



Offer #2024-07769

PhD Position F/M Framework for the Efficient Coupling Parallel Numerical Solvers and Deep Neural Networks on Supercomputers

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

The Centre Inria de l'Université de Grenoble groups together almost 600 people in 22 research teams and 8 research support departments.

Staff is present on three campuses in Grenoble, in close collaboration with other research and higher education institutions (Université Grenoble Alpes, CNRS, CEA, INRAE, ...), but also with key economic players in the area.

The Centre Inria de l'Université Grenoble Alpes is active in the fields of high-performance computing, verification and embedded systems, modeling of the environment at multiple levels, and data science and artificial intelligence. The center is a top-level scientific institute with an extensive network of international collaborations in Europe and the rest of the world.

Context

The candidate will join the [DataMove](#) INRIA team located on the campus of the Univ. Grenoble Alpes near Grenoble. The DataMove team is a friendly and stimulating group with a strong international visibility, gathering Professors, Researchers, PhD and Master students all pursuing research on High Performance Computing.

This work is part of a joint collaboration between INRIA and IFPEN. Expect close collaborations with IFPEN and very likely week-long visits there.

Hiring date is flexible. We expect to hire the candidate in October 2024, but we have the possibility to start the contract sooner if we find a good candidate or even later to accommodate some specific situations.

The city of Grenoble is surrounded by the Alps mountains, offering a high quality of life and where you can experience all kinds of mountain related outdoors activities and more.

Assignment

Context

The field of AI4science is rapidly growing, focusing on the integration of deep learning-based techniques into traditional numerical simulations. This PhD position aims to investigate software solutions that enable the efficient and flexible coupling of deep models with highly parallel numerical simulation codes on supercomputers.

Deep Models are data-driven Neural Networks (NN) trained to approximate physical, chemical, or biological processes. Various deep model architectures have emerged, including classical CNNs and MLPs, as well as more advanced models like GNNs, DeepONets, Transformers, and Flow Networks. These models can incorporate physics knowledge through different modalities, such as data from observations, synthetic data generated by simulation models, or PDE equations as regularizer terms in the loss function. The goals of such deep models are diverse, ranging from error reduction through super-resolution, capturing uncertainties, to supporting the resolution of inverse problems. The deep model needs to be trained first, and then it can be used in inference mode, either standalone or coupled with a traditional solver.

Research Objectives and Challenges

The objective of this PhD position is to investigate software solutions that enable the efficient and

flexible coupling of deep models with highly parallel numerical simulation codes on supercomputers. The research questions to be addressed include:

- Resource Sharing: How to share CPU and GPU resources on supercomputers effectively?
- Coupling Complexity: How to enable coupling without significantly increasing the application software complexity?
- Generalization: How to design a solution that is generic enough to adapt to various scenarios and codes, without requiring extensive customization efforts?

To address these challenges, the candidate will evaluate related work and existing coupling solutions, such as multiphysics code coupling, in-situ data processing, and workflow engines. The message passing parallel programming model supported by MPI is the standard approach for numerical simulation code parallelization, but the candidate will also consider alternative models, such as task-based and actor-based approaches.

The candidate will also consider different granularities of coupling, from one Deep Model instance being deployed at the level of each compute node up to having a single instance shared by all the code. The choice is often dictated by the type and locality of data that need to exchange the simulation code and the Deep Model.

Previous work in the team investigated how to couple MPI simulation code with task based Python frameworks like Ray and Dask. Dask is a popular open-source library for parallel computing, which has been used for coupling MPI simulation codes (<https://theses.hal.science/tel-04194958>). Ray, a framework adopted by the ML community, already support different types of hybridation with the deep learning frameworks JAX, Pytorch as well as the MPI model, and is used for coupling as in the actor/critics model for reinforcement learning. We will also investigate inference specific solutions as supported by ONNX for instance.

Skills

Candidates should have a Master in computer science, engineering degree or equivalent. PhDs in France are 3 years long and so the candidate will benefit from a 3 year contrat with INRIA to pursue her/his PhD and will get her/his PhD title from the Univ. Grenoble Alpes.

Candidates should have a strong taste for research with high technical and scientific skills, knowledge in machine learning, distributed/parallel computing. Good programming skills (C/C++, Python, Linux) that will enable them to develop prototypes, design and run large scale experiments on supercomputers to demonstrate the qualities of their scientific contributions. For that purpose you will have access to various academic supercomputers such as the Jean-Zay CPU/GPU machine (<http://www.idris.fr/eng/jean-zay/cpu/jean-zay-cpu-hw-eng.html>).

Good communication skills are expected as candidates will have to prepare quality scientific publications in english, and oral communications at international conferences and venues. A good level of English (written, oral) is thus required. French is not mandatory and INRIA will provide French classes if needed.

Please send with your curriculum, any element that will help us to better assess your skills, like internship or master reports, git code repository, as well as a few references to persons we can contact to get some feedback on your qualities.

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 (90 days / year) and flexible organization of working hours (except for internship)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage under conditions

Remuneration

1st and 2nd year: 2082 euros gross salary /month

3rd year: 2190 euros gross salary / month

General Information

- **Theme/Domain :** Distributed and High Performance Computing
Scientific computing (BAP E)
- **Town/city :** Saint-Martin d'Herès
- **Inria Center :** [Centre Inria de l'Université Grenoble Alpes](#)
- **Starting date :** 2024-10-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2024-07-10

Contacts

- Inria Team : [DATAMOVE](#)
- PhD Supervisor :
Raffin Bruno / bruno.raffin@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.

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.

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). 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.