

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de Ingeniería

Introducción a Cómputo Paralelo con CUDA
C/C++

LABORATORIO DE INTEL Y CÓMPUTO DE ALTO
DESEMPEÑO

Elaboran: Ariel Ulloa Trejo
Jaime Beltrán Rosales
Revisión: Ing. Laura Sandoval Montaña

Temario

1. Antecedentes
2. El GPU
3. Modelo de programación
CUDA
4. Manejo de matrices
5. Memoria compartida



2 El GPU

Arquitectura

FPU o ALU: Es la unidad que ejecuta las operaciones sobre los datos. El FPU está optimizado para hacer operaciones de coma flotante.

Unidad de control: Permite la sincronización y correcto funcionamiento de cada FPU o ALU. Puede tener uno o varios.

Memoria interna: Al igual que en un procesador, las tarjetas gráficas cuentan con diferentes niveles de memoria:

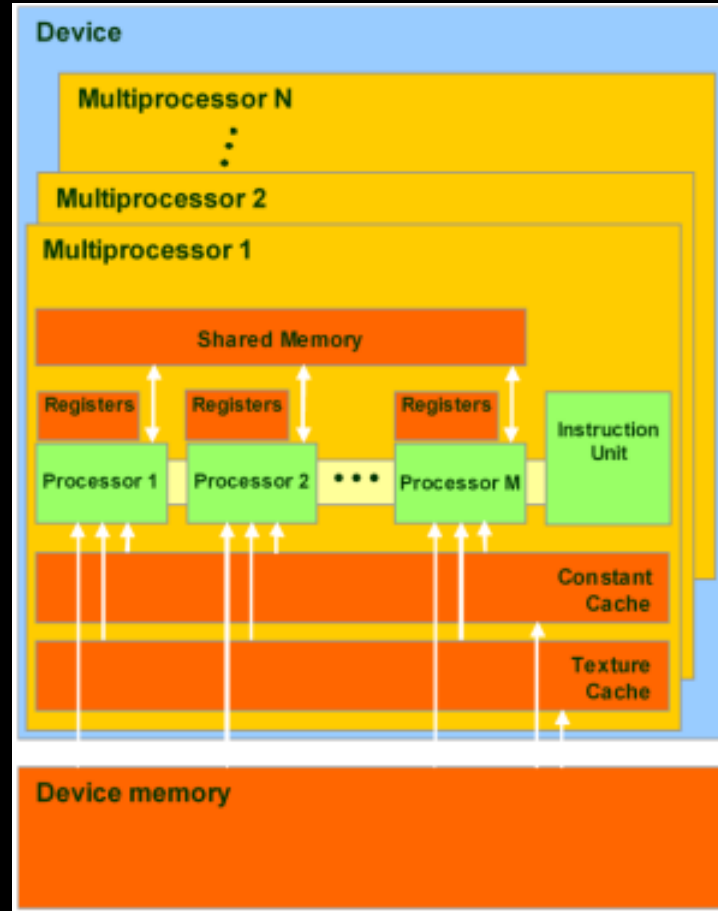
A nivel ALU o FPU: Registros.

A nivel bloque: Un grupo determinado de unidades de procesamiento pueden acceder a este nivel de memoria.

A nivel general: Todas las unidades de procesamiento y sus respectivos bloques tienen acceso a este espacio de memoria.

¿No se parece al CPU?

2 EI GPU



2 El GPU

CPU

- Propósito general.
- Pocos núcleos.
- Alta frecuencia de reloj.
- Gran cantidad de memoria caché.
- MIMD

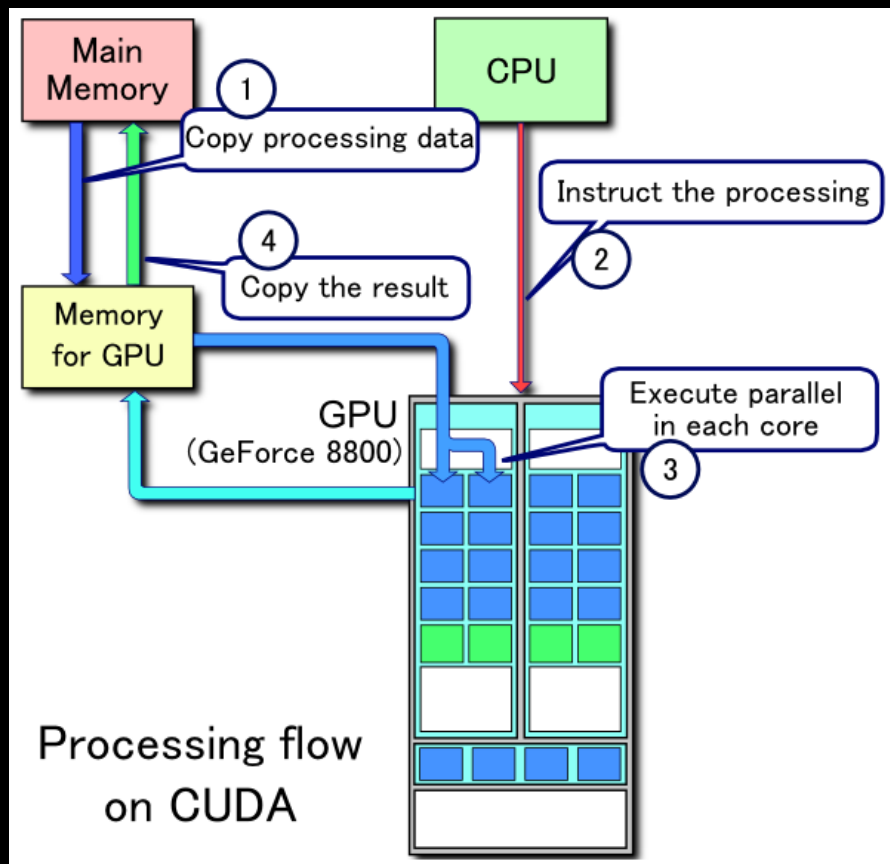
GPU

- Propósito específico (operaciones de coma flotante).
- Miles de núcleos.
- Menor frecuencia de reloj.
- Ideal para tareas repetitivas e independientes.
- SIMD

2 El GPU

Ejecución

1. Copiar datos a procesar.
2. indicar instrucciones.
3. Ejecutar en paralelo en cada núcleo.
4. Copiar el resultado.



2 EI GPU

