NVIDIA CUDA RESEARCH CENTER

APLICACIONES Y OPORTUNIDADES

Pedro Antonio Varo Herrero

Estudiante de Master MULCIA

Universidad de Sevilla

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

Contenido

- Bibliotecas Científicas
- Aplicaciones en ciencia e ingeniería
- Lenguajes de programación

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

HPC en vuestras investigaciones:

- **Big-data**.
- **OPTIMIZACIÓN SISTEMAS BIOINSPIRADOS** ...
- **Grafos**: "Representaciones ortogonales de grafos, generación masiva de grafos y cálculo de parámetros, filogenética computacional.
- Python de problemas de **electromagnetismo computacional.**
- **Procesamiento de Lenguaje Natural**
- **Aprendizaje automático**.
- **Spiking neural P systems** (y variantes) para la simulacion.
- **Procesado de imagen y video.**
- **Diseño de circuitos** electrónicos y en microelectrónica analógica. Frecuentemente requiero de recursos computacionales muy altos **en tareas de optimización y de simulación eléctrica."**
- **Aplicaciones a la robótica.** De momento se ha hecho uso de la nube.
- **Simulación por ordenador y teoría de fluidos complejos.**
- **Computación evolutiva**
- **simulaciones del comportamiento a fatiga de componentes mecánicos** mediante **modelos de elementos finitos.**
- **En problemas de electroconvección en líquidos.** Estoy interesado en hacer **simulaciones en 3D** que requieren mucha potencia de cálculo
- **Computación multiagentes**
- **Redes complejas, optimización y simulación estocástica.**
- **Proyectos de secuenciación y en comparación 3D de biomoléculas**
- **Procesamiento de imágenes, aprendizaje automático, gráficos 3D.**
- **Dinámica molecular**
- **Algoritmos de procesado de imagen y video**. Particularmente, video de alto rango dinámico (HDR) en tiempo real.
- **Sistemas embebidos**.
- **Simulación de propiedades magnéticas de materiales.**
- **Problemas de optimización con funciones de caja negra** (no convexas, no lineales) con restricciones no convexas y gran número de variables continuas.
- **Realizamos cálculo no lineales** de sistemas de varios millones de grados de libertad, para el **análisis de estructuras históricas.**
- **Información topológico algebraica global de imágenes médicas de resonancia mágnetica funcional 3D+t**
- **Calculos DFT en sistemas sólidos**, con condiciones periódicas de periodicidad. Usamos onda planas como funciones de base, lo que requiere que **parte del cálculo dependa de transformadas de Fourier 3D**. El programa que usamos es **VASP**.

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

- Bibliotecas Científicas

https://developer.nvidia.com/gpu-accelerated-libraries

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

- Bibliotecas Científicas

https://developer.nvidia.com/gpu-accelerated-libraries

- **Métodos numéricos**
- **Operaciones algebraicas**
- **Redes Neuronales**
- **Transformadas de Fourier**
- **Operaciones con señales e imágenes**
- **Soporte para LAPACK y BLAS**
- **Generación aleatoria de números**
- **Visualización en tiempo real de simulaciones**
- **Geometría computacional**
- **Análisis de secuencias de ADN**

Bibliotecas Cientificas

AmgX

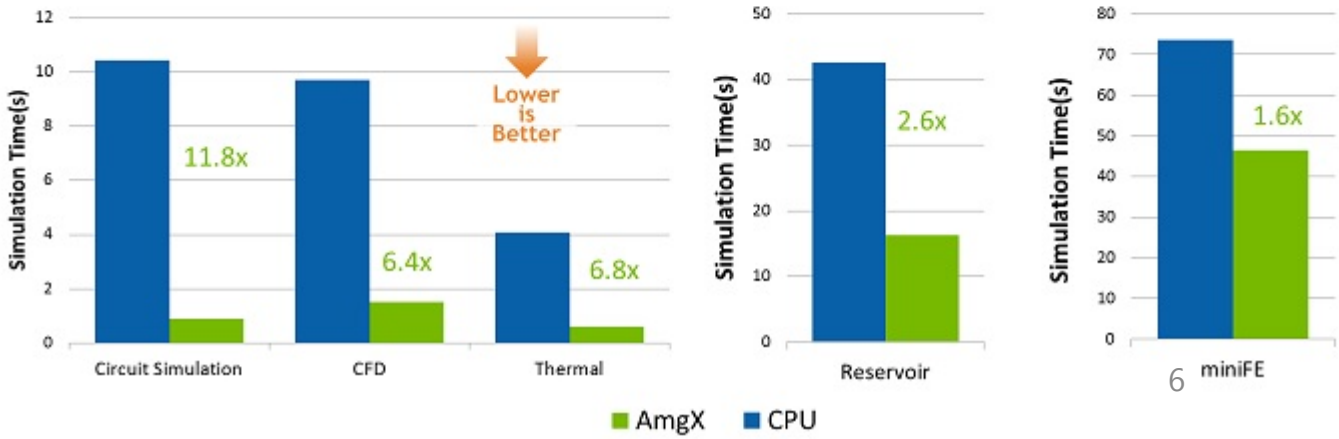
A simple path to accelerated core solvers, providing up to 10x acceleration in the computationally intense linear solver portion of simulations, and is very well suited for implicit unstructured methods.

https://developer.nvidia.com/amgx

Key Features

- Flexible configuration allows for nested solvers, smoothers, and preconditioners
- Ruge-Steuben algebraic multigrid
- Un-smoothed aggregation algebraic multigrid
- Krylov methods: PCG, GMRES, BiCGStab, and flexible variants
- Smoothers: Block-Jacobi, Gauss-Seidel, incomplete LU, Polynomial, dense LU
- Scalar or coupled block systems
- MPI support
- OpenMP support
- Flexible and simple high level C API

The AmgX Performance Benefit



Bibliotecas Cientificas

cuDNN

NVIDIA cuDNN is a GPU-accelerated library of primitives for deep neural networks, it is designed to be integrated into higher-level machine learning frameworks.

https://developer.nvidia.com/cuDNN

Key Features

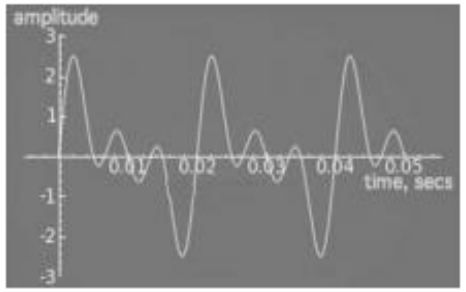
- Forward and backward convolution routines designed for convolutional neural nets, tuned for NVIDIA GPUs
- Always optimized for latest NVIDIA GPU architectures
- Arbitrary dimension ordering, striding, and subregions for 4d tensors means easy integration into any neural net implementation
- Forward and backward paths for many other common layer types (ReLU, Sigmoid, Tanh, pooling, softmax)
- Context-based API allows for easy multithreading

Get Started with cuDNN Today

cuDNN provides tuned implementations of routines frequently used in deep neural network applications, including:

- Convolution forward and backward, including cross-correlation
- Pooling forward and backward (Maximum and Average)
- Softmax forward and backward
- Neuron activations forward and backward (Rectified Linear, Sigmoid, Hyperbolic Tangent)
- Tensor transformation functions

Bibliotecas Cientificas



cuFFT

NVIDIA CUDA Fast Fourier Transform Library (cuFFT) provides a simple interface for computing FFTs up to 10x faster, without having to develop your own custom GPU FFT implementation.

Key Features

- 1D, 2D, 3D transforms of complex and real data types
- 1D transform sizes up to 128 million elements
- Flexible data layouts by allowing arbitrary strides between individual elements and array dimensions
- FFT algorithms based on Cooley-Tukey and Bluestein
- Familiar API similar to FFTW Advanced Interface
- Streamed asynchronous execution
- Single and double precision transforms
- Batch execution for doing multiple transforms
- In-place and out-of-place transforms
- Flexible input & output data layouts, similar tp FFTW "Advanced Interface"
- Thread-safe & callable from multiple host threads

https://developer.nvidia.com/cufft

Bibliotecas Cientificas

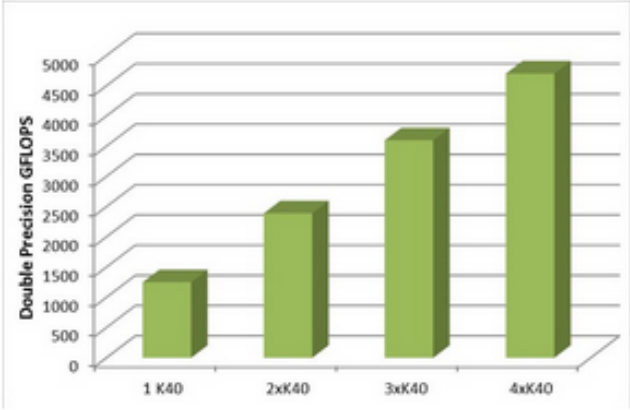
cuBLAS-XT

cuBLAS-XT is a set of routines which accelerate Level 3 BLAS (Basic Linear Algebra Subroutine) calls by spreading work across more than one GPU.

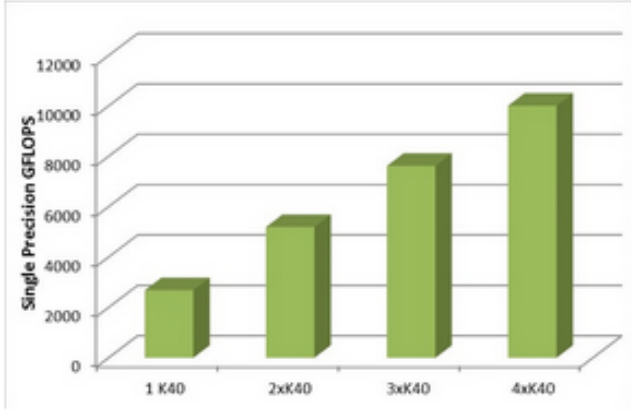
https://developer.nvidia.com/cublasxt

Performance

Review the latest CUDA 6.5 performance report to learn how much you could accelerate your code.



Performance of cuBLAS-XT DGEMM for a 16,384-by-16,384 matrix.



Performance of cuBLAS-XT SGEMM for a 16,384-by-16,384 matrix,.

Performance measured on 1-4 K40 cards with ECC enabled, connected via PCI-E Gen 3 to a dual-socket Intel(R) Xeon(R) CPU E5-2650@ 2.00GHz.

Bibliotecas Cientificas

NPP

NVIDIA Performance Primitives is a GPU accelerated library with a very large collection of 1000's of image processing primitives and signal processing primitives.

https://developer.nvidia.com/npp

Key Features

- **Eliminates unnecessary copying of data to/from CPU memory**
  - Process data that is already in GPU memory
  - Leave results in GPU memory so they are ready for subsequent processing
- **Data Exchange and Initialization**
  - Set, Convert, Copy, CopyConstBorder, Transpose, SwapChannels
- **Arithmetic and Logical Operations**
  - Add, Sub, Mul, Div, AbsDiff, Threshold, Compare
- **Color Conversion**
  - RGBToYCbCr, YcbCrToRGB, YCbCrToYCbCr, ColorTwist, LUT_Linear
- **Filter Functions**
  - FilterBox, Filter, FilterRow, FilterColumn, FilterMax, FilterMin, Dilate, Erode, SumWindowColumn, SumWindowRow
- **JPEG**
  - DCTQuantInv, DCTQuantFwd, QuantizationTableJPEG
- **Geometry Transforms**
  - Mirror, WarpAffine, WarpAffineBack, WarpAffineQuad, WarpPerspective, WarpPerspectiveBack , WarpPerspectiveQuad, Resize
- **Statistics Functions**
  - Mean_StdDev, NormDiff, Sum, MinMax, HistogramEven, RectStdDev

Bibliotecas Cientificas

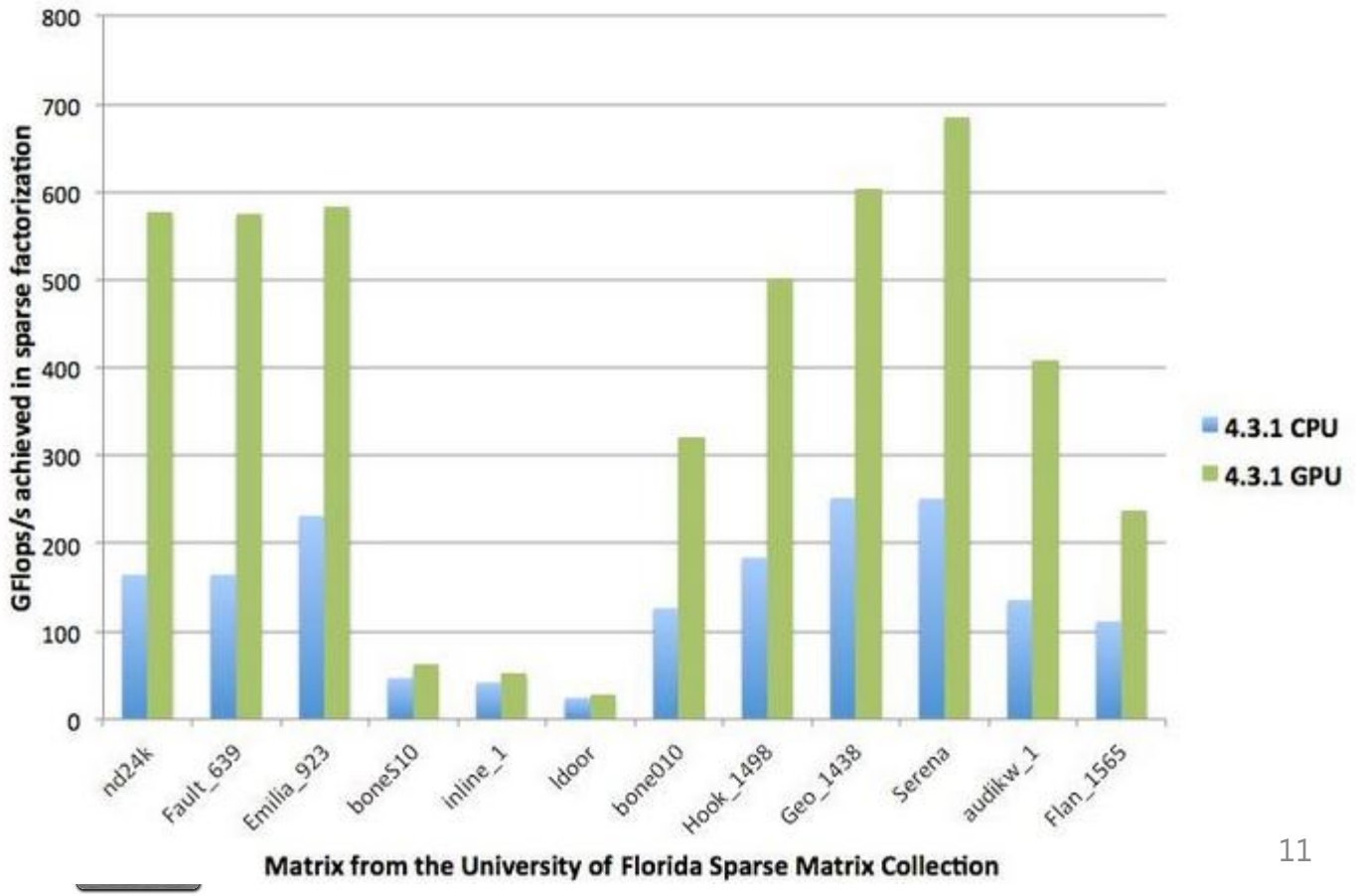
CHOLMOD

GPU-accelerated CHOLMOD is part of the SuiteSparse linear algebra package by Prof. Tim Davis. SuiteSparse is used extensively throughout industry and academia.

https://developer.nvidia.com/cholmod

Testing was performed using:

- CPU: Dual-socket Xeon E5-2690 v2 (Ivy Bridge) @ 3.00 Ghz.
- GPU: Nvidia Tesla K40m with ECC off and clocks set at full boost (3004,875)
- SuiteSparse compiled with Intel Composer XE 2013, Metis 4.0 and NVIDIA CUDA 6.0.



Bibliotecas Cientificas

MAGMA

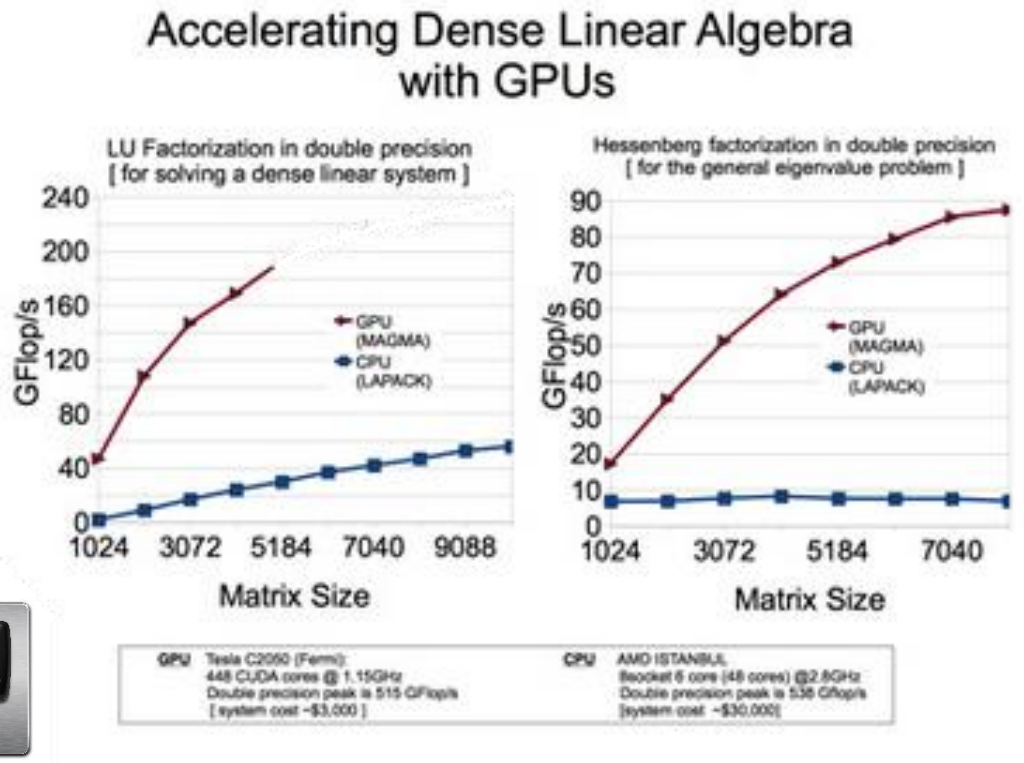
ICL UT

MAGMA

A collection of next gen linear algebra routines. Designed for heterogeneous GPU-based architectures. Supports current LAPACK and BLAS standards.

Key Features

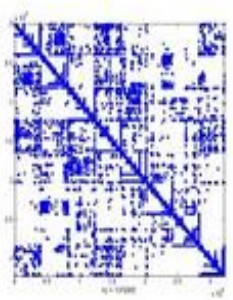
- Excellent performance and high accuracy (LAPACK compliant)
- Multiple precision arithmetic support (S/D/C/Z)
- Hybrid algorithms using both multicore CPUs and GPUs



https://developer.nvidia.com/magma

NVIDIA CUDA RESEARCH CENTER

Bibliotecas Cientificas



cuSPARSE

NVIDIA CUDA Sparse (cuSPARSE) Matrix library provides a collection of basic linear algebra subroutines used for sparse matrices that delivers over 8x performance boost.

Key Features

- Supports dense, COO, CSR, CSC, ELL/HYB and Blocked CSR sparse matrix formats
- Level 1 routines for sparse vector x dense vector operations
- Level 2 routines for sparse matrix x dense vector operations
- Level 3 routines for sparse matrix x multiple dense vectors (tall matrix)
- Routines for sparse matrix by sparse matrix addition and multiplication
- Conversion routines that allow conversion between different matrix formats
- Sparse Triangular Solve
- Tri-diagonal solver
- Incomplete factorization preconditioners ilu0 and ic0

https://developer.nvidia.com/cusparse

Bibliotecas Cientificas

ArrayFire

Comprehensive GPU function library, including functions for math, signal and image processing, statistics, and more. With interfaces for C, C++, Java, R and Fortran."

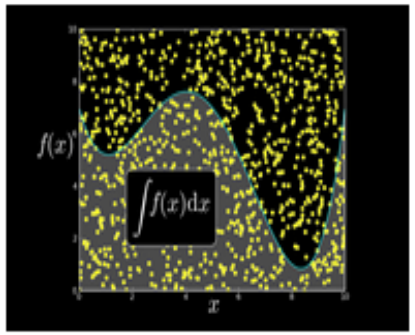
https://developer.nvidia.com/arrayfire

Key Features

- It contains excellent GPU implementations of hundreds of matrix, signal, and image processing routines that enable it outperform CPU libraries like IPP, MKL, Eigen, Armadillo, and more.
- It is optimized for any CUDA-enabled GPU. The same code will run on laptops, desktops, or servers.
- It includes thousands of lines of highly-tuned device code.
- It performs run-time analysis of your code to increase arithmetic intensity and memory throughput while avoiding unnecessary temporary allocations.
- It combines and enhances all the best CUDA libraries available, including the fastest FFT, BLAS, and LAPACK implementations.
- A simple array notation you can learn in minutes.
- A few lines of ArrayFire code accomplishes what would have taken 10-100X lines in raw CUDA.
- It is easier than templated programming and goes farther than simple directive-based approaches (and outperforms those approaches too).
- It supports easily scaling to take advantage of multiple GPUs.
- It can be used in C/C++ applications by itself or integrated with your existing CUDA code.
- It has hundreds of functions you need to make your code faster including arithmetic, linear algebra, statistics, signal processing, image processing, and related algorithms (see more).
- It supports single and double-precision floating point values, complex numbers, and booleans (see more).
- It supports manipulating vectors, matrices, and N-dimensional arrays (see more).
- It can execute loop iterations in parallel with gfor (see more).

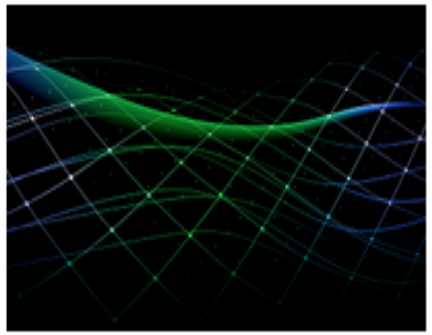
Developers have experienced from 2X to 100X speedups, depending on the data-parallelism inherent in the application.

Bibliotecas Cientificas



cuRAND

The CUDA Random Number Generation library performs high quality GPU-accelerated random number generation (RNG) over 8x faster than typical CPU only code.

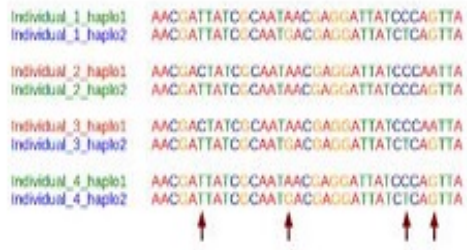


CUDA Math Library

An industry proven, highly accurate collection of standard mathematical functions, providing high performance on NVIDIA GPUs.

Thrust

A powerful, open source library of parallel algorithms and data structures. Perform GPU-accelerated sort, scan, transform, and reductions with just a few lines of code.



NVBIO

A GPU-accelerated C++ framework for High-Throughput Sequence Analysis for both short and long read alignment.

Bibliotecas Cientificas

Sundog Software

Triton Ocean SDK

Triton provides real-time visual simulation of the ocean and bodies of water for games, simulation, and training applications.

NVIDIA VIDEO CODEC SDK

Accelerate video performance with this complete set of NVIDIA video codec tools, which includes the NVENC H.264 hardware encoding API as well as NVCUVID CUDA decoding API.

[HiPLAR]

HiPLAR

HiPLAR (High Performance Linear Algebra in R) delivers high performance linear algebra (LA) routines for the R platform for statistical computing using the latest software libraries for heterogeneous architectures.

OpenCV

OpenCV

OpenCV is the leading open source library for computer vision, image processing and machine learning, and now features GPU acceleration for real-time operation.

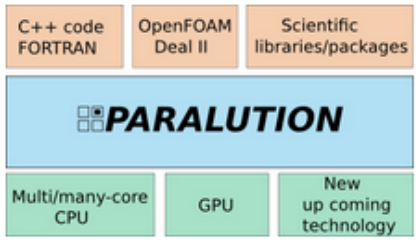
UNIVERSIDAD DE SEVILLA

NVIDIA CUDA RESEARCH CENTER

Bibliotecas Cientificas

Geometry Performance Primatives(GPP)

GPP is a computational geometry engine that is optimized for GPU acceleration, and can be used in advanced Graphical Information Systems (GIS), Electronic Design Automation (EDA), computer vision, and motion planning solutions.



Paralution

A library for sparse iterative methods with special focus on multi-core and accelerator technology such as GPUs.

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

- Aplicaciones CUDA en Ciencia e Ingeniería

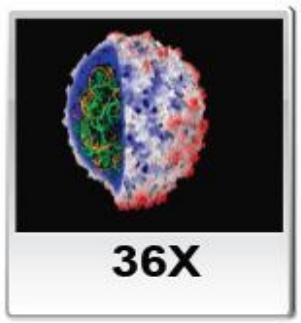
http://www.nvidia.es/object/gpu-computing-applications-es.html

Aplicaciones CUDA en Ciencia e Ingeniería

146X

Imágenes biomédicas

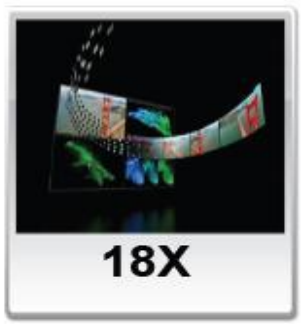
Univ. Utah



36X

Dinámica molecular

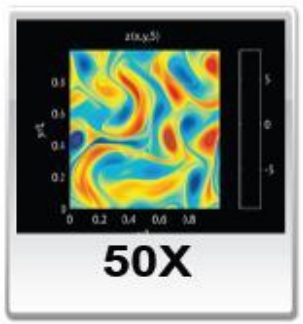
Univ. Illinois, Urbana



18X

Transcoding de vídeo

Elemental Tech



50X

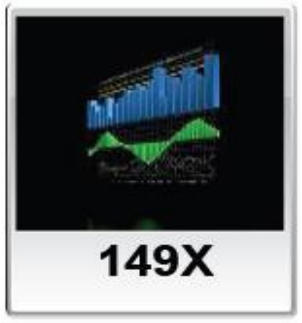
Computación Matlab

AccelerEyes

100X

Astrofísica

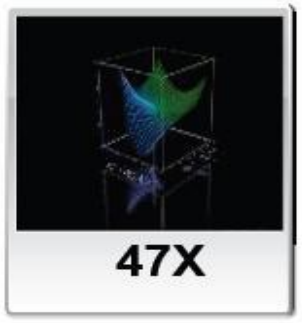
RIKEN



149X

Simulación financiera

Oxford



47X

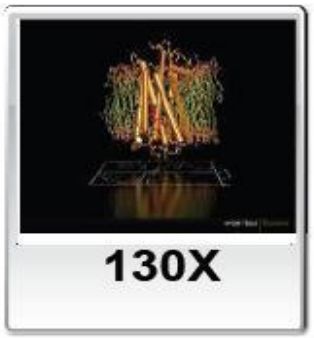
Algebra lineal

Univ. Jaume I

20X

Ultrasonidos 3D

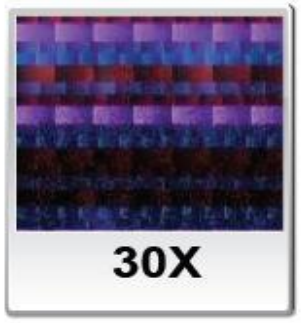
Techniscan



130X

Química cuántica

Univ. Illinois, Urbana



30X

Secuenciación genética

Univ. Maryland

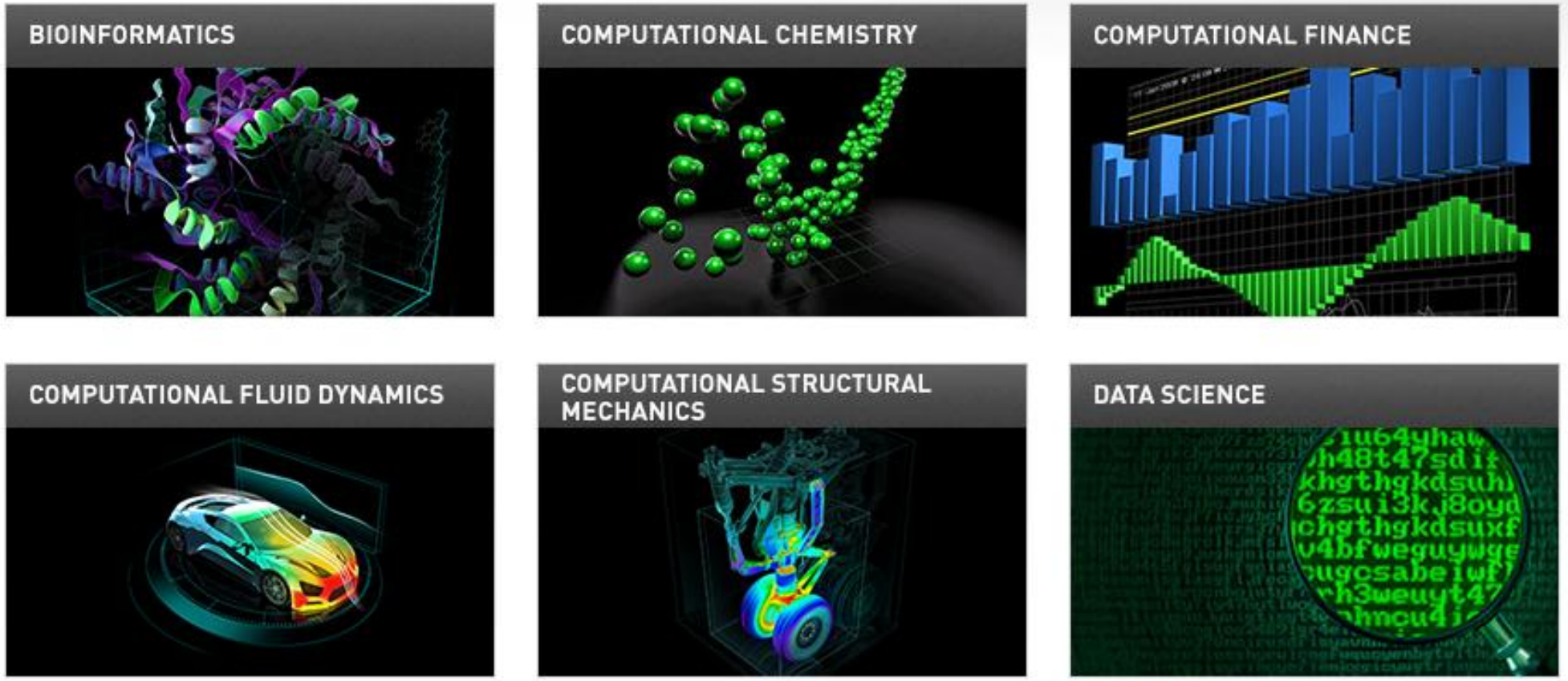
Aplicaciones CUDA en Ciencia e Ingeniería

- **Si comunicas a NVIDIA tus aplicaciones y resultados, las publican en su web.**
- **Abrimos nuevo campo donde publicar: High Performance Computing**

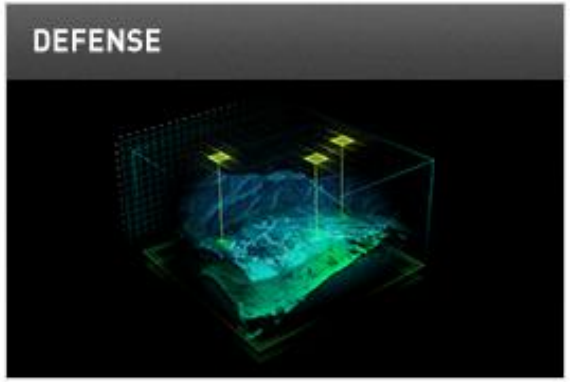
http://www.nvidia.es/object/tesla-case-studies-es.html

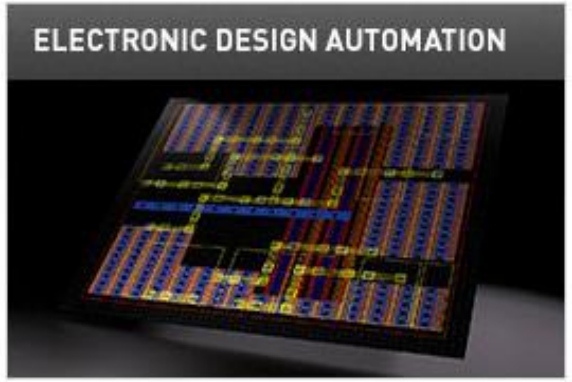
http://www.nvidia.com/object/gpu-applications-domain.html

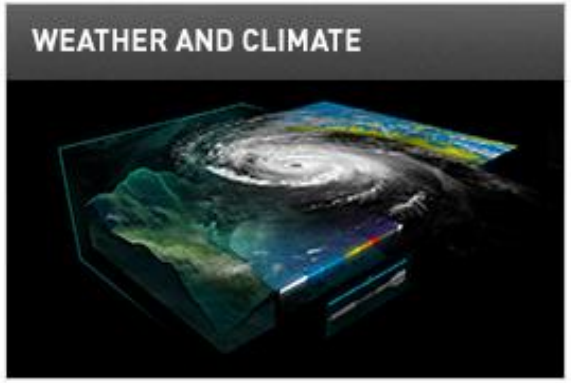
Aplicaciones CUDA en Ciencia e Ingeniería

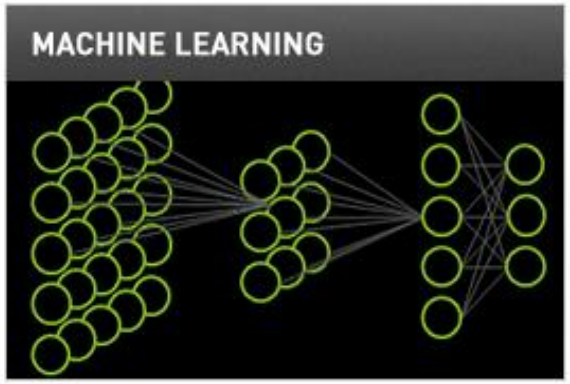


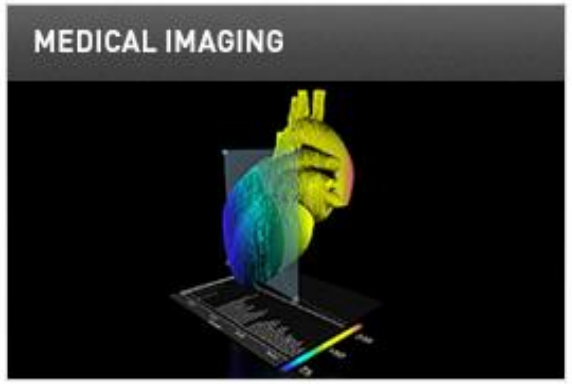
Aplicaciones CUDA en Ciencia e Ingeniería

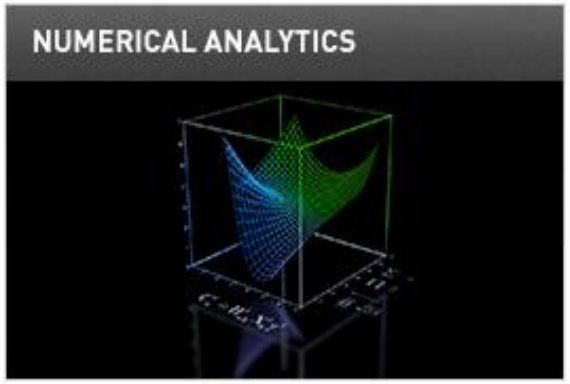




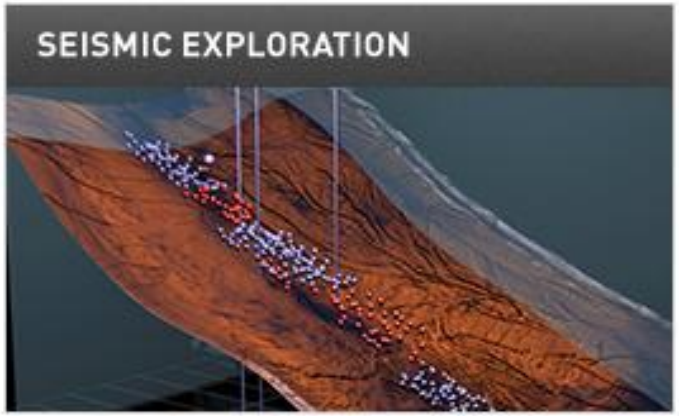


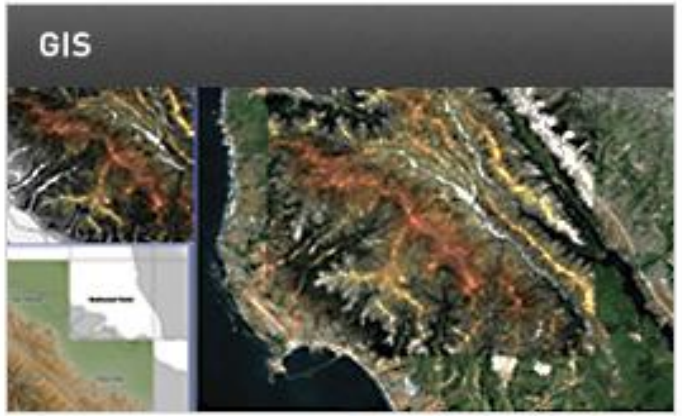


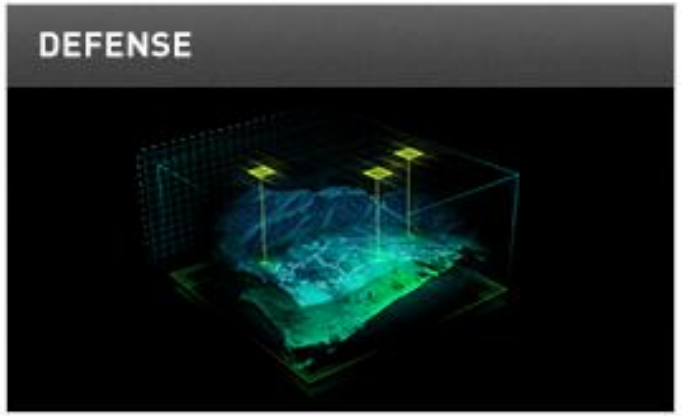




Aplicaciones CUDA en Ciencia e Ingeniería







NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

- Lenguajes de Programación

  https://developer.nvidia.com/language-solutions

  http://gpgpu.org/

Lenguajes de Programación

https://developer.nvidia.com/language-solutions

- **Desarrollo original en: C y C++**
- **Pero y el resto:**
  - **Python, C#, Java, .NET….**

CUDA Toolkit

Provides a comprehensive environment for C/C++ developers building GPU-accelerated applications.

Lenguajes de Programación

CUDA Toolkit

Provides a comprehensive environment for C/C++ developers building GPU-accelerated applications.

OpenACC

Directives for parallel computing, is a new open parallel programming standard designed to enable all scientific and technical programmers.

PGI Accelerator Fortran and C Compilers

Accelerate applications on GPU platforms by adding compiler directives to existing code.

The PGI CUDA C/C++ compiler for x86

Compile and optimize their CUDA applications to run on x86-based workstations, servers and clusters.

CUDA FORTRAN

Enjoy GPU acceleration directly from your Fortran program using CUDA Fortran from The Portland Group.

Anaconda Accelerate

Enables acceleration on your GPU or multi-core processor using Python.

PyCUDA

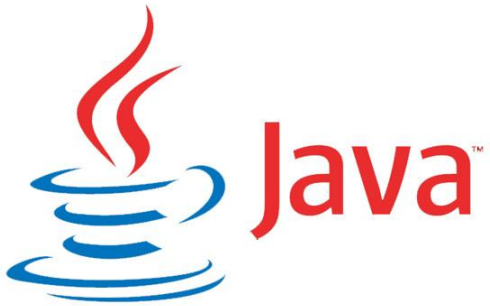
Gives you access to CUDA fuctionality from your Python code.

Altimesh Hybridizer™

An advanced productivity tool that generates vectorized C++ (AVX) and CUDA C code from .NET assemblies (MSIL) or Java archives (bytecode)

https://developer.nvidia.com/language-solutions

Lenguajes de Programación



+

CUDA Toolkit

Provides a comprehensive environment for C/C++ developers building GPU-accelerated applications.

http://devblogs.nvidia.com/parallelforall/next-wave-enterprise-performance-java-power-systems-nvidia-gpus/

https://developer.nvidia.com/language-solutions

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

- En Resumen:
  - Aplicaciones CUDA en Ciencia e Ingeniería

    http://www.nvidia.es/object/gpu-computing-applications-es.html
    http://www.nvidia.es/object/tesla-case-studies-es.html
    http://www.nvidia.com/object/gpu-applications-domain.html

  - Publicaciones e investigación

    http://hgpu.org/
    Universidad de Illinois - http://www.gpucomputing.net/
    Universidad de Cambridge - http://www.many-core.group.cam.ac.uk/projects/

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

- En Resumen:
  - Bibliotecas Científicas
    https://developer.nvidia.com/gpu-accelerated-libraries
  - Recursos para desarrolladores:
    https://developer.nvidia.com/language-solutions
    http://gpgpu.org/
    http://stackoverflow.com/

NVIDIA CUDA RESEARCH CENTER – Univ. De Sevilla

- Qué podemos hacer
  - **Comunidad de interesados en la tecnología GPU**
  - **Comunidad de desarrolladores para Arq. GPU**
  - **Sesiones temáticas para discutir problemas en Arq. GPU**
  - **Sesiones de iniciación para desarrollo en Arq. GPU**
  - **Cualquier tipo de colaboración**

  …..

  - **Lo que propongáis y esteis dispuestos**

Muchas Gracias Preguntas, sugerencias….

Pedro Antonio Varo Herrero

pevahe@gmail.com

Tw: @pevahe91