2021-04221 - Handling code variability in HPC application: application to a bioinformatic application

Contract type: Fixed-term contract
Level of qualifications required: PhD or equivalent
Function: Temporary Research Position

About the research centre or Inria department
Grenoble Rhône-Alpes Research Center groups together a few less than 650 people in 37 research teams and 8 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context
Numerical simulation is a key technology for many application domains. Thanks to the democratization of high performance computers (HPC), complex physics and more generally complex systems can now be simulated routinely. Numerical simulation is considered as the third pillar of sciences (with experiment and theory) and is critical to gain competitive position. HPC aims at maximizing performance of an application on a given architecture.

The classical optimization work flow is based on identifying hot spots or bottlenecks and on optimizing them. Natural design of HPC applications reflects this model where a set of basic building blocks is identified for its performance penalties if not optimized.

However, such an approach does not ensure performance portability except for well identified kernels (1). It does not address issues related to the whole software stack where several serious limitations occur. First, the complexity of hardware and applications imposes coordination of experts to select the best combination of algorithms and hardware for a given problem. Second, an algorithm may also impose new representations of data type which can impose time consuming code refactoring and data layout which could also be dependent of the architecture. Third, composition of several basic building blocks may deliver low performance due to bad resources sharing at runtime.

To achieve portable high performance, efficiently chaining algorithms and/or data structures becomes unavoidable, but it is very complex and takes a lot of time (of refactoring of years or decades old of legacy application). Moreover, co-design of software stack and application targeting parallel architectures to reach high performance raises several challenges. At the bottom level, architectures become more and more complex to program because micro-architecture requires more and more in-depth knowledge to be understood. At the middle level, the distribution of the applications on massively parallel hardware becomes unmanageable. At the top level, composing algorithmic variants of modern codes to enable more efficient applications is still an open challenge.

Assignment
The recruited person will contribute to a methodology to answer the question of how to productively achieve high level of performance of complex software in presence of code evolutions and architecture variations for domain developers. A major scientific objective of the project is then to propose a programming model that enables flexible and maintainable HPC applications to make them robust to code evolutions and architecture variations. This model could be based on an extension of the COMET component model [2] which enables to express and manipulate the application structure but with a limited expressivity. Moreover, we anticipate that such a model will need to extend features from existing runtime models such as OpenMP. Hence, a second scientific objective is to improve OpenMP-like runtime models to better report execution performance and behavior.

The recruited person will work in an stimulating environment in the AVALON team located in the LIP at ENS de Lyon. She will be in connection with the BEAGLE team that brings the Aevol case study [3].

Main activities
- Propose a model superseding COMET to support the abstractions needed for Aevol
- Propose improvement to OpenMP-like runtimes to support such a model
- Develop a proof of concept of the proposed approach
- Conduct an evaluation of the proof of concept

Skills
Technical skills and level required:
- very good knowledge of HPC programming models and runtimes (e.g. OpenMP)
- good programming skills (Python and C++)
- knowledge in (meta-)modeling will be appreciated

Benefits package

General Information
- Theme/Domain: Distributed and High Performance Computing
- Scientific computing (BAP E)
- Town/City: Lyon
- Inria Center: CRI Grenoble - Rhône-Alpes
- Starting date: 2022-02-01
- Duration of contract: 12 months
- Deadline to apply: 2022-03-31

Contacts
- Inria Team: AVALON
- Recruiter: Perez Christian / christian.perez@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.

Instruction to apply
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). Authorization 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.

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.
- 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