

Cours de physique théorique de Saclay

Lundi 30/09/2024, 14:15-17:00

Orme des Merisiers Salle Claude Itzykson, Bât. 774

Introduction to CUDA programming for physicists

François Gélis

Le site de l'IPhT

Abstract:

Many numerical computations in physics require applying repetitively the same operations on large arrays, a type of task likely to receive a substantial gain in speed when parallelized. Traditionally, parallelism has been achieved by increasing the number of CPUs in a computer, or the number of computing cores in each CPU.

Graphical Processing Units (GPU) may be viewed as a specialized type of processor with thousands of computing cores. Thanks to a combination of hardware and software developments, their use has expanded from the domain of computer graphics into that of general purpose computing. Besides commercial softwares written to take advantage of the computing power of GPUs, several tools exist to develop computer programs that offload parts of a computation to one or more GPUs.

Among these programming frameworks, CUDA (Compute Unified Device Architecture) is one of the most widely employed. CUDA is an extension of the C++ language, but can be used also in C, FORTRAN, Python and many other programming languages. The goal of this course is to expose the concepts of CUDA programming from the ground up, in order to write simple programs that use GPUs to accelerate critical sections of code.

Architecture of a GPU (computing units, memory, interplay with the host computer). What type of tasks are GPUs good for? Memory operations (allocation, data transfer to and from a GPU). GPU shared memory. CUDA kernels. Streams, synchronization between dependent tasks. Reduction operations on a GPU. Available libraries: cuFFT, cuBLAS, cuSPARSE, cuRAND.

Pour toute information, contacter ipht-lectures@cea.fr