Introduction to GPU computing with CUDA

Computing lab 1 - presentation

Pierre Kestener

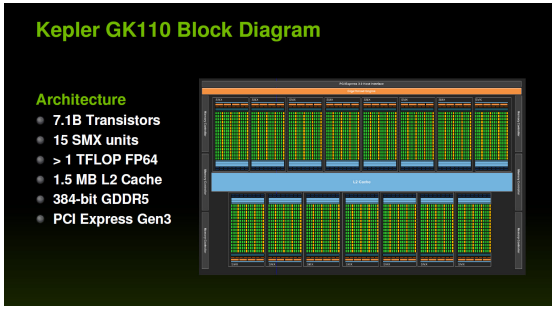
CEA-Saclay, DSM, France
Maison de la Simulation

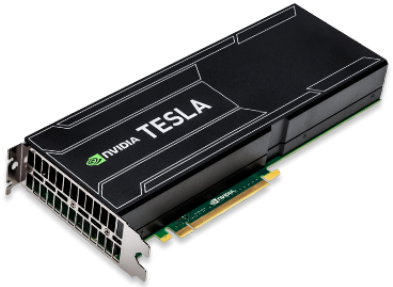
INFIERI, July 15th, 2014

HPC @ Maison de la Simulation

- **Maison de la Simulation** is a joint laboratory CEA-CNRS-Inria-UPS-UVSQ
- **Research and Service Unit**, CNRS USR 3441, High-Performance Computing
  - **HPC oriented multi-disciplinary research lab** (e.g. large scale parallel linear algebra algorithms, ...)
  - **Service unit** offering expertise and support to the HPC users community, especially for high-end software development.
  - **Center for education, training** and scientific animation around HPC

What is the GPU Programming lecture about ?



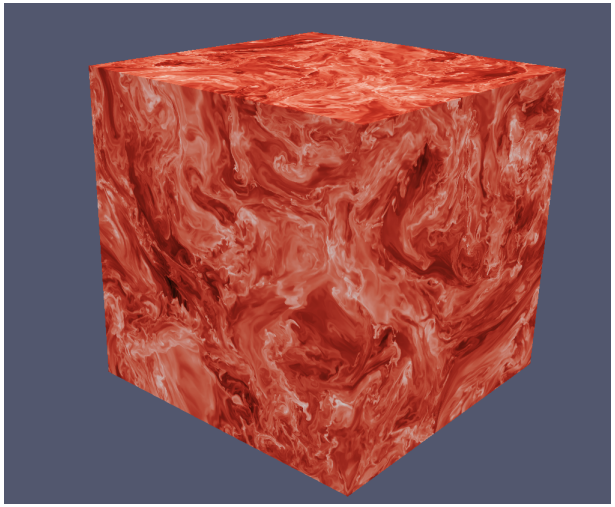


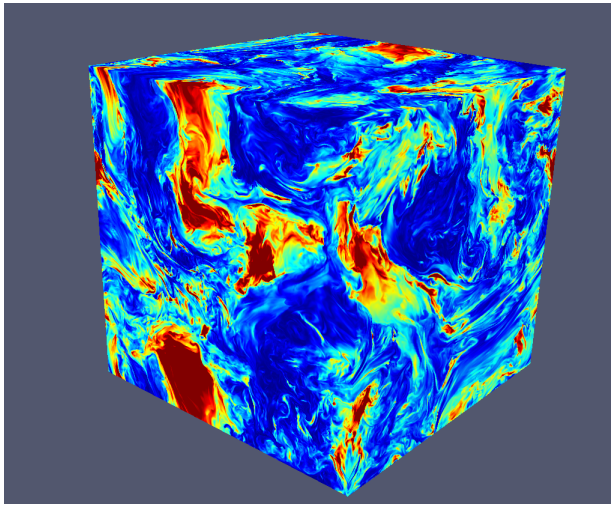
- **GPU : Graphics processing units**
- **CUDA architecture in 2007**: starting point of General-Purpose computing on GPU
- **A short historical overview of GPU:** from highly specific to general purpose processors
- **Today's GPU hardware architecture:** how different is a GPU from a CPU: the hardware point of view
  - **CPU are latency-oriented architectures** (large memory cache, complex control logic, ...)
  - **GPU are throughput-oriented architeures** (massive data parallelism to hide memory latencies, ...)

What is the GPU Programming lecture about ?

- **What is CUDA ?**
  - **a hardware architecture** (an execution model, memory hierarchy)
  - **a programming model** with a C-based programming language with extension specific to massive data parallelism
- Use a simple example to perform **CUDA code walk-through**
- Understand what type of algorithms will better benefit from a GPU implementation
- Give a short list of additionnal material for advanced knowledge of GPU

What is the GPU Programming lecture about ?

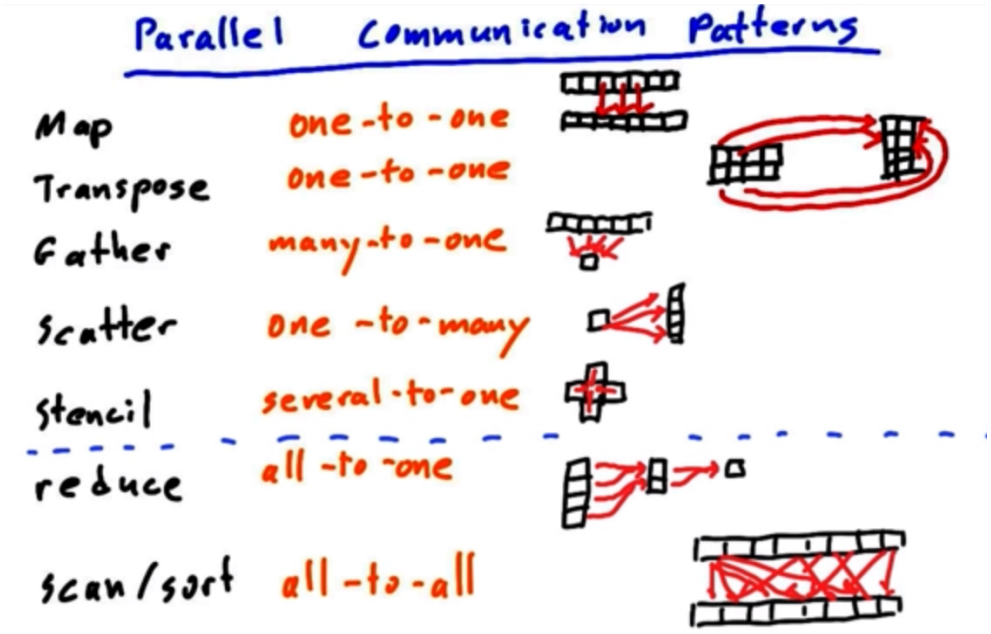




- **Illustrate GPU computing in astrophysics context for HPC applications**: characterizing turbulence in the interstellar medium
  - code RamsesGPU: perform very high resolution compressible MHD simulation on a cluster of GPU (PRACE/CURIE, OLCF/TITAN)
  - Large scale simulation issues: efficient use of GPU, performing parallel I/O with large files, in-situ visualization ...
  - High performance, which metric ? FLOPS/s, GBytes/s, parallel filesystem, ...
  - performing high Mach MHD simulations 2016^3 using 486 GPUs
  - **Scaling RamsesGPU performance on OLCF/TITAN: up to 4096 GPUs**

GPU Programming Computing Lab

- **Date: July 24th**
- Provide access to a remote compute workstation equipped with a high-end NVIDIA K20 GPU.
- Give some **basic exercises/tutorial** to get familiar with the development tools:
  - use of the `nvcc` compiler
  - know your GPU hardware
  - Basics of the programming model: `threadIdx`, `blockIdx`, ..
  - Understanding memory-bound / `compute-bound` algorithm implementation in GPU context



- **Illustrate one** `parallel patterns` **implementation on GPU:** the `stencil` pattern using a **heat equation solver** algorithm.