del ejercicio anterior (Ejercicio 1.1.3) si el ciclo de reloj fuese únicamente un 10 % mayor para la ALT2?

Ejercicio 1.1.5. Considere un procesador no segmentado con una arquitectura de tipo **LOAD/STORE** en la que las operaciones sólo utilizan como operandos registros de la CPU. Para un conjunto de programas representativos de su actividad se tiene que el 43 % de las instrucciones son operaciones con la ALU (3 CPI), el 21 % **LOADs** (4 CPI), el 12 % **STOREs** (4 CPI) y el 24 % **BRANCHs** (4 CPI). Se ha podido comprobar que un 25 % de las operaciones con la ALU utilizan operandos en registros que no se vuelven a utilizar. Compruebe si mejorarían las prestaciones si, para sustituir ese 25 % de operaciones, se añaden instrucciones con un dato en un registro y otro en memoria. Tengan en cuenta en la comprobación que para estas nuevas instrucciones el valor de CPI es 4 y que añadirlas ocasiona un incremento de un ciclo en el CPI de los **BRANCH**, pero no afectan al ciclo de reloj.

Ejercicio 1.1.6. Se ha diseñado un compilador para la máquina **LOAD/STORE** del problema anterior (Ejercicio 1.1.5). Ese compilador puede reducir en un 50 % el número de operaciones con la ALU, pero no reduce el número de **LOADs**, **STOREs**, y **BRANCHs**. Suponiendo que la frecuencia de reloj es de 50 Mhz, ¿Cuál es el número de MIPS y el tiempo de ejecución que se consigue con el código optimizado? Compárelos con los correspondientes del código no optimizado.

Ejercicio 1.1.7. En un programa que se ejecutan en un procesador no segmentado que funciona a 100 MHz, hay un 20 % de instrucciones **LOAD** que necesitan 4 ciclos, un 15 % de instrucciones **STORE** que necesitan 3 ciclos, un 40 % de instrucciones con operaciones en la ALU que necesitan 6 ciclos, y un 25 % de instrucciones de salto que necesitan 3 ciclos.

1. Si en las instrucciones que usan la ALU el tiempo en la ALU supone 4 ciclos, determine cuál es la máxima ganancia que se puede obtener si se mejora el diseño de la ALU de forma que se reduce su tiempo de ejecución a la mitad de ciclos.
2. Con qué porcentaje de instrucciones con operaciones en la ALU se podría haber obtenido en los cálculos del apartado una ganancia mayor que 2? Razone su respuesta.

Ejercicio 1.1.8. Suponga que en los programas que constituyen la carga de trabajo habitual de un procesador las instrucciones de coma flotante consumen un promedio del 13 % del tiempo del procesador.

1. Ha aparecido en el mercado una nueva versión del procesador en la que la única mejora con respecto a la versión anterior es una nueva unidad de coma flotante que permite reducir el tiempo de las instrucciones de coma flotante a tres cuartas partes del tiempo que consumían antes. ¿Cuál es la máxima ganancia de velocidad que puede esperarse en los programas que constituyen la carga de trabajo si se utiliza la nueva versión del procesador?
2. ¿Cuál es la máxima ganancia de velocidad con respecto a la versión inicial del procesador que, en promedio, puede esperarse en los programas debido a mejoras en la velocidad de las operaciones en coma flotante?
3. ¿Cuál debería ser el porcentaje de tiempo de cálculo con datos en coma flotante en los programas para esperar una ganancia máxima de 4 en lugar de la obtenida en el apartado 2?
4. ¿Cuánto debería reducirse el tiempo de las operaciones en coma flotante con respecto a la situación inicial para que la ganancia máxima sea 2 suponiendo que en la versión inicial el porcentaje de tiempo de cálculo con coma flotante es el obtenido en el apartado 3?

Ejercicio 1.1.9. Suponga que, en el código siguiente, `a[]` es un array de números de 32 bits en coma flotante y `b` un número de 32 bits en coma flotante y que debería ejecutarse en menos de 0,5 segundos para $N = 10^9$:

```
for (i=0; i<N; i++)
    a[i+2]=(a[i+2]+a[i+1]+a[i])*b;
```

1. ¿Cuántos GFLOPS se necesitan para poder ejecutar el código en menos de 0,5 segundos?
2. Suponiendo que este código en ensamblador tiene $7N$ instrucciones y que se ha ejecutado en un procesador de 32 bits a 2 GHz. ¿Cuál es el número medio de instrucciones que el procesador tiene que poder completar por ciclo para poder ejecutar el código en menos de 0,5 segundos?
3. Estimando que el programa pasa el 75 % de su tiempo de ejecución realizando operaciones en coma flotante, ¿cuánto disminuiría como mucho el tiempo de ejecución si se redujesen un 75 % los tiempos de las unidades de coma flotante?

Ejercicio 1.1.10. Un compilador ha generado un código máquina optimizado para el siguiente programa

```
par=0; impar=0;
for (i=0; i<N; i++)
    if ((i%2) == 0)
        par=par+c*x[i];
    else
        impar=impar-c*x[i];
```

sin utilizar instrucciones de salto dentro de las iteraciones del bucle (porque se ha usado la técnica de desenrollado el bucle que veremos en el Seminario 4): el código tiene un número de iteraciones de $N/2$, 7 instrucciones fuera del bucle (2 de almacenamiento en memoria, 5 instrucciones para inicializar registros), 9 instrucciones dentro del bucle (4 instrucciones para implementar el bucle for: incremento de la variable de control i , comparación, salto condicional y un salto incondicional; 4 instrucciones coma flotante y 2 instrucciones de carga desde memoria a registro -se leen dos componentes de x). El computador donde se ejecuta dispone de:

- Un procesador superescalar de 32 bits a 2 GHz capaz de terminar dos instrucciones de coma flotante por ciclo y dos instrucciones de cualquier otro tipo por ciclo, excepto instrucciones de carga, cuyo tiempo depende de si hay o no fallo de cache (si no hay fallo de cache suponen 1 ciclo), y las instrucciones de almacenamiento que suponen 1 ciclo.
- Dos caches integradas en el chip de procesamiento (una para datos y otra para instrucciones) de 512 KBytes cada una, mapeo directo, política de actualización de postescritura, líneas de 32 bytes, y latencia de un ciclo de reloj.
- Una memoria principal con latencia de 30 ns. y ciclos burst 6-1-1-1 a través de un bus de memoria de 200 MHz con 64 bits.

Conteste a las siguientes cuestiones:

1. ¿Cuál es la velocidad pico del procesador (en GFLOPS)?
2. ¿Cuál es el tiempo mínimo que tarda en ejecutarse el programa para $N = 211$?
3. ¿Cuántos MFLOPS alcanza el programa?

Observación. Considere que el vector x se almacena en memoria en una dirección múltiplo del tamaño de una línea de cache y que ningún componente está en cache cuando se referencia; N , i estarán en registros de enteros, `par`, `impar`, `c`, y `x[]` son números de 32 bits en coma flotante; dentro del bucle `c`, `par` e `impar` estarán en registros.

Cuestión 1.1.1. Indique cuál es la diferencia fundamental entre una arquitectura CC-NUMA y una arquitectura SMP.

Cuestión 1.1.2. ¿Cuándo diría que un computador es un multiprocesador y cuándo que es un multicomputador?

Cuestión 1.1.3. ¿Un CC-NUMA escala más que un SMP? ¿Por qué?

Cuestión 1.1.4. Indique qué niveles de paralelismo implícito en una aplicación puede aprovechar un PC con un procesador de 4 cores, teniendo en cuenta que cada core tiene unidades funcionales SIMD (también llamadas unidades multimedia) y una microarquitectura segmentada y superscalar. Razone su respuesta.

Cuestión 1.1.5. Si le dicen que un ordenador es de 20 GIPS ¿puede estar seguro que ejecutará cualquier programa de 20000 instrucciones en un microsegundo?

Cuestión 1.1.6. ¿Aceptaría financiar/embarcarse en un proyecto en el que se plantease el diseño e implementación de un computador de propósito general con arquitectura MISD? (Justifique su respuesta).

Cuestión 1.1.7. Deduzca la expresión que se usa para representar la ley de Amdahl suponiendo que se mejora un recurso del procesador, que hay una probabilidad f de no utilizar dicho recurso y que la mejora supone un incremento en un factor de p de la velocidad de procesamiento del recurso.

Cuestión 1.1.8. ¿Es cierto que si se mejora una parte de un sistema (por ejemplo, un recurso de un procesador) se observa experimentalmente que, al aumentar el factor de mejora, llega un momento en que se satura el incremento de velocidad que se consigue? (Justifique la respuesta)

Cuestión 1.1.9. ¿Es cierto que la cota para el incremento de velocidad que establece la ley de Amdahl crece a medida que aumenta el valor del factor de mejora aplicado al recurso o parte del sistema que se mejora? (Justifique la respuesta).

Cuestión 1.1.10. ¿Qué podría ser mejor suponiendo velocidades pico, un procesador superescalar capaz de emitir cuatro instrucciones por ciclo, o un procesador vectorial cuyo repertorio permite codificar 8 operaciones por instrucción y emite una instrucción por ciclo? (Justifique su respuesta).

Cuestión 1.1.11. En la Lección 2 de AC se han presentado diferentes criterios de clasificación de computadores y en el Seminario 0 de prácticas se ha presentado atcgrid. Clasifique atcgrid, sus nodos, sus encapsulados y sus núcleos dentro de la clasificación de Flynn y dentro de la clasificación que usa como criterio el sistema de memoria. Razone su respuesta.

Cuestión 1.1.12. En la Lección 1 de AC se han presentado diferentes criterios de clasificación del paralelismo implícito en una aplicación y en el Seminario 0 de prácticas se ha presentado atcgrid. ¿Qué tipos de paralelismo aprovecha atcgrid? Razone su respuesta.

1.2. Programación paralela