



Offer #2025-08941

Post-Doctorant F/H Preuves de programmes flottants sur GPU

The offer description below is in French

Contract type : Fixed-term contract

Renewable contract : Yes

Level of qualifications required : PhD or equivalent

Fonction : Post-Doctoral Research Visit

Level of experience : Recently graduated

About the research centre or Inria department

Le centre de recherche Inria de Saclay a été créé en 2008. Sa dynamique s'inscrit dans le développement du plateau de Saclay, en partenariat étroit d'une part avec le pôle de l'Université Paris-Saclay et d'autre part avec le pôle de l'Institut Polytechnique de Paris. Afin de construire une politique de site ambitieuse, le centre Inria de Saclay a signé en 2021 des accords stratégiques avec ces deux partenaires territoriaux privilégiés.

Le centre compte 39 équipes-projets, dont 27 sont communes avec l'Université Paris-Saclay ou l'Institut Polytechnique de Paris. Son action mobilise plus de 600 personnes, scientifiques et personnels d'appui à la recherche et à l'innovation, issues de 54 nationalités.

Le centre Inria Saclay - Île-de-France est un acteur essentiel de la recherche en sciences du numérique sur le plateau de Saclay. Il porte les valeurs et les projets qui font l'originalité d'Inria dans le paysage de la recherche : l'excellence scientifique, le transfert technologique, les partenariats pluridisciplinaires avec des établissements aux compétences complémentaires aux nôtres, afin de maximiser l'impact scientifique, économique et sociétal d'Inria.

Context

Avec l'Intelligence Artificielle, les GPUs deviennent de plus en plus spécialisés pour l'accélération de Réseaux de Neurones Profonds. De ce fait, une grande partie de la capacité de calcul de GPUs récents est basée est réservée aux calculs de produits de matrices.

L'objectif est de mener une recherche sur les garanties que l'on peut obtenir lorsque des programmes flottants sont exécutés sur des GPUs.

Des déplacements en conférence sont prévus, les frais de déplacements seront pris en charge dans la limite du barème en vigueur.

Assignment

Avec l'aide de Sylvie Boldo, la personne recrutée sera amenée à essayer de prouver des programmes sur GPU en partant des propriétés qu'on peut inférer du hardware. La principale difficulté est le manque de standards. Les formats de calculs utilisés par ces applications (FP8, FP16, BF16) ne sont pas tous décrit par le standard IEEE et donc ne sont pas contraints. Un nouveau standard pour les tout petits flottants est en cours d'être rédigé, mais sera probablement assez laxo sur les questions de précision. Et de toute façon, l'industrie avance plus vite que IEEE, avec de nombreuses modifications en parallèle et est souvent secrète sur la précision et les algorithmes utilisés.

[1] IEEE, 754-2019 - IEEE Standard for Binary Floating-Point Arithmetic, juin 2019.

[2]N. Wang, J. Choi, D. Brand, C.-Y. Chen et K. Gopalakrishnan, "Training Deep Neural Networks with 8-Bit Floating Point Numbers," in 32nd International Conference on Neural Information Processing Systems (NIPS), 2018.

[3]Intel, BFLOAT16 – Hardware Numerics Definition Revision 1.0, nov. 2018.

[4]I. S. P. W. Group, Interim Report on Binary Floating-point Formats for Machine Learning,

<https://github.com/P3109/Public/blob/main/Shared%20Reports/IEEE%20WG%20P3109%20Interim%20Report.pdf>, sept. 2023.

[5]O.-C. Project, OCP 8-bit Floating Point Specification (OFP8),

<https://www.opencompute.org/documents/ocp-8-bit-floating-point-specification-ofp8-revision-1-0-2023-12-01-pdf-1>, juin 2023.

[6]O.-C. Project, OCP Microscaling Formats (MX) Specification,

<https://www.opencompute.org/documents/ocp-microscaling-formats-mx-v1-0-spec-final-pdf>, sept. 2023.

Main activities

Le candidat devra étudier les GPU, à la fois en lisant les articles et brevets, mais aussi possiblement en faisant du rétro-engineering pour deviner les précisions et calculs effectués.

Cette analyse fera ensuite l'objet d'une formalisation, permettant son usage dans la vérification de programmes.

Skills

- bon niveau d'anglais (écrit, oral)
- une connaissance des GPU ou de Rocq est souhaitée

Benefits package

- Restauration subventionnée

- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

Remuneration

Salaire : 2.788 euros brut/mois

General Information

- **Theme/Domain** : Proofs and Verification
Software engineering (BAP E)
- **Town/city** : Gif sur Yvette
- **Inria Center** : [Centre Inria de Saclay](#)
- **Starting date** : 2025-10-01
- **Duration of contract** : 11 months
- **Deadline to apply** : 2025-09-30

Contacts

- **Inria Team** : [TOCCATA](#)
- **Recruiter** :
Boldo Sylvie / Sylvie.Boldo@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

Il est nécessaire d'avoir de la curiosité et de l'appétence sur les aspects à la fois matériels et formels de l'arithmétique des ordinateurs.
Une thèse dans l'un de ces domaines serait un atout.

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.