

PROGRAMA DE ASIGNATURA

“Computación paralela en GPU”

1. INFORMACIÓN GENERAL

1.1 Nombre de la asignatura	: Computación paralela en GPU
1.2 Código	: <i>por definir</i>
1.3 Créditos	: 3
1.4 Periodo académico	: II semestre
1.5 Tipo de asignatura	: electiva
1.6 Horas Teóricas	: 1
1.7 Horas Prácticas	: 1
1.8 Horas no presenciales	: 8
1.8 Cupo	: 20
1.9: Prerrequisitos	: INFO335 (HPC)
1.10 Prof. Responsable	: Cristóbal Navarro
1.11 Prof. (es) Colaborador(es)	: Hector Ferrada

2. DESCRIPCIÓN DE LA ASIGNATURA

La asignatura busca profundizar los conceptos modernos de la computación paralela por GPU. Se verán los principales aspectos sobre eficiencia que se deben tener en consideración al momento de diseñar y programar un algoritmo paralelo en GPU. La asignatura también contempla revisar y utilizar las últimas tecnologías asociadas al modelo de programación en GPU, tales como *tensor cores*, *ray-tracing cores* y *dynamic parallelism*.

3. OBJETIVOS

Ser capaz de diseñar e implementar un algoritmo para GPU usando el modelo de programación que entrega CUDA y OpenCL y las últimas tecnologías provenientes de los núcleos de propósito específico, siempre teniendo en cuenta los aspectos que son críticos en el rendimiento de todo algoritmo GPU.

4. CONTENIDOS

Unidad 1: Introducción al modelo de programación masivamente paralelo:

- Kernel, Grid, Bloques, Hilos
- Saxpy
- Multiplicación de Matrices
- *Coalesced Memory*
- Reducción paralela con shuffle

Unidad 2: Conceptos Avanzados sobre GPU Computing

- Shared Memory
- Streams de ejecución asíncronos
- Kernels Concurrentes

Unidad 3: Conceptos avanzados sobre GPU Computing

- Dynamic Parallelism.
- Mapeo de Hilos en espacios complejos.
- *Tensor Cores* y *RT cores*.
- Concepto de *Mixed Precisión*.
- Multi-GPU.
- Cooperative Threads.

5. METODOLOGÍA DEL TRABAJO

El curso sigue la metodología de tutoría de investigación. El desarrollo de los contenidos se realizará con reuniones cada 14 días, en las que se dedicará una tarde para cubrir un tópico. Gran porcentaje del curso es trabajo de exploración/investigación de los estudiantes, en los que ellos buscarán un tópico en particular y tratarán de dominarlo para presentarlo a sus compañeros.

6. EVALUACIÓN

La nota se calculará en función del trabajo y avances mostraron a lo largo de todas las reuniones.

7. BIBLIOGRAFÍA

- CUDA Programming Guide, Nvidia, 2018.
- C.A. Navarro, N. Hitschfeld, L. Mateu, A survey on parallel computing and its applications in data-parallel problems using GPU architectures, Comm. Comp. Physics, 2014.