Biomanycores, a repository of interoperable open-source code for many-cores bioinformatics

Jean-Stéphane Varré  Stéphane Janot  Mathieu Giraud
LIFL - UMR CNRS 8022, INRIA, Université Lille 1
jean-stephane.varre@lifl.fr, stephane.janot@lifl.fr, mathieu.giraud@lifl.fr

URL: http://www.biomanycores.org
Licences: Various open-source licences

Graphic processing units (GPUs) are a first step towards massively many-cores architectures, and recent trends blur the line between such GPUs and multi-core processors. Those architectures enable efficient parallel processing at a very low cost. The current GPUs offer coarse-grained level parallelism (work-groups or blocks of independent computations) as well as SIMD-like parallelism (work-items or threads).

GPUs have been used in bioinformatics since 2005, at the beginning tweaking graphics primitives [3, 8]. The CUDA libraries, first released in 2007 [2], have deeply simplified the development on GPUs. In the two past years a growing number of applications have been proposed [13, 10, 7, 14, 5, 12, 9]. The new OpenCL standard [1] should increase this research field and improve the portability of many-cores applications. However, lots of those references are “proof-of-concept” papers, and do not get actually used.

We present Biomanycores, a collection of many-cores bioinformatics tools, designed to bridge the gap between researches in high-performance-computing and usual bioinformaticians and biologists. The goal is both to gather many-cores programs and to propose interfaces to Bio* projects. The language of choice should be OpenCL, but, while no public implementation of OpenCL is available, CUDA projects are included.

The project is still in an early stage of development, but already includes 3 different applications: Smith-Waterman [10], pKnotsRG [11] and Position-Weight-Matrix scan [5], with interfaces to Biojava 1.6 [6], Bioperl 1.52 [15], and Biopython 1.50b [4]. We wish to open as much as possible Biomanycores to other high-performance bioinformatics applications and to better integrate to Bio* projects.

References