Offer #2024-07877

Energy consumption management and optimization in a high performance cardiac electrophysiology application

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : Temporary scientific engineer

Level of experience : From 3 to 5 years

About the research centre or Inria department

The Inria center at the University of Bordeaux is one of the nine Inria centers in France and has about twenty research teams.. The Inria centre is a major and recognized player in the field of digital sciences. It is at the heart of a rich R&D and innovation ecosystem: highly innovative SMEs, large industrial groups, competitiveness clusters, research and higher education players, laboratories of excellence, technological research institute...

Context

Context

Electricity is a cornerstone of life. Neurons generate electric currents across their cell membrane to transport signals, skeletal muscle cells do so to trigger their contraction, and cardiac muscle cells do both, synchronizing electrically to coordinate contraction across the heart muscle. Malfunctions in cellular electrical systems can have disastrous effects on the organism. Electrical disorders are responsible for half of all cardiac diseases, the most frequent cause of death in the world.

Cellular electric activity is generated by the interaction of many different proteins that are embedded in the cell membrane. Numerical models are vital to understand these complex systems. Yet, realistic simulations of cardiac electrophysiology in structurally abnormal tissue require discretization of the individual cells and their interconnections. This means at least a factor 10^5 increase in model size compared to the usual homogenized models, as well as a different model formulation. This does not only require exascale or larger supercomputers but also a joint effort of biomedical engineers, mathematicians, and computer scientists to build a platform that runs effectively on these machines.

The EuroHPC project MICROCARD and its follow-up MICROCARD2 (2024–2027, 30M) are building such a platform, a true digital twin of the cardiac muscle at the micrometer scale in line with the vision of the European Commission's Virtual Human Twin Initiative. Our lighthouse application will be called ☐CARP. It is a branch of the openCARP code that is dedicated to exascale cell-by-cell simulations. Where possible, our developments will be merged back as improvements to openCARP.

Objectives

Powerful time- and energy-efficient solvers preconditioners for large systems of linear equations are crucial for the exploitation of upcoming exascale and post-exascale computing resources. MICROCARD2 leverages a multidisciplinary collaboration between applied mathematicians and HPC scientists to deliver algorithms and codes that are tailored to these scales, the particularities of our lighthouse exascale application and those of heterogeneous accelerator-based architectures, striving for high performance

and energy efficiency. For example, space and time steps for integration of the membrane models can be increased when no activation wavefront is nearby. Previous work has shown that it is hard to use this adaptivity to improve performance on a parallel computer, due to the overhead of work redistribution. However, increased step size can still be used to improve energy efficiency. This will be handled in the integration methods for membrane models and in the linear system solvers and preconditioners. In the membrane models, also individual slowly-changing variables.

Energy efficiency will also be handled by the runtime system at the node level. On heterogeneous architectures, choosing the right computing unit to perform each computation can reduce energy consumption. Depending on the task's characteristics, allowing a small performance degradation can lead to large energy savings. Also data movement has an important effect on both performance and energy consumption. This cost should be considered when placement of computations implies data movement. Powercapping may also help to reduce energy usage. In some cases using more nodes while enforcing a powercap provides better performance than using fewer nodes at their maximum limit. The power budget for each task must therefore be determined. In order to dynamically set the powercap, we will benefit from the performance models already provided by the StarPU task-based runtime system developed by Inria team STORM, and augment them with powercap impact, after studying the impact of power capping on the performance models.

Assignment

Load imbalance as incurred by imperfect mesh partitioning, algebraic adaptivity, direct solver fill-in, or solution-dependent solver iteration count can be exploited to reduce power consumption and energy use. **The first part of the mission** will be to derive rough predictions of the computational work to be done on each device up to the next synchronization barrier, and then to implement throttling techniques such as enforcing a lower frequency scaling (when available), a lower parallelism degree, or the selection of kernel variants with lower energy consumption profiles.

The technical details of hardware computing platforms at supercomputing centers vary significantly due to their use of different combinations of technical solutions, vendors and generations. **The second part of the mission**, will be the adaptation of the computing kernels to the target hardware platform. This part of the work will therefore focus on porting the hardware-dependent part of the project's code generation and runtime systems onto the targeted computing platforms to ensure the application can access the full extent of the available computing resources – such as the CPU SIMD instruction sets and accelerator devices – and can access the monitoring and control services – such as the energy consumption metrics reporting and the power capping controls – in an abstract and portable manner, when available. In parallel, this second part of the work will leverage task scheduling, kernel performance modeling and variant selection, and optionally worker thread frequency scaling techniques if available to unprivileged users on the target platforms, to optimize the energy consumption level and enforce selected power capping profiles at the compute node level. The aim is to obtain an efficient trade-off between the potentially conflicting goals of obtaining short response times and minimal environmental impact.

Main activities

Main activities (5 maximum) :

- Implement energy consumption modeling and throttling techniques to actively manage the application energy consumption at the task-based runtime system level.
- Implement hardware adaptation techniques to tailor the application kernels and the runtime system layer to the target hardware platform.
- Ensure the continuous integration and the quality control of software developments.
- Conduct validation and performance evaluation experiments.
- Participate to software documentation, timely deliverable preparation and reporting, and dissemination & transfer efforts within the project and the community

Skills

Fluent skills are required in the following activities:

- C / C++ programming
- Software development in the Unix environment
- Parallel and distributed programming
- Remote development and operating using ssh
- Development using a version control system such as Git
- Continuous integration using a system such as GitLab-CI

Good technical and scientific skills in both written and oral English are necessary to successfully conduct the mission, interact with the project partners and prepare reporting documents.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

The gross monthly salary will be between 2692€ and 3404€, depending on your qualifications and professional experience (before social security contributions and monthly witholding tax).

General Information

- **Theme/Domain :** Distributed and High Performance Computing
  Scientific computing (BAP E)
- **Town/city :** Talence
- **Inria Center :** Centre Inria de l'université de Bordeaux
- **Starting date :** 2024-10-01
- **Duration of contract :** 2 years
- **Deadline to apply :** 2024-08-31

Contacts

- **Inria Team :** STORM
- **Recruiter :**
  Aumage Olivier / Olivier.Aumage@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

The key to success in realizing this mission is the ability of the candidate at international, multi-disciplinary team-work. MICROCARD2 involves 10 partner institutions from 4

European countries with people from academics and industrial entities, including cardiac physiology specialists, applied mathematics specialists and high performance computing specialists. It is therefore essential to be able to adapt to this level of diversity and to constructively interact with people of significantly different backgrounds.

<div style="border: 1px solid red; color: red;">

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

</div>

Instruction to apply

If you are interested by this job, please could you apply on website jobs.inra with the following documents :

- cv
- cover letter

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.