



FaMAF | **GPGPU**
Computing Group



Santificarás la atomicidad

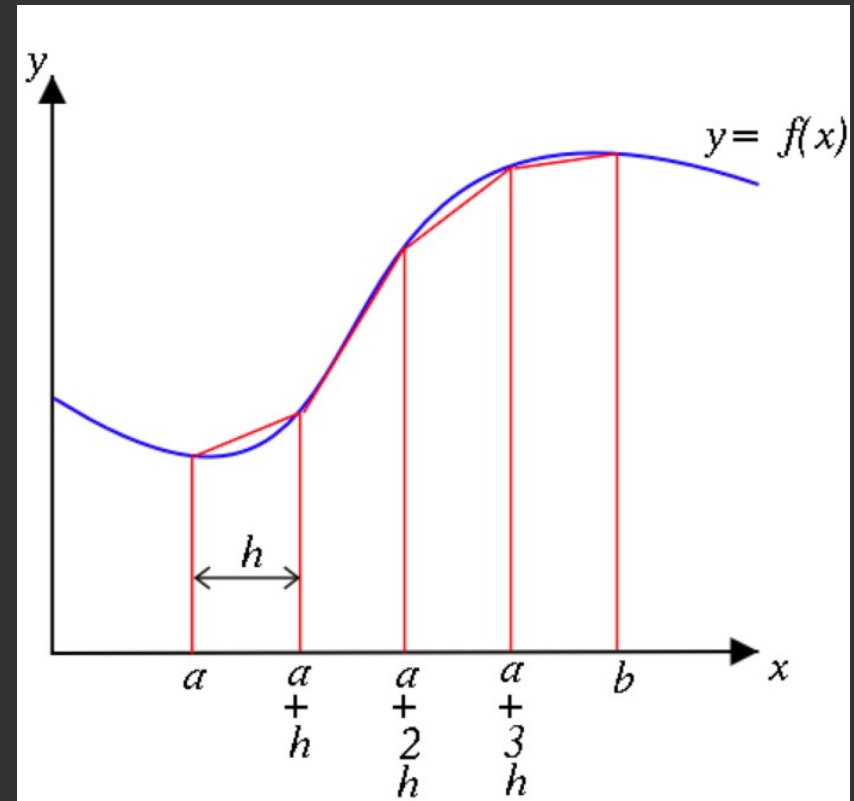
Nicolás Wolovick

Intentemos mejorar Simpson

Computar los $f(a)$, $f(a+h)$... y luego tirar todo al tacho no tiene sentido.

Dos kernels secuenciales:

- $f(a+i*h)$
- Sumar los datos



Simpson mejorado

1E6 intervalos, primero CPU y luego GPU.

Medimos **walltime**.

Miramos **resultados**.

Verificamos **corrección**.

```
nicolasw@ufo:~/MiniCursoGPGPU/Simpson$ ./simpson_cuda
WallTime secs: 0.282380
WallTime secs: 0.113798
2.00000000000, 0.0001376200
simpson_cuda: simpson_cuda.cu:71: int main(int, char**): Assertion `cpu==gpu'
failed.
Aborted
```

Simpson ¿mejorado?

Problemas varios

- Velocidad
- Resultados no-deterministas
- Corrección

¿Alguna idea?

Vamos a las bases

$\{x=0\}$

$x:=x+1$

$x:=x+1$

$\{x=2\}$

¿Cuánto vale x cuando terminan los hilos?

Vamos a las bases, y al duro metal

$\{x=0\}$

$a1:=x$	$a2:=x$
$inc(a1)$	$inc(a2)$
$x:=a1$	$x:=a2$

$\{1 \leq x \leq 2\}$

¿Cuánto vale x cuando terminan los hilos?

Program Topology en CUDA

Lanzo 2^{20} hilos divididos en bloques de 2^8 threads.

Todos incrementan la variable compartida x.

¿Cuanto valdrá x?

```
nicolasw@ufo:~/MiniCursoGPGPU/NonAtomicInc$ ./non_atomic_inc
Usage: ./non_atomic_inc <2^i threads> <2^j threadsPerBlock> [<2^k
blocksPerGrid>]
nicolasw@ufo:~/MiniCursoGPGPU/NonAtomicInc$ ./non_atomic_inc 20 8
Secs: 0.001420
GMops: 5.907470
GFlops: 0.000000
X: 54
```

SIMD, ma non troppo: SIMT

- Extremadamente SIMD
- De $x=1048576$ pasamos a $x=54+/-4$
- Aclaremos que pasaba con nuestro Simpson
- Soluciones:
 - AtomicInc: nope, solo enteros (en CUDA 3.0)
 - Butterfly-like summing up, $\log(N)$ fases.

Problemas de performance

Última aplicación: Poor Man's FoF.

- Tengo puntos en R^3 y quiero saber cuantos caen dentro de un radio ε del origen.
- Mismo problema del incremento atómico, solución: `atomicAdd(&(hits->x),1);`

Diferencias de Performance

```
nicolasw@ufo:~/MiniCursoGPGPU/FoFExplore$ ./fof
Usage: ./fof <2^i points> <2^j threadsPerBlock> [<2^k blocksPerGrid>]
nicolasw@ufo:~/MiniCursoGPGPU/FoFExplore$ ./fof 24 9
Secs: 0.005769
GMops: 58.163342
GFlops: 17.449003
Hits: 1571
```

```
nicolasw@ufo:~/MiniCursoGPGPU/FoFExplore$ ./fof 24 9
Secs: 0.018236
GMops: 18.400105
GFlops: 5.520032
Hits: 70229
```

Resumen

- Con millones de hilos se magnifican los problemas de:
 - Concurrencia (corrección)
 - Contención (velocidad)
- Juan Pablo y Ezequiel, van a mostrar como se **acumula** de manera correcta y rápida en la última clase.