

# Autotuning on Heterogeneous Systems

Adrián Pérez Diéguez

VIII Seminario de Invierno CAPAP-H

Elche - 30 de Enero de 2017



# Líneas de investigación HPC

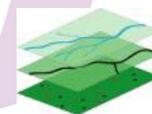
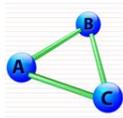


**Cluster/Cloud Computing**

**Middleware de  
Comunicación para Java**

**Desarrollo y optimización de  
aplicaciones distribuidas**

**Sistemas de Información  
Geográfica (GIS), Visualización  
LiDAR, GPUs**



## Heterogeneous Programming Library (HPL)

- Librería sobre OpenCL orientada a ...
  - *Facilitar la programación sobre OpenCL*

- Interfaz intuitivo
- Evita la inicialización, compilaciones, manipulación de buffers y transferencias, etc.
- Integración kernels en aplicación host
- Soporte de múltiples dispositivos
- Overhead reducido (<3%) sobre OpenCL

## Heterogeneous Programming Library (HPL)

- Librería sobre OpenCL orientada a ...
  - *Facilitar la portabilidad de rendimiento*

- Generación en runtime de los kernels OpenCL mediante el lenguaje embebido en C++
- Permite generar el código adaptándolo al dispositivo, características de los datos, etc.
- Ejemplo: se superó en rendimiento a cBLAS y ViennaCL
- En desarrollo: optimizador just-in-time para códigos HPL

- Disponible con documentación en <http://hpl.des.udc.es> y repositorio en GitHub

## Butterfly Processing Library for GPUs (BPLG)

Librería desarrollada en CUDA y C++

Resuelve: *Sistemas Tridiagonales, Scan, Sorting, FFT* y otras transformadas

Optimizada para diferentes arquitecturas  
*NVIDIA: Fermi, Kepler, Maxwell*

Engloba entornos GPU, Multi-GPU y Multi-Nodo

## Butterfly Processing Library for GPUs (BPLG)

Sigue una metodología de optimización en 2 fases:

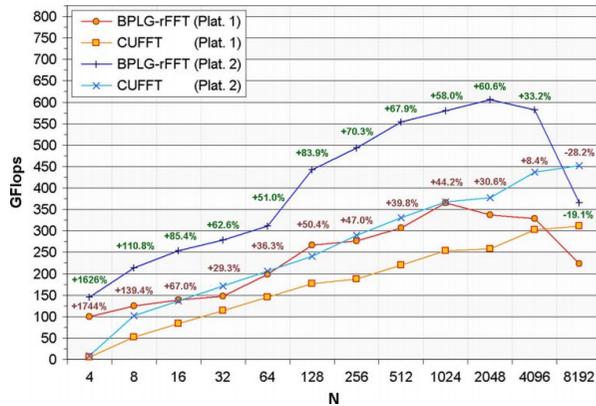
Obtención de los parámetros de rendimiento de la GPU basado en 'Premisas'

*Building blocks* para implementar los kernels:

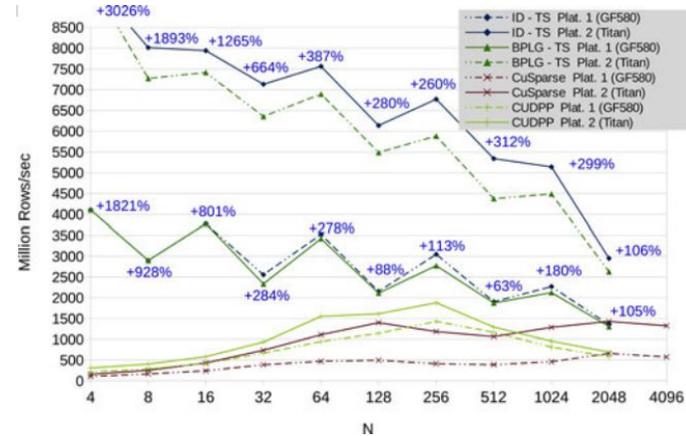
- *Template metaprogramming*
- *Generic Programming*
- *Compile-time optimizations*
- *Operations in any memory space hierarchy*

Web: <http://bplg.des.udc.es>

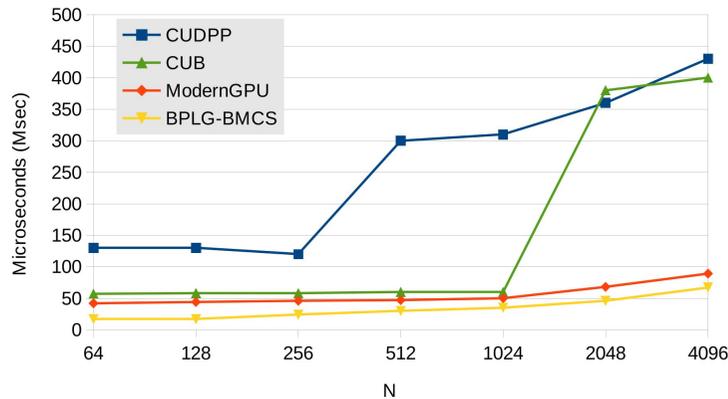
# Butterfly Processing Library for GPUs (BPLG)



*BPLG: A Tuned Butterfly Processing Library for GPU Architectures.* J. Lobeiras, M. Amor, R. Doallo. *International Journal of Parallel Programming*. 43(1), pp 1078 - 1102. 2015



*Designing Efficient Index-Digit Algorithms for CUDA GPU Architectures.* J. Lobeiras, M. Amor, R. Doallo. *IEEE Transactions on Parallel and Distributed Systems*. pp 1331 - 1343. 2016



*BPLG-BMCS: A GPU-Sorting Algorithm using Tuning Skeleton Library.* A. P. Diéguez, M. Amor, R. Doallo. *The Journal of Supercomputing*. 73(1), pp 4 - 16. 2015

