

# Curso 2022-2023

PRAC06: Versión CUDA del Proyecto

28/11/2022

Prácticas de Programación Concurrente y Paralela

Pablo Revuelta Sanz – José Ranilla

**Tipo de Práctica:** Abierta

**Entrega:** 21 diciembre 2022

**Duración:** 2 sesiones

## 1. Introducción

Última práctica de la asignatura, cuyo objetivo es resolver el proyecto planteado mediante una versión SIMT para procesadores gráficos compatibles CUDA. Es una práctica con un peso específico importante en la nota de la asignatura.

## 2. La Práctica

Partiendo de los códigos suministrados y del conocimiento adquirido por el alumnado en las prácticas anteriores, resolver el problema planteado en *PRAC04* usando GPUs, es decir, implementando *kernels* para calcular el fractal, promedio y binarizado. Para el cálculo del promedio se plantean dos posibilidades:

- 1) Implementar un *kernel* que calcule el promedio, es decir, una operación de reducción.
- 2) Usar funciones de la API de CUDA (concretamente *cublasDgemv*).

Al objeto de poder modificar el número de hilos por bloque sin necesidad de recompilar, se debe añadir un nuevo argumento a la línea de comandos del script de python *Fractal.py*, que ahora se llamará *FractalGPU.py*. Esto es, cambiar la forma de ejecutar de:

*Fractal.py xmin xmax ymin yres maxiter imagen-salida*

a:

*FractalGPU.py xmin xmax ymin yres maxiter hilos-por-bloque imagen-salida*

En el proceso de evaluación se usará dicha nueva sintaxis para ejecutar las prácticas de los alumnos.

En */opt/PracticasPCP22\_23/Practica06/* del servidor *di119.edv.uniovi.es* el alumnado dispone de un fichero comprimido (de nombre *PRAC06.tgz*) con lo imprescindible para iniciar el desarrollo de la práctica (funciones, prototipos, *Makefile*, etc.). En esta ocasión, también se provee un *shellscript* para lanzar correctamente los trabajos a las colas.

## 3. El entorno de Producción

Al igual que en la práctica anterior, los nodos de cálculo tienen una GPU NVIDIA GeForce GTX 1660 Ti, cada una de ellas con 6 GB GDDR6, 1536 Cuda cores, 5.5 TFLOPs y Cuda Compute Capability 7.5

## 4. Enfoques y Alternativas

El alumnado dispone de amplias opciones:

- 1) Homogéneo vs Heterogéneo.
- 2) Uso de memoria convencional, unificada o *pinned*.
- 3) Uso, o no, de memoria *shared*.
- 4) Esquemas de bloques 1D, 2D o 3D.
- 5) Promedio calculado a mano (y distintas estrategias posibles).
- 6) Etc.

El alumnado puede elegir la combinación que estime oportuna, desde la más sencilla, práctica *baseline* para obtener una nota “media” siempre y cuando la documentación esté a la altura y los programas ejecuten correctamente, hasta implementar todas las opciones/variantes, lo que conllevará una nota incrementada proporcionalmente a la complejidad de lo resuelto.

## 5. Experimentación.

Finalizada la codificación se debe ejecutar, usando SGE con el *wrapper* **ColaGPU** y computación en GPU, y completar con los tiempos de la función *mandel* la Tabla 1, para distintas resoluciones. **No grabar nunca imágenes en la experimentación.** Las posiciones sobre las que calcular el fractal, así como *maxiter* deben ser los mismos que en la PRAC04.

yres	Tiempo opción 1	Tiempo opción 2	...
1024			
2048			
4096			
8192			
10240			

**Tabla 1** Experimentación a realizar con la GPU.

Si se han implementado correctamente varios enfoques (ver sección 4 de este documento) se deben rellenar tantas columnas de la Tabla 1 como enfoques implementados.

## 6. ¿Qué y dónde entregar?

### 6.1. Códigos

Los ficheros del alumnado deben estar en la carpeta ***\$HOME/PRAC06***. Se copiará el fichero *mandelGPU.cu*, en el que deben estar todas las implementaciones realizadas (tanto funciones *extern* como los propios kernels). Si hay más de una función para resolver el mismo problema, las demás deberán llamarse de otra manera para permitir la compilación con el *Makefile* dado.

**El código fuente (en *\$HOME/PRAC06*) será copiado automáticamente a las 23:55:00 horas del día 21 de diciembre de 2022.**

### 6.2. Autocorrección

Debido a que el problema planteado es, por definición caótico (variaciones infinitesimales de las condiciones de partida pueden generar cambios cualitativos en el estado de salida), se provee de una nueva librería compartida, *mandelProfGPU.so*, que resuelve el mismo problema en GPU. Si se intenta comparar la solución de los kernels implementados con lo codificado en

*python* o en C, aparecerán errores debidos a la implementación de los tipos *double* en CPU y en CUDA. Por ello, se recomienda comparar siempre los resultados obtenidos con la nueva librería dada.

### 6.3. Documentación

La documentación se debe subir al Campus Virtual como un único fichero en formato PDF. Esta documentación constará de 2 partes: a) La primera, la correspondiente a la documentación de la práctica PRAC04 y b) la segunda, la correspondiente a la práctica PRAC06.

**Fecha límite para subir la documentación al CV: 23:55:00 horas del 21 de diciembre de 2022.**  
**La tarea del CV a utilizar es: Actividades de Evaluación → Prácticas 4 y 6.**

El alumnado debe seguir las siguientes indicaciones a la hora de articular su documentación, respetando el orden siguiente:

- **Práctica PRAC04.** Dar respuesta a todas las cuestiones planteadas en la sección 6 del enunciado de dicha práctica, es decir: a) análisis teórico de la complejidad temporal y espacial, tanto secuencial como paralela, calculando, al menos, la eficiencia, para cada solución propuesta; b) explicar por qué ha seleccionado los mecanismos OpenMP usados en cada función; c) Las tablas correspondientes a los experimentos de PRAC04; d) comparar el plano teórico con el empírico de cada solución (secuencial y paralela), entorno de ejecución y solución; e) todas aquellas comparativas adicionales que cada cual considere convenientes; f) conclusiones.
- **Práctica PRAC06.** Esta práctica empezará explicando los aspectos más relevantes del diseño/implementación realizado/a (cálculo del número de bloques, usos de memoria utilizados, mecanismo de uso de *shared* -si aplicable-, gestión de carga si heterogéneo...) para cada solución desarrollada. A continuación, una hoja con la Tabla 1 de este guion. Luego, la comparativa entre los distintos enfoques GPU implementados, si es que existen. Seguidamente, la comparación entre los tiempos de GPU con los de CPU, tanto secuencial como paralela. Calcular la aceleración que ha supuesto el uso de GPUs (contra la mejor implementación en CPU). Todas aquellas aportaciones adicionales que el alumnado considere conveniente y las conclusiones.