Integrated Project (IP)

Code analysis and CUDA programming

Tasks

Develop the following tasks for each proposed theme (list below):

1. Read and characterize the supplied code, converting it into C code if necessary; analyse qualitative and quantitatively the performance of this version (using profilers, including PAPI); identify those parts of the algorithm/code that are better suited for optimization and/or parallel computing.
2. Design and implement a shared-memory parallel version of the critical parts of the code with OpenMP (no need to struggle for optimum code); test and evaluate the resulting performance on one node in the SeARCH cluster.
3. Design and implement a hybrid code that combines a CPU core (the host portion of the code) with a CUDA-enabled GPU device (the kernel with the critical parts; no need to struggle for highly efficient code); test and evaluate performance on one node in the SeARCH cluster.

Methodologies, reports and presentations

1. Teams:
   a. 8 teams, 2 MEI students each (includes ERASMUS student);
   b. One theme per team; each team should apply for a project theme preferably before Monday 19th December;
   c. Themes will be allocated to teams before the end of 2011.
2. Deliverables: extended abstracts, presentations and reports in English; when and what:
   a. An extended abstract (max. 2 A4 pages) before Thursday 26th January;
   b. An individual oral presentation and discussion (20 min) on Tuesday 31st January;
   c. A final report (produced by the team, no longer than 8 pages plus plots and/or annexes) by Tuesday 7th February.

List of themes

1. **A finite volume case study from an industrial application.**
   (*co-Advisor: Stéphan Clain*)
   The goal of the study is the implementation of algorithms for share or distributed memory, powered by GPU technology in the context of the finite volume element method. The application we shall address is a numerical simulation of pollutant in an estuary (2D problem) and of an industrial area (3D problem).
   Note: This project is reserved for the student with a Math research scholarship

2. **3D Poisson-Boltzmann in molecular modelling.**
   (*co-Advisor: Nuno Micaêlo*)
   The main goal of the proposal is to explore the successive over-relaxation (SOR) algorithm to solve the Poisson-Boltzmann (PB) equation when applied in bio-molecular simulation studies, namely the multi-threaded and CUDA versions of SOR. The PB equation is quite relevant to compute the electrostatic properties of bio-molecules in bio-molecular simulation studies, namely in enzymatic catalysis, protein:ligand interaction, redox processes, or protonic equilibrium.

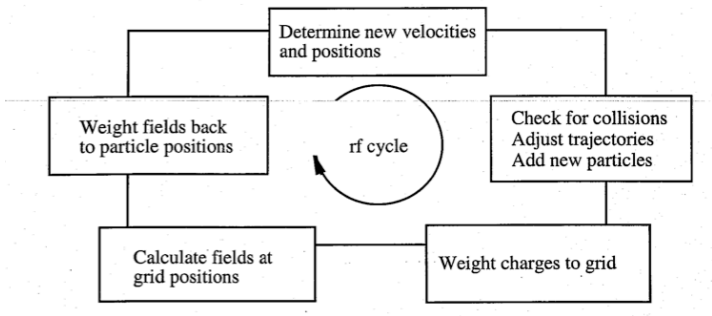
3. **A medical image processing filter.**
   (*co-Advisor: Carlos Manta Oliveira*)
   ... to be further detailed...

4. **Particle collision detection code.**
   (*co-Advisors: António Onofre and Miguel Oliveira*)
   One of the big challenges ahead of the High Energy Physics (HEP) community is the capability of optimizing the codes used by the several research groups to performed the analysis of the huge amount of data collected at the Large Hadron Collider (LHC) at CERN. Given the data rate being fed by the LHC, it is likely a significant gain in data processing speed and ti is of utmost importance for many research groups. Using the current C++ analysis code, LipCbrAnalysis, two fundamental questions may be addressed in this project:

   (i) In the search for signs of new physics, how to significantly improve the processing performance by adapting a version of one of the core pieces of the code;
   (ii) How to improve the efficiency of the I/O handling within the framework of the analysis code (order of magnitude of read data could be as high as few TB).
   In order to perform objective tests, a standalone version of the C++ code LipCbrAnalysis will be made available, together with the full source code and necessary libraries, plus a Monte Carlo simulated data set.

5. **Optimization of Particle In Cell code for plasma device simulation.**
   (*co-Advisor: Luís Silvino Marques*)
   Plasma simulation is the computational modelling of the interaction of charged particles with electric and magnetic fields. Even a low density plasma has trillions of charged particles per cubic centimetre, so we must be very clever to model the statistics of a physical system within the current computational limitations. In particle-in-cell (PIC) codes the plasma is

described using discrete particles (super particles), each representing many charged particles. The super particles interact with fields defined at discrete locations in space (i.e, on a mesh) using interpolation schemes to compute the forces on the particles. The particles generate the current and charge density source terms for Maxwell's equations by interpolation from the particle locations to the mesh. Eventually charged particles make collisions with neutral particles, which is handled by a *direct Monte-Carlo* scheme.



Algorithm for one time step of the simulation

6. **Unstructured 2D finite volume code to model the flow of Newtonian and non-Newtonian fluids.**
(*co-Advisor: Miguel Nóbrega*)
This work aims to analyse a numerical code previously developed to solve the energy conservation equation in 2D, using unstructured meshes. The work should address the porting to a CUDA-enabled GPU and include a quantification of the obtained speedups.

7. **Particle (light) transport Monte Carlo Simulation.**
(*co-Advisor: Eduardo Nunes-Pereira*)
A particle (light) transport Monte Carlo simulation code is presented for the radiative transport of energy in a 3D cylindrical geometry. The scientific goal is to obtain the transmission profile in conditions known as anomalous diffusion. These could correspond to the radiative transport in common fluorescent lamps. The Monte Carlo algorithm is implemented as a simulation that mimics the physical process of radiative transport in an actual sample: the code launches several particle (photon) trajectories inside the cylinder keeping track of where the particle is, as a function of the number of steps.
There are two aspects that make the Monte Carlo simulation demanding in terms of computation power. First, it is well known that the convergence statistics for Monte Carlo simulation codes is slow (converges in the order of the $\sqrt{N}$, with N the number of trajectories; a high number of trajectories is needed to assure practical convergence). Also, the case of anomalous diffusion corresponds to a self-similar fractal without characteristic scale. The probability distribution function is a heavy tailed power law, in which very improbable events are of paramount importance (and therefore a high number of trajectories is needed also to be able to capture these improbable events).
The Monte Carlo code corresponds to a serial implementation and uses modular design (modules group variables each computational task is clearly signaled in the code). It is written in Fortran 90 and makes uses of several utility functions and routines from Press at al., Numerical Recipes, 2nd Ed., Cambridge (2006). It uses a portable random number generator (initialized explicitly) to generate uniform random deviates in the unit interval. These random numbers are used to generate trajectories (generate analytical 2 spherical coordinates angles, and generate numerically the single step size, for each jump in the trajectories).

8. **Reaching for the electronic structure of nanomaterials with millions of carbon atoms.** (*co-Advisor: Manuel Melle Franco*)
   Carbon nanomaterials like carbon nanotubes or fullerenes represent one of the cornerstones of nanotechnology. We have added quantum mechanical capabilities to the molecular modelling software suite tinker (http://dasher.wustl.edu/ffe/) that currently allows to calculate systems with atoms up to 10000 carbon atoms in modest workstations in less than 24 hours. Note that the fastest simulations to date treat no more than 1000-2000 carbon atoms. The more ambitious objective of this project is to try different HPC strategies compatible with the SeARCH cluster to reach for up to 10-x100 larger materials. This will allow to explicitly search for new physics in finite, i.e., non crystalline materials like graphene flakes, carbon nanotubes and nanocones that are currently unreachable. This project is also expected to produce the largest quantum chemical calculation with explicit electrons ever. For the curious, some chemistry/physics papers produced with this methodology are:
   http://onlinelibrary.wiley.com/doi/10.1002/chem.200801818/abstract;jsessionid=63FF0933369E3A07D12
   http://prb.aps.org/abstract/PRB/v71/i15/e153403
   http://pubs.acs.org/doi/abs/10.1021/ja039918r

Additional comments

Each proposed theme is based on an invited talk; the speakers were asked to provide more details on their themes. If any is missing is due to lack of a reply.
On the other hand, some invited speakers are open to offer more themes.