

Arquitectura de los Computadores. Primera convocatoria 2018

1. (3 puntos) Sobre rendimiento

La empresa (Wireless Multimedia Sensor Networks) está diseñando microprocesadores especializados (Wireless Multimedia Sensor Processors) (WMSP) y software especializado, para integrarlos en pequeños sensores inteligentes multimedia que implementan protocolos de comunicación y funciones de compresión/descompresión. Estas funciones de compresión/descompresión pueden ser ejecutadas, siguiendo el modelo SIMD, mediante GPUs integradas en los WMSP. Se han diseñado diferentes modelos del sistema WMSP que integran diferentes modelos de GPU (GPU 1,2,3,4) con diferentes prestaciones. Se ha realizado un estudio de las partes del software del sistema WMSP que son paralelizables mediante modelo SIMD y ejecutables en las GPUs integradas y se han realizado pruebas de ejecución en los diferentes sistemas. En la imagen se observan los tiempos de ejecución del sistema completo, incluyendo la parte “no paralelizable” ejecutada en la CPU y la “paralelizable” SIMD ejecutada en las GPUs.

	No Paralelizable CPU	Paralelizable
CPU	20s	40s
GPU 1	20s	10s
GPU 2	20s	2s
GPU 3	20s	0,67s
GPU 4	20s	0,15 s

a) (1 punto) Calcula la aceleración global y la aceleración mejorada respecto a la opción CPU para cada una de las GPUs

	Tiempo no paralelizable	Tiempo paralelizable	Tiempo total = Tiempo no paralelizable + Tiempo paralelizable	A mejorada respecto a la opción CPU de cada una de las GPUs= Tiempo paralelizable CPU / Tiempo paralelizable GPUs	Aceleración global respecto a la opción CPU de cada una de las GPUs= Tiempo total CPU / Tiempo total GPUs
CPU	20,00	40,00	60,00	40/40	1,000
GPU1	20,00	10,00	30,00	40/10	2,000
GPU2	20,00	2,00	22,00	40/2	2,727
GPU3	20,00	0,67	20,67	40/0,67	2,903
GPU4	20,00	0,15	20,15	40/0,15	2,978

Ejercicio inspirado en la página 30 del tema 2 de teoría

b) (1 punto) El equipo de diseño hardware ha comprobado que es capaz de mejorar la “GPU 1” reduciendo a la mitad los tiempos de GPU (GPU 1 mejorada). Se desea saber si el equipo de desarrollo software puede incrementar el porcentaje de paralelización (fracción mejorada), para mejorar el rendimiento manteniendo la versión actual de la “GPU 1”. ¿Qué incremento en el porcentaje de paralelización se necesitará para obtener la misma ganancia de rendimiento que el equipo hardware?

	No Paralelizable CPU	Paralelizable
CPU	20s	40s
GPU1	20s	10s
GPU1 mejorada por el equipo hardware	20s	5s

La aceleración global que puede conseguir el equipo hardware es:

$$A_{g \text{ hardware}} = \frac{20 + 40}{20 + 5} = \frac{60}{25} = 2,4$$

El equipo de desarrollo software utiliza la “GPU1” no mejorada, cuya aceleración mejorada es:

$$A_{\text{mejorada}} = \frac{40}{10} = 4$$

El equipo de desarrollo software se plantea la misma aceleración global que el equipo hardware “2,4”, utilizando la “GPU1” no mejorada cuya aceleración mejorada es “4”, incrementando el porcentaje de paralelización o fracción mejorada:

$$A_{g \text{ hardware}} = 2,4 = \frac{1}{(1 - f_{m \text{ software}}) + \frac{f_{m \text{ software}}}{a_m}} = \frac{1}{(1 - f_{m \text{ software}}) + \frac{f_{m \text{ software}}}{4}} \rightarrow f_{m \text{ software}} = 0,78$$

La fracción mejorada inicial (siempre definida sobre la opción peor) es:

$$f_{\text{mejorada inicial}} = \frac{\text{Tiempo paralelizable CPU}}{\text{Tiempo total CPU}} = \frac{40}{40 + 20} = \frac{40}{60} = 0,67$$

El incremento en el porcentaje de paralelización (fracción mejorada) para obtener la misma ganancia de rendimiento que el equipo hardware (2,4) es:

$$\Delta_{f_m} = |f_{m \text{ software}} - f_{m \text{ inicial}}| = |0,78 - 0,67| = 0,11$$

Ejercicio inspirado en la página 40 del tema 2 de teoría

- c) (1 punto) La empresa WMSN está planteando optimizar el rendimiento de la CPU de su WMSP. Se ha realizado un estudio de la mezcla de instrucciones de la CPU ejecutando el software del sistema. El estudio refleja los siguientes resultados de frecuencias de instrucciones y CPIs:

Instrucción	Frecuencia	CPI	CPI nuevo
ALU	45%	1	1
LOAD	12%	2	2
STORE	24%	2	2
SALTO	14%	1	1
PORT	5%	3	4

La CPU utiliza la instrucción PORT para escribir en los puertos de comunicación. Esta instrucción PORT debe ejecutar siempre una instrucciones STORE previa para almacenar en memoria los datos que finalmente se escriben al puerto de comunicación. La empresa está pensando en realizar una modificación para que PORT escriba directamente los datos en los puertos de comunicación, sin necesidad de realizar antes una STORE. Esto permitiría transformar parejas de instrucciones STORE/PORT en la nueva instrucción PORT.

STORE Reg, Mem → PORT Reg
 PORT Mem

Supongamos que esta modificación del repertorio de instrucciones incrementa en 1 el número de ciclos de reloj para la instrucción PORT, pero sin afectar a la duración del ciclo de reloj. **Calcula la aceleración de la versión supuestamente mejorada respecto a la anterior.**

Tiempo de ejecución arquitectura A:

$$T_{ejecución A} = RI_A * CPI_A * clk_A$$

$$CPI_A = 0,45 * 1 + 0,12 * 2 + 0,24 * 2 + 0,14 * 1 + 0,05 * 3 = 1,46$$

$$T_{ejecución A} = RI_A * 1,46 * clk_A$$

Tiempo de ejecución arquitectura mejorada B:

En la arquitectura B, dado que la nueva instrucción PORT escribe directamente los datos en los puertos de comunicación, sin necesidad de realizar antes una STORE, es posible eliminar las instrucciones STORE asociadas a instrucciones PORT. Es decir las parejas STORE; PORT de la arquitectura A se rempazan por PORTn. En esta nueva situación la mezcla de instrucciones y CPIs sería forma:

Instrucción	Frecuencia	CPI	Frecuencia modificada	Frecuencia sobre 100%	CPI
ALU	45%	1	45%	47,4%	1
LOAD	12%	2	12%	12,6%	2
STORE	24%	2	24% - 5% = 19%	20%	2
SALTO	14%	1	14%	14,7%	1
PORT	5%	3	0%	0%	0
PORTn	0%	0	5%	5,3%	4
	100%		100% - 5% = 95%	100%	

$$T_{ejecución B} = RI_B * CPI_B * clk_B$$

$$CPI_B = \frac{0,45 * 1 + 0,12 * 2 + 0,19 * 2 + 0,14 * 1 + 0,05 * 4}{1 - 0,05} = 0,474 * 1 + 0,126 * 2 + 0,2 * 2 + 0,147 * 1 + 0,053 * 4 = 1,48$$

$$clk_B = clk_A$$

$$RI_B = 0,95 * RI_A$$

$$T_{ejecución B} = RI_B * CPI_B * clk_B = 0,95 * RI_A * 1,48 * clk_A = 1,41 * RI_A * clk_A$$

$$A = \frac{T_{ejecución A}}{T_{ejecución B}} = \frac{1,46 * RI_A * clk_A}{1,41 * RI_A * clk_A} = \frac{1,46}{1,41} = 1,0354$$

Ejercicio inspirado en la página 75 del tema 2 de teoría