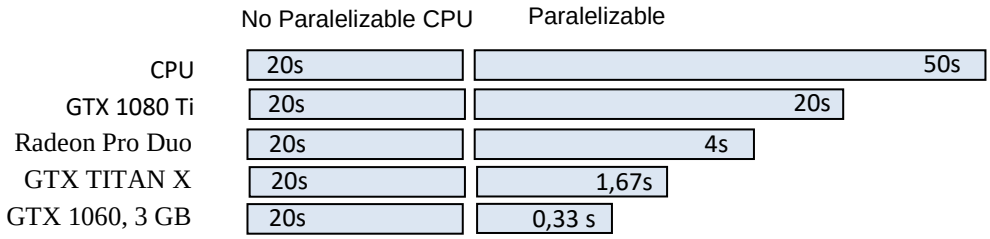
Nombre:_____

Arquitectura de los Computadores. Primera convocatoria 2017

1. **(2 puntos)** La empresa CVApps (Computer Vision Applications) está diseñando un sistema de guiado de vehículos mediante metodologías basadas en "deep learning", cuyos requerimientos de rendimiento vienen impuestos por las altas velocidades de los vehículos. Se ha hecho un estudio de las partes del código del sistema de guiado que son paralelizables y ejecutables en GPUs. El código paralelizable se ha ejecutado en modelos de GPU con diferentes prestaciones:

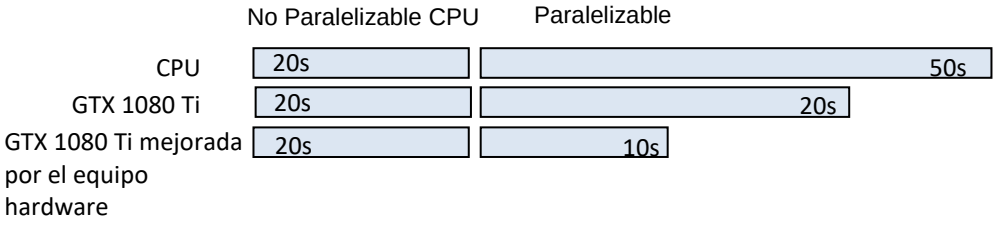


a) (1 punto) Calcula la aceleración global y la aceleración mejorada respecto a la opción CPU para cada una de las GPUs

	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	50,00	70,00	50/50	1,000	70/70	1,000
GTX 1080 Ti	20,00	20,00	40,00	50/20	**2,500**	70/40	**1,750**
Radeon Pro Duo	20,00	4,00	24,00	50/4	**12,500**	70/24	**2,917**
GTX TITAN X	20,00	1,67	21,67	50/1,67	**29,940**	70/21,67	**3,230**
GTX 1060, 3 GB	20,00	0,33	20,33	50/0,33	**151,515**	70/20,33	**3,443**

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 "GTX 1080 Ti" reduciendo a la mitad los tiempos de GPU. 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 "GTX 1080 Ti". ¿Qué incremento en el porcentaje de paralelización se necesitará para obtener la misma ganancia de rendimiento que el equipo hardware?



La aceleración global que puede conseguir el equipo hardware es:

$$A_{g\ hardware} = \frac{50+20}{20+10} = \frac{70}{30} = 2{,}33$$

El equipo de desarrollo software utiliza la "GTX 1080 Ti" no mejorada, cuya aceleración mejorada es:

$$A_{mejorada} = \frac{50}{20} = 2{,}5$$

El equipo de desarrollo software se plantea la misma aceleración global que el equipo hardware "2,33", utilizando la "GTX 1080 Ti" no mejorada cuya aceleración mejorada es "2,5", incrementando el porcentaje de paralelización o fracción mejorada:

$$A_{g\ hardware} = 2{,}33 = \frac{1}{(1 - f_{m\ software}) + \frac{f_{m\ software}}{a_m}} = \frac{1}{(1 - f_{m\ software}) + \frac{f_{m\ software}}{2{,}5}} \rightarrow f_{m\ software} = 0{,}95$$

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

$$f_{mejorada\ inicial} = \frac{Tiempo\ paralelizable\ CPU}{Tiempo\ total\ CPU} = \frac{50}{50+20} = \frac{50}{70} = 0{,}7143$$

El incremento en el porcentaje de paralelización (fracción mejorada) para obtener la misma ganancia de rendimiento que el equipo hardware (2,33) es:

$$\Delta_{fm} = |f_{m\ software} - f_{m\ inicial}| = |0{,}95 - 0{,}7143| = 0{,}238$$

Ejercicio inspirado en la página 40 del tema 2 de teoría

Nombre:_____

2. (1,5 puntos) La misma empresa CVApps (Computer Vision Applications) está diseñando un procesador especializado para un sistema de visualización realista, donde es necesario un alto rendimiento en aplicaciones de generación de gráficos. En este procesador y ejecutando estas aplicaciones la mezcla de instrucciones y CPIs son:

Instrucción	Frecuencia	CPI
ALU	40%	1
LOAD	11%	2
STORE	29%	2
SALTO	15%	1
IMP	5%	4

La máquina, debe realizar siempre instrucciones STORE para almacenar los datos que utiliza la instrucción IMP para imprimir en pantalla. La empresa está pensando en realizar una modificación para que IMP cargue directamente los datos a imprimir en pantalla, sin necesidad de realizar antes una STORE. Supongamos que este repertorio extendido de instrucciones incrementa en 1 el número de ciclos de reloj para la instrucción IMP, pero sin afectar a la duración del ciclo de reloj.

a) Calcula la aceleración de la versión supuestamente mejorada respecto a la anterior

Tiempo de ejecución arquitectura A:

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

$$CPI_A = 0{,}4 * 1 + 0{,}11 * 2 + 0{,}29 * 2 + 0{,}15 * 1 + 0{,}05 * 4 = 1{,}55$$

$$T_{ejecución\ A} = RI_A * 1{,}55 * clk_A$$

Tiempo de ejecución arquitectura mejorada B:

En la arquitectura B, dado que la nueva instrucción IMPn carga directamente los datos a imprimir en pantalla sin necesidad de realizar antes una STORE, es posible eliminar las instrucciones STORE asociadas a instrucciones IMP. Es decir las parejas STORE; IMP de la arquitectura A se remplazan por IMPn. En esta nueva situación la mezcla de instrucciones y CPIs queda de la siguiente forma:

Instrucción	Frecuencia	CPI	Frecuencia modificada	Frecuencia sobre 100%	CPI
ALU	40%	1	40%	42,1%	1
LOAD	11%	2	11%	11,6%	2
STORE	29%	2	29% - 5% = 24%	25,3%	2
SALTO	15%	1	15%	15,8%	1
IMP	5%	4	0%	0%	0
IMPn	0%	0	5%	5,3%	5
	100%		100% - 5% = 95%	100%	

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

$$CPI_B = \frac{0{,}4 * 1 + 0{,}11 * 2 + 0{,}24 * 2 + 0{,}15 * 1 + 0{,}05 * 5}{1 - 0{,}05} = 0{,}421 * 1 + 0{,}116 * 2 + 0{,}253 * 2 + 0{,}158 * 1 + 0{,}053 * 5 = 1{,}58$$

$$clk_B = clk_A$$

$$RI_B = 0{,}95 * RI_A$$

$$T_{ejecución\ B} = RI_B * CPI_B * clk_B = 0{,}95 * RI_A * 1{,}58 * clk_A = 1{,}5 * RI_A * clk_A$$

$$A = \frac{T_{ejecución\ A}}{T_{ejecución\ B}} = \frac{1{,}55 * RI_A * clk_A}{1{,}5 * RI_A * clk_A} = \frac{1{,}55}{1{,}5} = 1{,}0333$$

Ejercicio inspirado en la página 75 del tema 2 de teoría